6-14-2016 12:00 AM

Joint Modelling in Liver Transplantation

Elizabeth M. Renouf, *The University of Western Ontario*

Supervisor: David Bellhouse, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

In the setting of liver transplantation, clinical trials and transplant registries regularly collect repeated measurements of clinical biomarkers which may be strongly associated with a time-to-event such as graft failure or disease recurrence. Multiple time-to-event outcomes are routinely collected. However, joint models are rarely used. This thesis will describe important considerations for joint modelling in the setting of liver transplantation. We will focus on transplant registry data from the United States. We develop a new tool for joint modelling in the context where a critical health event can be tracked in the longitudinal biomarker and often presents as a non-linear trajectory with a sharp jump. We capture this non-linearity with a single change-point longitudinal component that is linked to the survival model via random effects in a way that incorporates the size of this change, which is a novel way to use a sharp change in the subject-specific random effect as a linkage in a joint model. We also propose an alternative to time dependent analysis of treatment effects by using a joint survival outcome model with a time-to-drug-failure event and a terminal event in graft failure that is more appropriate to use in drug effectiveness studies where subjects are discontinued from an immunosuppressant (in favour of alternative treatment) due to health reasons. Modelling drug regime failures as a time-to-event process has not been previously considered in transplant studies. We show that this method shows a significant association of time-to-drug-failure with time-to-graft-failure, whether applied with a longitudinal component or on its own in a joint survival outcome model.

Co-Authorship Statement

Paper Title: A joint model for change-point longitudinal data with critical events.
Publication: In preparation.
List of authors: Elizabeth Renouf, C.B. Dean, David Bellhouse, Vivian McAlister
The issue of drug efficacy after liver transplantation as a topic of study for this thesis was proposed by Dr. McAlister, Dr. Bellhouse and myself as an extension of previous work together. The joint modelling approach was suggested by Dr. Dean and Dr. Bellhouse. I was responsible for all data analyses using the joint modelling approach. Dr. Dean suggested using Bayesian methodology and a change-point for the longitudinal data. Further methodology such as the value of using the sharp jump approach was developed with Dr. Dean and Dr. Bellhouse. Dr. McAlister provided input related to the application to liver transplantation. I wrote the paper with input and suggested edits from the other co-authors.

Paper Title: Lost in transplantation: sources of bias in the analysis of transplant registry data.
Publication: In preparation.
List of authors: Elizabeth Renouf, C.B. Dean, David Bellhouse, Vivian McAlister
The issue of drug efficacy after liver transplantation as a topic of study for this thesis was proposed by Dr. McAlister, Dr. Bellhouse and myself as an extension of previous work together. The joint modelling approach using a time-to-event process for drug failure was proposed by Dr. Dean. Dr. McAlister provided input related to the application to liver transplantation. I was responsible for all data analyses using the joint modelling approach. I wrote the paper with input from the other co-authors.

Paper Title: Drug failure analysis in transplant registry data
Publication: Accepted by: Journal of Statistics in Medical Research
List of authors: Elizabeth Renouf, C.B. Dean, David Bellhouse, Vivian McAlister
The issue of drug efficacy after liver transplantation as a topic of study for this thesis was proposed by Dr. McAlister, Dr. Bellhouse and myself as an extension of previous work together. The joint modelling approach using a time-to-event process for drug failure was proposed by Dr. Dean. Dr. McAlister provided input related to the application to liver transplantation. I suggested the addition of the longitudinal component to the two-outcome joint model. I was responsible for all data analyses using the joint modelling approach. I wrote the paper with input from the other co-authors.

For Kayla Maybee (2002-2011)

Acknowledgements

This work would not have been possible without the endless patience and support of my family throughout this long journey. Thank you for helping me look forward to my goal while still enjoying the present smiling hour.

Thank you to those professors who kept me focused on the doctoral path, to David Bellhouse for listening to every idea and finding the good ones; and to Vivian McAlister who encouraged this project ab ovo usque ad mala. Thank you to Charmaine Dean for giving shape to my thesis with direction that allowed the project to reach its goal. I am grateful to have had the opportunity to write my thesis under the direction of these three very different, yet all thoughtful, patient and encouraging, supervisors.

Thanks also to those in the department of Statistical & Actuarial Sciences for countless opportunities to expand my experience over the years, and to the hardworking support staff, Jennifer and Jane. Thank you to NSERC for the generous support of my doctoral work. Thank you to the SRTR for allowing me the privilege of working with their vast data collection. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

Keywords: Joint models, longitudinal, observational, change-point, survival, liver transplantation, time-to-event

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are many areas of medicine where enrolment in a registry occurs as part of the health care process. Data elements collected by registries often include both longitudinal and time-to-event measurements. For example, in Canada, the United States and Europe, every patient receiving a transplanted organ is documented in detailed registry databases that include baseline information, adverse events, and regular follow up information from the time of wait-listing for transplant until death. In Canada, the Canadian Organ Replacement Register (CORR) collects detailed information on every person waitlisted for transplant, and follows them through their transplant for the rest of their life. In the United States, the Scientific Registry for Transplant Recipients (SRTR) fulfills a similar role. These databases include baseline information and repeated measurements on health status and other information at specific times for data collection. Several of these organizations allow researchers to request data sets for research projects and analysis. The Ontario Cancer Registry, the B.C. Cancer Agency Registry, and the New York State Cancer Registry are three examples of mandatory enrolment databases in oncology that capture every patient who has been diagnosed with cancer (excluding non-melanoma skin cancer). The Ontario Mental Health Reporting System is a mandated database that collects information on patients admitted to mental health beds in the province of Ontario including follow up every three months for patients who are still in hospital. The New York State Car-

diac Surgery Reporting System collects data on all patients undergoing coronary artery bypass grafts and percutaneous coronary intervention in non-federal New York state hospitals.

Registry data offer a rich source of real-world information. Physicians and other decision-makers appreciate the external validity gained through analysis of such administrative data. Randomized clinical trials often exclude patients with challenging comorbidities and as a consequence, conclusions reached may be restricted to a specific population group. Analyses of observational data from registries may complement the findings from clinical trials, adding valuable insight.

In transplantation research, the use of joint outcome modelling techniques seems appropriate for many situations, especially since there is much more data collected and available than is commonly utilized beyond a typical survival analysis. And yet, in a field where so much has been published on aspects of survival post-transplantation, surprisingly little work has been done in transplant analyses using a joint outcome model to simultaneously incorporate both longitudinal information available and the time-to-event data. Only Liu et al. [2004] examined recurrent hospitalization events jointly with survival in kidney transplant candidates using data from the SRTR in a paper that was influential in the field of joint outcome modelling. Longitudinal covariate history can afford much insight on the progression of graft status after liver transplant, and statistical methods are available which can incorporate the trajectory for the longitudinally collected data into the hazard function of a survival event such as graft failure or death. There are many other time-to-event applications in transplantation research where the use of joint outcome modelling techniques could prove useful, such as drug exposure and its relationship to other key variables (e.g., time to new onset diabetes after transplant or time to cancer recurrence or occurrence, or viral load and time to recurrence of Hepatitis C). The primary goal of this thesis is to establish that joint outcome modelling techniques are appropriate and useful for transplant registry data analysis.

Alarm bells have been raised for some methods used in observational data. In particular, some research examining overlapping data from the SRTR have come to opposing conclusions.

There is a need to understand these issues, and potential sources of bias should be critically examined. We propose that joint outcome modelling, while avoiding violation of statistical assumptions, gains strength from employing more of the wealth of data that is available to infer on survival outcome.

After liver transplantation, many laboratory values are assessed on a regular basis to give information on the well-being of the patient. Physicians caring for these patients develop an intuitive sense, based on research and practice, of the relationship of these longitudinal lab values to the well-being of their patient. However, the collection of these biomarkers is widely spaced. Often in the intervening time between data collection, a health event triggers a sharp jump in the biomarker. We intend to assess whether we can use the information from such sharp changes to better predict outcome. Interest lies in both understanding and interpreting the meaning of within-subject patterns of change, as well as assessing the relationship between the presence of a sharp jump in the biomarker and the risk of graft failure.

We are also interested in characterising the relationship between the two event processes of graft failure and drug failure. When graft failure occurs, the subject either dies or immediately receives a new liver transplant. Drug failure refers to the event where the initial immunosuppressive regime is discontinued in favour of another one. Determining efficacy of immunosuppressive treatment is not straightforward. Some treatments have a much higher rate of discontinuation and some regimes may be given to a subset of patients with more serious comorbidities.

In Chapter 2 we use a two-stage approach to investigate the relationship of a sharp change in biomarker levels with the risk of graft failure in a joint model with a longitudinal and survival outcome. In Chapter 3 we apply a joint survival outcome model to the problem of drug failure, which takes into account duration of treatment and allows us to include unmeasured confounders through a frailty component. In Chapter 4 we further build on this joint survival outcome approach by adding in a longitudinal component. The goal is to determine whether a joint survival outcome model, in drug failure and graft failure, can add value to the analysis of

drug efficacy. To answer these questions we will explore current techniques in joint modelling using data on liver transplant recipients.

Chapter 1 Bibliography

Liu, Lei and Wolfe, Robert A and Huang, Xuelin. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3): 744-756, 2004.

Chapter 2

A joint model for change-point longitudinal data with critical events

2.1 Introduction

Techniques for joint modeling and its application have been the subject of intensive research in statistical methodology over the last twenty years. The field had its origins in the study of the informative dropout process in the analysis of repeated outcomes, where the time of dropout occurrence is modeled together with the repeated outcome. Important developments grew out of the application to data from HIV/AIDS trials where survival was jointly modeled with a longitudinal laboratory value, CD4 counts. These early models are described in detail by Tsiatis and Davidian [2004], Molenberghs and Verbeke [2005] and Verbeke and Molenberghs [2009]. The important developments can be traced from the two-stage approaches of Self and Pawitan [1992] and others, to the likelihood based approaches of De Gruttola and Tu [1994] and Henderson et al. [2000], as well as the latent class approach of Lin et al. [2002]. With each year, the number of articles published on this topic grows enormously as new research techniques flourish. Extensions to early approaches include multivariate models allowing for multiple time-dependent covariates, first discussed by Song et al. [2002] and Lin et al. [2002]. In an important paper, Proust-Lima et al. [2009] described using a latent class approach to model the link between the longitudinal covariates and the time-to-event. The predictive ability of the longitudinal marker and tools for assessing accuracy of prediction have been studied by

Pauler and Finkelstein [2002], Garre et al. [2008], Rizopoulos [2011], Rizopoulos [2012a], Taylor et al. [2013], and Proust-Lima et al. [2014]. Various approximations to achieve faster computational algorithms have been studied by Rizopoulos [2012b] and Barrett et al. [2015] among others.

Early use of change-point models in the longitudinal component of joint models include Faucett et al. [2002] and Pauler and Finkelstein [2002]. Both papers use a piecewise linear mixed model with random change-point and a Cox proportional hazards model. In Jacqmin-Gadda et al. [2006] the authors employed a piecewise polynomial mixed model with a random change-point for the longitudinal marker and a log-normal model depending on the random change-point for the time-to-event model. Here, the structure of three latent classes is used to account for the correlation between the longitudinal trajectory (cognitive test scores) and the risk of an event (dementia). Garre et al. [2008] proposed a change-point longitudinal component in a joint model examining time to graft failure in kidney transplant recipients. A Bayesian approach was taken with three latent classes and they linked the trajectory of reciprocal serum creatinine for each latent class to the survival model using Cox proportional hazards. The authors noted the computational difficulty of this time-consuming approach using WinBUGS (Lunn et al. [2000]). In a more recent paper, Ghosh et al. [2011] used a longitudinal process with multiple change-points to model viral RNA data and a proportional hazards model for the time to study dropout, clustering subjects according to estimated change-points, and then linking the survival model based on each cluster. After applying a smooth transition function to non-linear longitudinal data, Tapsoba et al. [2011] used a polynomial change-point model with both a correction score and a conditional score method to link the proportional hazards survival model.

These approaches to joint modeling with change-points use a combination of the pre-change-point intercept random effect and the slope random effect before and after the change-point as association parameters in the survival model. Both Faucett et al. [2002] and Pauler and Finkelstein [2002] use the pre-change-point intercept random effect and the pre- and post-

change-point slope random effects in their models. Jacqmin-Gadda et al. [2006] uses pre-change-point intercept along with the pre-change-point slope and further slope coefficients from a piecewise polynomial after the change-point. Garre et al. [2008] simplifies to only the pre-change-point intercept random effect and the post-change-point slope. Ghosh et al. [2011] and Tapsoba et al. [2011] use the pre-change-point intercept random effect and the change in slope at the change-point, with Ghosh et al. [2011] also employing the pre-change-point slope random effect. None of these models examine the post-change-point intercept random effect, or the difference in intercept random effects before and after a change-point. This is of interest in biomarkers where sharp jumps followed by stability can occur. An important element in many studies where health suffers a sharp change which affects survival is the subject-specific measure of the change in a patient's 'benchmark' biomarker level after a crisis as a critical predictor of survival, especially in situations where the longitudinal biomarker tends to be stable around a unique baseline for each subject - a baseline which is determined by measured and unmeasured (and possibly unknown) characteristics of each subject. The adverse event can be tracked via marked non-linearity with a single change-point longitudinal component in a joint model.

We take a novel approach by including the intercept random effect after the change-point, and a measure of the jump at the change-point. In our model, not only is this size of change considered, but also whether the change in the longitudinal trajectory is a health improvement or a health deterioration. Since many biomarkers change sharply during the time of a crisis event and the crisis can have a lingering effect on the biomarker level, we believe that this modeling framework is particularly useful for extracting relevant information from the longitudinal data. We link the survival component through the random effects in a way that incorporates the size of the change in the subject-specific effect, along with an indicator of health improvement or deterioration which can be determined from the biomarker. This is a novel way to use a sharp change in the subject-specific random effect as a linkage in a joint model and an important measure for a clinical context. This framework for analysis will be broadly useful in clinical

practice settings such as transplant follow-up, where patients have frequent data collection of longitudinal biomarkers. We demonstrate the ability of this framework to perform well on long-term registry data where time-points are much further apart. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government. In Section 2.2 we describe the development of our single change-point joint model. In Section 2.3 we motivate the need for accounting for these sharp changes in random effects with an application to patient survival after liver transplantation. We end with a final section containing a discussion and future work in this area.

2.2 Model development

2.2.1 Joint longitudinal and survival submodel

In many medical longitudinal settings we see a sharp jump in previously stable levels of a marker. While we may not know the precise cause of the jump, we can observe whether the change resulted in a health improvement or a deterioration. We can draw information on the magnitude of that change for an individual using a model which incorporates subject-specific random effects before and after the change. This model allows the trajectory of the longitudinal marker to differ before and after the change-point, while assessing the degree of change on an individual basis and how it affects survival.

The linear random effects model with one discontinuous change-point, having n subjects with n_i time-points ($j = 1, ...n_i$) for subject $i = 1, .., n$, with a single subject-specific change-point τ_i for the longitudinal marker y_{ij}, is specified as

$$y_{ij} = \begin{cases} \beta_{01} + \beta_{11} m_{ij} + \boldsymbol{x}'_{ij}\boldsymbol{\beta}_1 + b_{01i} + b_{11i} m_{ij} + \epsilon_{ij} & j = 1, ..., \tau_i \\ \beta_{02} + \beta_{12} m_{ij} + \boldsymbol{x}'_{ij}\boldsymbol{\beta_2} + b_{02i} + b_{12i} m_{ij} + \epsilon_{ij} & j = \tau_i + 1, ..., n_i \end{cases} \tag{2.1}$$

where β_{01} and β_{11} are, respectively, the intercept and slope before the change-point and β_{02} and β_{12} are the new intercept and slope after the change-point; m_{ij} is the time at which the longitudinal measurement is taken; $\boldsymbol{x_{ij}}$ are a set of covariates with respective vectors of regression parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$, before and after the change-point. We assume that the random effects $\boldsymbol{b_i}$ are normally distributed both before and after a change-point, but we allow for a change in these subject-specific random effects after the change-point, i.e. $\boldsymbol{b_i} = (b_{01i}, b_{11i}, b_{02i}, b_{12i}) \sim N(\boldsymbol{0}, \boldsymbol{D})$. The variance-covariance matrix, $\boldsymbol{D} = (\nu_1, \nu_2, \nu_3, \nu_4) I_{4x4}$, where I is a 4x4 identity matrix. The measurement error $\epsilon_{ij} \sim N(0, \sigma^2)$ is assumed i.i.d. and independent of the random effects.

The Weibull (α, λ) model is commonly used for modelling time-to-event data and has probability density function

$$f(t|\lambda, \alpha) = \lambda \alpha t^{\alpha - 1} \exp(-\lambda t^{\alpha}) \tag{2.2}$$

where t is the survival time, and α and λ are the shape and scale parameters respectively. For subject i we assume lifetime has a Weibull (α, λ_i) such that

$$\log(\lambda_i) = \boldsymbol{z}_i' \boldsymbol{\gamma} + W_i \tag{2.3}$$

where $\boldsymbol{z_i}$ and $\boldsymbol{\gamma}$ are covariates and corresponding regression coefficients. W_i are random effects which link the survival to the longitudinal model. We specify them as

$$\begin{aligned} W_i =& \rho_1 \hat{b}_{01i} + \rho_2 \hat{b}_{02i} \\ &+ \rho_3 \delta_{improve} \left\{ (\hat{\beta}_{02} + \hat{b}_{02i}) - (\hat{\beta}_{01} + \hat{b}_{01i}) + \tau_i \left[(\hat{\beta}_{12} + \hat{b}_{12i}) - (\hat{\beta}_{11} + \hat{b}_{11i}) \right] \right\} \\ &+ \rho_4 \delta_{deteriorate} \left\{ (\hat{\beta}_{02} + \hat{b}_{02i}) - (\hat{\beta}_{01} + \hat{b}_{01i}) + \tau_i \left[(\hat{\beta}_{12} + \hat{b}_{12i}) - (\hat{\beta}_{11} + \hat{b}_{11i}) \right] \right\} \\ &+ \rho_5 \hat{b}_{11i} + \rho_6 \hat{b}_{12i} \end{aligned} \tag{2.4}$$

where the vector $\boldsymbol{\rho}$ are association parameters measuring the relationship between the vector of random effects $\boldsymbol{b_i}$ and time to graft failure. Note that the parameter ρ_3 measures the association between survival and the change in random effect for those individuals with a change-point and an associated improvement in health status where $\delta_{improve}$ is an indicator for improvement. Improvement is defined as a decrease in the mean biomarker level after the change-point. We discuss this in more detail in the next section. Similarly, ρ_4 measures the association between survival and the jump at the change-point for those individuals with a change-point and associated deterioration in health status, where $\delta_{deteriorate} = 1 - \delta_{improve}$. The indicators δ are similar to those adopted by Pauler and Finkelstein [2002], who include an indicator variable for whether a change-point occurred in an analysis of PSA biomarkers and their relationship to cancer recurrence. We extend this to two indicators specifying whether the subject experienced a health improvement or a health deterioration after the change-point. If no change-point occurred, both indicators are zero.

2.2.2 Estimation and model comparison

It is challenging in a joint analysis to estimate a change-point in the longitudinal component. For this reason, we are proposing a naive two-stage approach, where the existence and placement of a change-point is determined ahead of time through a separate analysis of the longitudinal component and these are fixed for inference in the joint model. We fit the longitudinal model first and assume the change-point and whether or not there was an improvement or deterioration in health are fixed as estimated from the longitudinal analysis, and then we use this framework in a joint analysis. The change-point was estimated using the methods described in Killick and Eckley [2014]. Hence, the first stage analysis fits only the longitudinal component and yields $\hat{\tau}_i$, $\hat{\delta}_{improve}$, and $\hat{\delta}_{deteriorate}$ and these are fixed in the joint analysis.

No change-point was allowed between the first and second longitudinal observations to assure identifiability of parameters in the model. Also, since the first measurement is taken pre-transplant and the second measurement is six months after transplant, the effect of having

a transplant is such that almost everything has a change-point and it is difficult to separate out the effect of the drug during this time-frame. Thus, subjects who die early will not experience a change-point. We account for this by including the δ indicator variables of the presence of a change-point in our survival model as covariates. This means that we treat those who have no change-point the same way as those who die before a change-point can be estimated. While there are differences between these two groups, we cannot account for this in this model without more frequent longitudinal observations. In the situation where everyone deteriorates, ρ_3 cannot be estimated. Thus if no improvements are expected, the term involving ρ_3 can be omitted.

In the following mathematical descriptions, it is assumed that τ_i, $\delta_{improve}$, and $\delta_{deteriorate}$ are known quantities. In practice, these are estimated through the first stage analysis of only the longitudinal data. The likelihood from the joint model of both the longitudinal values and the survival model is described here. Using the model for the longitudinal component in equation (2.1), the density function of the longitudinal observations, the random effects and the covariates is

$$g(y_{ij}|\boldsymbol{b}_i, \sigma^2, \boldsymbol{x}_{ij})k(\boldsymbol{b}_i|\boldsymbol{D}) \tag{2.5}$$

where $g(\cdot)$ and $k(\cdot)$ are normal density functions. For the survival component, the conditional density function $f(t_i|\lambda_i, \alpha)$ is Weibull with hazard function $h_i(\cdot)$ for the i^{th} individual given by

$$h_i(t|\boldsymbol{b}_i, y_{ij}, \boldsymbol{z}_i) = \lambda_i \alpha (\lambda_i t_i)^{\alpha-1} \tag{2.6}$$

where $\lambda_i = \exp(\boldsymbol{z}_i^{'}\boldsymbol{\gamma} + W_i)$ as described in equation (2.3). Letting δ_i be the censoring indicator, equalling 1 if the event occurs and 0 otherwise, and letting T_i be the observed time-to-event variable, which is the minimum of the event time and censoring time, the joint likelihood can then be expressed as,

$$L \propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} g(y_{ij}|\boldsymbol{b}_i, \sigma^2, \boldsymbol{x}_{ij}) f(t_i|\lambda_i, \alpha)\ k(\boldsymbol{b}_i|\boldsymbol{D}). \tag{2.7}$$

Estimation was carried out with a Bayesian approach and a Markov chain Monte Carlo (MCMC) algorithm in R (R Core Development Team [2014]) and JAGS (Plummer et al. [2003]) to obtain estimates of the posterior distributions, with the priors specified in JAGS as non-informative: $\beta_{1k} \sim N(0, 10000)$, $\beta_{2k}(k = 1, 2) \sim N(0, 10000)$, the components of $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are each random variables of independent $N(0, 10000)$; similarly, the components of $\boldsymbol{\gamma} \sim N(0, 10000)$, $\alpha \sim \Gamma(shape = 1, rate = 0.001)$, and $\rho_1 ... \rho_6$ are i.i.d. $N(0, 10000)$. The prior distributions for the random effects $\boldsymbol{b_i}$ were specified as $N(\mathbf{0}, \boldsymbol{D})$ with the ν_i as independent random variables with priors $\sim N(0, \sigma_b)$ and σ_b was specified as $\Gamma(shape = 0.001, rate = 0.001)$.

We ran 3 chains for 200,000 iterations with 100,000 burn-in. Convergence was judged by the Brooks-Gelman-Rubin (BGR) convergence diagnostic (Brooks and Gelman [1998]). We used the deviance information criterion (DIC) for model comparison. The DIC was proposed by Spiegelhalter et al. [2002] as a Bayesian model selection tool which gives a measure of goodness of fit while penalizing for the complexity of the model. Covariates were removed from the model if the credible interval contained zero.

2.3 Application

The data we use to illustrate the utility of the model are taken from the Scientific Registry for Transplant Recipients (SRTR), for those patients receiving their first cadaveric liver transplant in the United States between January 1, 2000 and December 31, 2000. The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the U.S., submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. A detailed description of the database is available in Dickinson et al. [2003]. Longitudinal variables collected include international normalized ratio (INR), bilirubin, albumin, creatinine, alkaline phosphatase, aspartate aminotransferase (SGOT), and alanine aminotransferase (SGPT). The

laboratory values are endogenous time-dependent variables. Endogenous variables are typically biomarkers measured on the subject under study. A more detailed definition of an endogenous covariate is provided in Rizopoulos [2012a] and Diggle et al. [2002]. The features of these variables, which are usually measured with error and at different times for each individual, dictate special considerations for analysis, making the use of a survival model with a time dependent covariate inappropriate. Scientific considerations suggest the use of creatinine as our longitudinal biomarker here since, besides being an indicator of kidney function and thus a marker for the potentially toxic effect of the immunosuppressive regime followed by transplant recipients for life, creatinine can serve as a surrogate for a measure of general health.

Data time-points for collection of longitudinal creatinine (mg/dl) data are: immediately pre-transplant (day of), six months post-transplant, one year post-transplant, and then yearly on the anniversary of the transplant date. Creatinine values are collected on the day of graft failure or patient death, however we do not include measurements taken at these times. Using these values would introduce bias since the joint model assumes that the longitudinal measurements are taken at uninformative time-points (Liu et al. [2007]). The joint model does include immediate (day 0) pre-transplant lab measurements for all subjects.

We are interested in characterising the relationship between the event process and the trajectories of the repeated measures, specifically whether there is value in this relationship from a clinical perspective that could inform patient care (i.e., if there is a change-point detected, should this inform the care?). Since the longitudinal measures are likely to be related to the event of graft loss, we need to account for this in our analysis.

We assessed model fit on a separate validation data set also from SRTR (patients receiving their first cadaveric liver transplant in the United States between January 1, 2001 and December 31, 2001).

All subjects selected for this analysis were age 16 and older, receiving a first liver transplant with no other organs transplanted. There were 3207 subjects, with 1321 events (graft failure, defined as death or retransplant). Of these, 1128 patients died and 193 were retransplanted.

Note that, as is traditional in liver transplant studies, death from any cause with a functioning graft is counted as a graft failure. Maximum follow up time was 11.4 years post transplant and median follow up time was 8 years.

2.3.1 Results

In Figure 2.1 we show the Kaplan-Meier survival curve for the training dataset. In the longitudinal component modeling log(creatinine), 738 subjects (23%) experienced a change-point. Figures 2.2 through 2.4 show the longitudinal profile of log(creatinine) over time for randomly selected patients who have more than three data points. We show those without a change-point, those with a change-point and health improvement (i.e. a decrease in the biomarker) and those with a change-point and health deterioration (an increase in the biomarker).

To select the variables for the components of the joint model we started with all covariates deemed significant by recent literature in liver transplant survival models (e.g., Kamath et al. [2001], Merion et al. [2005], Rana et al. [2013]) and proceeded with model selection. We considered immunosuppressive drugs that are commonly given after liver transplantation including tacrolimus (TAC), cyclosporine (CYA), and sirolimus (SRL). We also considered the combinations TAC + SRL and CYA + SRL as separate categories to allow for synergistic effects, and also the evaluation of SRL on its own without contamination of effects of calcineurin inhibitors. We included only baseline immunosuppression, i.e. the drug recorded as maintenance immunosuppression on the day of transplant. If this field was empty, we used the drug at time of hospital discharge. We did not include use of steroids or induction therapies or concomitant mycophenolic mofetil.

The Weibull (α, λ) model (2.3) in the survival component was formulated where the vector of covariates included Hepatitis C virus (HCV) status, hepatocellular carcinoma (HCC) status, immunosuppressive drug at baseline, standardized age, standardized donor age, race (African American or other), diabetic status, gender, whether a split liver was received, previous (re-transplant) malignancy, use of a non heart-beating donor, whether the subject was in fulminant

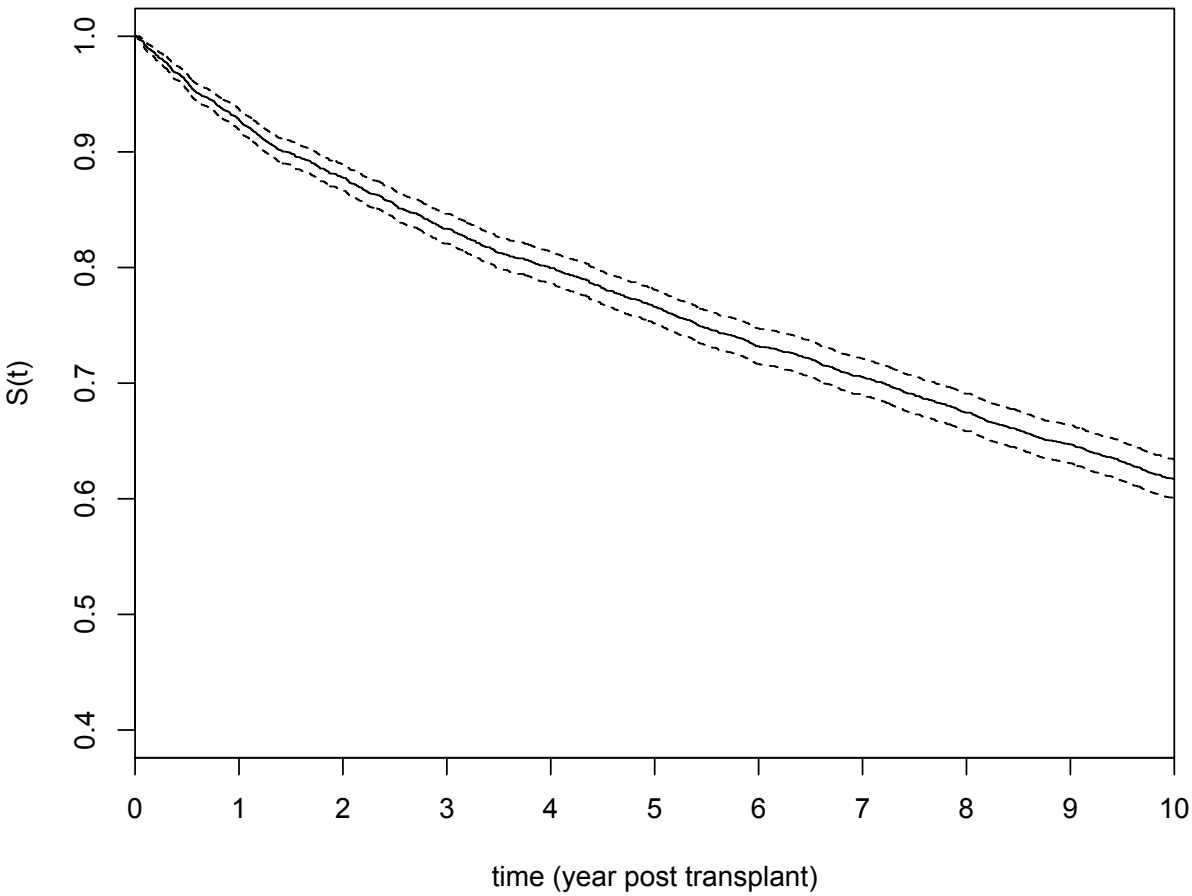

Figure 2.1: Kaplan-Meier survival curve (graft failure) for the training data.

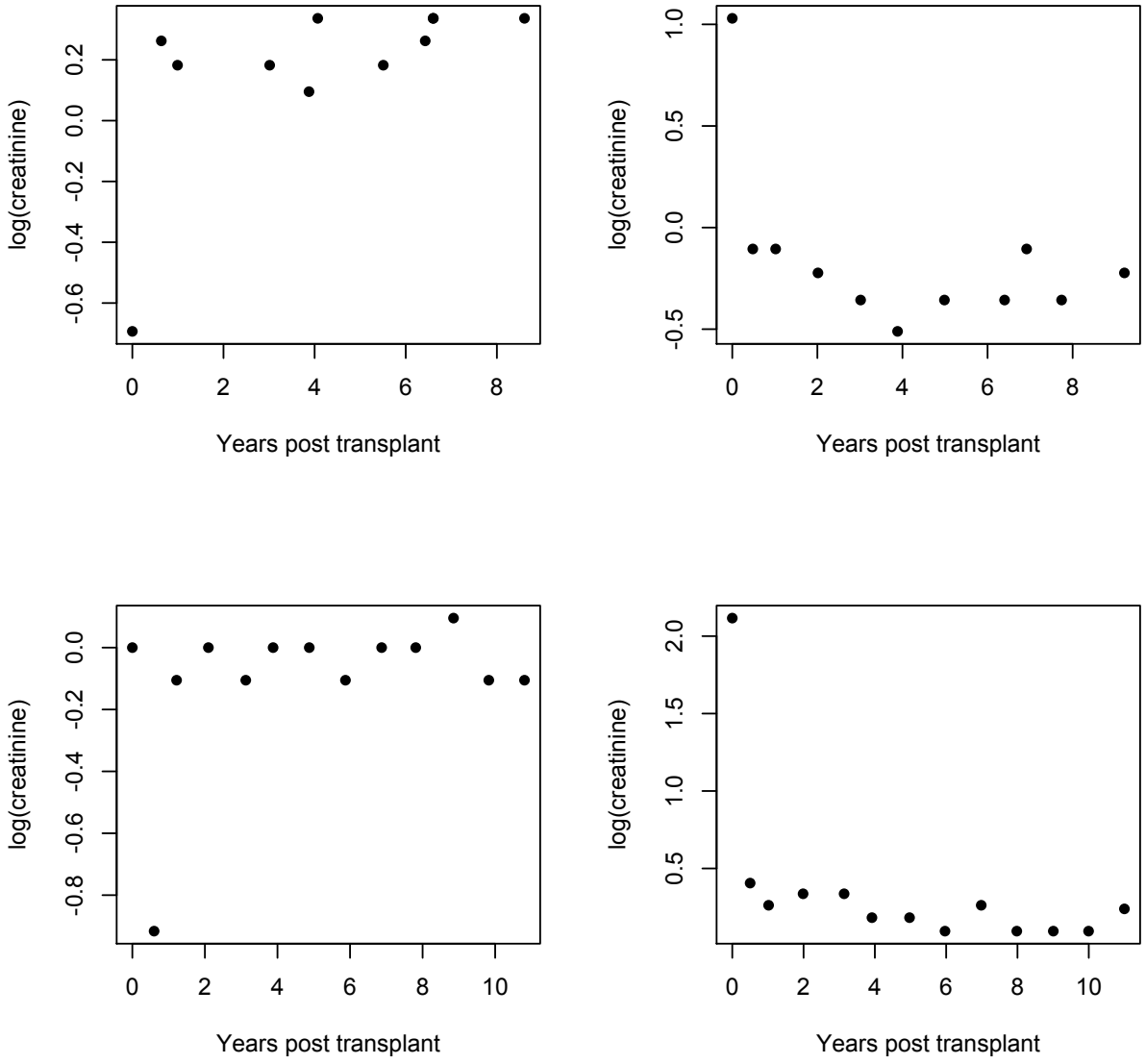

Figure 2.2: Longitudinal plots of log(creatinine) for four randomly selected patients who did not experience a change-point.

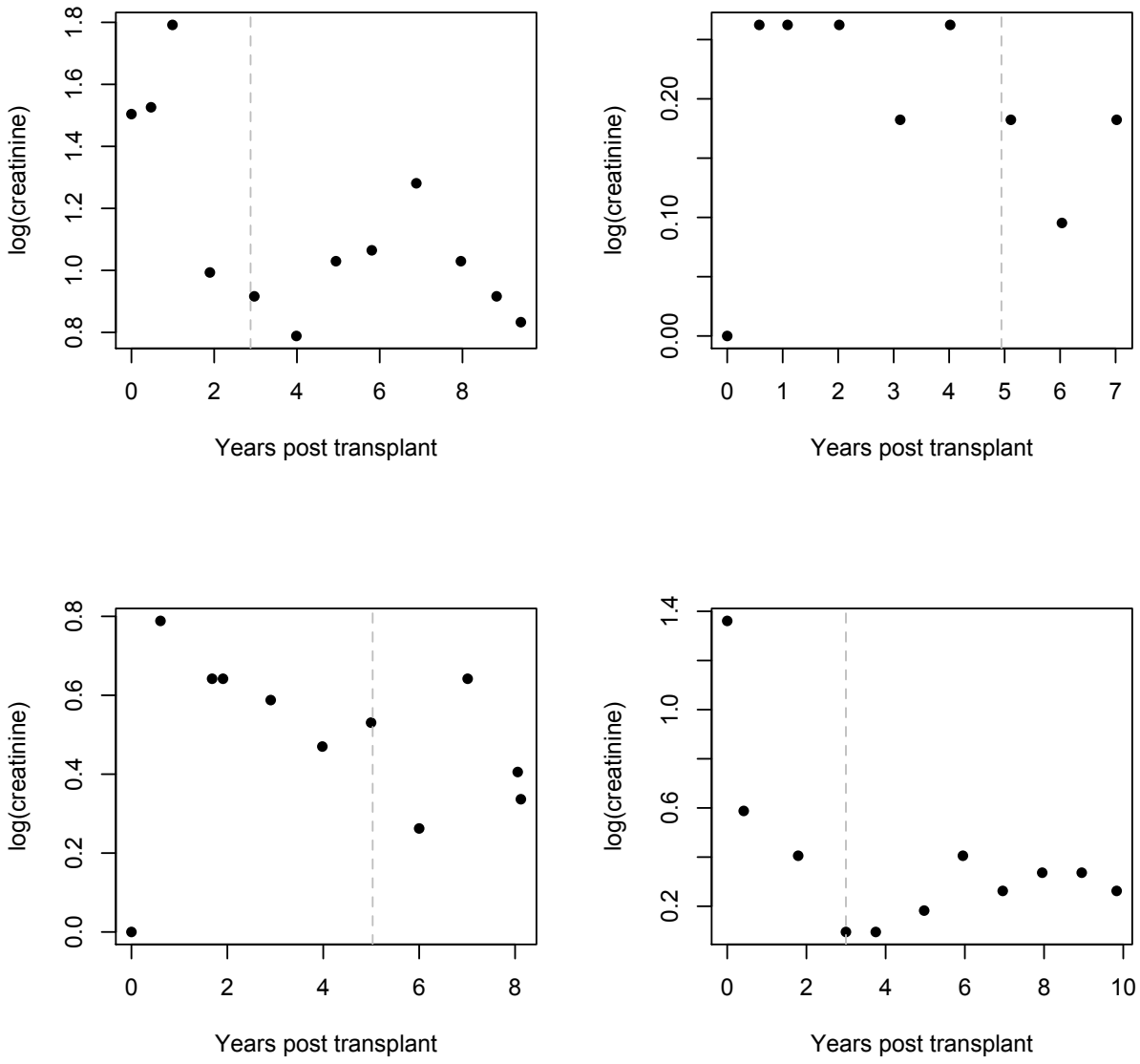

Figure 2.3: Longitudinal plots of log(creatinine) for four randomly selected patients with a change-point and improvement.

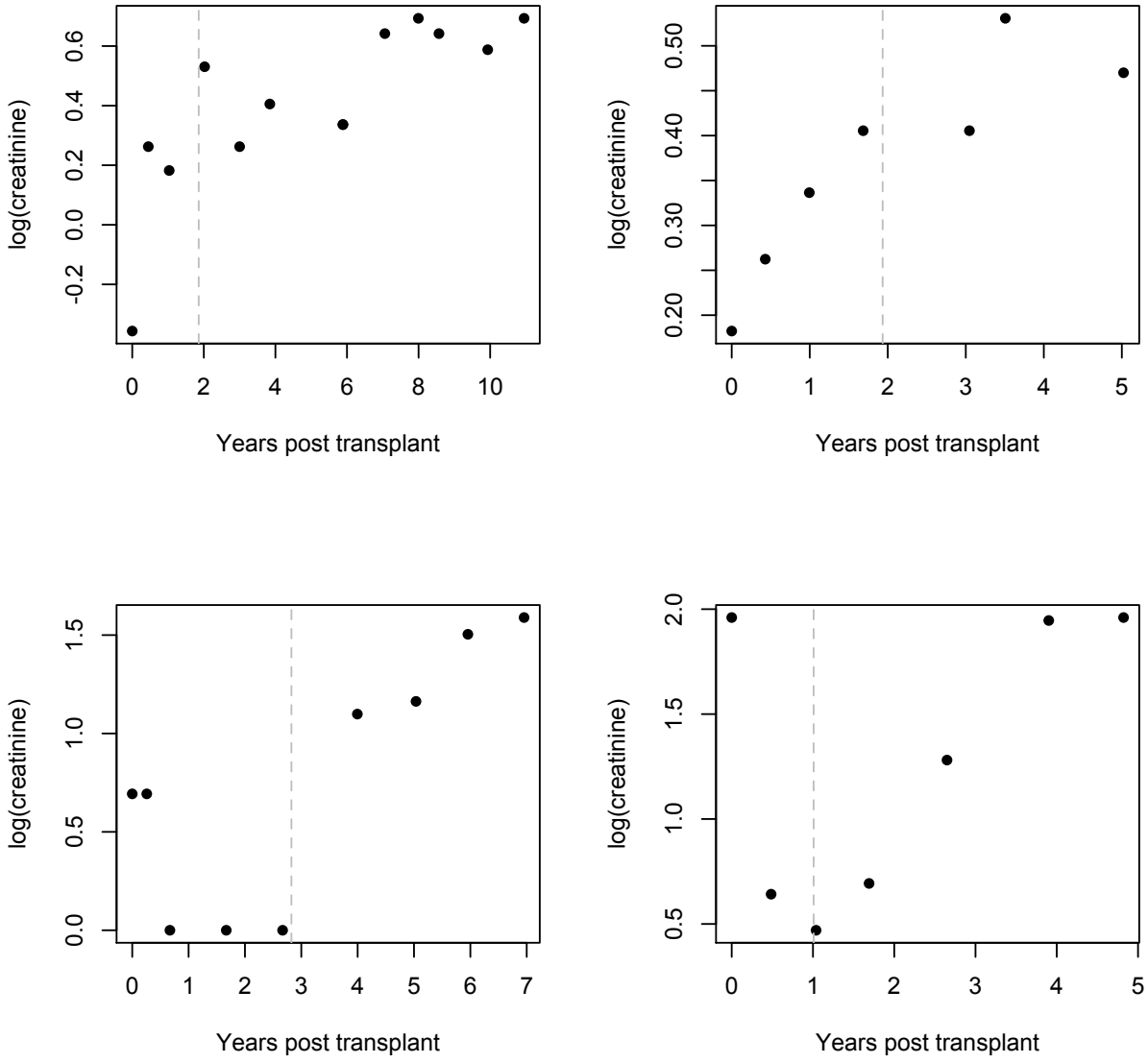

Figure 2.4: Longitudinal plots of log(creatinine) for four randomly selected patients with a change-point and deterioration.

hepatic failure at the time of transplant, and indicators for improvement or deterioration in health status ($\delta_{improve}$ and $\delta_{deteriorate}$).

In the longitudinal component (2.1) the vector of covariates before the change-point includes the same covariates as in the survival component except for the addition of an indicator of whether the subject was in fulminant hepatic failure at the time of transplant because of its significant effect on log(creatinine). Donor age was omitted since it did not significantly affect levels of log(creatinine).

We did not find that any baseline covariates continued to have a significant effect on log(creatinine) trajectory after a change-point. However, we only consider immunosuppressive drug at baseline, i.e., we did not adjust for changes in immunosuppressive treatment. Table 2.1 shows the DIC values for some of the better-fitting survival sub-models, where the longitudinal model was specified as in (2.1). A lower DIC is preferable. We see much higher DIC values when the slope random effects are included, and similarly for the indicator for health improvement ($\delta_{improve}$), indicating the added complexity from including these variables does not benefit the model.

We found that a model without slope random effects and where the size of the jump is estimated by $(\hat{\beta}_{02} + \hat{b}_{02i}) - (\hat{\beta}_{01} + \hat{b}_{01i})$ fit the validation model better and had the lowest DIC. In our example, this is likely because the slope parameters are so small. We also found that the parameters ρ_2 and ρ_3 were not significant in the model. Hence, W_i becomes

$$\begin{aligned} W_i = \rho_1 \, \hat{b}_{01i} \\ + \rho_4 \, \delta_{deteriorate} \left[(\hat{\beta}_{02} + \hat{b}_{02i}) - (\hat{\beta}_{01} + \hat{b}_{01i}) \right] \end{aligned} \tag{2.8}$$

Results from the best fitted joint model are shown in Table 2.3. All of the variables shown have a significant effect on the survival model; note that the credible interval for both gender and HCV contain zero, however they are significant in their interaction (i.e, there is increased hazard for females who are HCV positive). Only one of the immunosuppressive treatments

Linkage with survival sub-model	**DIC**
$\rho_1\ b_{01i} + \rho_2\ b_{02i} + \rho_3\ \delta_{improve}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$ $+\rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right] + \rho_5\ b_{11i} + \rho_6\ b_{12i}$	82581
$\rho_1\ b_{01i} + \rho_2\ b_{02i}$ $+\rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right] + \rho_5\ b_{11i} + \rho_6\ b_{12i}$	52187
$\rho_1\ b_{01i}+$ $+\rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right] + \rho_5\ b_{11i} + \rho_6\ b_{12i}$	47831
$\rho_1\ b_{01i} + \rho_2\ b_{02i} + \rho_3\ \delta_{improve}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$ $+\rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$	40200
$\rho_1\ b_{01i} + \rho_2\ b_{02i}$ $+\rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$	40542
$\rho_2\ b_{02i} + \rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$	39153
$\rho_1\ b_{01i} + \rho_4\ \delta_{deteriorate}\ \left[(\hat{\beta}_{02} + b_{02i}) - (\hat{\beta}_{01} + b_{01i})\right]$	38335

Table 2.1: Bayesian model selection for the Weibull joint model.

Table 2.2: Summary of initial drug therapies.

Initial treatment	n (%)
TAC	2610 (81.4%)
SRL + TAC	175 (5.5%)
SRL	52 (1.6%)
CYA	333 (10.4%)
SRL + CYA	37 (1.2%)

(SRL) is significantly different from the reference drug TAC in terms of impact on survival, however this result should be interpreted with caution since the numbers in this group are small (see Table 2.2). We see that hazard of graft failure is increased for those who are HCC positive, have a race of African American compared to all others, are diabetic, and increased donor and recipient age. We tested all interactions of HCV status with other covariates and found a significant interaction with donor age (higher donor age was associated with increased risk of graft loss for both HCV positive and negative, with an additional risk for HCV positive subjects), and gender (increased hazard for HCV positive subjects who are female compared to no significant difference in hazard by gender for HCV negative subjects). Hazard ratios for the survival coefficients are shown in Table 2.4.

To put the results into context we may consider a subject who experienced a change-point and a subsequent deterioration in health. Table 2.5 shows the minimum, mean and maximum random effect values from this group of subjects before and after the change-point. We see that the hazard of graft loss for subjects who experienced a change-point and a subsequent deterioration in health can be as high as 1.793, after taking into account the random effects (b_{01} and b_{02}), the association parameters (ρ_1 and ρ_4), and the change-point indicator $\delta_{deteriorate}$. For an individual i having the maximum values of the subject specific random effects, this is calculated as follows: $\exp(\rho_1\ b_{01i} + \rho_4[(\beta_{02} + b_{02i}) - (\beta_{01} + b_{01i})] + \delta_{deteriorate})$.

Parameter	**Mean**	**Standard error**	**CI (lower)**	**CI (upper)**
Longitudinal submodel				
intercept before c-p	-0.142	0.034	-0.214	-0.078
intercept after c-p	0.454	0.020	0.415	0.494
slope before c-p	0.00003	0.000002	0.00002	0.00003
slope after c-p	0.000006	0.000005	0.000003	0.000002
age (standardized)	0.007	0.001	0.006	0.009
diabetes	0.047	0.020	0.007	0.086
fulminant failure	0.082	0.028	0.026	0.135
gender (F)	-0.130	0.011	-0.152	-0.109
HCC	-0.082	0.031	-0.144	-0.023
HCV	-0.032	0.011	-0.053	-0.012
variance b_{01}	0.072		0.067	0.076
variance b_{02}	0.245		0.222	0.271
Survival submodel:				
α (shape parameter)	0.901	0.020	0.858	0.946
intercept	-9.075	0.259	-9.602	-8.615
HCV	-0.135	0.150	-0.438	0.146
HCC	0.277	0.142	0.016	0.539
age(standardized)	0.010	0.003	0.004	0.016
donor age (standardized)	0.009	0.002	0.004	0.013
ixn: donor age x HCV	0.009	0.003	0.003	0.015
race (AA)	0.498	0.093	0.311	0.680
diabetes	0.430	0.093	0.247	0.609
gender (F)	-0.138	0.083	-0.302	0.026
ixn: gender x HCV	0.319	0.120	0.088	0.555
previous malignancy	0.387	0.114	0.151	0.597
baseline treatment: TAC	–	–	–	–
SRL + TAC	0.041	0.120	-0.207	0.274 (ns)
SRL	-0.645	0.282	-1.223	-0.118
CYA	-0.021	0.092	-0.203	0.161(ns)
CYA + SRL	0.055	0.246	-0.203	0.161 (ns)
$\delta_{improve}$	-0.690	0.135	-0.975	-0.435
$\delta_{deteriorate}$	-0.765	0.177	-1.115	-0.435
Association parameters in the survival submodel:				
ρ_1	0.529	0.126	0.281	0.770
ρ_4	0.690	0.168	0.339	1.008

Table 2.3: Posterior means for the log hazard, standard deviations and quantiles (ixn = interaction).

Parameter	**Hazard ratio**
HCC	1.32
age (standardized)	1.01
donor age (standardized)	1.01
ixn: donor age x HCV	1.01
race (AA)	1.65
ixn: gender x HCV	1.38
diabetes	1.54
baseline trt - SRL	0.52
$\delta_{improve}$	0.50
$\delta_{deteriorate}$	0.47
association parameters	
ρ_1	1.70
ρ_4	1.99

Table 2.4: Hazard ratios for significant covariates from the survival component of the joint model. All elements except age and donor age are indicator variables, with absence being the reference level (ixn = interaction).

Random effect	**min**	**mean**	**max**
b_{01} (random effect before c-p)	-0.959	-0.025	0.838
b_{02} (random effect after c-p)	-1.327	0.195	1.720

Table 2.5: Summary of values of random effects before and after change-point for subjects with deterioration in health status.

2.3.2 Assessment of fit

We estimated each individual's random effects in the longitudinal model using all longitudinal data from the validation data set (a new data set of 3192 subjects, receiving a first liver transplant in 2001). In this validation set, 721 (23%) of subjects, experienced a change-point. Then, using the saved parameter estimates from the training model, we tested the fit of our model on the new data by estimating $S(t)$ for each subject in the test data set. Figure 2.5 shows the predicted mean survival curve for the test data using the joint model set versus its Kaplan-Meier curve. This model follows the trajectory of the 95% confidence interval bounds for the actual survival curve but is outside the upper bound. It does however, fit more closely than the predictions on the test data than Cox proportional hazards model, one of the most commonly used survival model in liver transplant research. The joint model fits better because it contains the added information from the longitudinal component. By including association parameters for the random effects and the adjustment indicators for the change-point, the joint model allowed for somewhat improved long-term fit. The authors recognize the limitation of this type of prediction in registry data, which requires many years of retrospective longitudinal measurements. Ideally, prediction would be applied in a clinical setting with much more frequent biomarker measurements, updated in real-time. We found that, in comparing covariate effects from the joint model and the Cox proportional hazards model, the hazard ratios for significant covariates were similar, interesting differences are seen in interactions with HCV. Being female was not significant in the joint model (except in the interaction with HCV), likely because the joint model accounts for creatinine levels which are lower in women than men, thus adding more information to the model. Since creatinine is an endogenous covariate, its use as a time-dependent covariate would violate the statistical assumptions of the Cox proportional hazards model, therefore a joint modeling approach is preferred. Table 2.6 shows a comparison of hazard ratios and standard errors for our joint model vs the Cox proportional hazards survival model.

We also show subject-specific comparisons of model fit using the validation data in Figures

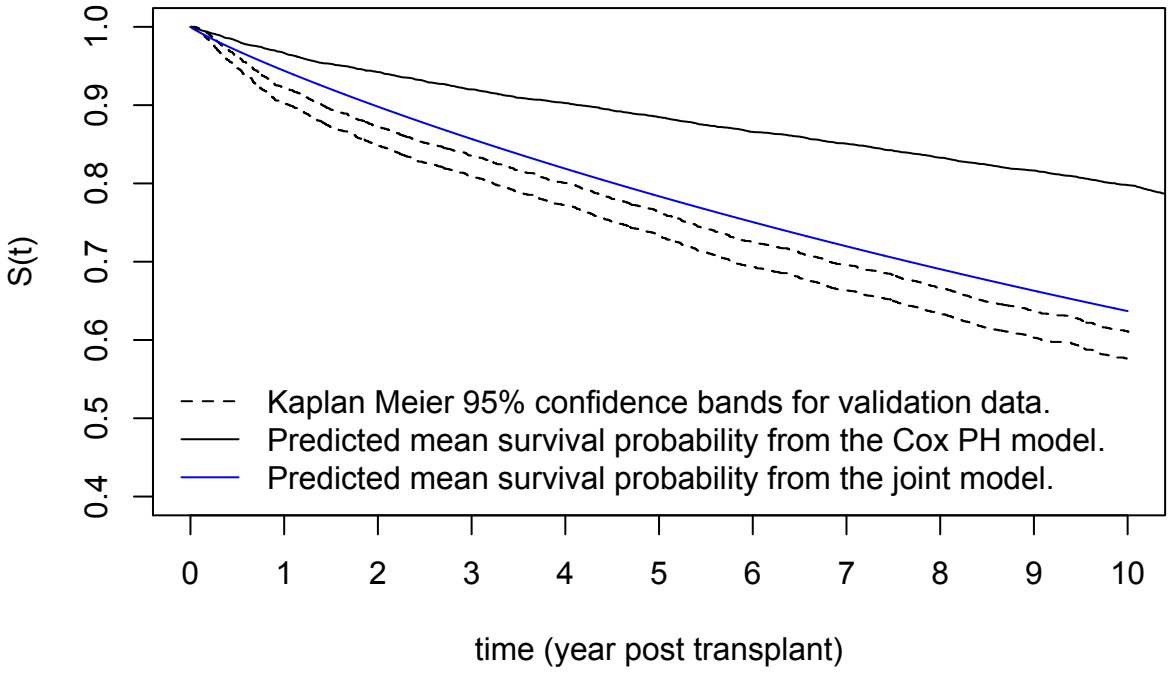

Figure 2.5: Mean predicted subject-specific survival curve for the validation dataset, showing predictions for both the joint model and the Cox proportional hazards model.

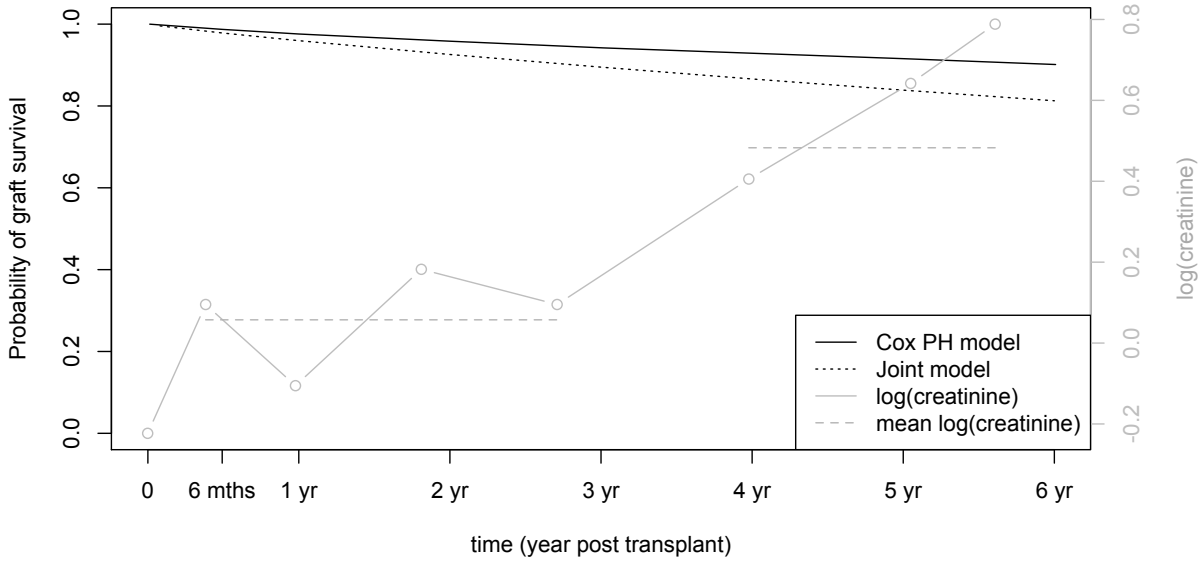

Figure 2.6: Subject-specific predictions of graft survival by model for a randomly selected subject with change-point and graft failure (actual graft failure at 5.7 years). Subject is female, HCV+, standardized age 0.51, standardized donor age -1.73, baseline treatment CYA.

2.6, 2.7, and 2.8. We selected our subject-specific comparisons randomly from subjects with greater than 5 longitudinal data points in four categories: those with change-point and graft failure, those with change-point and no graft failure, and those without a change-point and with graft failure. Other details of the randomly selected subject covariate values are included with the graph information. These curves were fitted using the random effects calculated from all of the longitudinal data in the validation data set. In both cases where graft failure occurred (Figures 2.6 and 2.8), the joint model predicts a lower probability of graft survival than the Cox model. In Figure 2.7, where the subject had a change-point but no graft failure, the joint model predicts a lower probability of survival than the Cox model at time of censoring. Here the joint model assigns a higher risk of graft failure due to a change-point occurrence at four years. By using the joint model with log(creatinine) in the longitudinal component, we have obtained better fit for long term survival after liver transplant. The joint model accounts for the increased risk to graft failure when there is a change-point, or crisis, indicated in the longitudinal biomarker.

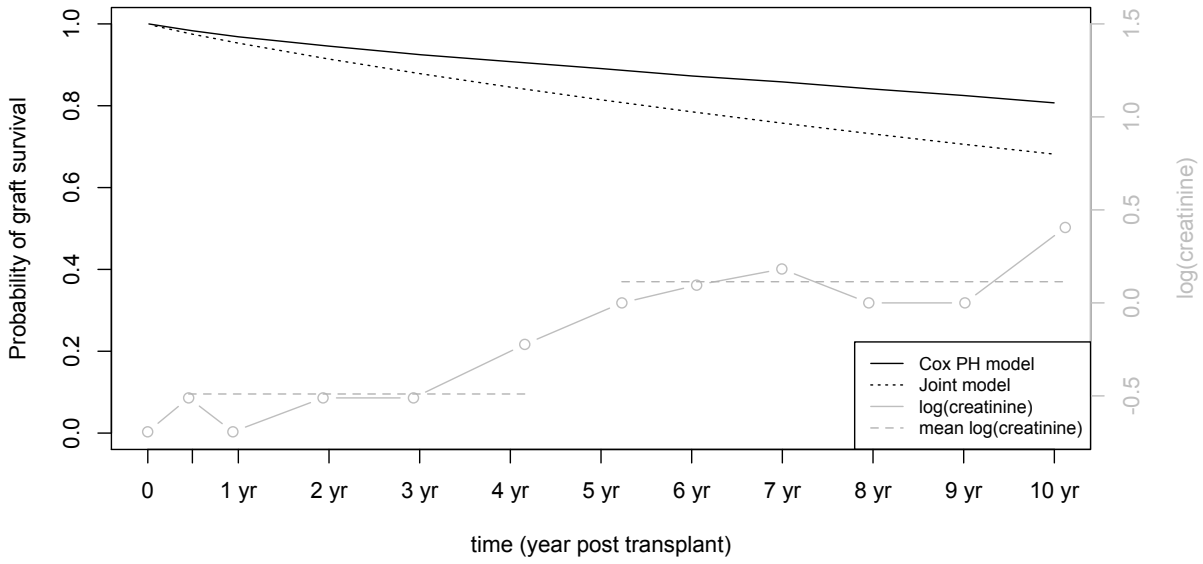

Figure 2.7: Subject-specific predictions of graft survival by model for a randomly selected subject with change-point and no graft failure (censored at 10 years). Subject is female, HCV-, African American, standardized age -0.45, standardized donor age -0.24, baseline treatment TAC.

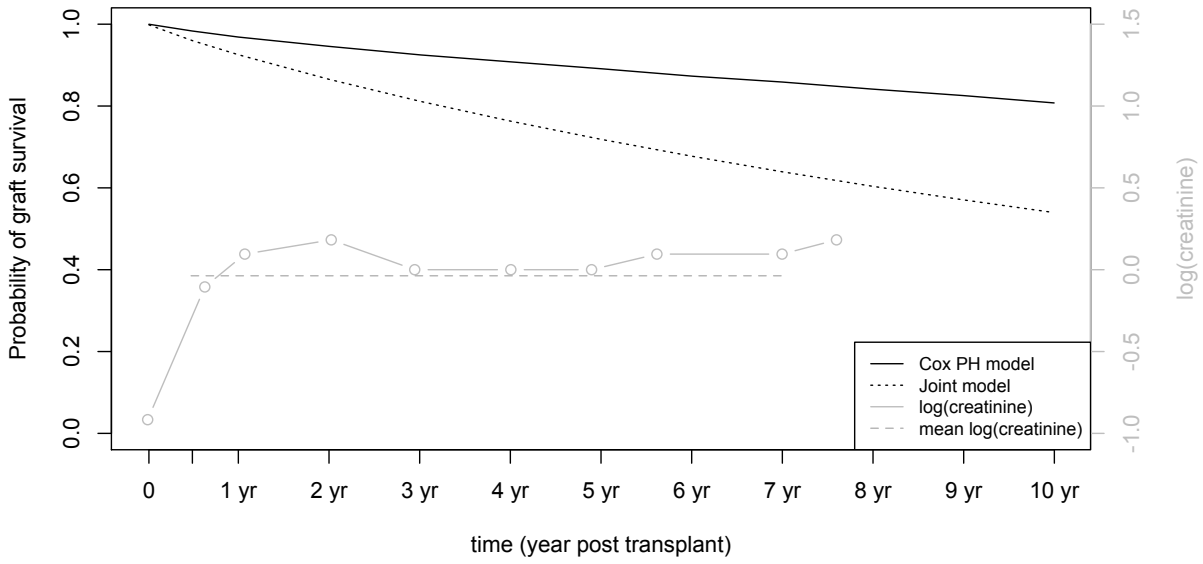

Figure 2.8: Subject-specific predictions of graft survival by model for a randomly selected subject with NO change-point and graft failure (actual graft failure at 9.5 years). Subject is male, HCV+, standardized age -0.45, standardized donor age -0.24, baseline treatment TAC.

2.4 Discussion

The fit of the survival model on the validation data shown in Figure 2.5 is a promising result for liver transplant survival studies, where adequate model fit has been challenging. Most survival analyses do not include testing model fit on a separate validation set, but where it does, results are adequate for the short term only (Burroughs et al. [2006], Rana et al. [2013]) and long-term validations are not attempted. By including longitudinal biomarkers in a joint model, we obtain a clearer picture of long-term prognosis for transplant recipients, as well as the effect on survival of a sharp change in creatinine value. While a proper joint analysis would estimate the existence of a change-point, it's direction, and perform the joint survival analysis simultaneously, this ad hoc approach appears to work reasonably well and is not computationally complex. Ideally, a joint model of this type would be applied in a clinical setting where much more frequent longitudinal values offer the ability for real-time dynamic analysis of subject-specific random effects. Individual-specific shifts in the longitudinal marker could be linked to the risk of graft failure in a way that is clinically relevant and effective.

The broader implication for both observational studies and clinical trials is that joint modeling should be the preferred method to determine parameter estimates and treatment effects. The collection of longitudinal data is standard and the inclusion of this additional data in the survival analysis provides valuable information that results in greater efficiency of parameter estimates (Klein et al. [2013]), even when data collection time-points are far apart, as in registries. Real-time monitoring in clinical practice or clinical trials, with more frequent longitudinal measurements than we have available in the registry data, would result in meaningful real-time random effect calculations. Incorporation of sharp changes in the longitudinal data allows the effect of adverse events to provide additional information to the parameter estimates and provides a better fit for long-term survival estimates.

The model would benefit from improvements to the fit early on, with the flexibility to accommodate a steeper decline initially for the predicted mean survival (cf. Figure 2.5). Our model certainly has limitations that cannot be resolved without more frequent data collection

points, which would allow us to estimate a change-point in the first month after transplant. With a finer time-scale we could distinguish between those who die before a change-point can be estimated, and those who survive a long period with a stable biomarker and no change-point. There are many avenues to explore investigating optimal change-point selection methods, and the incorporation of estimation of the change-point as another individual level random effect in the joint model. More frequent data collection points would allow for study of the criterion for change-point occurrence and the appropriate window size to inform clinical care. Smaller time windows of data analysis with more frequent creatinine measurements would allow the assessment of potential survival bias introduced by the fact that we require repeated measurements in order to assess a change-point here, and our measurements are widely spaced. Since researchers are often limited to using registry data with clinical trial or clinical care data not readily available, we shall in future consider using the pre-transplant creatinine values available in the registry, which are collected on a more frequent basis while the subject is on the waiting list, in a similar joint model. While it is difficult to predict technical failures or other complications at the time of transplant, it is possible that the use of pre-transplant creatinine in a joint model may provide a better fit for the first year post-transplant. Another avenue we intend to explore in a forthcoming paper is the use of a mixture of Weibull models to account for this operative mortality due to issues surrounding the transplant procedure that may not be captured by our model. A time-varying effect for the coefficient HCV is another avenue that warrants investigation.

With transplant data, many possibilities are available for extension to a multivariate longitudinal component. However, the high computational cost when random effects are of high dimension makes this challenging. We found that omitting the slope random effect and including only the intercept and difference random effects allow for a model that is computationally less complex. The direction of the slope is still captured by the use of indicator variables for whether the subject has a mean increase or decrease in laboratory value before and after a change-point. Our model was developed on a small data set, covering only one year of trans-

plant data. With further refinements and using the larger available data set, the results would provide more medical insight. Multiple change-points in joint models have been studied by Ghosh et al. [2011] and are a natural extension to this model, with a possibly cumulative effect on graft failure.

There are many complexities to consider when studying immunosuppression after transplant. For the sake of simplicity in developing the model we took a basic approach to immunosuppressive treatment here since we considered only the drug given at baseline and did not include changes in drug regime. An improved approach will be discussed in the next chapter.

When a liver transplant recipient experiences a serious adverse event such as organ rejection, infection, drug toxicity, or cancer occurrence, it is likely that these events have a negative effect on survival since they can be life-threatening. It follows that severe events exhibit a lingering effect on longitudinal lab profiles. We use log(creatinine) as a proxy for general health and we allow the occurrence of a sharp change in log(creatinine) to add more information to the association between the longitudinal biomarker and the risk of graft loss. Traditional random effects measure the subject-specific change from the mean slope and/or intercept, and we are incorporating more information on the size of the change from a subject's own baseline. This allows us to incorporate a feature of the biomarker trajectory into the linkage between the two processes that conveys more information than just the current level of the biomarker.

Variable	Cox PH		**Joint Model**	
	Hazard ratio	Std error	Hazard ratio	Std error
HCC	1.31	0.14	1.37	0.14
age (standardized)	1.01	0.003	1.01	0.003
donor age (standardized)	1.01	0.002	1.01	0.002
ixn: donor age x HCV	1.01	0.003	1.01	0.003
diabetes	1.54	0.09	1.54	0.09
race (AA)	1.70	0.09	1.65	0.09
gender (F)	0.80	0.08	–	–
ixn: gender x HCV	1.44	0.12	1.38	0.12
baseline trt - SRL	0.55	0.27	0.52	0.28
$\delta_{improve}$	–	–	0.50	0.14
$\delta_{deteriorate}$	–	–	0.47	0.18
ρ_1	–	–	1.70	0.13
ρ_4	–	–	1.99	0.17

Table 2.6: Hazard ratios and standard errors for significant covariates in the traditional Cox proportional hazards model vs a joint model with change-point (ixn = interaction).

Chapter 2 Bibliography

Jessica Barrett, Peter Diggle, Robin Henderson, and David Taylor-Robinson. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):131–148, 2015.

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

Andrew K Burroughs, Caroline A Sabin, Keith Rolles, Valerie Delvart, Vincent Karam, John Buckels, John G O'Grady, Denis Castaing, Jurgen Klempnauer, Neville Jamieson, et al. 3-month and 12-month mortality after first liver transplant in adults in Europe: predictive models for outcome. *The Lancet*, 367(9506):225–232, 2006.

V. De Gruttola and X.M. Tu. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, pages 1003–1014, 1994.

David M Dickinson, Mary D Ellison, and Randall L Webb. Data sources and structure. *American Journal of Transplantation*, 3(s4):13–28, 2003.

Peter Diggle, P. J. Heagerty, K.-Y. Liang, and S. L. Zeger. *The Analysis of Longitudinal Data*. Oxford University Press, 2002.

Cheryl L Faucett, Nathaniel Schenker, and JMG Taylor. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics*, 58(1):37–47, 2002.

Francisca Galindo Garre, Aeilko H Zwinderman, Ronald B Geskus, and Yvo WJ Sijpkens. A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):299–308, 2008.

Pulak Ghosh, Kaushik Ghosh, and Ram C Tiwari. Joint modeling of longitudinal data and informative dropout time in the presence of multiple changepoints. *Statistics in Medicine*, 30(6):611–626, 2011.

R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Hélène Jacqmin-Gadda, Daniel Commenges, and Jean-François Dartigues. Random change-point model for joint modeling of cognitive decline and dementia. *Biometrics*, 62(1):254–260, 2006.

P.S. Kamath, R.H. Wiesner, M. Malinchoc, W. Kremers, T.M. Therneau, C.L. Kosberg, G. D'Amico, E.R. Dickson, and W. Kim. A model to predict survival in patients with end-stage liver disease. *Hepatology*, 33(2):464–470, 2001.

Rebecca Killick and Idris Eckley. changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19, 2014.

John P. Klein, Hans C. van Houwelingen, and Joseph G. Ibrahim. *Handbook of Survival Analysis*. Chapman and Hall/CRC, 2013.

H. Lin, B.W. Turnbull, C.E. McCulloch, and E.H. Slate. Latent class models for joint analysis of longitudinal biomarker and event process data. *Journal of the American Statistical Association*, 97(457):53–65, 2002.

L. Liu, X. Huang, and J. O'Quigley. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, 64(3):950–958, 2007.

David J Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10 (4):325–337, 2000.

Robert M Merion, Douglas E Schaubel, Dawn M Dykstra, Richard B Freeman, Friedrich K Port, and Robert A Wolfe. The survival benefit of liver transplantation. *American Journal of Transplantation*, 5(2):307–313, 2005.

G. Molenberghs and G. Verbeke. *Models for discrete longitudinal data*. Springer, 2005.

Donna K Pauler and Dianne M Finkelstein. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine*, 21(24):3897–3911, 2002.

Martyn Plummer et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, 124:125–132, 2003.

C. Proust-Lima, P. Joly, J.F. Dartigues, and H. Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4):1142–1154, 2009.

C Proust-Lima, M Sène, JMG Taylor, and H Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research*, 23 (1):74–90, 2014.

Abbas Rana, Tun Jie, Marian Porubsky, Shahid Habib, Horacio Rilo, Bruce Kaplan, Angelika Gruessner, and Rainer Gruessner. The survival outcomes following liver transplantation (SOFT) score: validation with contemporaneous data and stratification of high-risk cohorts. *Clinical Transplantation*, 27(4):627–632, 2013.

D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

D. Rizopoulos. *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R*, volume 6. CRC Press, 2012a.

D. Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(3):491–501, 2012b.

S. Self and Y. Pawitan. Modeling a marker of disease progression and onset of disease. *AIDS Epidemiology: Methodological Issues*, pages 231–255, 1992.

X. Song, M. Davidian, and A.A. Tsiatis. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 3(4):511–528, 2002.

David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

Jean de Dieu Tapsoba, Shen-Ming Lee, and CY Wang. Joint modeling of survival time and longitudinal data with subject-specific changepoints in the covariates. *Statistics in Medicine*, 30(3):232–249, 2011.

Jeremy MG Taylor, Yongseok Park, Donna P Ankerst, Cecile Proust-Lima, Scott Williams, Larry Kestin, Kyoungwha Bae, Tom Pickles, and Howard Sandler. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213, 2013.

R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing; 2012. *Open access available at: http://cran. r-project. org*, (version 3.0.2), 2014.

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

G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, 2009.

Chapter 3

Lost in transplantation: sources of bias in the analysis of transplant registry data

3.1 Introduction

In 2002, a randomized clinical trial examining use of hormone therapy in menopausal women was terminated early because it failed to confirm findings of previous studies (using observational data) showing a protective effect of hormone therapy (Prentice et al. [2005]). The clinical trial, surprisingly, showed the opposite: an increased risk of coronary heart disease for those taking hormone therapy. Hernán et al. [2008] suggest that the source of inconsistent findings stemmed in part from the assumption that there were no unmeasured confounders related to treatment discontinuation. Methodological issues like this make assessment of statistical results challenging, and comparison of results across studies frustrating. In transplant studies analyzing observational data for efficacy of immunosuppressive regimes, there is a similar problematic situation. Added to this is the great disparity in the number of discontinuations by regime. We are concerned that time-to-drug-discontinuation is incorporated inappropriately in a typical survival analysis. We define the term time-to-drug-failure as the time when the initial immunosuppressive regime was discontinued. We propose a two outcome joint survival model, where we consider jointly time-to-drug-failure and time-to-graft-failure. Typically, duration of treatment is not used in an analysis and this approach allows us to incorporate this useful data.

The use of frailty effects in a joint modelling approach will account for some unmeasured

confounders in treatment discontinuation. Frailties are the term for random effects in the setting of survival analysis, used to address unobserved heterogeneity. In fact, Wienke [2010] suggests that the first use of a frailty was by Beard [1959] who used what he called a 'longevity factor' to account for unobserved covariates in assessment of duration of life. While we cannot measure unobserved covariates, frailties can be used to account for the effect on time-to-event of some of the unmeasured covariates and so reduce bias in the estimates (Klein and Moeschberger [2003]). In a joint modelling context, we can use frailties to account for the correlation between two time-to-event processes, such as time-to-drug-failure and time-to-death. By measuring the individual frailty in one process, and using it as a link in a second (terminal) process we can obtain a measure of the effect of the first, non-terminal, time-to-drug-failure on the terminal event, time-to-graft-failure. In doing so we account for heterogeneity caused by unmeasured covariates (Wienke [2010]).

After liver transplant, immune suppression is necessary to prevent rejection of the grafted organ. There are two main immunosuppressive agents used, tacrolimus (TAC) and cyclosporine (CYA). Since the 1990s many randomized clinical trials have compared the performance of the two anti-rejection drugs, with TAC usually performing better in preventing rejection of the transplanted liver and in reducing patient mortality, however resulting in more cases of new onset diabetes post transplant (Shrum et al. [2016]). Accompanying the data from clinical trials are transplant data registries that are maintained in many countries around the world. These registries contain data for every patient who receives a transplanted organ, with detailed follow up information until death. Statistical analyses comparing the effectiveness of immunosuppressive agents after transplant using observational data have utilized a variety of approaches, and further challenges have presented themselves since the arrival of a controversial new drug sirolimus (SRL) in the late 1990s. SRL demonstrated immunosuppressive, anti-fungal and anti-tumor activity through a mechanism of action that is different from TAC and CYA (Sehgal [2003]). Some analyses have used a standard Cox proportional hazards survival model with treatment changes included as a time-dependent covariate (Hadley et al. [1995], Watt et al.

[2012]). However, assumptions of this model require that changes in the time-dependent covariate be unrelated to the outcome in a survival analysis. In transplantation this is not the case, as one drug may be discontinued and an alternative chosen due to graft rejection or the declining health of the patient for some other reason. Naive comparisons over time across non-randomized treatment groups are also commonplace (Toso et al. [2010], Watt et al. [2012]). In some cases (Toso et al. [2010]) only subjects remaining on a single treatment without switching are compared, therefore discarding much of the information in the data and leading to a biased selection of healthy subjects. In addition, conflicting results have been reported on the effectiveness of immunosuppressive treatment with SRL after liver transplant using either randomized clinical trial (RCT) data or registry data and it remains a contentious issue (Wiesner et al. [2003], Shah et al. [2015]). Questions have been raised about effectiveness of SRL for subjects who are transplanted for either hepatitis C virus (HCV) or hepatocellular carcinoma (HCC). Using registry data, Watt et al. [2012] found that SRL, whether given at baseline or initiated later, was detrimental to patient survival in HCV positive subjects. Other studies such as Wagner et al. [2010], McKenna et al. [2011] and Teperman et al. [2013] show (using various data sources) that SRL improves survival in HCV positive subjects, limiting fibrosis occurrence (the scarring of liver tissue that leads to cirrhosis). Resolving this question is of interest from a clinical perspective since HCV (which often leads to HCC) is a leading indicator for liver transplant. In this analysis approximately 55% of those transplanted for HCC are also infected with HCV.

The generalizability and external validity offered by the 'real-world' data of transplant registries is important and valued by the transplant community. Registries contain data from patients with comorbidities that often exclude them from a clinical trial. Analysis of registry data complements the findings of RCTs and can provide insight for new research. However, conflicting results and inconsistent methodology can be a source of frustration, and can lead to dismissal of clinically relevant findings. Hernán et al. [2008] notes that when findings between RCTs and observational studies disagree, it is tempting to blame lack of randomization

in observational data. Hernán et al. [2008] have further shown that observational studies can produce conflicting results depending on the analytic approach employed. The data in the Scientific Registry for Transplant Recipients (SRTR) is a frequent source of influential research papers that have the power to change medical practice (e.g., Selck et al. [2008], who raised important questions surrounding organ allocation for re-transplantation in recipients of organs donated after cardiac death; and Locke et al. [2014] who demonstrated successful strategies for the transplantation of HIV infected kidney transplant recipients). The divergent opinions generated by analysis of data with strong overlap have also had an influence on practice, with the results of Toso et al. [2010] demonstrating anti-cancer properties in SRL resulting in increased use of SRL, followed by Watt et al. [2012] showing increased risk in HCV positive subjects resulting in subsequent decreased use of SRL. The development and application of unbiased statistical methods for analysis of this detailed and sizeable observational data set is very important.

The goal of our investigation is to develop a method for assessing effectiveness of immunosuppression after liver transplantation which can be employed for large datasets as occurs with observational data from registries. We will identify risk factors for graft failure while taking into account the primary initial immunosuppression and time-to-failure of each commonly used drug combination in a joint survival outcome model. By jointly modelling time-to-drug-failure and time-to-graft-failure, we can account for the association between time on initial drug therapy and graft survival. We also include baseline biomarker values as covariates. We hypothesize that shorter time-to-drug-failure increases the risk of graft failure since more stable patients do not change baseline treatment and thus experience improved outcomes. Time-to-drug-failure serves as a surrogate indicator for many possible health events since someone who is doing well after transplant is less likely to experience a change in immunosuppression. Using two time-to-event processes for drug failure and graft survival has not been applied in analyses of survival after liver transplant in registry data. Since the change in drug may be directly related to health status, we propose that treating it as a time-to-event process in a joint

model examining graft failure will provide less biased results than a Cox proportional hazards model with treatment as a time-dependent covariate. A detailed review of the literature on joint survival models can be found in Chapter 2.

3.2 Method

3.2.1 Data and variables

To illustrate our joint frailty model we used data from the Scientific Registry for Transplant Recipients (SRTR), for those patients age 16 and older, receiving a first cadaveric liver transplant in the United States between January 1, 2000 and December 31, 2006. We identified 23,980 subjects without missing data in the covariates under consideration, with 4,523 (19%) events (graft failure, defined as death or retransplant) in the first three years post transplant. The primary endpoint was graft survival. We focus on graft survival rather than patient survival, since graft survival is an important outcome for the limited resource of organ transplantation. Immunosuppressive therapy is recorded at baseline and discharge from hospital after transplant, then at six months, and yearly. Exact dates for treatment changes are not given, so the data are interval censored. To simplify the analysis we took the midpoint of the interval as the date of treatment change. If a subject also died during the interval where a treatment change was recorded, we took the midpoint of the start of the interval and the date of death as the day of treatment change. If subjects were listed as receiving both TAC and CYA at baseline (an impossible combination) and we could not verify baseline treatment, the subject was removed from the analysis. The SRTR collects immunosuppressive therapy for only the first five years after transplant. We chose to analyze baseline immunosuppression by combination (rather than as any exposure or not, as in Watt et al. [2012] or Toso et al. [2010]), since this accounts for any interaction or synergistic effects between SRL and the calcineurin inhibitors (CYA or TAC). This also allows us to see the effects of SRL uncontaminated by other drugs.

Variables considered include induction immunosuppression (given only at baseline) includ-

ing anti-CD25 antibody and thymoglobulin, adjuvant immunosuppression such as steroids, mycophenolate mofetil (MMF) and azathioprine, again only whether or not they were given as initial therapy.

Baseline covariates included were hepatocellular carcinoma (HCC) status, immunosuppressive drug regime at baseline (maintenance, induction and adjuvant), recipient age, donor age, race (African American or other), diabetic status (insulin dependent Y/N), recipient gender, donor gender, sex mismatch (Y/N), whether a split liver was received, whether the recipient was in fulminant hepatic failure at the time of transplant, whether a non heart-beating donor was used, recipient blood type, previous malignancy (prior to transplant), renal insufficiency (defined as dialysis within the week prior to transplant (Y/N)), creatinine on day 0, bilirubin on day 0, INR on day 0, and donor body mass index (BMI) .

3.2.2 Statistical methods

Here we present a joint survival outcome model with two time-to-event processes each of which we model as Weibull as given by equation 2.2, but allowing different parameters and covariates in each submodel. If a subject experiences a graft failure event, they are censored in the drug failure model. For the ith individual, let t_{ki} represent lifetime, $k = 1, 2$, with t_{1i} representing time-to-drug-failure with shape parameter α_1, and t_{2i} representing time-to-graft-failure with shape parameter α_2. The scale parameter λ_{ki} can vary across individuals and event types. We re-parameterize λ_{ki} as

$$\log(\lambda_{ki}) = z_{ki}^{'}\boldsymbol{\gamma}_k + \phi_k c_i \quad k = 1, 2 \tag{3.1}$$

where z_{ki} and $\boldsymbol{\gamma}_k$ are the covariates and corresponding regression coefficients and where $\phi_1 = 1$ and $\phi_2 = \phi$. The c_i are subject-specific frailty terms which link the two responses. In the graft failure component, ϕ measures the association between c_i and time to graft failure.

Estimation was carried out with a Bayesian approach and a Markov chain Monte Carlo

(MCMC) algorithm in R (R Core Development Team [2010]) and JAGS (Plummer et al. [2003]) to obtain estimates of the posterior distributions, and with the priors specified in JAGS as non-informative where the components of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are $\sim N(0, 10000)$, $\alpha_k \sim \text{Unif}(0, 1)$ and $c_i \sim N(0, 1/\sigma_c)$, with $\sigma_c \sim \text{Unif}(0, 10000)$. We ran 3 chains for 250,000 iterations with 220,000 burn-in, which took approximately 40 hours. Convergence was judged by the Brooks-Gelman-Rubin (BGR) convergence diagnostic Brooks and Gelman [1998]. We used the deviance information criterion (DIC) for model comparison. Covariates were removed from the model if the 95% credible interval contained zero. The final model for each cohort was chosen after considering DIC as well as Cox Snell residuals for the graft survival component.

We compared the results of our joint model to traditional multivariate Cox proportional hazards models for survival after liver transplant. We used a stepwise approach and included the same covariates tested in the joint model. We also compared our results from the joint survival model to a Cox model using propensity scores to account for use of SRL. Details of this analysis are provided later. Lastly, we compare our results to a Cox proportional hazards model where only those who remained on the same immunosuppression protocol for at least six months after transplant were analyzed, as done in Toso et al. [2010], for example.

The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government. The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the U.S., submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. A detailed description of the database is available in Dickinson et al. [2003].

3.3 Results

3.3.1 Joint frailty model

Because of observed interactions between HCV and other covariates, we split the data into HCV positive and negative cohorts. We know that HCV positive subjects fare more poorly than the HCV negative cohort, due to increasing fibrosis in the transplanted organ due to recurrent HCV. Recurrence of HCV is universal after transplant however the speed at which it progresses depends on different factors such as viral load, donor age and other risks that are not completely understood, although there is some evidence that this may include immunosuppressive regime (McKenna et al. [2011], Howell et al. [2014]). There were 10,417 subjects (2,357 (22.6%) graft failure events) who were HCV positive and 13,563 subjects (2,166 (16.0%) graft failure events) who were HCV negative. Of these 23,980 total subjects identified for our analysis, 5,542 (23%) changed their initial therapy within the first three years post-transplant. Table 3.1 shows the number of subjects on each drug combination and how many changed therapies by year 3. Table 3.2 shows the number of subjects on each immunosuppressive regime by HCV status. These tables shows the great disparity in treatment switching among the various combinations. A subject started on TAC alone is much less likely to discontinue baseline immunosuppression. A subject started on the SRL + TAC combination is approximately 4 times more likely to discontinue this combination than someone started on TAC alone. Treating drug failure as a time-to-event process is important to acknowledge the information provided by this event and to account for it in the analysis.

Figure 3.1 shows Kaplan-Meier curves for graft survival by initial drug therapy, separated by HCV cohort. We can see that the HCV negative cohort enjoys better 3 year survival overall. We show the same data by treatment, with 95% confidence intervals, in Figure 3.2. The SRL as well as SRL in any drug combination have wide confidence intervals reflecting their sample sizes. Figure 3.3 shows Kaplan-Meier curves for time to drug failure for each therapy. It is interesting to compare patterns of drug changes by cohort. Here we see that SRL alone or in

Table 3.1: Summary of the number of individuals (n) on initial drug therapies and the number and percent that changed (n changed (%)) from initial therapy within three years post-transplant.

Initial treatment	n	n changed (%)
TAC	20,543	3,756 (18.3%)
SRL + TAC	882	676 (76.6%)
SRL	365	166 (45.5%)
CYA	1977	783 (39.6%)
SRL + CYA	213	161(75.6%)

Table 3.2: Summary of initial drug therapies by HCV status.

HCV positive cohort	
Initial treatment	n (%)
TAC	8964 (86.1%)
SRL + TAC	363 (3.5%)
SRL	160 (1.5%)
CYA	841 (8.0%)
SRL + CYA	89 (0.9%)
HCV negative cohort	
Initial treatment	n (%)
TAC	11579 (85.4%)
SRL + TAC	519 (3.8%)
SRL	205 (1.5%)
CYA	1136 (8.4%)
SRL + CYA	124 (0.9%)

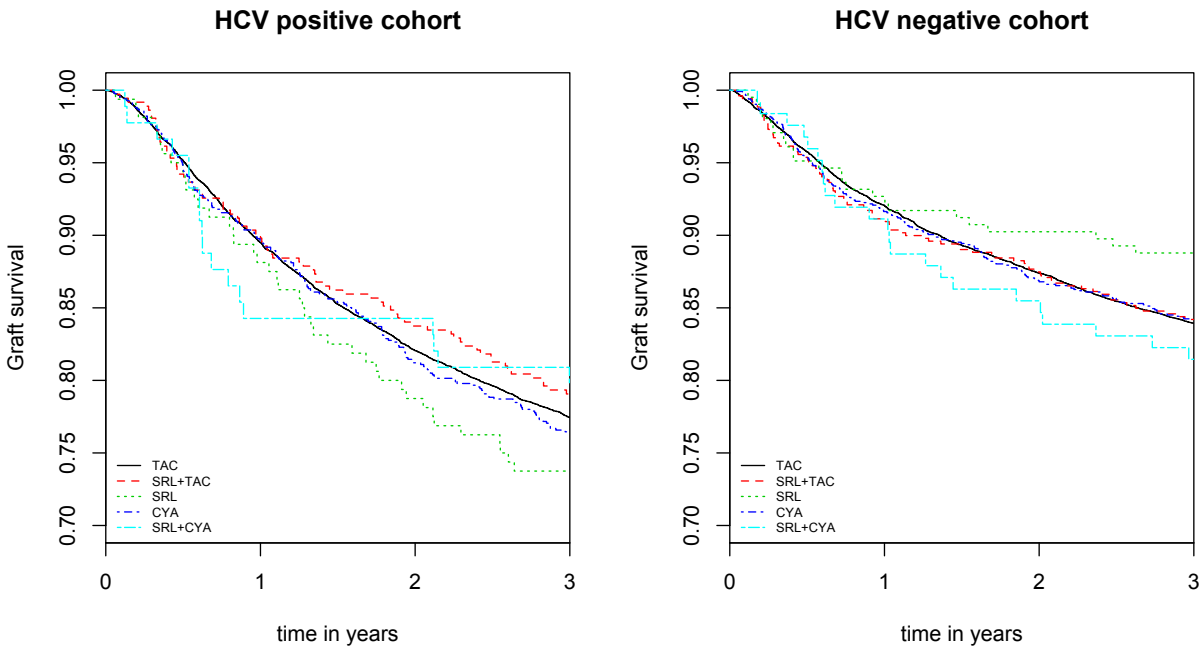

Figure 3.1: Time-to-graft-failure, by initial immunosuppression.

combination (SRL + TAC, SRL + CYA) has the shortest time to drug failure. Of the 5,542 who terminated initial therapy within the 3 year study period, 2190 (40%) did so in the first year post-transplant.

Results from the best fitting joint model for the HCV positive cohort are shown in Table 3.3. Most of the covariates listed in Table 3.3 have a significant effect on graft survival or time to drug failure. We include non-significant information for certain important variables to show their credible intervals. A Bayesian 95% credible interval expresses the uncertainty of our estimate and gives us the most plausible interval for our estimate, with a probability of 95% (Lesaffre and Lawson [2012]).

Baseline treatment was very significant in the time-to-drug-failure submodel. We see that being on any initial treatment other than tacrolimus significantly shortens the time to change of initial immunosuppressive therapy. The greatly increased risk of drug failure from initial immunosuppressive regime fits with the degree of treatment switching seen. Use of anti-CD25 an-

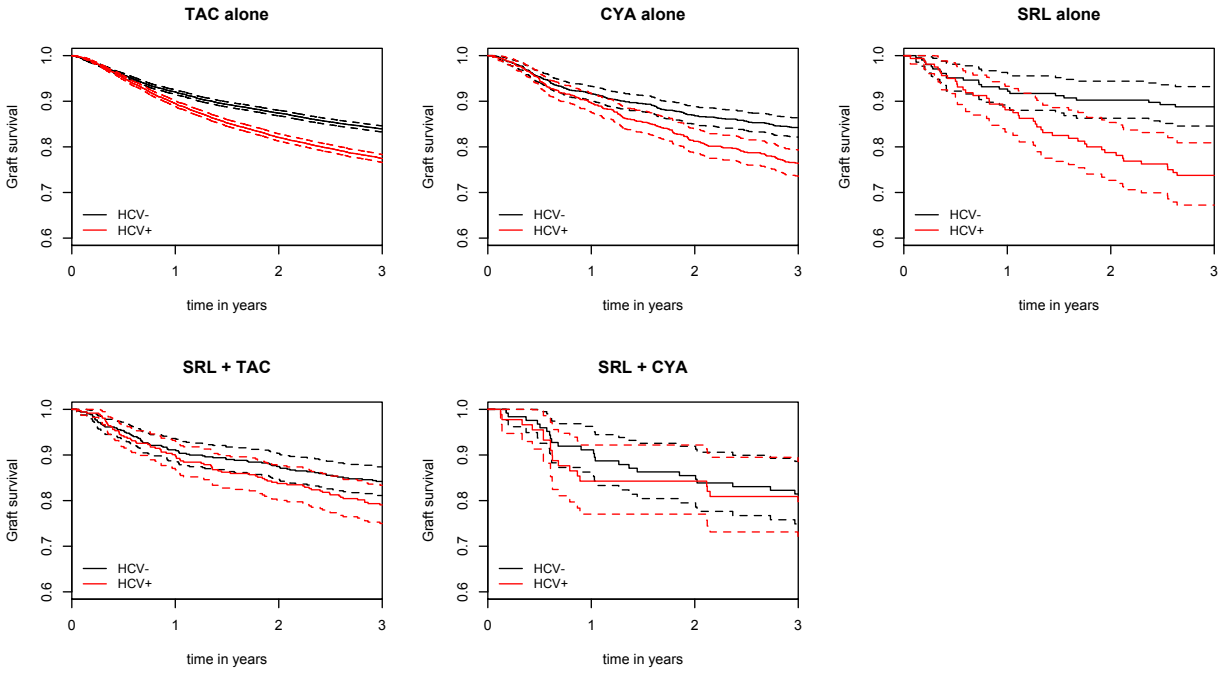

Figure 3.2: Time-to-graft-failure, by initial immunosuppression, showing confidence intervals for each cohort.

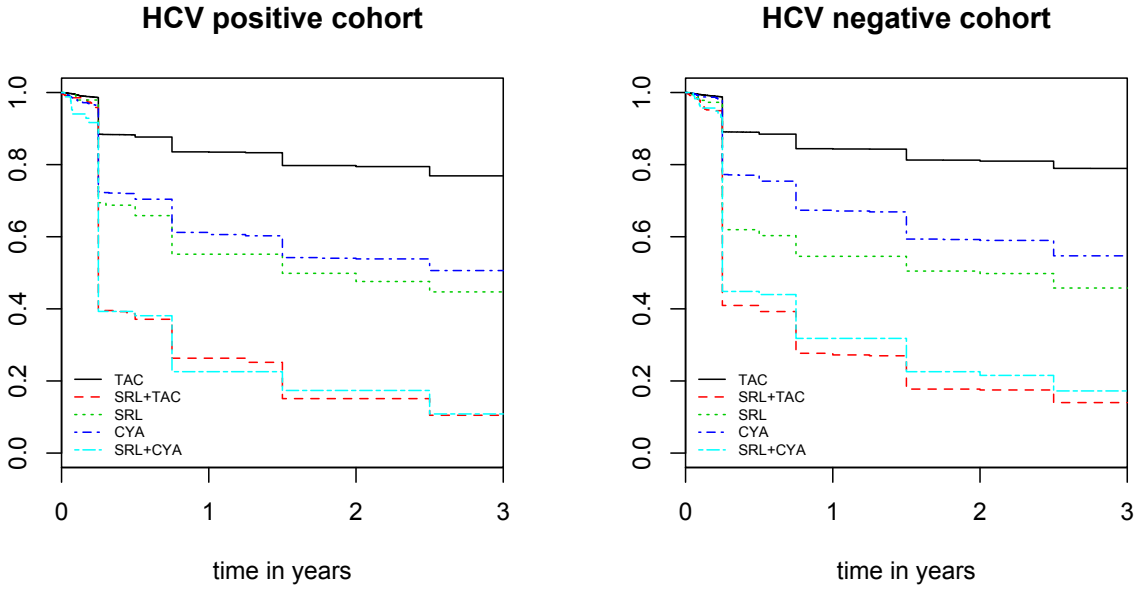

Figure 3.3: Time-to-drug-failure, by initial immunosuppression.

Table 3.3: Posterior means for the log hazard, hazard, std err and quantiles from the HCV positive cohort joint model. (ns=not significant; AA = African American; F = female)

Parameter	**Mean**	**Hazard**	**Std err**	**CI (lower)**	**CI (upper)**
Time-to-drug-failure submodel:					
intercept	-9.445		0.092	-9.623	-9.268
log(bilirubin) day 0	-0.208		0.067	-0.346	-0.080
log(INR) day 0	0.467		0.194	0.118	0.861
anti-CD25 antibody	-0.345		0.145	-0.631	-0.062
baseline trt: TAC alone (ref)	—		—	—	—
SRL + TAC	7.066		0.291	6.487	7.646
SRL	3.167		0.366	2.448	3.878
CYA	2.711		0.175	2.372	3.062
CYA + SRL	7.033		0.532	5.988	8.086
σ_c	12.658			11.905	13.333
Time-to-graft-failure submodel:					
α (shape parameter)	0.923		0.019	0.890	0.923
intercept	-8.717		0.180	-9.084	-8.393
HCC	0.159	1.172	0.057	0.045	0.267
recipient age (decades)	0.0005	1.0005	0.003	-0.004	0.005 (ns)
donor age (decades)	0.018	1.018	0.001	0.016	0.021
race (AA)	0.416	1.516	0.064	0.289	0.540
gender (F)	0.208	1.231	0.048	0.109	0.300
diabetic	0.220	1.246	0.087	0.051	0.388
anti-CD25 antibody	-0.145	0.865	0.065	-0.272	-0.016
log(creatinine) day 0	0.160	1.174	0.044	0.077	0.246
non heart-beating donor	0.392	1.480	0.104	0.188	0.589
previous malignancy	0.186	1.204	0.074	0.040	0.334
renal insufficiency	0.261	1.298	0.118	0.023	0.480
baseline trt: TAC (ref)	–		–	–	–
SRL + TAC	-0.122		0.116	-0.359	0.110 (ns)
SRL	0.186		0.157	-0.127	0.478 (ns)
CYA	0.022		0.073	-0.121	0.171 (ns)
CYA + SRL	-0.141		0.249	-0.656	0.310 (ns)
ϕ	0.065	1.067	0.008	0.050	0.080

tibody induction therapy significantly lengthens time to initial drug failure (hazard ratio 0.70). High INR and high bilirubin at time of transplant shortens time-to-drug-failure.

In the time-to-graft-failure component, we found that none of the immunosuppressive treatments are significantly different from the reference drug TAC in terms of impact on graft survival, after taking into account covariates and the time-to-drug-failure frailty effect. The model is structured so that drug effects are tested in comparison to tacrolimus, the most commonly used treatment. Risk of graft failure is increased for female subjects, those who are diabetic at transplant, those who are of African American race compared to all other races, for those who had any previous malignancy, for those with renal insufficiency, for those who receive an organ from a non heart-beating donor, for higher day 0 creatinine, for those with HCC, and for increased donor age. The most dominant effects are race, whether a non-heart-beating donor was used, and renal insufficiency. Recipient age was kept in the graft failure model because it improved DIC.

The individual frailty c_i from the time-to-drug-failure component has a significant link (ϕ) between the two time-to-event submodels in the HCV positive cohort, where it increases survival time for those with negative log frailties, and decreases survival time for those with positive log frailties. Negative log frailties are associated with lower risk of a drug failure event. Therefore, a change in baseline therapy is associated with shorter graft survival. This is medically sensible since a change in treatment is often precipitated by an adverse event such as rejection. Histograms of individual log frailties grouped by whether a drug failure occurred are shown in Figure 3.4, and plots grouped by initial treatment are shown in Figure 3.5. The value of the frailty ranges from approximately -7 to +9, so the estimated coefficient for the individual frailty in the survival model, while small at 0.065 (hazard: $\exp(0.065 \times c_i)$), can have a large effect depending on the value of the subject specific frailty, with a hazard ratio ranging from a beneficial 0.63, to an increased hazard of 1.8. The variance of the subject-specific frailty is 12.7, which shows there is a great deal of unobserved heterogeneity between subjects. To put the results into context, Table 3.4 shows the estimated log frailties for each

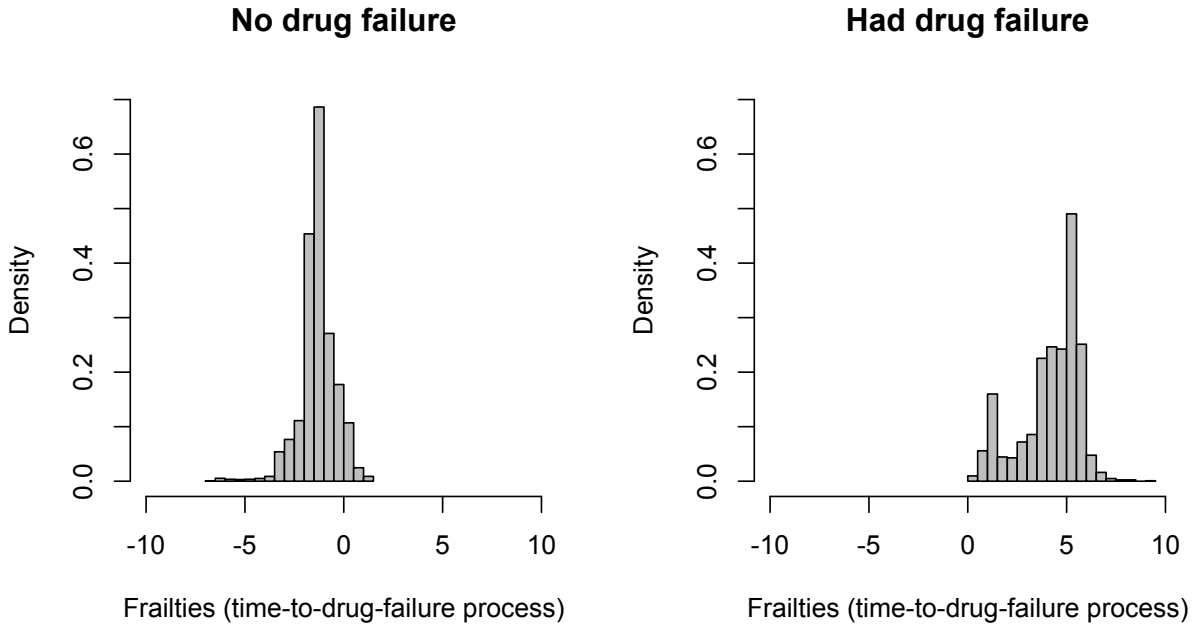

Figure 3.4: HCV positive cohort: Histograms of individual log frailties from time-to-drug-failure model, by whether initial treatment was changed. Positive log frailties are associated with increased risk of graft failure.

treatment combination, comparing the mean, minimum and maximum frailties of subjects who did not change therapy to those who did. For the group on baseline therapy of TAC alone who did change drug, the mean frailty shows the highest risk of graft failure (mean hazard ratio: 1.4), suggesting that subjects who do not do well on the 'gold standard' TAC are at greatest risk of graft failure.

Results from the best fitted joint model for the HCV negative cohort are shown in Table 3.5. As in the HCV positive cohort, baseline treatment was very significant in the time-to-drug-failure submodel. We see that being on any initial treatment other than TAC significantly shortens the time to failure of initial immunosuppressive therapy. Interestingly, use of thymoglobulin induction therapy significantly shortens time to inital drug failure (hazard 1.5). Anti-CD25 antibody, INR and bilirubin at time of transplant were not significant in the HCV negative cohort for time-to-drug-failure.

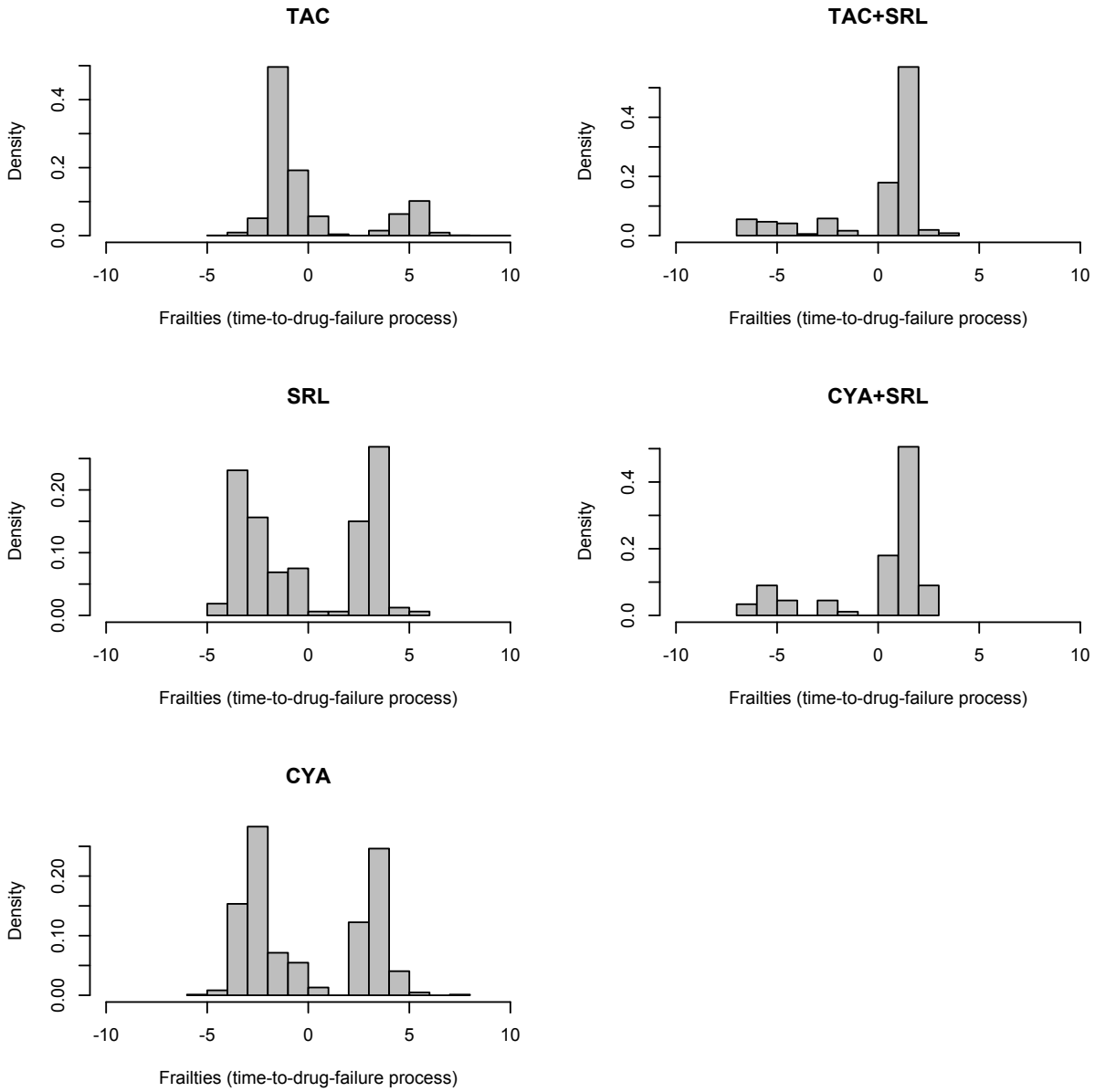

Figure 3.5: HCV positive cohort: Histograms of individual log frailties from time-to-drug-failure model, by initial immunosuppression. Positive log frailties are associated with increased risk of graft failure.

Table 3.4: Summary of frailty effects on time-to-graft-failure by initial drug therapy for the HCV positive cohort.

Initial trt	No drug failure: Mean (min, max)	Hazard (min, max)	Had drug failure: Mean (min, max)	Hazard (min, max)
TAC	-1.3 (-4.6, 1.4)	(0.7, 1.1)	5.0 (3.2, 9.2)	(1.2, 1.8)
SRL + TAC	-4.4 (-6.9, -1.4)	(0.6, 0.9)	1.3 (0.2, 3.9)	(1.0, 1.3)
SRL	-2.5 (-4.3, 0.1)	(0.8, 1.0)	3.1 (2.0, 5.2)	(1.1, 1.4)
CYA	-2.4 (-5.1, 0.5)	(0.7, 1.0)	3.4 (2.1, 7.6)	(1.1, 1.6)
SRL + CYA	-4.6 (-6.2, -1.9)	(0.7, 0.9)	1.4 (0.5, 2.9)	(1.0, 1.2)

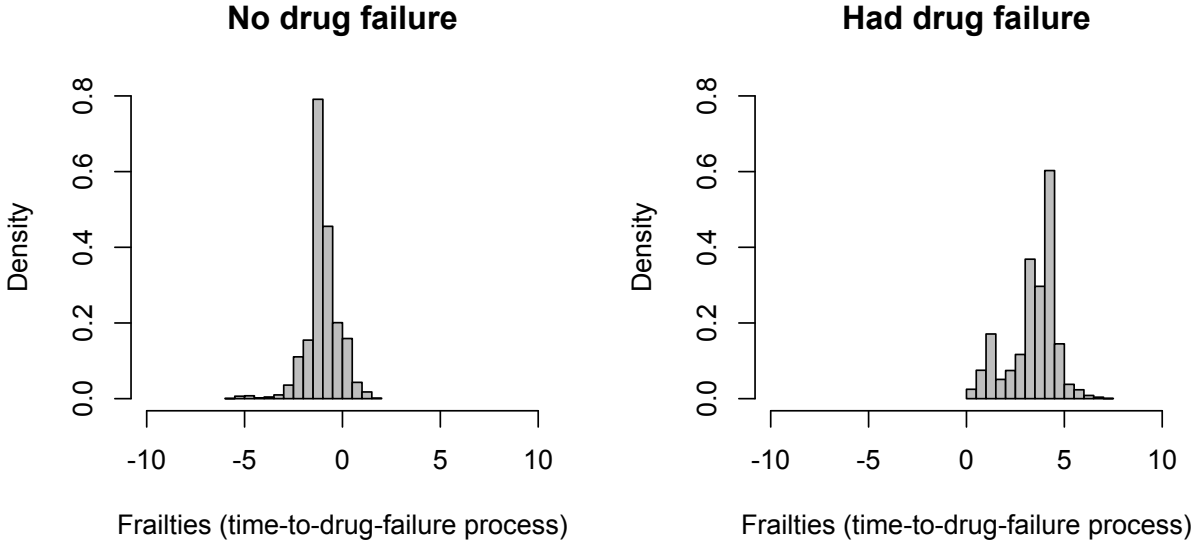

Figure 3.6: HCV negative cohort: Histograms of individual log frailties from time-to-drug-failure model, by whether initial treatment was changed. Positive log frailties are associated with increased risk of graft failure.

Table 3.5: Posterior means for the log hazard, hazard, std err and quantiles from the HCV negative cohort joint model. (ns=not significant; AA = African American; F = female)

Parameter	**Mean**	**Hazard**	**Std err**	**CI (lower)**	**CI (upper)**
Time-to-drug-failure submodel:					
intercept	-9.716		0.053	-9.820	-9.614
thymoglobulin	0.384		0.168	0.071	0.714
baseline trt: TAC alone (ref)	—		—	—	—
SRL + TAC	6.291		0.229	6.491	7.371
SRL	3.480		0.320	2.855	4.111
CYA	2.415		0.141	2.138	2.688
CYA + SRL	6.510		0.434	2.138	2.688
σ_c	12.346			11.765	12.987
Time-to-graft-failure submodel:					
α (shape parameter)	0.817		0.015	0.784	0.847
intercept	-8.336		0.146	-8.624	-8.064
HCC	0.394	1.483	0.064	0.264	0.519
recipient age (decades)	0.008	1.008	0.002	0.005	0.012
donor age (decades)	0.010	1.010	0.001	0.007	0.012
race (AA)	0.394	1.483	0.074	0.249	0.535
gender (F)	-0.246	0.782	0.048	-0.341	-0.151
diabetes	0.241	1.273	0.087	0.072	0.409
non heart-beating donor	0.324	1.383	0.047	0.232	0.415
previous malignancy	0.236	1.266	0.089	0.059	0.410
renal insufficiency	0.454	1.575	0.102	0.248	0.650
baseline trt: TAC	–		–	–	–
SRL + TAC	0.041		0.112	-0.184	0.257 (ns)
SRL	-0.443	0.642	0.219	-0.910	-0.045
CYA	0.011		0.083	-0.154	0.168 (ns)
CYA + SRL	0.185		0.214	-0.248	0.584 (ns)
ϕ	0.085	1.089	0.008	0.071	0.101

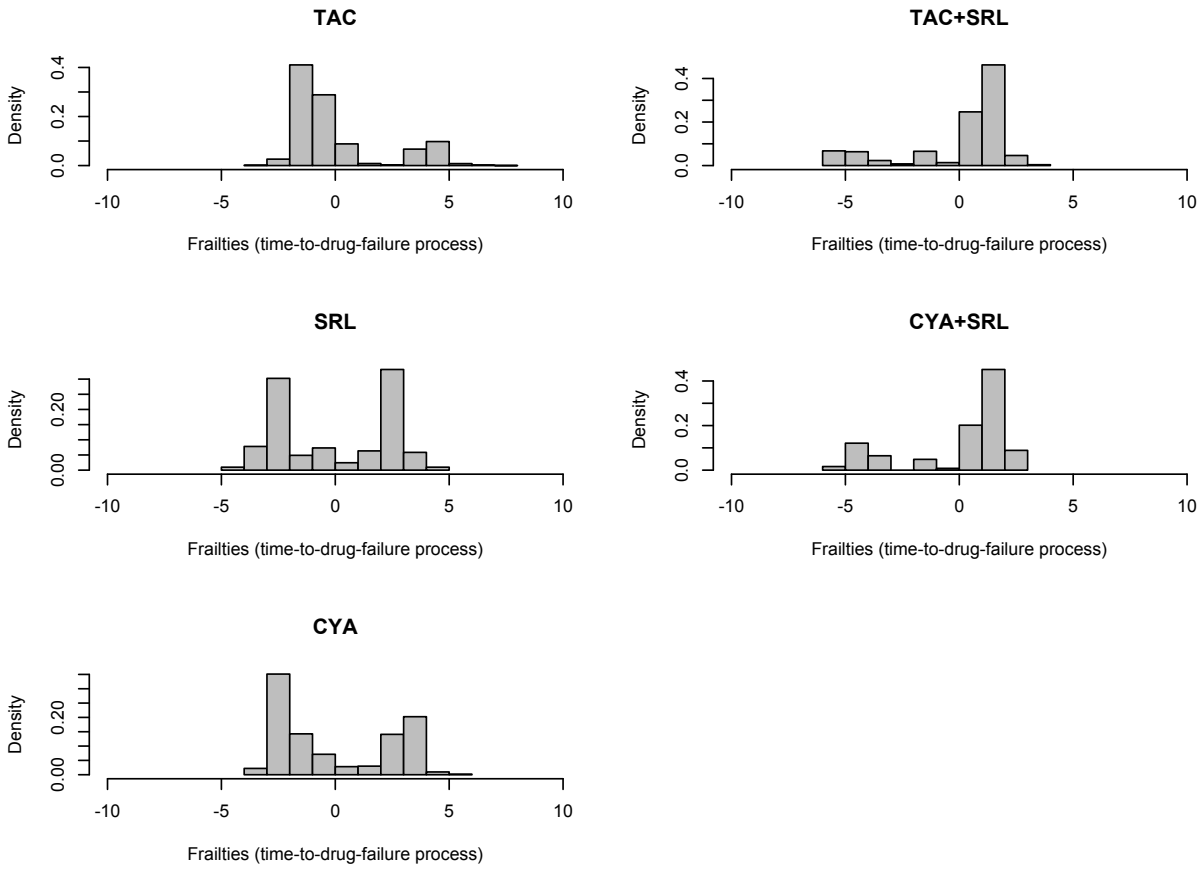

Figure 3.7: HCV negative cohort: Histograms of individual log frailties from time-to-drug-failure model, by initial immunosuppression. Positive log frailties are associated with increased risk of graft failure.

In the time-to-graft-failure component, we found that being on SRL was significantly better (hazard: 0.642) than the reference drug TAC in terms of impact on graft survival, after taking into account covariates and the time-to-drug-failure frailty effect (posterior mean: -0.443; 95% credible interval: -0.910,-0.045). It is a beneficial dominant effect even when taking into account frailty terms and all other variables in the model. For example, if we randomly select an HCV negative cohort subject on SRL as initial therapy who did not change drug, we find they have a frailty $(c_i) = -3.3$ and the total effect of SRL on graft failure has a hazard of $\exp(-0.443 + 0.085 \times c_i) = 0.49$. If we do the same for a subject started on SRL who did have a drug failure event, we find they have $c_i = 3.2$ and the effect of SRL on graft failure has a hazard of 0.84.

Similar to the HCV positive cohort, risk of graft failure is increased for those who are diabetic at transplant, those who are of African American race compared to all other races, for those who had any previous malignancy, for those with renal insufficiency, for those who receive an organ from a non heart-beating donor, for those with HCC, and for increased donor age. Recipient age also increased risk of graft failure in this cohort. Being female decreased risk of graft failure (hazard ratio: 0.782), an opposite effect to that seen in the HCV positive cohort (hazard ratio: 1.231). We found that anti-CD25 antibody was not significantly protective against risk of graft failure in this cohort. Creatinine on day of transplant also not significant. The most dominant effects in the HCV negative cohort were renal insufficiency, HCC and race.

The individual frailty c_i from the time-to-drug-failure component also has a significant link (ϕ) between the two time-to-event submodels in the HCV negative cohort, where negative log frailties are associated with lower risk of a drug failure event. The posterior mean was 0.085, larger than in the HCV positive cohort. Histograms of individual log frailties for the HCV negative cohort, grouped by whether a drug failure occurred are shown in Figure 3.6, and plots grouped by initial treatment are shown in Figure 3.7. The value of the frailty ranges from approximately -6.5 to +8.5, so the estimated coefficient for the individual frailty in the survival model, while small at 0.085 (hazard: $\exp(0.085 \times c_i)$), can have a large effect depending on

Table 3.6: Summary of frailty effects on graft survival by initial drug therapy for the HCV negative cohort.

Initial trt	No drug failure: Mean (min, max)	Hazard (min, max)	Had drug failure: Mean (min, max)	Hazard (min, max)
TAC	-1.1 (-5.1, 1.6)	(0.7, 1.1)	5.1 (3.7, 8.5)	(1.4, 2.1)
SRL + TAC	-4.3 (-6.4, -1.6)	(0.6, 0.9)	1.4 (0.4, 3.9)	(1.0, 1.4)
SRL	-2.7 (-4.8, 0.6)	(0.7, 1.1)	3.1 (1.9, 4.8)	(1.2, 1.5)
CYA	-2.2 (-4.8, 0.8)	(0.7, 1.1)	3.6 (2.3, 6.5)	(1.2, 1.7)
SRL + CYA	-4.4 (-6.0, -1.3)	(0.6, 0.9)	1.5 (0.5, 2.9)	(1.0, 1.3)

the value of the subject specific frailty, with a hazard ratio ranging from a beneficial 0.58, to an increased hazard of 2.06. The variance of the subject-specific frailty is 12.3, which shows there is a great deal of unobserved heterogeneity between subjects (similar to the HCV positive cohort). Table 3.6 shows these effects for each treatment combination, comparing the mean, minimum and maximum frailties of subjects who did not change therapy to those who did. The situation is similar to the HCV positive cohort, with a slightly higher hazard seen here in the negative cohort for the group on baseline therapy of TAC alone who have the highest risk of graft failure if they also experience drug failure (mean hazard ratio: 1.5).

There are some differences between the two cohorts. In the HCV positive cohort, anti-CD25 antibody appears to extend the time-to-drug-failure and also has a beneficial effect on graft survival, however neither of these effects are seen in the HCV negative cohort. Use of thymoglobulin appears to speed up time-to-drug-failure in the HCV negative cohort whereas it is not significant for HCV positive subjects. HCV negative subjects also see a beneficial effect on graft survival from the use of SRL that is not seen in the HCV positive cohort. The risk to graft failure from HCC and renal insufficiency is greater in the HCV negative cohort.

3.3.2 Standard Cox model

We compared our joint frailty method to one using a standard Cox proportional hazard model with no time-to-drug-failure component. Table 3.7 shows the results from the standard Cox

Table 3.7: Results from a standard Cox proportional hazards model for the HCV positive cohort.

Parameter	Coefficient	Hazard	StdError	p
HCC	0.156	1.169	0.057	0.006
donor age (decades)	0.019	1.019	0.001	< 0.001
log(creatinine) day 0	0.160	1.174	0.044	< 0.001
race (AA)	0.390	1.476	0.062	< 0.001
gender (F)	0.200	1.221	0.048	< 0.001
diabetes	0.219	1.245	0.086	0.011
non heart-beating donor	0.381	1.464	0.106	< 0.001
previous malignancy	0.199	1.220	0.075	0.008
renal insufficiency	0.250	1.284	0.116	0.031
anti-CD25 antibody	-0.136	0.873	0.065	0.036
sex mismatch	0.101	1.106	0.047	0.032
baseline trt: TAC (ref)	–	–	–	–
SRL + TAC	-0.107	0.899	0.118	0.364 (ns)
SRL	0.183	1.201	0.156	0.241 (ns)
CYA	0.031	1.032	0.074	0.676 (ns)
CYA + SRL	0.031	0.897	0.237	0.646 (ns)

model for the HCV positive cohort. None of the immunosuppressive treatments are significantly different from TAC, which matches the results we obtained in the joint model. The standard Cox model identified sex mismatch as significantly increased hazard of graft failure, which was not seen in the joint model.

The results from the HCV negative cohort for the standard Cox model without a joint component are shown in Table 3.8. Again, they are similar to the results from the joint model in terms of covariate effects. We see a beneficial effect of SRL alone on graft survival in the HCV negative cohort, which matches the results from the joint model. The hazard is similar (0.642 in the joint model vs 0.660 in the standard Cox model). Other research papers that treat 'any SRL' as one category would see this significant effect washed out by the contamination of the calcineurin inhibitors CYA and TAC. Receiving MMF at baseline, being in fulminant hepatic failure at time of transplant, and high creatinine at baseline (day 0) were significant predictors of graft failure in the standard Cox model but not in the joint model.

Table 3.8: Results from a standard Cox proportional hazards model for the HCV negative cohort.

Parameter	Coefficient	Hazard	StdError	p
donor age (decades)	0.010	1.010	0.001	< 0.001
log(creatinine) day 0	0.091	1.095	0.047	0.038
gender (F)	-0.217	0.805	0.047	< 0.001
recipient age (decades)	0.007	1.007	0.002	< 0.001
HCC	0.394	1.483	0.065	< 0.001
non heart-beating donor	0.655	1.925	0.094	< 0.001
MMF at baseline	-0.102	0.903	0.044	0.020
race (AA)	0.406	1.501	0.075	< 0.001
diabetic	0.237	1.267	0.086	0.006
fulminant hepatic failure	-0.230	0.793	0.114	0.043
previous malignancy	0.230	1.259	0.089	0.009
renal insufficiency	0.386	1.472	0.107	< 0.001
baseline trt: TAC (ref)	–		–	–
SRL + TAC	0.045	1.047	0.113	0.688 (ns)
SRL	-0.416	0.660	0.210	0.048
CYA	0.016	1.016	0.079	0.838 (ns)
CYA + SRL	0.193	1.213	0.210	0.357 (ns)

For both cohorts, we were expecting to see more differences between the standard Cox model and the joint model, however the results are very similar. The Cox model gives almost the same estimates and the standard errors are comparable.

3.3.3 Cox model with immortal time bias

We compared the results of our joint model to what the stable therapy method used by Toso et al. [2010]. The method involves selection of only subjects who remain on the same immunosuppressive regime over a specified time period. This method can be criticized for including a more healthy subset of patients resulting in a so-called immortal time bias (Di Martino et al. [2015]), since subjects must survive for a 6 month time period without a drug failure event in order to be included in the analysis. Toso et al. [2010] examined subsets of HCC positive subjects versus HCC negative, but for the sake of comparison to our joint model, we will repeat the analysis with our HCV cohorts. These groups have much overlap since those who have

Table 3.9: Results from a Cox proportional hazards model for the HCV positive cohort showing immortal time bias.

Parameter	Coefficient	Hazard	StdError	p
donor age (decades)	0.018	1.018	0.001	< 0.001
log(creatinine) day 0	0.186	1.204	0.048	< 0.001
gender (F)	0.181	1.199	0.054	< 0.001
HCC	0.164	1.178	0.063	0.009
non heart-beating donor	0.397	1.487	0.117	< 0.001
thymoglobulin	0.171	1.186	0.085	0.044
steroids at baseline	0.192	1.212	0.091	0.035
race (AA)	0.447	1.563	0.068	< 0.001
previous malignancy	0.165	1.179	0.085	0.054
sex mismatch	0.128	1.137	0.047	0.06
baseline trt: TAC (ref)	–		–	–
SRL + TAC	-0.343	0.709	0.198	0.084 (ns)
SRL	0.131	1.140	0.191	0.493 (ns)
CYA	-0.007	0.993	0.090	0.938 (ns)
CYA + SRL	-0.501	0.606	0.448	0.263(ns)

HCV often go on to develop HCC before transplant (Simonetti et al. [1992]). The results from the stable Cox proportional hazards model for the HCV positive cohort are shown in Table 3.9. We see that the results do not differ dramatically from either the joint model or the standard Cox model, with some variables such as diabetic status and anti-CD25 antibody losing significance, and others such as steroid and thymoglobulin use at baseline becoming significant predictors. In Table 3.10, we do see the effect of the immortal time bias in the HCV negative cohort, where the benefit of being on SRL alone is exaggerated (hazard 0.392) and the p value is much smaller (0.006).

3.3.4 Other sources of bias

We were initially unable to confirm results found in Watt et al. [2012] who found that SRL presented increased risk in the HCV positive cohort. In our investigations we discovered three more statistical issues that cannot be ignored when analyzing transplant data: contamination of treatment effects, failure to set one treatment as the reference level (control), and identifying

Table 3.10: Results from a Cox proportional hazards model for the HCV negative cohort showing immortal time bias.

Parameter	Coefficient	Hazard	StdError	p
donor age (decades)	0.011	1.011	0.001	< 0.001
log(creatinine) day 0	0.112	1.118	0.048	0.020
diabetic	0.304	1.356	0.096	0.002
gender (F)	-0.227	0.797	0.053	< 0.001
age in decades	0.008	1.008	0.002	< 0.001
HCC	0.414	1.512	0.072	< 0.001
non heart-beating donor	0.648	1.912	0.107	< 0.001
MMF at baseline	-0.114	0.893	0.049	0.019
split liver	0.619	1.857	0.216	0.004
race (AA)	0.458	1.580	0.083	< 0.001
previous malignancy	0.211	1.235	0.100	0.035
Renal insufficiency	0.441	1.554	0.115	< 0.001
baseline trt: TAC (ref)	–		–	–
SRL + TAC	-0.159	0.853	0.188	0.399 (ns)
SRL	-0.936	0.392	0.334	0.006
CYA	-0.110	0.896	0.096	0.252 (ns)
CYA + SRL	-0.638	0.528	0.448	0.155 (ns)

living donor organs versus deceased donor organs.

In our attempt to replicate the results from Watt et al. [2012] we used a standard Cox Proportional hazards model and a propensity score to account for non-randomized drug selection. We calculated the propensity score in the same way as Watt et al. [2012] where a logistic regression model was employed to determine probability of being assigned an SRL-based treatment given known covariate information, and subsequently this probability was included as a covariate in the survival model. We also removed hepatitis B positive subjects, and those in fulminant hepatic failure at the time of transplant, and we stratified the analysis by transplant centre as the authors of the study did. Our analyses thus far have excluded subjects receiving a transplanted organ from a living donor, due to the significant benefits to survival for this subset. We were able to reproduce the results from Watt et al. [2012] only if we included recipients transplanted from both living and deceased donors in our analysis (10,936 subjects with 2,444 graft failures). However, it is important to note that Watt et al. [2012] did not include a covariate to

account for living versus deceased donor.

Without requiring a propensity score, and without stratification by centre, we can reproduce the result showing increased hazard of treatment with SRL for HCV positive subjects found in Watt et al. [2012] by including both living and cadaveric donors without a covariate to adjust for this distinction. We maintained our classification by drug combination rather than adopt the one in Watt et al. which contaminates the effect of treatment with SRL alone. Using the covariates found significant in Watt et al. [2012], we show these results below in Table 3.11. However, these results are biased since the covariate for living versus cadaveric donor has a significant effect on the model (see Table 3.12) and should be included. Table 3.12 also shows an interaction trending toward significance between donor type and some baseline treatments.

This raises the question of drug effectiveness in recipients transplanted from living donors only. We performed a subset analysis of living donors in the HCV positive cohort from the same time period (771 subjects with 144 graft failure events), and using a standard Cox proportional hazards model we do find results which merit further investigation, specifically a greatly increased hazard for those subjects receiving a transplant from a living donor, treated with SRL alone at baseline (hazard 3.19, p value 0.013) and also for those on CYA alone (hazard (1.74, p value 0.034). No other treatment effects were significant. In the HCV negative cohort for living donors only (1403 subjects, 199 graft failure events), the only significant treatment effect seen was an increased hazard for those on CYA alone (hazard 1.81, p value 0.019). However, these subset results for living donors must be interpreted with caution since there were very small numbers in the SRL and SRL combination groups. A larger scale study of living donor data is warranted here.

3.4 Discussion

When analyzing registry data, one is often presented with challenging issues. Faced with the disparity in treatment switching and with non-randomized treatment allocation, these chal-

Table 3.11: Results from a standard Cox proportional hazards model for the HCV positive cohort with both living and cadaveric donors included but not accounted for as a covariate. This analysis uses covariates found significant in Watt et al. [2012].

Parameter	**Coefficient**	**Hazard**	**StdError**	**p**
donor age (decades)	0.018	1.018	0.001	< 0.001
log(creatinine) day 0	0.200	1.222	0.041	< 0.001
diabetic	0.256	1.292	0.084	0.002
gender (F)	0.198	1.219	0.046	< 0.001
HCC	0.218	1.244	0.056	< 0.001
baseline trt: TAC (ref)	–		–	–
SRL + TAC	-0.036	0.964	0.111	0.742(ns)
SRL	0.296	1.345	0.148	0.045
CYA	0.068	1.070	0.072	0.347 (ns)
CYA + SRL	-0.027	0.973	0.231	0.905 (ns)

Table 3.12: Results from a standard Cox proportional hazards model for the HCV positive cohort with both living and cadaveric donors, accounted for as a covariate.

Parameter	**Coefficient**	**Hazard**	**StdError**	**p**
living donor (LD)	-0.204	0.815	0.101	0.043
donor age (decades)	0.018	1.018	0.001	< 0.001
log(creatinine) day 0	0.195	1.216	0.003	< 0.001
diabetic	0.250	1.284	0.084	0.003
gender (F)	0.202	1.224	0.046	< 0.001
HCC	0.216	1.241	0.056	< 0.0001
baseline trt : TAC (ref)	–		–	–
SRL + TAC	-0.069	0.934	0.117	0.560(ns)
SRL	0.234	1.263	0.156	0.134 (ns)
CYA	0.034	1.034	0.076	0.656 (ns)
CYA + SRL	-0.492	0.949	0.237	0.825 (ns)
LD x TAC (ref)	–		–	–
LD x SRL + TAC	0.388	1.475	0.351	0.269 (ns)
LD x SRL	0.829	2.290	0.484	0.087 (ns)
LD x CYA	0.476	1.610	0.255	0.063 (ns)
LD x CYA + SRL	0.492	1.636	1.032	0.633 (ns)

lenges can seem difficult to overcome. Some use propensity scores to account for non-randomization, where a logistic regression model is employed to determine probability of being assigned a certain treatment based on known covariate information, and subsequently including this probability as a covariate in the survival model. However, the use of propensity scores is contentious and may present a biased result if confounding or latent variables are not accounted for in the model (Pearl [2003], Freemantle et al. [2013]). It also does not add any measure of time-on-drug to the model. In the field of transplantation, where treatment assignment is subject to an unmeasured assessment of the physician, use of propensity scoring in observational data is ill-advised. Selection for treatment is likely to be based on unobservable factors such as the intuitive sense of the transplant physician. These facts and the additional evidence that SRL is less nephrotoxic than TAC and CYA led to its early use in circumstances where subjects were burdened with poorer prognoses from the start, complicating analysis of observational data. For these reasons, a joint outcome model such as the one we propose where latent factors may be captured by individual frailties is preferable. Analyzing treatment failures as a time-to-event process is a preferred approach in observational data analysis since it avoids discarding data, or the violation of model assumptions. Analysis of the two time-to-event outcomes of drug failure and graft failure using a joint model can account for dependence between the two processes without making strong assumptions. We allow the important information contained in the time-to-drug-failure component to influence the hazard of the time-to-graft failure component. Transplant registry data, with 100% enrolment and follow up until death, is a valuable and readily available data source that can provide insight into factors affecting health outcomes after liver transplant. The time-to-drug-failure acts as a surrogate for time to any adverse event such as infection, rejection, cancer occurrence or other, that may be unreliably or not at all collected in the registry data. This two outcome joint model describes the data structure well. We believe that use of a straightforward joint survival outcome model is appropriate for two reasons: first, clinical practice tells us that the association is strong between the drug failure and graft failure processes, and second, less complex joint modelling techniques are more likely to

be adopted in practice in this field. Joint modelling is not commonplace in analysis of registry data for transplantation, yet despite readily available and appropriate data, it has only been used in one application (Liu et al. [2004] for kidney transplant waiting list data).

Other sources of bias include use of treatment categories that do not allow for distinct treatment groups to be examined, failure to account for living vs cadaveric donor organs, and use of subjects who must survive in a stable treatment pattern in order to be included in the analysis (immortal time bias). Use of improper methods in analyzing registry data leads to anti-registry bias and also, more alarmingly, to changes in treatment patterns that may deprive patients of valuable and beneficial treatment options.

A limitation of this data set is the lack of exact dates for treatment changes, and use of the midpoint approximation is not ideal, however RCT data (with exact dates) are not readily available for public analysis. Statistical bias related to the use of the midpoint is discussed in Rücker and Messerer [1988] and Odell et al. [1992]. To add recurrent events to the drug failure model would involve the use of doubly interval censored data that are also correlated. Furthermore, data on subjects who change treatment more than once during a data collection interval is not collected with sufficient granularity in the SRTR to allow for a joint outcome analysis with recurrent events in drug failure. We did not consider adjuvant drug therapies as additional time-to-event processes; however this would be an interesting and clinically relevant extension of this work. A prospective study, or retrospective with drug data review would be ideal so that exact dates of termination can be evaluated. To include multiple time-to-event processes with interval censored drug termination (the current state of SRTR immunosuppressive data) would be unwieldy in a joint model. An application of joint modelling techniques to more detailed RCT data with exact dates for drug data would provide greater understanding of the problem. With collection of exact dates, many other time-to-event processes could be examined for their association with graft failure, such as time to post-transplant diabetes, or time to cancer occurrence or recurrence. Additionally, greater statistical power can be obtained by considering more than just the first drug failure in a recurrent event setting similar to Huang et al. [2011].

The purpose of this paper was to address the source of bias behind differing results when certain traditional statistical methods are used, and to offer a novel approach which shows promise. We also proposed a novel approach for drug efficacy using a joint survival outcome model, and we show that the method can provide additional information in the assessment of efficacy for what is a complex situation in transplantation.

Chapter 3 Bibliography

Robert E Beard. Appendix: Note on some mathematical mortality models. In *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing)*, volume 5, pages 302–311. Wiley Online Library, 1959.

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

Mirko Di Martino, Ursula Kirchmayer, Nera Agabiti, Lisa Bauleo, Danilo Fusco, Carlo Alberto Perucci, and Marina Davoli. The impact of time-window bias on the assessment of the long-term effect of medication adherence: the case of secondary prevention after myocardial infarction. *BMJ Open*, 5(6):e007866, 2015.

David M Dickinson, Mary D Ellison, and Randall L Webb. Data sources and structure. *American Journal of Transplantation*, 3(s4):13–28, 2003.

Nick Freemantle, Louise Marston, Kate Walters, John Wood, Matthew R Reynolds, and Irene Petersen. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*, 347, 2013.

Susan Hadley, Matthew H Samore, W David Lewis, Roger L Jenkins, Adolf W Karchmer, and Scott M Hammer. Major infectious complications after orthotopic liver transplantation and comparison of outcomes in patients receiving cyclosporine or FK506 as primary immunosuppression. *Transplantation*, 59(6):851–858, 1995.

Miguel A Hernán, Alvaro Alonso, Roger Logan, Francine Grodstein, Karin B Michels, Meir J Stampfer, Walter C Willett, JoAnn E Manson, and James M Robins. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology (Cambridge, Mass.)*, 19(6):766, 2008.

J Howell, P Angus, and P Gow. Hepatitis C recurrence: the Achilles heel of liver transplantation. *Transplant Infectious Disease*, 16(1):1–16, 2014.

Xin Huang, Gang Li, Robert M Elashoff, and Jianxin Pan. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*, 17(1):80–100, 2011.

John P Klein and Melvin L Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, 2003.

Emmanuel Lesaffre and Andrew B Lawson. *Bayesian Biostatistics*. John Wiley & Sons, 2012.

Lei Liu, Robert A Wolfe, and Xuelin Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

Jayme E Locke, Shikha Mehta, Rhiannon D Reed, Paul MacLennan, Allan Massie, Anoma Nellore, Christine Durand, and Dorry L Segev. A national study of outcomes among HIV-infected kidney transplant recipients. *Journal of the American Society of Nephrology*, pages ASN–2014070726, 2014.

GJ McKenna, JF Trotter, E Klintmalm, N Onaca, R Ruiz, LW Jennings, M Neri, JG O'Leary, GL Davis, MF Levy, et al. Limiting hepatitis C virus progression in liver transplant recipients using sirolimus-based immunosuppression. *American Journal of Transplantation*, 11(11): 2379–2387, 2011.

Patricia M Odell, Keaven M Anderson, and Ralph B D'Agostino. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, pages 951–959, 1992.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge Univ Press, 2003.

Martyn Plummer et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, 124:125–132, 2003.

Ross L Prentice, Mary Pettinger, and Garnet L Anderson. Statistical issues arising in the Women's Health Initiative. *Biometrics*, 61(4):899–911, 2005.

Gerta Rücker and Dorle Messerer. Remission duration: An example of interval-censored observations. *Statistics in Medicine*, 7(11):1139–1145, 1988.

SN Sehgal. Sirolimus: its discovery, biological properties, and mechanism of action. *Transplantation Proceedings*, 35(3):S7–S14, 2003.

Fred W Selck, Eric B Grossman, Lloyd E Ratner, and John F Renz. Utilization, outcomes, and retransplantation of liver allografts from donation after cardiac death: implications for further expansion of the deceased-donor pool. *Annals of Surgery*, 248(4):599–607, 2008.

M Shah, A Shankar, I Gee, K Nash, M Hoare, P Gibbs, S Davies, and GJM Alexander. A retrospective 15-year review: survival advantage after switching to sirolimus in hepatitis C virus infected liver graft recipients. *Alimentary Pharmacology & Therapeutics*, 41(4): 379–392, 2015.

B Shrum, EM Haddad, VC McAlister, E Renouf, R Malthaner, MS Kjaer, and Gluud LL. Cyclosporin versus tacrolimus for liver transplanted patients, updated. *Cochrane Database of Systematic Reviews*, in press, 2016.

Rosa Giovanna Simonetti, Calogero Camma, Felice Fiorello, Mario Cottone, Maria Rapicetta, Liliana Marino, Germana Fiorentino, Antonio Craxì, Annarita Ciccaglione, Roberto Giuseppetti, et al. Hepatitis C virus infection as a risk factor for hepatocellular carcinoma in patients with cirrhosis. a case-control study. *Ann Intern Med*, 116(2):97–102, 1992.

R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (version 3.0.2) 2010.

Lewis Teperman, Dilip Moonka, Anthony Sebastian, Linda Sher, Paul Marotta, Christopher Marsh, Baburao Koneru, John Goss, Dennis Preston, and John P Roberts. Calcineurin inhibitor–free mycophenolate mofetil/sirolimus maintenance in liver transplantation: The randomized spare-the-nephron trial. *Liver Transplantation*, 19(7):675–689, 2013.

Christian Toso, Shaheed Merani, David L Bigam, AM Shapiro, and Norman M Kneteman. Sirolimus-based immunosuppression is associated with increased survival after liver transplantation for hepatocellular carcinoma. *Hepatology*, 51(4):1237–1243, 2010.

Doris Wagner, Daniela Kniepeiss, Silvia Schaffellner, Estrella Jakoby, Helmut Mueller, Astrid Fahrleitner-Pammer, Philipp Stiegler, Karl-Heinz Tscheliessnigg, and Florian Iberer. Sirolimus has a potential to influence viral recurrence in HCV positive liver transplant candidates. *International Immunopharmacology*, 10(8):990–993, 2010.

Kymberly D Watt, Ross Dierkhising, Julie K Heimbach, and Michael R Charlton. Impact of sirolimus and tacrolimus on mortality and graft loss in liver transplant recipients with or without hepatitis C virus: an analysis of the Scientific Registry of Transplant Recipients database. *Liver Transplantation*, 18(9):1029–1036, 2012.

Andreas Wienke. *Frailty Models in Survival Analysis*. CRC Press, 2010.

Russell H Wiesner, Michael Sorrell, and Federico Villamil. Report of the first International Liver Transplantation Society expert panel consensus conference on liver transplantation and hepatitis C. *Liver Transplantation*, 9(11):S1–S9, 2003.

Chapter 4

Drug failure analysis in transplant registry data

4.1 Introduction

In some fields of medicine, the dynamics of treatment and treatment failures require special consideration in statistical analysis. Transplantation is one of these fields. Transplant recipients must maintain a lifelong immunosuppressive regime, and a failure of one regime requires a change to an alternative drug. A failure in immunosuppressive treatment regime is on the causal pathway to graft failure. Some analyses have used a standard Cox proportional hazards survival model with treatment changes included as a time-dependent covariate (Hadley et al. [1995], Watt et al. [2012]). However, assumptions of this model require that changes in the time-dependent covariate be unrelated in this way to the survival outcome. Ignoring this assumption may lead to biased results (Fisher and Lin [1999]). In transplantation, one drug may be discontinued and an alternative chosen due to graft rejection or the declining health of the patient for some other reason and so drug failures are typically indeed related to the survival outcomes. The primary objective of this paper is to develop a method for analyzing the efficacy of immunosuppressive treatment after liver transplant that accounts for potential dependencies between drug failures and survival. Since the change in drug may be directly related to health status, we propose that treating it as a time-to-event process in a joint outcome model with time-to-graft-failure will provide less biased results. Modelling drug failures as a

time-to-event process has not been considered in analyses of survival after organ transplant in registry data.

We build on the method proposed in Chapter 3 by adding in the longitudinal component from Chapter 2. While the existence of a change-point is an important aspect of the longitudinal analysis, we do not include it here on account of the complexity it poses and the need for a simpler trajectory for the longitudinal component. We are considering three year graft survival and this limits the number of time-points available for building the longitudinal component.

The joint modelling of two time-to-event processes can be used for drug studies where the duration on a particular drug and the survival of a patient on that drug is of interest, especially in situations where the termination of the initial treatment drug is driven by the deteriorating health status of the patient. There is an association between failures of immunosuppressive treatment and graft survival, since treatment failures are likely triggered by another event that increases the risk of death or graft failure (e.g., organ rejection, or cancer occurrence). We suggest that jointly modelling graft survival and time-to-drug-failure in a joint survival outcome model with frailty terms linking the two outcomes is preferable in this situation. We consider an individual frailty to account for unobserved heterogeneity between patients. In addition, we are able to add a longitudinal component to the joint model with the variable creatinine, which is collected every year after transplant. Creatinine is important both as a general health indicator, as well as an indicator of treatment drug toxicity which may also lead to treatment switching.

Much of the research in modelling two time-to-event processes started with the joint analysis of recurrent events and a terminating event. These models allow for a recurrent and a terminal event that are not independent. An important early paper in this area is Liu et al. [2004] who proposed a Cox proportional hazards model with shared frailty for recurrent events and a terminal event in an MCMC approach. The unobserved frailty in the model measures the latent health status of the patient and it is related to both the recurrent event and the terminal event. A shared frailty joint model accommodating multivariate longitudinal and bivariate time-to-event

data with extension to a multivariate survival component was proposed by Chi and Ibrahim [2006] in the setting of cure fractions. Other important work in the area of multiple time-to-event processes includes Rondeau et al. [2007], where the authors propose a joint frailty model using a non-parametric likelihood method. Both Liu et al. [2004] and Rondeau et al. [2007] offer a thorough review of the history of research in this area from the 1990s to the mid 2000s. Another important paper by Rondeau et al. [2008] uses two additive shared frailties to model trial and treatment heterogeneity in a meta-analysis. Liu and Huang [2009] is one of the few papers to consider repeated longitudinal events and two or more time-to-event processes, using a shared random effects model. The hazard of the terminal event (death) depends on both the longitudinal random effects (CD4 counts) and the frailty term from the recurrent event (infection). More recently, Musoro et al. [2015] proposed a shared frailty model for multiple longitudinal outcomes and multiple repeated events (infections) without a terminal event using an MCMC approach for inference.

The use of a joint modelling approach for modelling two event processes which did not include a recurrent event, but one that must come before the other, was first undertaken by Elashoff et al. [2007a] in a competing risks setting. Methods for multiple failure times in the setting of competing risks and semi-competing risks data have become very popular, with further papers from Elashoff et al. [2007b] followed by Williamson et al. [2008] and many others. Most recently, the illness-death models of Xu et al. [2010], Han et al. [2014] and Rouanet et al. [2015] have been applied to semi-competing risks data.

In this paper we utilize a joint model with a longitudinal component in log(creatinine), and a bivariate survival model comprised of a time-to-drug-failure process and a time-to-graft-survival process. This is a novel approach to drug efficacy analysis in transplantation. While Liu and Huang [2009] considered a longitudinal outcome with a recurrent process and a terminal event, a model with two survival events and a longitudinal component has not been previously considered. We hypothesize that higher creatinine levels are correlated with a greater risk of drug failure and also with greater risk of failure of the transplanted liver. In examining

three year survival, we do not use a change-point in the longitudinal process as in the previous chapter as too few time-points are available in the longitudinal trajectory. We also hypothesize that having a change from initial drug therapy is associated with a greater risk of graft failure. In Section 4.2 we describe the development of our two time-to-event process joint model. In Section 4.3 we motivate the need for this type of model in the analysis of graft survival after liver transplantation.

4.2 Statistical methods

4.2.1 Joint longitudinal and survival submodels

The longitudinal component related to modelling the trajectory of log(creatinine) is a longitudinal mixed effects model that is linked to both survival models via random effects. We use a simplified version (without change-point) of the longitudinal component from Chapter 2, examining graft failure up to three years post transplant. Let y_{ij} represent the longitudinal marker log(creatinine) for subject i, $i = 1, ..., n$, at time-point m_{ij}, $j = 1, ...n_i$. The mixed effects model is

$$y_{ij} = \beta_0 + \beta_1 m_{ij} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + b_{0i} + b_{1i} m_{ij} + \epsilon_{ij} \qquad j = 1, ..., n_i; i = 1, ..., n \tag{4.1}$$

where β_0 and β_1 are the intercept and slope; $\boldsymbol{x_{ij}}$ are a set of covariates with respective vectors of regression parameters $\boldsymbol{\beta}$. The random effects b_{0i}, b_{1i} are independent and normally distributed. Let $\boldsymbol{b_i} = (b_{0i}, b_{1i})$. The variance-covariance matrix of $\boldsymbol{b}_i$, $\boldsymbol{D} = \text{diag}(\nu_1, \nu_2)$. Then we assume the measurement error $\epsilon_{ij} \sim N(0, \sigma^2)$ is independent of the random vector $\boldsymbol{b_i}$.

The longitudinal process influences two time-to-event processes: time-to-drug-failure and time-to-graft-failure. For the ith individual, let t_{ki} represent lifetime, $k = 1, 2$, with t_{1i} representing time-to-drug-failure with shape parameter α_1, and t_{2i} representing time-to-graft-failure with shape parameter α_2. The scale parameter λ_{ki} can vary across individuals and events types.

We re-parameterize λ_{ki} as

$$\log(\lambda_{ki}) = z_{ki}^{'}\boldsymbol{\gamma}_k + W_{ki} + \phi_k c_i \quad k = 1, 2 \tag{4.2}$$

where z_{ki} and $\boldsymbol{\gamma}_k$ are the covariates and corresponding regression coefficients and where $\phi_1 = 1$ and $\phi_2 = \phi$. The W_{ki} link the random effects from the longitudinal model to the time-to-drug-failure model, where $W_{1i} = \zeta_1\ \hat{b}_{0i} + \zeta_2\ \hat{b}_{1i}$, and the time-to-graft-failure model ($W_{2i} = \rho_1\ \hat{b}_{0i} + \rho_2\ \hat{b}_{1i}$). The $\boldsymbol{\zeta}$ and $\boldsymbol{\rho}$ are association parameters measuring the relationship between the vector of random effects $\boldsymbol{b_i}$ and the time-to-event processes. The c_i are subject-specific frailty terms which link the two responses. In the graft failure component, ϕ measures the association between c_i and risk of graft failure. We assume that the repeated measures of creatinine are correlated through the random effects $\boldsymbol{b}_i$, and the hazard of the terminal event depends on the longitudinal component and the time-to-drug-failure component through $\boldsymbol{b}_i$ and c_i respectively. The frailty term c_i is assumed independent of $\boldsymbol{b}_i$.

The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government. The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the U.S., submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. A detailed description of the database is available in Dickinson et al. [2003].

4.3 Application: A joint model for three year graft survival after liver transplant using registry data

The goal of our investigation is to apply a new method for assessing effectiveness of immunosuppression after liver transplantation, and to identify risk factors for graft failure while taking into account the initial immunosuppression and time on a particular drug or drug combination. Through joint modelling of the time-to-drug-failure and time-to-graft-failure processes, we can account for the association between time on initial drug therapy and graft survival. Time-to-drug-failure can be extended to a recurrent event format following the methods of Rondeau et al. [2007], Liu and Huang [2009] or others, however we provide an example with the time to first occurrence only, to provide a simplified model that establishes whether there is an association between creatinine, time on initial drug, and graft survival. We hypothesize that higher creatinine values (i.e. larger random effects) and a shorter time-on-initial-drug (i.e. larger frailty) increase the risk of graft failure. The time-to-drug-failure frailty serves as a surrogate indicator for many possible health events since someone who is doing well after transplant is less likely to experience a change in immunosuppressive regime.

Estimation was carried out with a Bayesian approach and a Markov chain Monte Carlo (MCMC) algorithm in R (R Core Development Team [2010]) and Just Another Gibbs Sampler (JAGS, Plummer et al. [2003]) to obtain estimates of the posterior distributions, and with the priors specified in JAGS as non-informative where the components of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are $\sim N(0, 10000)$, $\alpha_k \sim \text{Unif}(0, 1)$, and ρ_1, ρ_2 and ζ_1, ζ_2 are $N(0, 10000)$. The $c_i \sim N(0, 1/\sigma_c)$, with $\sigma_c \sim \text{Unif}(0, 10000)$. The prior distributions for the random effects $\boldsymbol{b_i}$ were specified as $N(\mathbf{0}, \boldsymbol{D})$ with the diagonal components of $\boldsymbol{D} \sim \text{Unif}(0, 10000)$. We ran 3 chains for 500,000 iterations with 400,000 burn-in, which took approximately 30 hours. Convergence was judged by the Brooks-Gelman-Rubin (BGR) convergence diagnostic (Brooks and Gelman [1998]). We used the deviance information criterion (DIC) for model comparison. Covariates were removed from the model if the credible interval contained zero.

We analyzed those patients age 16 and older, receiving a first cadaveric liver transplant in the United States between January 1, 2000 and December 31, 2002. There were 10,015 subjects, with 1,757 (17.5%) events (graft failure, defined as death or retransplant) in the first three years post-transplant. Immunosuppressive therapy is recorded at baseline and discharge from hospital after transplant, then at six months, and yearly. Exact dates for treatment failures are not given, so the data are interval censored. To simplify the analysis we took the midpoint of the interval as the date of treatment change. If a subject also died during the interval where a treatment change was recorded, we took the midpoint of the start of the interval and the date of death as the day of treatment change. We do not use the data collected on the date of death or retransplant, since this would introduce bias as only those who die have a measurement at this unspecified time-point. If we could not determine baseline treatment, the subject was removed from the analysis. We chose to analyze baseline immunosuppression by combination (rather than as any exposure or not, as in Watt et al. [2012] or Toso et al. [2010]), since this accounts for any interaction or synergistic effects between SRL and the calcineurin inhibitors (CYA or TAC). This also allows us to see the effects of SRL uncontaminated by other drugs.

4.3.1 Results

The SRTR collects immunosuppressive therapy for only the first five years after transplant. Most of the changes from initial therapy occur within the first year. Of the 10,015 subjects without missing data, 2,468 (25%) changed their initial therapy within the first three years post-transplant. Table 4.1 shows the number of subjects on each drug combination and how many changed therapies by year 3. This table shows the great disparity in treatment switching among the various regimes. Treating drug failure as a time-to-event process is important to acknowledge the large amount of drug switching taking place in every group except TAC and to account for it in the analysis. We only examine primary immunosuppressive therapy in this model. For the sake of simplicity, induction therapy such as thymoglobulin or anti-CD25 antibody, and adjuvant immunosuppression such as steroids, azathioprine, or mycophenolate

Table 4.1: Summary of the number of individuals (n) on initial drug therapies and the number and percent that changed (n changed (%))from initial therapy within three years post-transplant.

Initial treatment	n	n changed (%)
TAC	8218	1554 (18.9%)
SRL + TAC	533	405 (76.0%)
SRL	149	26 (47.7%)
CYA	1026	416 (40.5%)
SRL + CYA	89	67 (75.3%)

mofetil, are not considered at this time.

Figure 4.1 shows Kaplan-Meier curves for graft survival by initial drug therapy. Those started on CYA + SRL, SRL alone, or TAC alone enjoy the best survival. Figure 4.2 shows Kaplan-Meier curves for time to drug failure for each therapy. Here we see that SRL alone or in combination (SRL + TAC, SRL + CYA) has the shortest time to drug failure. Of the 2,468 who terminated initial therapy, 80% did so in the first year post-transplant.

We analyzed HCV positive and HCV negative subjects separately on account of non-proportional hazards in this variable. There were 4,513 subjects (952 graft failure events) who were HCV positive and 5,502 subjects (805 graft failure events) who were HCV negative. Recurrence of HCV is universal after transplant, however the speed at which it progresses depends on different factors such as viral load, donor age and other risks that are not completely understood. There is growing evidence that this includes immunosuppressive regime (McKenna et al. [2011], Howell et al. [2014]). Table 4.2 shows the number of subjects on each immunosuppressive regime in each cohort.

Other variables considered included recipient and donor age in decades, recipient and donor gender, gender mismatch between recipient and donor, donor BMI, whether a non-heart-beating donor was used, hepatocellular carcinoma (HCC), previous malignancy (pre-transplant), recipient race, recipient blood type, whether a split liver was received, and whether the recipient was in fulminant hepatic failure at the time of transplant. Donor BMI was missing in three subjects so the average was substituted. We also tested for covariate interactions with

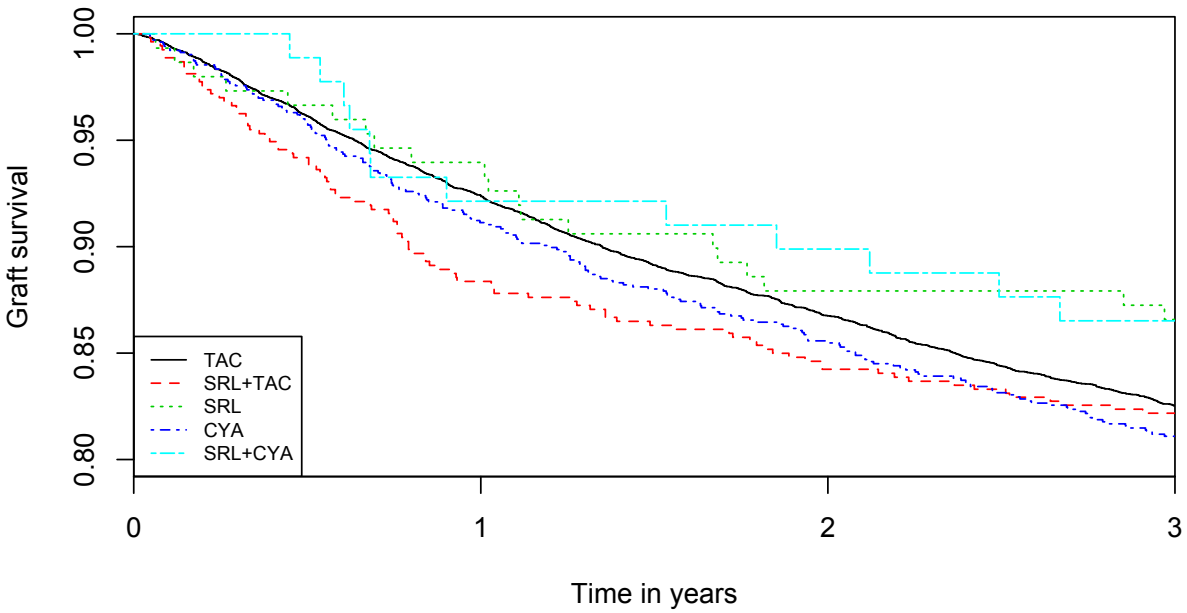

Figure 4.1: Graft survival by initial immunosuppression.

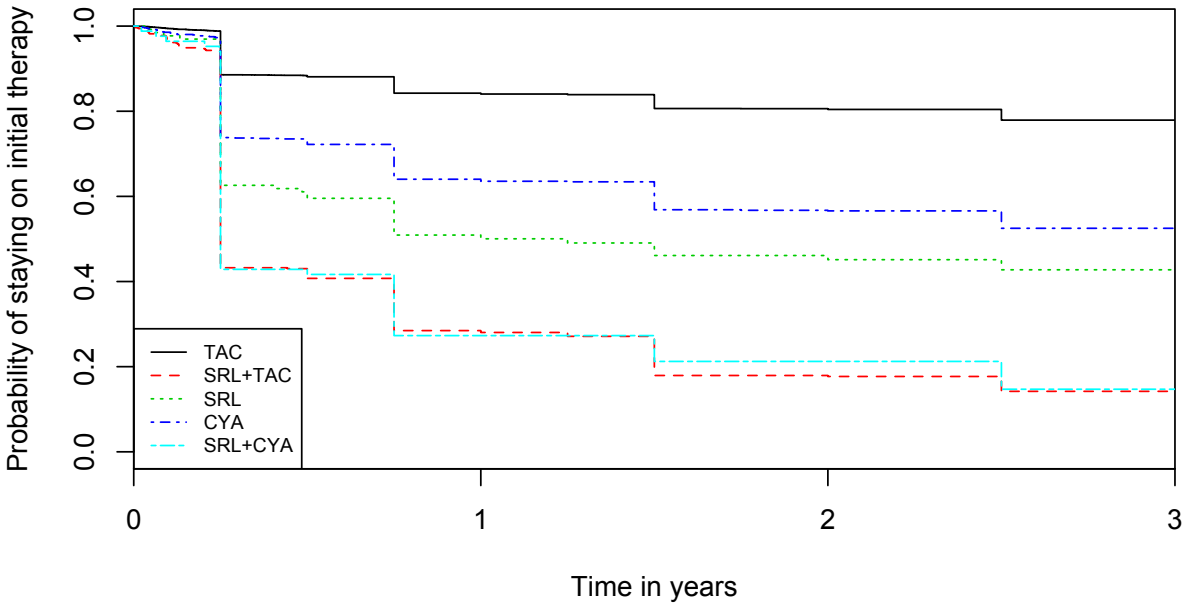

Figure 4.2: Time-to-drug-failure, by initial immunosuppression.

Table 4.2: Summary of initial drug therapies by HCV status.

HCV positive cohort	
Initial treatment	n (%)
TAC	3741(82.9%)
SRL + TAC	223 (4.9%)
SRL	69 (1.5%)
CYA	440 (9.7%)
SRL + CYA	40 (0.9%)
HCV negative cohort	
Initial treatment	n (%)
TAC	4477 (81.3%)
SRL + TAC	310 (5.6%)
SRL	80 (1.5%)
CYA	586 (10.7%)
SRL + CYA	49 (0.9%)

HCC status. We incorporated into our analysis longitudinal values of creatinine up to three years post-transplant. Missing values were allowed in creatinine since the longitudinal trajectory can still be estimated. There were a maximum of 5 longitudinal values (day of transplant, 6 months, year 1, year 2, year 3). The mean number of creatinine observations per subject was 3.3 and the median was 3.

We found a lower DIC when we did not include the slope random effect from the longitudinal component in both time-to-event processes, and so the models for both cohorts included only an intercept random effect in the longitudinal component of the model. This could be because the value of the overall slope β_1 and its estimated random effects are very small (0.006) in the longitudinal model, or possibly because in our three year analysis there are not enough repeated measurements of creatinine to improve the model when this random effect is included. The final model for each cohort was chosen after considering DIC as well as Cox Snell residuals for the graft survival component.

Results from the best fitting joint model for the HCV positive cohort are shown in Table 4.3. Most of the covariates listed in Table 4.3 have a significant effect on time-to-graft-failure

or time-to-drug-failure. The covariate gender was kept in the longitudinal model because it improved the DIC even though the credible interval contained zero.

The random effect b_{0i} was an important linkage term between the longitudinal and the graft failure submodels. In the longitudinal component, we saw a lower log(creatinine) over time in female subjects, while a higher log(creatinine) was seen in older subjects, those in fulminant failure at time of transplant, those who were diabetic at transplant, and those with a higher donor BMI.

The only covariate significant in the time-to-drug-failure submodel was treatment. Being on any initial treatment other than TAC significantly shortens the time to change of initial immunosuppressive therapy. The greatly increased risk of drug failure from initial immunosuppressive regime fits with the degree of treatment switching seen. The linkage parameter ζ was not significantly associated with to time-to-drug-failure.

In the time-to-graft-failure component, we found that none of the immunosuppressive treatments are significantly different from the reference drug TAC in terms of impact on graft survival, after taking into account covariates and the time-to-drug-failure frailty effect. The model is structured so that drug effects are tested in comparison to TAC, the most commonly used treatment. Risk of graft failure is increased for those who are of African American race compared to all other races, for those who had any previous malignancy, and for increased donor age. We found no significant effect for HCC status in the model, even with the variable for previous malignancy removed, and no interaction between initial drug therapy and HCC status.

The subject-specific intercept random effect b_{0i} from the longitudinal component has a significant link with graft survival through the association parameter (ρ_1), with a hazard ratio of 1.567 for each one unit change in the random effect. The random effects themselves range from -0.69 to 1.43, and so the hazard can be as beneficial as 0.73 for those with lower than average creatinine levels, and as high as 1.90. When grouped by drug the random effects do not show any distinct patterns (see Figure 4.3).

The individual frailty c_i from the time-to-drug-failure component has a significant link

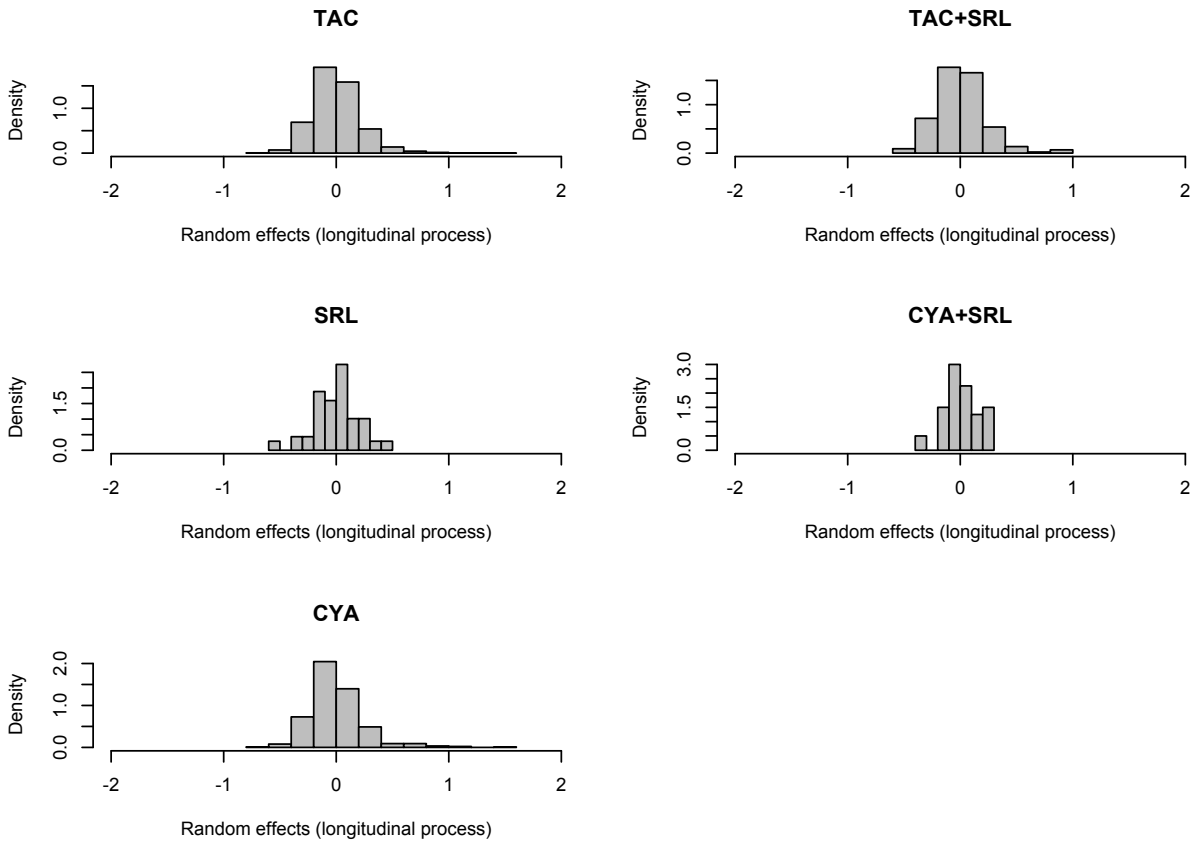

Figure 4.3: HCV positive cohort: Histograms of subject-specific intercept random effects from the longitudinal model, by initial treatment. Positive random effects are associated with increased risk of graft failure.

through the association parameter ϕ to the graft survival model in the HCV positive cohort, where it increases survival time for those with negative log frailties, and decreases survival time for those with positive log frailties. Negative log frailties are associated with a lower risk of a drug failure event. Therefore, failure of initial therapy is associated with shorter graft survival. This is medically sensible since a change in treatment is often precipitated by an adverse event such as rejection. Histograms of individual log frailties grouped by whether a drug failure occurred are shown in Figure 4.4, and plots grouped by initial treatment are shown in Figure 4.5. The value of the frailty can be quite large, with a range of approximately -6 to +7, so the estimated coefficient for the individual frailty in the survival model, while small at 0.046 (hazard: $\exp(0.046 \times frailty)$), can have a large effect depending on the value of the subject specific frailty, with a hazard ratio ranging from a beneficial 0.76, to an increased hazard of 1.38. Note however, that there is quite a wide range of individual frailties for all drug therapies (see Figure 4.5). The variance of the subject-specific frailty is 10.5, which shows there is a great deal of unobserved heterogeneity between subjects. To put the results into context, Table 4.4 shows these effects for each treatment combination, comparing the mean, minimum and maximum frailties of subjects who did not change therapy to those who did. For the group on initial therapy of TAC alone who did experience a drug failure event, the mean frailty shows the highest risk of graft failure (mean hazard ratio: 1.25), suggesting that subjects who do not do well on the 'gold standard' TAC are at greatest risk of graft failure.

Results from the best fitting joint model for the HCV negative cohort are shown in Table 4.5. The random effect for the intercept b_{0i} was seen as an important linkage term between the longitudinal and the graft failure submodels (but again, not in the time-to-drug-failure submodel). We saw a lower log(creatinine) over time in female subjects and in subjects on SRL + TAC, while a higher log(creatinine) was seen in older subjects, those in fulminant failure at time of transplant, those who were diabetic at transplant, and those with a higher donor BMI. This last result, while the effect is small, is nevertheless interesting and could be investigated

Table 4.3: Posterior means for the log hazard, hazard, standard error and quantiles from the HCV positive cohort joint model. (ns=not significant; AA = African American; F = female)

Parameter	**Mean**	**Hazard**	**Std err**	**CI (lower)**	**CI (upper)**
Longitudinal submodel:					
intercept	-0.056		0.026	-0.105	-0.006
slope	0.006		0.0002	0.006	0.006
age in decades	0.004		0.0005	0.003	0.005
fulminant failure	0.361		0.116	0.138	0.589
gender (F)	-0.116		0.010	-0.137	-0.096 (ns)
diabetes	0.068		0.018	0.032	0.102
race (AA)	0.032		0.006	0.022	0.044
baseline trt: TAC (ref)	—		—	—	—
SRL + TAC	0.012		0.021	-0.028	0.055 (ns)
SRL	0.021		0.036	-0.047	0.089 (ns)
CYA	0.040		0.015	0.010	0.069
CYA + SRL	-0.038		0.047	-0.128	0.054 (ns)
ν_1 (variance of b_{0i})	0.064			0.060	0.067
Time-to-drug-failure submodel:					
intercept	-9.761		0.097	-9.952	-9.580
baseline trt: TAC alone (ref)	—		—	—	—
SRL + TAC	7.522		0.401	6.755	8.310
SRL	3.705		0.609	2.545	4.895
CYA	3.022		0.251	2.527	3.514
CYA + SRL	9.523		1.102	7.436	11.715
σ_c (variance of c_i)	14.929			13.682	16.374
Time-to-graft-failure submodel:					
α (shape parameter)	0.933		0.028	0.879	0.986
intercept	-8.681		0.215	-9.088	-8.263
donor age (decades)	0.016	1.016	0.002	0.012	0.020
race (AA)	0.182	1.200	0.036	0.114	0.248
gender (F)	0.128		0.075	-0.021	0.272 (ns)
previous malignancy	0.451	1.568	0.107	0.239	0.660
baseline trt: TAC (ref)	–		–	–	–
SRL + TAC	-0.101		0.158	-0.426	0.190 (ns)
SRL	-0.462		0.321	-1.146	0.142 (ns)
CYA	0.040		0.108	-0.184	0.243 (ns)
CYA + SRL	-0.210		0.373	-0.977	0.493 (ns)
ρ_1	0.447	1.564	0.160	0.131	0.770
ϕ	0.046	1.047	0.011	0.023	0.068

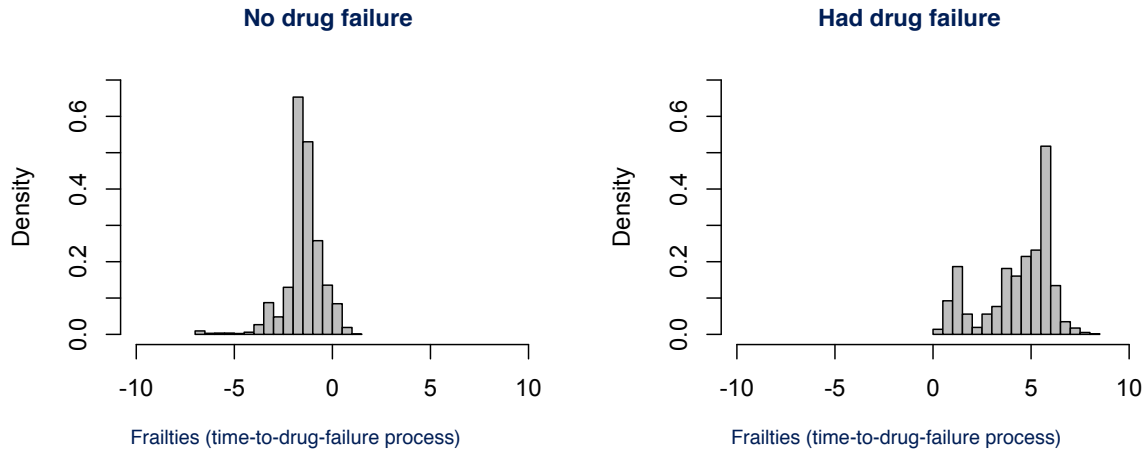

Figure 4.4: HCV positive cohort: Histograms of individual log frailties from time-to-drug-failure model, by whether initial treatment was changed. Positive log frailties are associated with increased risk of graft failure.

Table 4.4: Summary of frailty effects on graft failure by initial drug therapy for the HCV positive cohort.

Initial trt	No drug failure: Mean (min, max)	Hazard ratio (min, max)	Had drug failure: Mean (min, max)	Hazard ratio (min, max)
TAC	-1.1 (-4.3, 1.3)	0.8, 1.1	4.9 (3.4, 7.1)	1.2, 1.4
SRL + TAC	-3.9 (-6.0, -1.3)	0.8, 0.9	1.2 (0.2, 3.5)	1.0, 1.2
SRL	-2.2 (-3.9, 0.1)	0.8, 1.0	2.7 (1.7, 4.5)	1.1, 1.2
CYA	-2.2 (-4.6, 0.5)	0.8, 1.0	3.0 (1.9, 4.9)	1.1, 1.3
SRL + CYA	-4.5 (-5.7, -2.7)	0.8, 0.9	0.6 (-0.02, 2.0)	1.0, 1.1

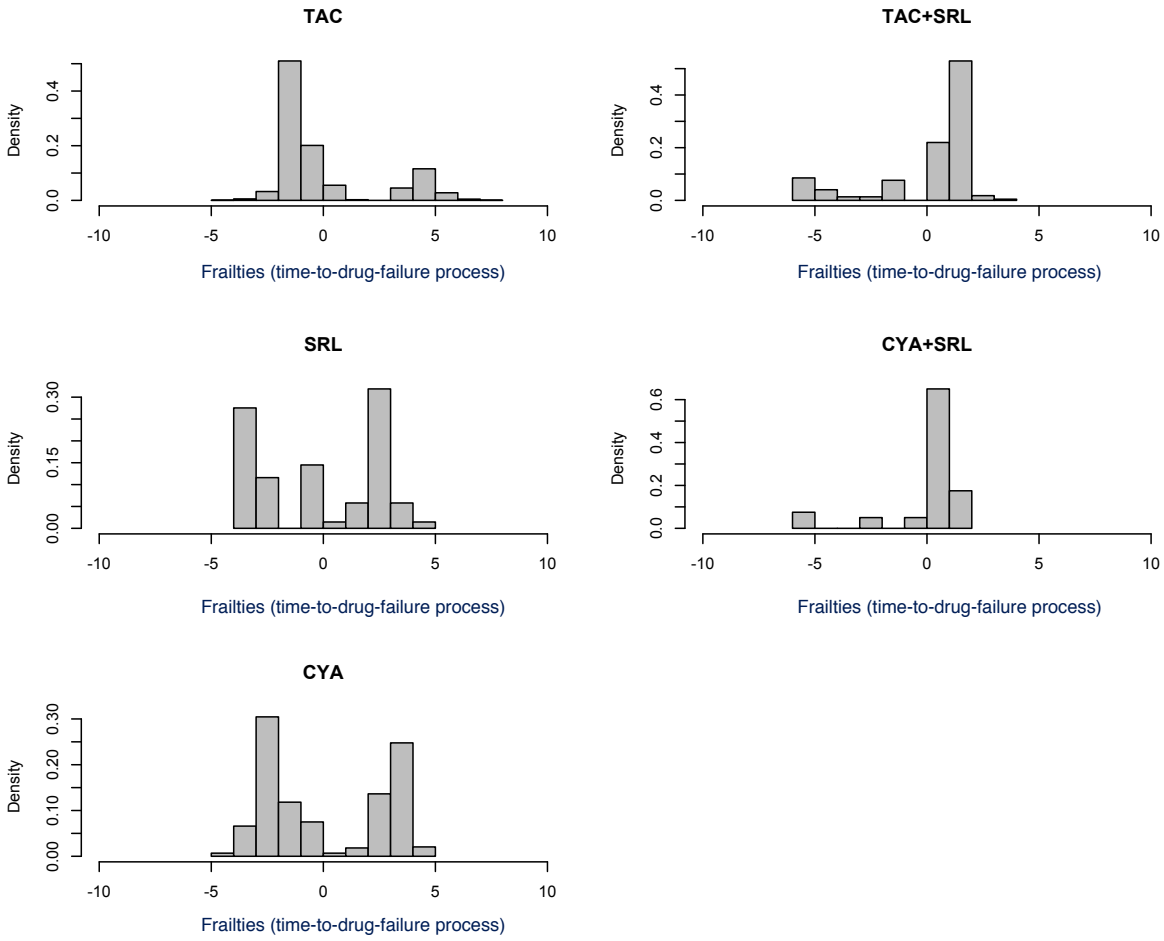

Figure 4.5: HCV positive cohort: Histograms of individual log frailties from time-to-drug-failure model, by initial immunosuppression. Positive log frailties are associated with increased risk of graft failure.

further. It is possible that a donor with a higher BMI has more fatty tissue in the liver, resulting in more difficult postoperative recovery and higher creatinine. Fatty liver can also lead to more inflammation and therefore higher doses of immunosuppression. It is also interesting that this effect was not seen in the HCV positive cohort, possibly because HCV positive subjects are often counselled against accepting a marginal donor organ with increased risk due to fatty liver.

Similar to the HCV positive cohort, the only covariates significant in the time-to-drug-failure submodel were baseline treatment. Being on any initial treatment other than tacrolimus significantly shortens the time to failure of initial immunosuppressive therapy.

In the graft survival component, no immunosuppressive treatment was significantly different from the reference drug TAC in terms of impact on graft survival. Risk of graft failure is increased for those who have HCC, for those who are African American, and for increased donor age. Gender was kept in the model because it improved DIC. We also tested all variables for interaction with HCC status and did not find any significant interaction.

Another interesting feature of this analysis is the strong effect of the longitudinal random effect for log(creatinine) in the graft survival submodel. A one unit increase in log(creatinine) has a hazard ratio of 1.881 (compared to 1.567 in the HCV positive cohort). The subject-specific random effect ranges from -0.971 to 1.469, so the hazard can be reduced for negative random effects (hazard ratio 0.541) or increased to as much as 2.530 for the largest random effect. It is possible that this HCV negative cohort, having a wider variety of indications for liver transplant compared to the more homogeneous HCV positive cohort, comprise a group with diseases that involve more renal decompensation. This leads us to conclude that our approach taken in Chapter 2 where both cohorts were analyzed together may not be the most suitable method compared to analyzing them separately as we do here. When grouped by drug (see Figure 4.6), the intercept random effects are similar to the HCV positive cohort, i.e. no distinct pattern by drug (compare to Figure 4.3).

The individual frailty c_i from the time-to-drug-failure component also has a significant

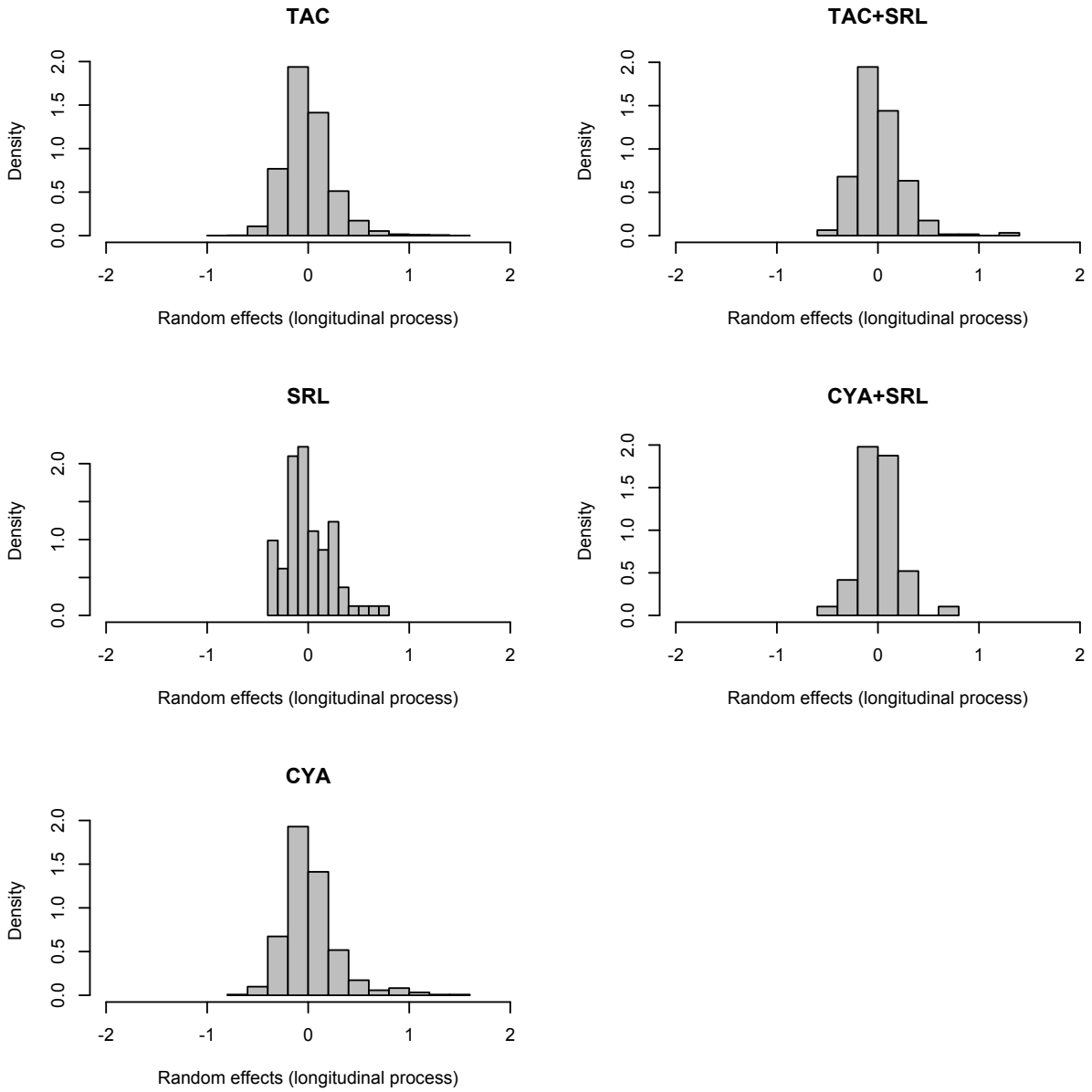

Figure 4.6: HCV negative cohort: Histograms of subject-specific intercept random effects from the longitudinal model, by initial treatment. Positive random effects are associated with increased risk of graft failure.

linkage through the association parameter ϕ in the time-to-graft-failure model for the HCV negative cohort, where it increases survival time for those with negative log frailties, and decreases survival time for those with positive log frailties. This result is slightly larger than in the HCV positive cohort, with a higher risk seen in the negative cohort. Histograms of individual log frailties grouped by whether a drug change occurred are shown in Figure 4.7, and plots grouped by initial treatment are shown in Figure 4.8. The value of the frailty also had a wider range than the HCV positive cohort (-6.5 to +9). Figure 4.8 shows the wide range of individual frailties for all drug therapies. The variance of the subject-specific frailty was similar to the positive cohort, showing there is a great deal of unobserved heterogeneity between subjects.

Table 4.6 shows the frailty effects for each treatment combination, comparing the frailties of subjects who did not change therapy to those who did in the HCV negative cohort. Again we see the greatest mean frailty in those who had baseline drug failure on TAC.

The overall picture emerging here is that subjects who must discontinue baseline immunosuppressive therapy indeed have a poorer outcome, as expected, due to events precipitating drug failure. The calcineurin-sparing combinations of SRL + CYA and SRL + TAC stand out (in both cohorts) for having the largest negative mean frailty in those who did not experience drug failure, and the smallest positive mean frailty in those who did have a drug failure event. This translates to a lower risk of graft failure regardless of failures in treatment. The effect of initial drug regime on graft survival must be considered in the context of the time-to-drug-failure, in order to understand the overall risk to each subject. Stable subjects who do not need to change baseline treatment have the best outcome, regardless of initial regime. For those who do have to switch treatments, those at highest risk seem to be the small group who must switch from TAC alone to some other treatment. This suggests an interesting new area of research into the dynamics of drug failure on TAC compared to any other treatment, exploring what risks are specific to this group.

We compared our model to a model using the often employed standard Cox proportional

Table 4.5: Posterior means for the log hazard, hazard, standard error and quantiles from the HCV negative cohort joint model.

Parameter	**Mean**	**Hazard**	**Std err**	**CI (lower)**	**CI (upper)**
Longitudinal submodel:					
intercept	-0.128		0.019	-0.165	-0.091
slope	0.005		0.0002	0.005	0.005
age in decades	0.007		0.0004	0.006	0.008
fulminant failure	0.104		0.018	0.068	0.140
gender (F)	-0.154		0.009	-0.173	-0.137
diabetes	0.052		0.016	0.021	0.084
donor BMI	0.0002		0.00007	0.0001	0.0003
baseline trt: TAC alone (ref)	—		—	—	—
SRL + TAC	-0.041		0.018	-0.078	-0.006
SRL	-0.005		0.036	-0.074	0.063 (ns)
CYA	0.015		0.014	-0.011	0.104 (ns)
CYA + SRL	0.014		0.045	-0.073	0.104 (ns)
ν_1 (variance of b_{0i})	0.075			0.071	0.079
Time-to-drug-failure submodel:					
intercept	-9.949		0.081	-10.110	-9.792
baseline trt: TAC alone (ref)	—		—	—	—
SRL + TAC	7.249		0.336	6.594	7.916
SRL	4.112		0.565	3.029	7.916
CYA	2.643		0.220	2.206	3.075
CYA + SRL	5.844		0.735	4.437	7.356
σ_c (variance of c_i)	15.152			14.085	16.129
Time-to-graft-failure submodel:					
α (shape parameter)	0.848		0.029	0.789	0.906
intercept	-8.134		0.226	-8.584	-7.683
HCC	0.527	1.694	0.135	0.253	0.790
donor age (decades)	0.010	1.011	0.002	0.006	0.014
gender	-0.212	0.809	0.077	-0.355	-0.060
baseline trt: TAC (ref)	–		–	–	–
SRL + TAC	0.233		0.145	-0.067	0.516 (ns)
SRL	-0.199		0.321	-0.836	0.380 (ns)
CYA	0.167		0.112	-0.059	0.388 (ns)
CYA + SRL	-0.707		0.516	-1.777	0.211 (ns)
ρ_1	0.632	1.881	0.153	0.337	0.929
ϕ	0.064	1.066	0.012	0.039	0.088

Table 4.6: Summary of frailty effects on graft survival by initial drug therapy for the HCV negative cohort.

Initial trt	No drug failure: Mean (min, max)	Hazard ratio (min, max)	Had drug failure: Mean (min, max)	Hazard ratio (min, max)
TAC	-1.3 (-5.0, 1.4)	0.7, 1.1	5.7 (4.4, 9.1)	1.3, 1.8
SRL + TAC	-4.8 (-6.5, -1.9)	0.7, 0.9	1.6 (0.6, 4.3)	1.0, 1.3
SRL	-3.3 (-4.9, -0.7)	0.7, 1.0	3.3 (2.0, 4.7)	1.1, 1.4
CYA	-2.6 (-5.4, 0.5)	0.7, 1.0	4.0 (2.7, 6.8)	1.2, 1.5
SRL + CYA	-4.3 (-5.5, -1.3)	0.7, 0.9	2.3 (1.3, 3.4)	1.1, 1.2

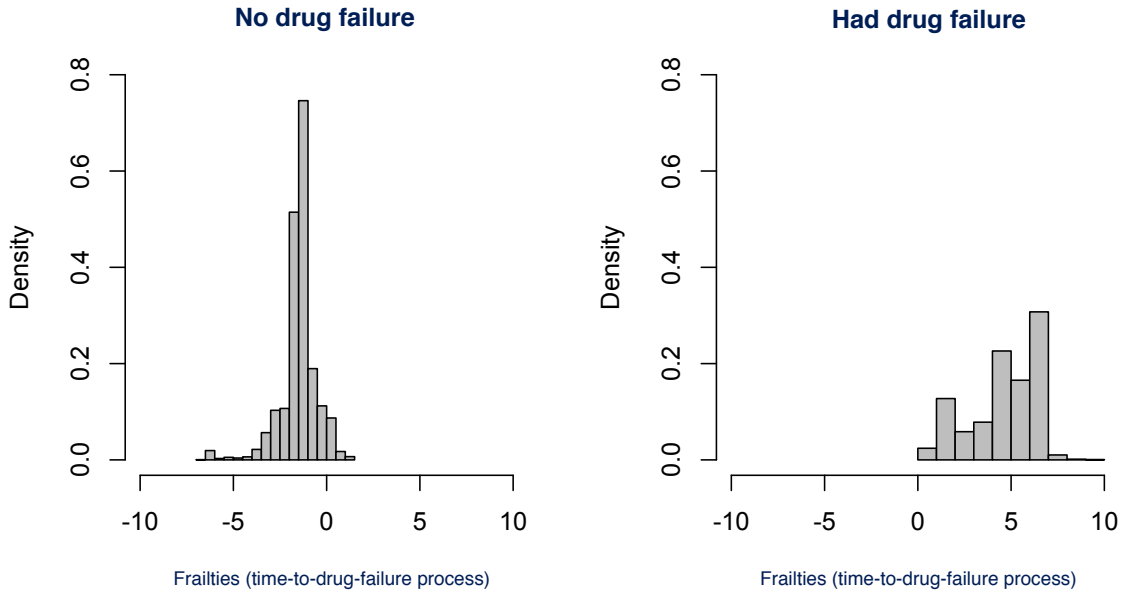

Figure 4.7: HCV negative cohort: Histograms of individual log frailties from time-to-drug-failure model, by whether drug failure occurred. Positive log frailties are associated with increased risk of graft failure.

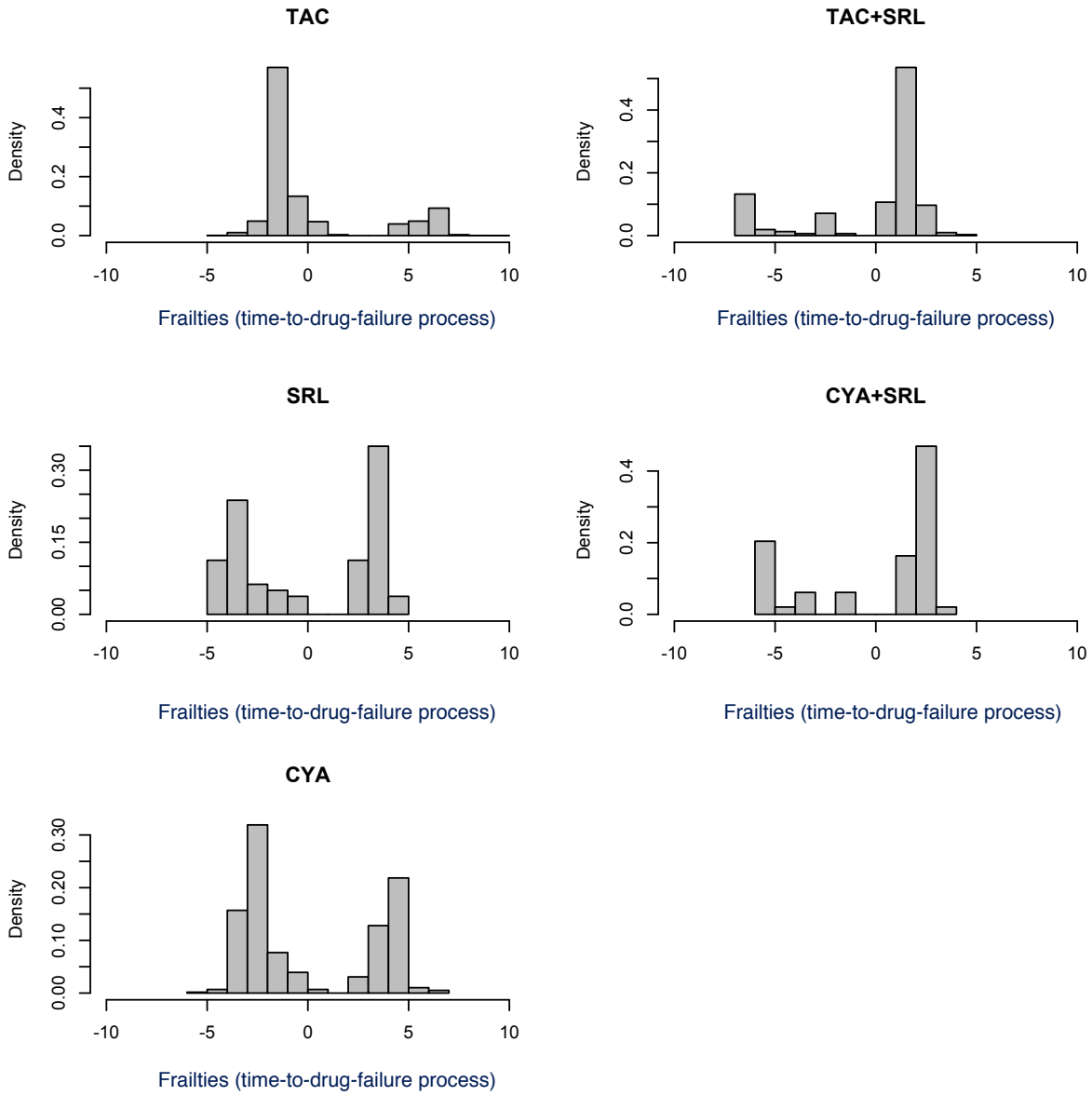

Figure 4.8: HCV negative cohort: Histograms of individual log frailties from time-to-drug-failure model, by initial immunosuppression. Positive log frailties are associated with increased risk of graft failure.

Table 4.7: Results from a standard Cox proportional hazards model for the HCV positive cohort.

Parameter	**Coefficient**	**Hazard**	**StdError**	**p**
donor age (decades)	0.016	1.016	0.002	< 0.001
log(creatinine) day 0	0.220	1.246	0.070	0.002
gender (F)	0.141	1.152	0.074	0.057
previous malignancy	0.385	1.470	0.111	< 0.001
HCC	0.197	1.217	0.112	0.078
race (AA)	0.163	1.177	0.034	< 0.001

hazards model, with initial treatment as a baseline covariate. We also included log(creatinine) from the day of transplant as a covariate (86 were missing and so the mean was substituted for these cases). Using a stepwise procedure, the covariates that remain significant in the model are similar to both of the joint models presented above, although not identical. The significant covariates from this model for the HCV positive cohort are presented in Table 4.7. There were no significant differences in graft survival by baseline treatment for the HCV positive cohort. The hazards for the significant covariates are very similar to the joint model, and the standard errors are almost the same. The Cox proportional hazards model has a smaller effect for log(creatinine) compared to the random effects in the joint model, and the values for log(creatinine) are smaller (range: -2.3 to 2.9) than the values for the random effects and so this translates to a smaller effect overall compared to the intercept random effect for creatinine in the joint model. The joint model also has the added hazard from covariate ϕ and its association with the frailty c_i in the time-to-drug-failure model.

In addition, we compared our model to a Cox proportional hazards model that treats the use of SRL as a time-dependent covariate. Significant findings are shown in Table 4.8. We treated SRL as a time-dependent indicator variable where it takes a value of 1 if the subject is started on SRL (in combination or alone) at transplant, and changes to 0 when the subject changed treatment. Here we see that the effect for SRL is very significant (p = 0.029) and the hazard is 1.262 for a subject starting on SRL, and this is a result that we did not see in either of the previous models. This shows the biased results that are obtained when a time-

Table 4.8: Results from a Cox proportional hazards model with time-dependent SRL treatment covariate for the HCV positive cohort.

Parameter	**Coefficient**	**Hazard**	**StdError**	**p**
any SRL (time-dependent)	0.233	1.262	0.107	0.029
donor age (decades)	0.010	1.016	0.002	<0.001
log(creatinine) day 0	0.155	1.168	0.060	0.010
gender (female)	0.126	1.134	0.064	0.048
race (African American)	0.164	1.178	0.029	<0.001
any CYA at baseline	0.634	1.885	0.279	0.023
previous malignancy	0.388	1.474	0.094	<0.001

Table 4.9: Results from a standard Cox proportional hazards model for the HCV negative cohort.

Parameter	**Coefficient**	**Hazard**	**StdError**	**p**
donor age (decades)	0.009	1.009	0.002	< 0.001
log(creatinine) day 0	0.158	1.171	0.065	0.015
gender (F)	-0.172	0.842	0.075	0.022
race (AA)	0.128	1.137	0.042	0.002
previous malignancy	0.724	2.063	0.133	< 0.001

dependent covariate is used in a Cox proportional hazards model when a change in the time-dependent covariate is also related to outcome. Some of this bias could stem from the dramatic differences in treatment switching. Subjects who switch treatment early are those with adverse events who therefore experience more graft failure. Results from other covariates such as donor age, log(creatinine), gender and previous malignancy are similar to the previously illustrated models.

The comparisons to the Cox proportional hazards model for the HCV negative cohort show similar findings. In the HCV negative cohort, when we apply the standard Cox proportional hazards model to graft survival we get the significant results seen in Table 4.9. Again, baseline treatment is not significant factor for graft survival in any combination. The effects for donor age and gender are similar to the joint model.

In the Cox model with time dependent treatment effect for SRL, we again see a greatly

Table 4.10: Results from a Cox proportional hazards model with time-dependent SRL treatment covariate for the HCV negative cohort.

Parameter	Coefficient	Hazard	StdError	p
any SRL (time-dependent)	0.404	1.498	0.095	< 0.001
donor age (decades)	0.008	1.008	0.002	< 0.001
log(creatinine) day 0	0.166	1.180	0.055	0.003
gender (F)	-0.177	0.838	0.064	0.006
recipient age (decades)	0.005	1.005	0.003	0.057
HCC	0.226	1.253	0.119	0.059
previous malignancy	0.758	2.133	0.112	< 0.001
race (AA)	0.102	1.108	0.036	0.005
blood type AB	-0.309	0.734	0.149	0.038
diabetic	0.239	1.270	0.099	0.016
CYA at baseline	0.276	1.317	0.087	0.002

increased risk for any SRL exposure, with significant results shown in Table 4.10.

This analysis shows that using two time-to-event processes to analyze drug failure and graft failure, along with a longitudinal component in creatinine, is a valuable approach. The model has captured the increased risk to graft failure that is present with a sharp change in the biomarker, or with an adverse event that precipitates drug failure.

4.4 Discussion

Analyzing treatment changes as a time-to-event process is a preferred approach in observational data analysis since it avoids discarding data. It also avoids the violation of model assumptions such as when treating drug as a time-dependent covariate. Analysis of the two time-to-event outcomes of drug failure and graft survival using a joint model can account for dependence between the two processes without making strong assumptions. We allow the important information contained in the time-to-drug-failure component to influence the hazard of the time-to-graft failure component. Transplant registry data, with 100% enrolment and follow up until death, is a valuable and readily available data source that can provide insight into factors affecting health outcomes after liver transplant. The time-to-drug-failure process

acts as a surrogate for time to any adverse event such as infection, rejection, cancer occurrence or other, that may be unreliably collected in the registry data. This two outcome joint model describes the data structure well. We believe that use of a straightforward joint survival outcome model is appropriate for two reasons: first, clinical practice tells us that the association is strong between the drug failure and graft failure processes, and second, less complex joint modelling techniques are more likely to be adopted in practice in this field. Joint modelling is not commonplace in analysis of SRTR data for transplantation, yet despite readily available and appropriate data, it has only been used in one application of which we are aware (Liu et al. [2004] for kidney transplant data).

We did not find any significant differences between baseline treatments when compared to the 'gold standard' treatment TAC. However in the context of our joint outcome model, treatment effects should be interpreted on a subject-specific basis conditional on the frailty. We found great variability in the frailties. This is an indication that treatments alone are not good predictors of survival - generally, one treatment does provide better on average survival than another, yet there is great variability between treatments. When we condition on the frailty, we will have a better prediction of survival and better understanding of survival by treatment patterns.

Rather than discard information, we have presented a model that includes all available data in a way that makes the most scientific sense. The longitudinal component takes into account factors affecting creatinine level over time. It is a proxy for choice of initial treatment, mimicking the physician decision process by taking into account all covariates affecting the evolution of creatinine over time. Degree of renal impairment is an important factor due to nephrotoxicity of the immunosuppression treatment. Our model offers more insight into the medical process and makes scientific sense. The subject-specific frailties from the time-to-drug-failure model account for latent variables that have a significant effect on graft survival, thus accounting for more of the variance in the data and enhancing our understanding of the whole process while improving the fit of the model. The information contained in the frailty

covariate ϕ adds valuable dimension to the model that cannot be accounted for in traditional models.

A limitation of this data set is the lack of exact dates for treatment failures, and the use of the interval midpoint is not ideal, however randomized clinical trial data (with exact dates) is not readily available for public analysis. Statistical bias related to the use of the midpoint is discussed in Rücker and Messerer [1988] and Odell et al. [1992]. To add recurrent events to the drug failure model would involve the use of doubly interval censored data that are also correlated. Furthermore, data on subjects who experience treatment failure more than once during a data collection interval is not collected with sufficient granularity in the SRTR to allow for a joint outcome analysis with recurrent events in drug failure. An application of joint modelling techniques to detailed randomized clinical trial data would provide greater understanding of the problem, since clinical trials normally collect exact dates. With collection of exact dates, many other time-to-event processes could be examined for their association with graft survival, such as time to post-transplant diabetes, or time to cancer occurrence or recurrence.

Another limitation is that important information such as treatment dosage amounts and drug trough levels are missing from the registry data. Trough levels measure the amount of drug exposure per patient, which can vary on a subject-specific basis even when subjects are given the same amount of drug. This is another interesting avenue worth pursuing with transplant data in joint frailty models.

The authors acknowledge the limitations of this three year retrospective analysis using registry data. All models in this paper would benefit from more data in the SRL arms. In Chapter 2 we were limited in the size of our dataset because the inclusion of ten years of longitudinal data increased convergence time. In this chapter we have reduced the longitudinal time-frame but we suspect that the use of only three years of data in the longitudinal component is not enough to see a significant effect from the random effect from the ζ_k in the time-to-drug-failure component. However our goal was met which showed that joint modelling provides added

value to analysis of survival after liver transplantation.

Chapter 4 Bibliography

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

Yueh-Yun Chi and Joseph G Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, 2006.

David M Dickinson, Mary D Ellison, and Randall L Webb. Data sources and structure. *American Journal of Transplantation*, 3(s4):13–28, 2003.

Robert M Elashoff, Gang Li, and Ning Li. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in medicine*, 26(14):2813, 2007a.

Robert M Elashoff, Gang Li, and Ning Li. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3):762–771, 2007b.

Lloyd D Fisher and Danyu Y Lin. Time-dependent covariates in the cox proportional-hazards regression model. *Annual Review of Public Health*, 20(1):145–157, 1999.

Susan Hadley, Matthew H Samore, W David Lewis, Roger L Jenkins, Adolf W Karchmer, and Scott M Hammer. Major infectious complications after orthotopic liver transplantation and comparison of outcomes in patients receiving cyclosporine or FK506 as primary immunosuppression. *Transplantation*, 59(6):851–858, 1995.

Baoguang Han, Menggang Yu, James J Dignam, and Paul J Rathouz. Bayesian approach for flexible modeling of semicompeting risks data. *Statistics in Medicine*, 33(29):5111–5125, 2014.

J Howell, P Angus, and P Gow. Hepatitis C recurrence: the achilles heel of liver transplantation. *Transplant Infectious Disease*, 16(1):1–16, 2014.

Lei Liu and Xuelin Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58 (1):65–81, 2009.

Lei Liu, Robert A Wolfe, and Xuelin Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

GJ McKenna, JF Trotter, E Klintmalm, N Onaca, R Ruiz, LW Jennings, M Neri, JG O'Leary, GL Davis, MF Levy, et al. Limiting hepatitis C virus progression in liver transplant recipients using sirolimus-based immunosuppression. *American Journal of Transplantation*, 11(11): 2379–2387, 2011.

Jammbe Z Musoro, Ronald B Geskus, and Aeilko H Zwinderman. A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. *Biometrical Journal*, 57(2):Biometrical Journal, 2015.

Patricia M Odell, Keaven M Anderson, and Ralph B D'Agostino. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, pages 951–959, 1992.

Martyn Plummer et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, 124:125–132, 2003.

Virginie Rondeau, Simone Mathoulin-Pelissier, Hélène Jacqmin-Gadda, Véronique Brouste, and Pierre Soubeyran. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8(4):708–721, 2007.

Virginie Rondeau, Stefan Michiels, Benoit Liquet, and Jean-Pierre Pignon. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*, 27(11):1894–1910, 2008.

Anaïs Rouanet, Pierre Joly, Jean-François Dartigues, Cécile Proust-Lima, and Hélène Jacqmin-Gadda. Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia. *Biometrics*, 59:1–23, 2015.

Gerta Rücker and Dorle Messerer. Remission duration: An example of interval-censored observations. *Statistics in Medicine*, 7(11):1139–1145, 1988.

R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (version 3.0.2) 2010.

Christian Toso, Shaheed Merani, David L Bigam, AM Shapiro, and Norman M Kneteman. Sirolimus-based immunosuppression is associated with increased survival after liver transplantation for hepatocellular carcinoma. *Hepatology*, 51(4):1237–1243, 2010.

Kymberly D Watt, Ross Dierkhising, Julie K Heimbach, and Michael R Charlton. Impact of sirolimus and tacrolimus on mortality and graft loss in liver transplant recipients with or without hepatitis C virus: an analysis of the scientific registry of transplant recipients database. *Liver Transplantation*, 18(9):1029–1036, 2012.

PR Williamson, Ruwanthi Kolamunnage-Dona, Pete Philipson, and AG Marson. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine*, 27(30):6426–6438, 2008.

Jinfeng Xu, John D Kalbfleisch, and Beechoo Tai. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725, 2010.

Chapter 5

Conclusion and future work

We have presented three papers applying joint modelling techniques to transplant registry data. We have shown that the longitudinal biomarker creatinine is associated with graft survival, where higher creatinine levels and a sharp increase in biomarker trajectory are associated with greater risk of graft failure. We showed that inclusion of biomarkers in the longitudinal component can add insight to the post-transplant survival process. Ideally, future work would include a powerful computational framework such as parallel computing to handle a large dataset with multiple longitudinal biomarkers and a more detailed examination of change-point effects. Much more work could be done investigating the relationship between biomarker change-points and graft failure.

We have also shown that using a two outcome joint survival model to assess efficacy of immunosuppressive drugs is a valuable approach. A shorter time-to-drug-failure is associated with greater risk of graft failure in a joint model where individual frailties capture latent factors related to time on drug. Our novel application of joint time-to-event outcomes to the drug efficacy issue holds promise. More power could be added to statistical models by harnessing recurrent event data for drug failure. In addition, the opportunities for the application of joint outcome models for dynamic predictions in a clinical care setting are exciting. For example, the decision to switch drug regime is a significant decision-making point for the transplant physician. A prognosis application for treatment change decisions would be a powerful tool. Whether to switch treatment, or stay with the current one and wait for retransplant is an im-

portant question, since there is a risk inherent in a change of immunosuppression protocol and there is impact on the immune system of the patient. Risks related to treatment changes include post-transplant lymphoproliferative disease, cancer and death. The framework proposed in this thesis could be used to assess the risk of a treatment change on an individual basis. While a powerful computational analysis would be required initially to determine the best model, predictions for a single patient using the results of the model would be straightforward and simple in a clinical setting.

Furthermore, joint models are well suited to explore the wealth of information collected pre-transplant, while a subject is on the waiting list, and to test for the impact of covariates on the post-transplant process or even to determine ahead of the transplant the optimum treatment regime for each subject. Biomarker profiles are collected rigorously pre-transplant and this information can be used to make decisions about post-transplant care.

While we were limited by a number of factors, specifically widely spaced intervals for data collection and a lack of exact dates for drug regime failures, the significant associations discovered through joint outcome models using registry data has implications for clinical trials and clinical care. Joint models provide added value and more information to the analysis, and with data that is unhindered by wide collection intervals, or lack of exact dates for drug failures, the results would be interesting.

We have shown that there are pitfalls to avoid in the use of traditional methods for post-transplant survival analysis, and while the application of unbiased statistical methods is important in the analysis of registry data, there is still much work to be done, and the methods for registry data analysis should continue to borrow strength from discoveries made in the analysis of clinical trial data. It is important that researchers work to allow results and techniques developed in one area to enhance and enable important discoveries in others. We hope that by showing the value of joint modelling in transplant registry data it will open up opportunities to refine and develop new methods with more detailed clinical trial or clinical care data.

These results are not limited to transplantation, and can be applied generally to other reg-

istries in different areas of medicine such as cancer and mental health.

Curriculum Vitae

M. Elizabeth Renouf

Education

M.Sc., Statistics, 2008-2010

Collaborative Program in Biostatistics

Department of Statistical & Actuarial Sciences

University of Western Ontario

B.Sc., Statistics, 2004-2007

Department of Statistical & Actuarial Sciences

University of Western Ontario

M.A., Classics, 1993-1995

Department of Classical Studies

University of Western Ontario

B.A., Classical Studies and Latin, 1989-1993

Department of Classical Studies

University of Western Ontario

Honours and Awards

NSERC Vanier CGS D 2010-2013

NSERC CGS M, 2009

NSERC USRA, 2006

Publications and Presentations

Bellhouse DR, Renouf EM, Raut R, Bauer MA. (2009) De Moivre's Knowledge Community: an Analysis of the Subscription List to the Miscellanea Analytica. Notes and Records of the Royal Society 63, 137-162.

Stanford DA, Renouf EM, McAlister VC. (2008) Waiting for Liver Transplantation in Canada: Wait List History 2000-04 and Sensitivity Analysis for the Future. Health Care Management Science Jun.11(2) 1184-195.

McAlister VC, Davenport E, Renouf E. (2007) Cholecystectomy deferral in patients with endoscopic sphincterotomy. Cochrane Database Syst Rev. Oct 17;(4):CD006233.

Haddad EM, McAlister VC, Renouf E, Malthaner R, Kjaer MS, Gluud LL. (2006) Cyclosporin versus tacrolimus for liver transplanted patients. Cochrane Database Syst Rev. Oct 18;(4):CD005161.

McAlister VC, Haddad E, Renouf E, Malthaner RA, Kjaer MS, Gluud LL. (2006) Cyclosporin versus Tacrolimus as Primary Immunosuppressant After Liver Transplantation: A Meta-Analysis. American Journal of Transplantation 6(7): 1578-1585.

Renouf E. and McAlister VC. (2004) Old Drugs for new: Contemporary Immunosuppressant Development. Medicinal Chemistry Reviews Online 1, 1-5.

Arlen D, Cole E, Hebert D, Holland D, Jevnikar A, Knoll G, Zaltzman J, Renouf E, Feldman P. (2004) Study of Cardiovascular Outcomes in Renal Transplantation (SCORe). Presented at the Canadian Society of Transplantation Annual Scientific Meeting.

Renouf E, McMichael JP, McAlister VC. (2003) Predicting Sirolimus Doses and Levels from Prior or Concurrent Tacrolimus or Cyclosporine Dose/Level Using the Intelligent Dosing System. Presented at the American Transplant Congress, Washington, D.C.

Jevnikar A, Cole E, Zaltzman J, Arlen D, Knoll G, Hebert D, Renouf E, Sim D. (2001) Selection of Tacrolimus for Induction and Conversion Therapy: A Retrospective Analysis of the Clinical Indications and Effectiveness in Non-Study Renal Transplant Patients in Ontario, Canada. Presented at the 2nd International Congress on Immunosuppression, San Diego, California.

July 6, 2016