Methods for the Analysis of Complex Time-to-Event Data

Methods for the Analysis of Complex Time-to-Event Data

A dissertation presented

by

Catherine Lee

to

The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts

July 2017

Dissertation Advisors: Rebecca Betensky and Sebastien Haneuse

Catherine Lee

Methods for the Analysis of Complex Time-to-Event Data

Abstract

This dissertation work is motivated by two time-to-event data examples, where current statistical methods are inadequate in addressing the nuances of data and the corresponding scientific question(s) of interest.

In Chapter 1, we address a current issue regarding the analysis of time-varying biomarkers of Alzheimer's disease. Relating time-varying biomarkers of Alzheimer's disease (AD) to time-to-event using a Cox model is complicated by the fact that AD biomarkers are sparsely collected, typically only at study entry; this is problematic since Cox regression with time-varying covariates requires observation of the covariate process at all failure times. The analysis might be simplified by using study entry as the time origin and treating the time-varying covariate measured at study entry as a fixed baseline covariate. We first derive conditions under which using an incorrect time origin of study entry results in consistent estimation of regression parameters when the time-varying covariate is continuous and fully observed. We then derive conditions under which treating the time-varying covariate as fixed at study entry results in consistent estimation. We provide methods for estimating the regression parameter when a functional form can be assumed for the time-varying biomarker, which is measured only at study entry. We demonstrate our analytical results in a simulation study and apply our methods to data from the Rush Religious Orders Study and Memory and Aging Project, and data from the Alzheimer's Disease Neuroimaging Initiative.

In Chapter 2, we focus our attention on graft-versus-host disease (GVHD), a debilitating condition associated with significant morbidity, compromised quality of life and

mortality, that is a frequent complication of hematopoietic cell transplantation (HCT). For the most part, researchers investigating risk factors for acute GVHD, a sub-category that is diagnosed within 100 days of transplantation, have used standard survival analysis methods or logistic regression. Doing so, however, ignores two important clinical issues. First, patients who undergo HCT are at significant risk of death in the short-term; in our motivating data, from the Center for International Blood and Bone Marrow Transplant Research (CIBMTR), 100-day mortality among 9,651 patients who underwent HCT between 1999-2011 was 15%. Naïve treatment of death as a censoring mechanism (in either survival or logistic regression analyses), however, is problematic and can lead to erroneous conclusions. Second, acute GVHD is only diagnosed within 100 days of the transplant; beyond 100 days, a patient may be diagnosed with chronic GVHD for which treatment options/strategies generally differ. As such, in contrast to the typical assumption that the support for the event of interest is the positive part of the real line, patients who have undergone HCT are only eligible to experience the event of interest within a finite time interval. In this paper, building on the cure fraction and semi-competing risks literature, we propose a novel multi-state model that simultaneously: (i) accounts for mortality through joint modeling of acute GVHD and death, and (ii) explicitly acknowledges the finite time scale in which the event of interest can take place. The joint observed data likelihood is derived, with estimation and inference performed via maximum likelihood. The proposed framework is compared via comprehensive simulations to a number of alternative approaches that each acknowledge some but not all clinical aspects of acute GVHD. Finally, the methods are illustrated with an analysis of stem cell transplantation registry data from the CIBMTR.

In Chapter 3, we then consider risk prediction of both acute GVHD and death simultaneously. More generally, this work concerns joint risk prediction in the semi-competing risks setting. We propose to consider prediction through the calculation and evaluation of patient-specific *absolute risk profiles* for the acute GVHD and death. In particular, we note that at any given point in time after transplantation, a patient will: (1) have experienced both events; (2) be alive with a diagnosis of acute GVHD; (3) have died without acute GVHD; or (4) be alive without acute GVHD. Thus, in contrast to much of the

prediction literature, we consider the task of prediction as being one where we seek to classify patients at any given point in time into one of four categories based on a vector of probabilities that add to 1.0. We develop this framework utilizing the proposed model in Chapter 2. We then propose a framework for evaluation of predictive performance for risk profiles based on the *hypervolume under the manifold* (HUM) statistic, an extension of the well-known area-under-the-curve (AUC) statistic for univariate binary outcomes. As part of this, we propose a method for estimating the HUM statistic in the presence of potential verification bias which arises when the true outcome category is unknown. Throughout, we illustrate the proposed methods using data from the CIBMTR.

Contents

Acknowledgments

First, I would like to thank my advisors, Rebecca Betensky and Sebastien Haneuse, for your tireless efforts editing my drafts and endless patience. I have learned so much from you both beyond methods development.

I would also like to thank my committee members, Deborah Blacker and Dianne Finkelstein, for your helpful feedback and honest advice.

Finally, I would like to thank my husband, Mike, for your infinite support and love, putting your life on hold for five years so that I could advance my career, and being the best dad to our sweet boys, Max and Teddy.

Time-to-event data with time-varying biomarkers measured only at study entry, with applications to Alzheimer's disease

Catherine Lee

Department of Biostatistics

Harvard T.H. Chan School of Public Health

Rebecca A. Betensky

Department of Biostatistics

Harvard T.H. Chan School of Public Health

1.1 Introduction

Alzheimer's disease (AD) is the most common type of dementia in older adults, characterized by loss of cognition and functional impairment. Its prevalence is estimated to be 5.5 million in 2017 (Alzheimer's Association, 2017), with an estimated incidence of 476,000 in 2016 for people age 65 or older. Much of Alzheimer's research over the past decade has been dedicated to the identification and validation of biomarkers that reflect the underlying neuropathology of the disease. Biomarkers will help identify subjects early in their disease course, when therapeutic interventions might be successful. They also will help identify surrogate endpoints for progression, which will allow for clinical trials that do not require long follow-up to a clinical AD diagnosis.

Several models have been proposed for time-varying AD biomarkers such as PET-PIB, FDG-PET, and MRI measures (Jack et al., 2010; Capuano et al., 2016). Relating these biomarker trajectories to time-to-event using a Cox regression model would allow for the prediction of disease progression. However, such an analysis presents a challenge, as imaging biomarkers are typically sparsely collected. This is problematic since the Cox regression with time-varying covariates requires observation of the covariate process at all failure times. The analysis could be simplified by treating the time-varying covariate collected only at study entry as a fixed baseline covariate representative of a subject's disease state at the time of measurement. However, in longitudinal AD studies, the time at study entry tends to be arbitrary and the interpretability of such an analysis is tenuous (Cnaan and Ryan, 1989; Keiding and Knuiman, 1990). In a time-to-event analysis, a well-chosen time origin should align subjects with respect to risk, so that conditional on baseline covariates subjects are comparable (Cox and Oakes, 1984). In addition, the choice of time origin should address the goals of the analysis and provide meaningful estimates (Cnaan and Ryan, 1989; Fieberg and DelGiudice, 2009; Korn et al., 1997; Lamarca et al., 1998; Liestol and Andersen, 2002; Pencina et al., 2007; Sperrin and Buchan, 2013). A natural choice of time origin in Alzheimer's disease is pathological disease onset, a point at which individuals are comparable with respect to risk. However, disease onset precedes clinical symptoms by decades and thus the exact time of onset is unknown. As

an alternative, the time of first clinical diagnosis of mild cognitive impairment (MCI) is a reasonable and measurable time origin. Birth is a plausible time origin since disease risk increases with age, however individuals at a given age may not necessarily be comparable with respect to risk, even after conditioning on fixed covariates. In summary, while using study entry as the time origin would enable use of the time varying biomarker as a fixed covariate in the analysis, it would jeopardize the interpretability of the analysis.

Several authors have considered the use of an incorrect time origin of study entry in a proportional hazards model with a fixed-time covariate. Assuming the true time origin is birth, Korn et al. (1997) showed that fitting a proportional hazards model that uses study entry as the time origin does preserve the true regression coefficient corresponding to a fixed baseline covariate, provided that age at entry is included as a covariate in the model and the baseline hazard on the age scale follows an exponential form. Thiébaut and Bénichou (2004) and Pencina et al. (2007) investigated the practical ramifications of this result when the baseline hazard is not of exponential functional form (e.g., Weibull or piecewise Weibull), and found that the bias in the regression coefficient was not large for a wide range of true values.

In the case of time-varying covariates, some simulation studies have suggested an extension of the result from Korn et al. (1997), however this has not been established analytically. Cnaan and Ryan (1989) argued that using an incorrect time origin would preserve the regression parameter of interest provided that survival on the age scale is exponentially distributed. In a simulation study, Thiébaut and Bénichou (2004) found that use of study entry as the time origin results in unbiased estimates for a Cox model with constant baseline hazard function and a fully-observed time-varying binary covariate. Through simulations, Thiébaut and Bénichou (2004) and Griffin et al. (2012) illustrated that bias can result when an incorrect choice of time origin is used in a Cox model with time-varying covariates when the true baseline hazard is Weibull. There are no papers that consider estimation based on a time-varying covariate that is observed only at study entry, with the time origin taken either to be onset or study entry.

There is a considerable literature on dealing with sparse observation of time-varying covariates (Lin and Ying, 1993; Zhou and Pepe, 1995; Chen and Little, 1999; Wu et al.,

2012; Wang and Taylor, 2001; Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997) in survival regression models; standard methods include imputation, likelihood estimation, and joint modeling. These methods are not applicable to many current Alzheimer's studies due to the their very sparse observation of the time-varying imaging biomarker. Relevant to this setting, Sperrin and Buchan (2013) proposed a two-stage method for analyzing survival data with time-varying covariates measured at a single, arbitrary time point. They assume that the covariate process can be expressed as the sum of a population-level function and individual time-independent errors. The first stage involves estimating the population parameters of the covariate function and then calculating for each subject the residual error relative to the observed value. In the second stage, the estimated residual for each subject is included as a covariate in a survival model. They provided an analytical argument for the use of birth as the time origin, rather than study entry. They showed through simulations that their two-stage residual method has superior predictive ability, as measured by Brier and logarithmic scores, when birth is used as the origin and the estimated residuals and age are included in the model as covariates, compared to two competing models that treat the time-varying covariate as fixed: one from study entry that adjusts for age as a covariate and another from birth that does not adjust for age.

In this paper, we consider the analysis of time-to-event data in conjunction with a time-varying covariate that is observed only at study entry. We assume the true time origin for the risk model is known and precedes study entry. We extend the results of Korn et al. to the case of time-varying biomarkers, and derive the conditions for which treating a time-varying biomarker as fixed at study entry results in an unbiased regression coefficient estimator. We then extend the work of Sperrin and Buchan (2013), who proposed methods for relating a time-varying covariate with fixed-time residuals to time-to-event when the covariate is measured only at study entry, by incorporating time-varying residuals. Furthermore, whereas Sperrin and Buchan focused on predictive ability, we investigate bias and standard error estimation. We use the sigmoidal AD biomarker model (Jack et al., 2013) as the basis for this approach. We support our analytical results and methods with simulation studies. Finally, we apply our methods to data from the Rush Religious Orders Study and Memory and Aging Project and from the Alzheimer's Disease Neu-

roimaging Initiative (ADNI) and demonstrate the impact of different treatments of the time-varying biomarker.

1.2 Notation and models

For subject i, let U_i denote the time from disease onset to event, let C_i denote the time from disease onset to censoring, let E_i denote the time from disease onset to study entry, and let $Z_i(u)$ denote a time-varying covariate. Then $T_i = U_i - E_i$ represents the time from study entry to event. The observed time from onset to event is $Y_i = \min(U_i, C_i)$, and $\delta_i = I(Y_i \leq C_i)$ is the status indicator. We assume that the time-varying covariates are collected only at study entry. The observed data are thus $(E_i, Y_i, Z_i(E_i), \delta_i)$.

We assume that the time-varying covariate, or biomarker, $Z_i(u)$, after transformation via $f(\cdot)$, satisfies a general additive error model, $f(Z_i(u)) = f(S(u)) + \epsilon_i(u)$, where $S(u)$ is the population-level biomarker trajectory and $\epsilon_i(u)$ is an individual-level time-dependent deviation. We assume that the individual-level deviations have mean 0 and are defined by $\epsilon_i(u) = a_i + b_i u$, so that

$$f(Z_i(u)) = f(S(u)) + a_i + b_i u. \tag{1.1}$$

When f is the identity function, we refer to (1.1) as the *absolute error model* since the individual-level deviations are on the same scale as the population-level trajectory. When f is the identify function and $b_i \equiv 0$, the covariate model in (1.1) coincides with the covariate model provided by Sperrin and Buchan (2013). When $f(x) = \text{logit}(x/M)$, for some constant M, we refer to (1.1) as the *logit error model*.

In the setting of Alzheimer's disease, many population-level biomarker trajectories are plausibly modeled using a sigmoidal function (Jack et al., 2010)

$$S(u) = \frac{M}{1 + e^{-\lambda(u - T_0)}}, \tag{1.2}$$

where the upper bound M is assumed to be known, and the parameters λ and T_0, which dictate the shape of the sigmoid, are not known. When $T_0 = 0$, $S(0) = M/2$; the value of T_0 determines the lateral shift of this curve. The parameter λ determines the rate at which

the sigmoid accelerates to its maximum value with larger values indicating a faster rate of increase; at $u = T_0$ the slope of the sigmoid is $M\lambda/4$.

The covariate function $S(u)$ should be selected on the basis of prior studies and subject matter understanding. Scatter plots of the cross-sectional data may also give insight into to the functional form of the covariate process. Choice of the error function, $f(\cdot)$, also requires consideration. The absolute error model may not be appropriate for covariates that are constrained by a maximum measured value, such as PET amyloid, as it may lead to individual-specific trajectories that exceed the maximum value. The logit error model with time-varying deviations has an attractive interpretation when the population-level function is assumed to be the sigmoid in (1.2). The individual specific deviations modify the acceleration parameter, λ, and the shift parameter, T_0, so that each subject is allowed to have an individual-specific trajectory. In particular, inserting the sigmoidal function (1.2) into the logit error model (1.1) yields an individual specific sigmoidal covariate function, $Z_i(u) = M/[1 + e^{-\widetilde{\lambda}(u-\widetilde{T_0})}]$, where $\widetilde{\lambda_i} = \lambda + b_i$ and $\widetilde{T_{0,i}} = (\lambda T_0 - a_i)/(\lambda + b_i)$. The logit error model assumes that the time-varying covariate is linear on the logit scale, which may be appropriate only in some settings.

We assume that the true hazard function for the event of interest, with time measured from onset, is

$$\lambda_{U_i}(u|\overline{Z_i}(\cdot)) = \lambda_{0U}(u)e^{\beta g(Z_i(u))}, \tag{1.3}$$

for some function g, and baseline hazard function, $\lambda_{0U}(u)$. In this paper, we consider the case $g = f$. In general, g should be defined to best capture the nature of the biomarker trajectory and its relationship to the hazard of the event.

1.3 Implications of using study entry as the time origin for Cox models with time-varying predictors

When an appropriate time origin precedes study entry and the covariate of interest is not time-varying, Korn et al. (1997) provided sufficient conditions under which fitting a model that uses study entry as the time origin preserves the desired estimated regression coefficient corresponding to the fixed covariate. They assumed that while the true

survival model is given by

$$\lambda_{U_i}(u|x_i, E_i) = \lambda_{0U}(u)e^{\beta x_i},$$

where x_i is a fixed covariate, the data are fit instead to the model,

$$\lambda_{T_i}(t|x_i, E_i) = \lambda_{0T}(t)e^{\xi x_i + \gamma E_i}, \tag{1.4}$$

which uses study entry as the time origin. It is clear that $\xi = \beta$ provided that $\lambda_{0U}(u) = c\exp\{\gamma u\}$ and $\lambda_{0T}(t) = c\exp\{\gamma t\}$, for some positive constant c. Under these conditions, $\widehat{\xi}$ based on fitting the Cox model specified in (1.4) will be a consistent estimator of β.

We now extend the results of Korn et al. to the case of a continuous time-varying covariate, $w_i(u)$, and provide conditions under which using study entry as the time origin and/or treating a time-varying covariate as fixed will preserve the Cox regression parameter of interest. Suppose that the true survival model is

$$\lambda_{U_i}(u|\overline{w_i}(\cdot), E_i) = \lambda_{0U}(u)e^{\beta w_i(u)}. \tag{1.5}$$

We assume further that E and U are continuous random variables, and that the time-varying biomarker $w(u)$ is continuous and defined on the support of U. Without loss of generality, we assume that the support of E is contained in the support of U. In addition, we assume that $\beta \neq 0$, and the study entry times, E_i, and biomarkers, $w_i(u)$, are not constant across individuals.

We consider three possible models that might be fit to these data. First, we consider the model that uses study entry as the time origin and adjusts for study entry time as a covariate

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) = \lambda_{0T}(t)e^{\xi w_i(t+E_i) + \gamma E_i}. \tag{1.6}$$

In Theorem 1 part (a) below, we derive conditions on the true baseline hazard function under which the coefficient of $w_i(t + E_i)$ is equal to β. Assuming only that $\lambda_{0U}(u) = ce^{\gamma u}$, the Cox partial likelihood derived from (1.6) can be used for estimation of β. Note that this extends the fixed covariate case (Korn et al.).

Next we consider models that treat the time-varying predictor as fixed, using either onset or study entry as the origin, and derive conditions under which we obtain a consistent estimator of β. If the value of the time-varying covariate is known only at study

entry, it can be included in a Cox model as a fixed covariate in a second model with onset as the time origin:

$$\lambda_{U_i}(u|\overline{w_i}(\cdot), E_i) \;=\; \lambda_{1U}(u)e^{\alpha w_i(E_i)}. \tag{1.7}$$

Note that estimation of this model must adjust for delayed entry (left truncation). If time from onset to study entry is known, we can include E_i as a covariate in a third model using study entry as the time origin:

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) \;=\; \lambda_{2T}(t)e^{\zeta w_i(E_i)+\eta E_i}. \tag{1.8}$$

The case where there is no covariate adjustment for study entry time is included in this model when $\eta = 0$. Alternatively, the time of study entry could be adjusted for through stratification or through covariate adjustment by a non-linear function of E_i, which we do not consider here. In Theorem 1 parts (b) and (c), we derive conditions under which the coefficients of $w_i(E_i)$ in (1.7) and (1.8) are equal to β.

Theorem 1. *Suppose the true hazard regression model is given by (1.5).*

(a) (Korn et al. extension) Consider the model in (1.6): $\beta = \xi \iff \lambda_{0U}(u) = ce^{\gamma u}$ and $\lambda_{0T}(t) = ce^{\gamma t}$, for some positive constant c. In practice, if $\lambda_{0U}(u) = ce^{\gamma u}$, (1.6) can be used to estimate β without specifying that $\lambda_{0T}(t) = ce^{\gamma t}$.

(b) Consider the model in (1.7): $\beta = \alpha \iff w_i(u) = w_i(E_i) + g(u)$, for some continuous function g that is zero on the support of E. In practice, if $w_i(u) = w_i(E_i) + g(u)$, (1.7) can be used to estimate β in an analysis that adjusts for left-truncation.

(c) Consider the model in (1.8): $\beta = \zeta \iff w_i(t + E_i) - w_i(E_i) = k(t)$, $\lambda_{0U}(u) = ce^{\eta u}$ and $\lambda_{2T}(t) = ce^{\beta k(t)+\eta t}$, for $c > 0$. In practice, under these conditions on λ_{0U} and $w_i(t + E_i) - w_i(t)$, (1.8) can be used to estimate β even if $\lambda_{2T}(t) \neq ce^{\beta k(t)+\eta t}$.

The proof for Theorem 1 is given in the Appendix. Theorem 1 provides guidelines for implementing Cox regression with a time-varying covariate using study entry as the origin or when the time-varying covariate is measured only at study entry. Theorem 1

part (a) shows that study entry can be used as the origin if the covariate process is fully-observed or if its functional form is known, as long as the baseline hazard is of exponential functional form. These conditions are reasonable in some settings: in many instances, a covariate process can be modeled using subject-level knowledge (Fisher and Lin, 1999), and constant hazard functions have been empirically shown to model incidence in many epidemiological studies (Ingram et al., 1997), and exponential-form (Gompertz) baseline hazards are widely used to model mortality.

If the time-varying covariate is observed only at study entry and its functional form is known, then the covariate process can be treated as fixed in an analysis from onset, as long as the covariate function, $w_i(u)$, satisfies the condition in Theorem 1 part (b). An example of such a function is $w_i(u) = a_i + m \cdot (h(u) - h(L))1\{u > L\}$, where L is the upper bound of the support of E and h is continuous; this is a time-varying continuous covariate function that is constant on the support of E but can vary for values of u outside of the support of E. Since we are mainly concerned with $w(u)$ on the support of E, Theorem 1 part (b) indicates that treating a time-varying (i.e., non-constant) biomarker as fixed using onset as the origin is not valid.

Theorem 1 part (c) shows that if the time-varying covariate is observed only at study entry, an analysis using study entry as the origin is valid only for special cases of the baseline hazard and covariate process. As discussed above, the baseline hazard assumptions are realistic in some settings. One example of a time-varying covariate that satisfies the condition in Theorem 1 part(c) is $w_i(t) = a_i + a_1 t$, a linear function with a subject-specific intercept. Thus, in some instances it is reasonable to treat time-varying covariates as fixed in an analysis from study entry.

The focus of this paper is the impact of the time-varying covariate $w_i(u) = f(S(u)) + a_i + b_i u$, as defined in (1.1), on the hazard function through the Cox model defined in (1.5). Theorem 1 part (a) allows for valid estimation of β using study entry as the time origin provided that f, a_i and b_i are known, E_i is included in the model as a covariate, and $\lambda_{0U}(u) = c\exp\{\gamma u\}$. If the covariate process is known only at study entry, then it is not valid to treat the time-varying covariate as fixed in an analysis from onset because $w_i(u) - w_i(E_i) = f(S(u)) - f(S(E_i)) + b_i(u - E_i)$ is a function that depends on i and violates

the condition of part (b) of the Theorem.

The implications of Theorem 1 part (c) differ for the two examples of f introduced in Section 1.2. First consider the absolute error model with $w_i(u) = S(u) + a_i + b_i u$, where $S(u)$ is the sigmoid in (1.2). Theorem 1 part (c) establishes that using study entry as the origin and treating the time-varying covariate as fixed will yield biased estimates of β. To see this, Theorem 1 part (c) requires that $w_i(t + E_i) - w_i(E_i)$ is independent of i, however in this case $w_i(t + E_i) - w_i(E_i) = S(t + E_i) - S(E_i) + b_i t$ is not independent of i. This is the case even if $b_i \equiv 0$. Alternatively, for the logit error model with $w_i(u) = \text{logit}[S(u)/M] + a_i + b_i u$, if $b_i \equiv b$, using study entry as the origin and treating the time-varying biomarker as fixed is valid, provided that the baseline hazard is of an exponential form, since $w_i(t + E_i) - w_i(E_i) = (\lambda + b)t$ is independent of i. If $b_i \neq b$, this condition is not satisfied.

Note that the results of Theorem 1, parts (a) and (c), apply even when $\beta = 0$. In contrast, Theorem 1 part (b) does not apply to this case. If $\beta = 0$, then the true survival model is

$$\lambda_{U_i}(u|\overline{w_i}(\cdot), E_i) = \lambda_{0U}(u).$$

If we fit the model in (1.7), it is easy to see that $\alpha = 0 \iff \lambda_{0U}(u) = \lambda_{1U}(u)$. In practice, this implies that fitting (1.7) when $\beta = 0$ will yield $\widehat{\alpha} = 0$ without restrictions on $w_i(u)$.

1.4 Methods

Based on our analysis in Section 1.3, it is not advisable to treat time-varying biomarkers measured only at study entry as fixed covariates for the time origin of disease onset. And for the time origin of study entry, this can be done only under strong assumptions on the baseline hazard function and the covariate function. Thus, lacking full observation of the biomarkers over time, we require an approach that will allow us to estimate the full biomarker trajectory based on cross-sectional observation of it, and importantly, does not require restrictive assumptions on the baseline hazard function of the functional form of the covariate process.

Assuming that the time-varying biomarker follows the general error model in (1.1)

and the true hazard can be modeled as in (1.3) for $g = f$, we provide methods for estimating the regression parameter β when the fully-observed time-varying biomarker is measured only at study entry. We note that the components of the error, a_i and b_i, cannot be simultaneously estimated when only a single biomarker measurement is available for each subject. We thus consider the cases $a_i \equiv 0$ and $b_i \equiv 0$ separately.

1.4.1 Case: $b_i \equiv 0$

The covariate process in this case is defined as $f(Z_i(u)) = f(S(u; \lambda, T_0)) + a_i$, with individual-level deviations that are fixed over time. We assume that $E(a_i) = 0$. We assume that the value of M, which represents the maximum value of the sigmoidal predictor in (1.2), is known. In order to estimate the regression parameter β, we first estimate the parameters λ and T_0 of the population-level sigmoid $S(u; \lambda, T_0)$ using least squares to minimize

$$\sum_i [f(Z_i(E_i)) - f(S(E_i; \lambda, T_0))]^2$$

with respect to λ and T_0. We then estimate the a_i as $\widehat{a_i} = f(Z_i(E_i)) - f(S(E_i; \widehat{\lambda}, \widehat{T_0}))$, and the time-varying covariate process as $\widehat{f(Z_i(u))} = f(S(u; \widehat{\lambda}, \widehat{T_0})) + \widehat{a_i}$. The true hazard function $\lambda_{U_i}(u|f(\overline{Z_i}(\cdot)))$ can then be approximated as

$$\lambda_{U_i}(u|\widehat{f(Z_i(\cdot))}) = \lambda_{0U}(u) \exp\{\beta \cdot \widehat{f(Z_i(u))}\} = \lambda_{0U}(u) \exp\{\beta \cdot \Big[f(S(u; \widehat{\lambda}, \widehat{T_0})) + \widehat{a_i}\Big]\} = \widetilde{\lambda_{0U}}(u) \exp\{\beta \cdot \widehat{a_i}\}. \tag{1.9}$$

The model in (1.9) shows that the desired regression parameter β can be estimated by fitting a Cox model with $\widehat{a_i}$ included as a fixed-covariate.

Note that if $f(Z_i(u))$ is linear in u, as is the case with the sigmoid-based logit error model, where

$$f(Z_i(u)) = \text{logit}[Z_i(u)/M] = \lambda(u - T_0) + a_i,$$

then the model based on $f(\widehat{Z_i(t + E_i)}))$ is equal to that based on $f(Z_i(E_i))$. In this case,

$$f(\widehat{Z_i(t + E_i)}) = \widehat{\lambda}(t + E_i - \widehat{T_0}) + \widehat{a_i} = \widehat{\lambda} t + \widehat{f(Z_i(E_i))} = \widehat{\lambda} t + f(Z_i(E_i)),$$

by the definition of $\widehat{a_i}$. Thus, fitting a Cox model from study entry with $\widehat{f(Z_i(t + E_i))}$ as a time-varying covariate is equivalent to fitting the model from entry with the fixed

covariate $f(Z_i(E_i))$, since the population-level $\widehat{\lambda}t$ term will be absorbed into the baseline hazard function and will not affect estimation.

1.4.2 Case: $a_i \equiv 0$

The covariate process in this case is defined as $f(Z_i(u)) = f(S(u); \lambda, T_0) + b_i u$, with time-varying individual-level deviations. At each E_i it follows that

$$\frac{f(Z_i(E_i))}{E_i} = \frac{f(S(E_i; \lambda, T_0))}{E_i} + b_i, \tag{1.10}$$

and thus λ and T_0 can be estimated via least squares by minimizing

$$\sum_i \left[\frac{f(Z_i(E_i))}{E_i} - \frac{f(S(E_i; \lambda, T_0))}{E_i} \right]^2$$

with respect to λ and T_0. We then estimate b_i as

$$\widehat{b_i} = \frac{f(Z_i(E_i))}{E_i} - \frac{f(S(E_i; \widehat{\lambda}, \widehat{T_0}))}{E_i}.$$

Note that estimation of b_i uses only those entry times E_i that are nonzero; in some settings it is possible that a time origin, such as disease onset, may coincide with study entry for some subjects. We estimate the time-varying covariate process as $\widehat{f(Z_i(u))} = f(S(u; \widehat{\lambda}, \widehat{T_0})) + \widehat{b_i} u$. The true hazard function $\lambda_{U_i}(u | f(\overline{Z_i}(\cdot)))$ is approximated as

$$\begin{aligned} \lambda_{U_i}(u | \widehat{f(Z_i(\cdot))}) &= \lambda_{0U}(u) \exp\{\beta \cdot \widehat{f(Z_i(u))}\} \\ &= \lambda_{0U}(u) \exp\left\{\beta \cdot \left[f(S(u; \widehat{\lambda}, \widehat{T_0})) + \widehat{b_i} u \right]\right\} \\ &= \widetilde{\lambda_{0U}}(u) \exp\{\beta \cdot \widehat{b_i} u\}. \end{aligned} \tag{1.11}$$

The regression parameter β can be estimated using Cox regression with the time-varying covariate $\widehat{b_i} u$.

1.4.3 Adjusting sigmoid for disease severity

In the AD setting, individual sigmoidal trajectories might be explained by individual-level covariates (e.g., APOE4 allele status, gender) and not just simply by unexplained individual deviations. We propose to incorporate this into the sigmoidal model by allowing

the lateral shift of the population-level sigmoid, given by T_0, to be a function of covariates. Let w_i denote a vector of individual-specific covariates reflecting disease severity. We model the individual-specific lateral shifts as $T_{0,i} = \gamma' \mathbf{w}_i$, for a vector of coefficients γ. Then the sigmoid is redefined as

$$S(u, \mathbf{w}_i; \lambda, \gamma) = \frac{M}{1 + e^{-\lambda(u - \gamma' \mathbf{w}_i)}},$$

additional individual deviations are modeled as $f(Z_i(u, \mathbf{w}_i)) = f(S(u, \mathbf{w}_i; \lambda, \gamma)) + a_i + b_i u$, and the true hazard model is assumed to be

$$\lambda_{U_i}(u | f(\overline{Z_i}(\cdot))) = \lambda_{0U}(u) \exp\{\beta f(Z_i(u, \mathbf{w}_i; \lambda, \gamma))\}. \tag{1.12}$$

This can be implemented for any f. In the case of the logit error model (1.1),

$$\begin{aligned} \text{logit}\left[\frac{Z_i(u, \mathbf{w}_i)}{M}\right] &= \text{logit}\left[\frac{S(u, \mathbf{w}_i); \lambda, \gamma}{M}\right] + a_i + b_i u \\ &= \lambda(u - \gamma' \mathbf{w}_i) + a_i + b_i u, \end{aligned}$$

so that the individual-specific lateral shifts are given by $\gamma' \mathbf{w}_i$. The model in (1.12) then simplifies as:

$$\begin{aligned} \lambda_{U_i}(u | \text{logit}(\overline{Z_i}(\cdot)/M)) &= \lambda_{0U}(u) \exp\{\beta \left[\text{logit}(Z_i(u, \mathbf{w}_i; \lambda, \gamma)/M)\right]\} \\ &= \lambda_{0U}(u) \exp\{\beta \left[\lambda(u - \gamma' \mathbf{w}_i) + a_i + b_i u\right]\} \\ &= \widetilde{\lambda_{0U}}(u) \exp\{\beta[a_i + b_i u - \lambda \gamma' \mathbf{w}_i]\}. \end{aligned}$$

As discussed above, both error components a_i and b_i, cannot be estimated simultaneously. The estimation procedure is analogous to that described in Sections 1.4.1 and 1.4.2, and differs only with regard to the parameters of disease severity adjusted population-level sigmoid; least squares is used to estimate λ and γ. For example, for the case $b_i \equiv 0$, we use least squares to minimize

$$\sum_i \left[f(Z_i(E_i)) - f(S(E_i; \lambda, \gamma))\right]^2$$

with respect to λ and γ.

1.5 Simulations

We simulated left-truncated time-to-event data such that the time from onset to event, U_i, satisfies the hazard model in (1.3) and the transformed time-varying predictor, $f(Z_i(u))$, follows the logit error model in (1.1) with sigmoidal S in (1.2), upper bound M, and $f(x) = \text{logit}(x/M)$. Study entry times, E_i, were generated independently of the survival times, U_i. The time-varying covariates were available only from the times of study entry. The data, $(E_i, U_i, f(Z_i(E_i)))$, were generated to achieve specified truncation probabilities, denoted $\pi_{\text{trunc}} = P(U_i < E_i)$. Parameters of the covariate process were chosen to reflect PIB PET measurements of the AD biomarker, Aβ; hazard regression parameters were chosen so that time-to-AD diagnosis values align with those observed in AD studies. The data were generated in the following steps:

1. Select $\beta \in \{0, 0.25, 0.5, 1, 1.5\}$, $n \in \{100, 250\}$, $M = 3$, $\lambda \in \{0.1, 0.25, 0.4\}$, $T_0 = 40$, $\sigma_a \in \{0.5, 1\}$, $\sigma_b \in \{0.01, 0.05\}$.

2. Generate error components, $a_i \sim N(0, \sigma_a^2)$ and set $b_i = 0$ or generate $b_i \sim N(0, \sigma_b^2)$ and set $a_i = 0$.

3. Generate $V_i \sim \text{Unif}(0, 1)$ and solve for U_i in

$$\begin{aligned} V_i &= \exp\left\{ -\int_0^{U_i} \lambda_{U_i}(w)\, dw \right\} \\ &= \exp\left\{ -\int_0^{U_i} \lambda_{0U}(w) \exp\{\beta \cdot [f(S(w)) + a_i + b_i w]\}\, dw \right\}. \end{aligned} \tag{1.13}$$

   This can be easily done when the integral in (1.13) can be calculated in closed form, as discussed in Bender et al. (2005). This is the case under the logit error model (1.1) where $f(S(u)) = \lambda(u - T_0) + a_i + b_i u$. We considered three separate baseline hazard functions, $\lambda_{0U}(u)$, defined below: constant, exponential-form, normal density kernel. The constant and exponential-form baseline hazard functions appear in Theorem 1 and were chosen to confirm our theoretical results. The normal density kernel baseline hazard function was chosen because it permits a closed form

solution for U_i and violates the requirements of the theorem, and thus also confirms our theoretical results.

Baseline hazard functions:

i. (Constant baseline hazard) $\lambda_{0U}(u) = c$, where $c = 0.0154, 0.01, 0.004, 0.001, 0.001$ corresponding to $\beta = 0, 0.25, 0.5, 1, 1.5$,

ii. (Exponential-form baseline hazard) $\lambda_{0U}(u) = ce^{\psi u}$, where $c > 0$ and $\psi \neq 0$ with $c = 7 \times 10^{-8}$ and $\psi = 0.2$,

iii. (Normal density kernel baseline hazard) $\lambda_{0U}(u) = m \exp\{-c(u-a)^2\}$, where $(c, a, m) = (0.001, 60, 1)$ for the fixed-deviations case ($b_i \equiv 0$) and $(c, a, m) = (0.1, 70, 100)$ for the time-varying deviations case ($a_i \equiv 0$). The hazard function can be expressed as follows,

$$\lambda_{U_i}(u) = m \exp\left\{-c(u-a)^2 + \beta[\lambda(u - T_0) + a_i + b_i u]\right\} = \phi(u; \mu_i, \sigma^2) \cdot K_i,$$

where $\phi(u; \mu_i, \sigma^2)$ is the normal density with mean $\mu_i = a + \beta(\lambda + b_i)/(2c)$ and variance $\sigma^2 = 1/(2c)$, and

$$K_i = m \exp\left\{\beta\left[a(\lambda + b_i) + a_i + \frac{\beta(\lambda + b_i)^2}{4c} - \lambda T_0\right]\right\}\sqrt{2\pi\sigma^2}.$$

4. Generate study entry times $E_i \sim \text{Uniform}(0, R_{\text{trunc}})$, where R_{trunc} is chosen to achieve $P(U_i < E_i) = \pi_{\text{trunc}} = 0.2, 0.4$, and is a function of the parameters of the distribution of U. A table of R_{trunc} used can be found in Table 4.1 of the Supplemental Material section.

5. Retain the first n observations that satisfy $E_i < U_i$, i.e. incorporate delayed entry to this study (left-truncation).

6. Calculate $f(Z_i(E_i)) = f(S(E_i)) + a_i + b_i E_i = \lambda(E_i - T_0) + a_i + b_i E_i$, where either a_i or b_i is zero.

7. Estimate λ and T_0 using least squares as in Sections 1.4.1 and 1.4.2.

8. Estimate a_i or b_i using $\widehat{\lambda}$ and $\widehat{T}_0$ as in Sections 1.4.1 and 1.4.2.

9. Define $\widehat{f(Z_i(u))} = \widehat{\lambda}(u - \widehat{T}_0) + \widehat{a}_i$ or $\widehat{f(Z_i(u))} = \widehat{\lambda}(u - \widehat{T}_0) + \widehat{b}_i u$, as appropriate.

10. Fit the following models, adjusting for left truncation when the origin is onset:

$$\lambda_{U_i}(u|f(\overline{Z}_i(\cdot)), E_i) = \lambda_{0U}(u)e^{\beta_1 f(Z_i(u))} \tag{A}$$

$$\lambda_{U_i}(u|f(\overline{Z}_i(\cdot)), E_i) = \lambda_{0U}(u)e^{\beta_2 \widehat{f(Z_i(u))}} \tag{A'}$$

$$\lambda_{U_i}(u|f(Z_i(E_i)), E_i) = \lambda_{1U}(u)e^{\alpha f(Z_i(E_i))} \tag{B}$$

$$\lambda_{T_i}(t|f(Z_i(E_i)), E_i) = \lambda_{1T}(t)e^{\zeta_1 f(Z_i(E_i))} \tag{C}$$

$$\lambda_{T_i}(t|f(Z_i(E_i)), E_i) = \lambda_{2T}(t)e^{\zeta_2 f(Z_i(E_i)) + \eta E_i} \tag{D}$$

$$\lambda_{T_i}(t|f(\overline{Z}_i(\cdot)), E_i) = \lambda_{3T}(t)e^{\xi_1 \widehat{f(Z_i}(t+E_i))} \tag{E}$$

$$\lambda_{T_i}(t|f(\overline{Z}_i(\cdot)), E_i) = \lambda_{4T}(t)e^{\xi_2 \widehat{f(Z_i}(t+E_i)) + \gamma E_i} \tag{F}$$

Model (A) is the true survival model, where $f(Z_i(u)) = \lambda(u - T_0) + a_i + b_i u$. Model (A′) corresponds to the proposed method in Section 3.3 where $\widehat{f(Z_i(u))}$ is defined in Step 9. Model (B) corresponds to Theorem 1 part (b), and models (C) and (D) correspond to Theorem 1 part (c), where $f(Z_i(E_i))$ is from Step 3. Models (E) and (F) correspond to Theorem 1 part (a), where $\widehat{f(Z_i(t + E_i))} = \widehat{\lambda}(t + E_i - \widehat{T}_0) + \widehat{a}_i + \widehat{b}_i(t + E_i)$. Given the linear form of $f(Z_i(u))$ with $b_i \equiv 0$, estimation via Models E and F is equivalent to that via Models C and D, respectively.

11. Repeat 5000 times.

The statistical package `R` (R Core Team, 2014) version 3.1.2 was used for all data analyses. Nonlinear least squares minimization was executed using the `nls` function in the `stats` package (Brazzale, 2005) with convergence parameters set to `maxiter=1e4` and `minFactor=1e-10` and with the default tolerance `tol = 1e-5`. Hazard regression was implemented using the function `coxph` in the `survival` package (Therneau, 2015). Cox models with time-varying predictors were fit using the time-transform option of `coxph` or with the use of counting process style input. R code is available in the Supporting Information.

The simulation results are given in Table 1.1 for varying values of β, $\lambda = 0.25$, $\sigma_a = 0.5$ and $\sigma_b = 0.05$, and $\pi_{\text{trunc}} = 0.2$ for the fixed (a_i) and time-varying (b_i) individual-level deviations models, respectively. Means and standard deviations of simulated estimates are presented. Type I error and power were taken to be the proportion of simulations which rejected the test statistic with probability less than 0.05 for $\beta = 0$ and $\beta > 0$, respectively. Type I error and power corresponding to $\beta = 0.5$ are given in Tables 4.2 and 4.3 in the Supplemental Material section, respectively, for the parameter values used in Table 1.1. For the case of fixed deviations (a_i) and a normal density kernel baseline hazard, the parameters originally chosen for the baseline hazard function to align with the AD setting result in a baseline hazard function that closely resembles an exponential, and thus do not serve the purpose of illustrating results under deviations from the exponential. Thus, the parameters of the normal density kernel baseline hazard were then chosen to illustrate this contrast, $(c, a, m) = (0.05, 30, 1)$, and the results for this parameterization of the normal kernel density baseline hazard function in the fixed deviations case are given in Table 1.1.

The simulation results in column one of Table 1.1 confirm that fitting the true survival model, in (A), gives consistent estimates of β. The second column demonstrates that the proposed method described in Section 3.3, which corresponds to model (A′), performs nearly as well as fitting the true model in column one for all baseline hazard distributions considered (constant, exponential-form, normal density kernel), with slightly higher standard errors, due to estimation of the covariate process.

When the covariate process follows the general error model in (1.1), Theorem 1 part (b) guarantees that treating a time-varying biomarker as fixed at its value at study entry will always produce biased estimates of β if the true origin is used, for nonzero values of β. This analysis corresponds to fitting model (B) and the results are listed in column three of Table 1.1 for $\beta \neq 0$ for both fixed and time-varying deviations and for all three baseline hazards considered, and are as expected. When $\beta = 0$, we expect the estimated regression coefficient to be zero (see discussion following Theorem 1); this is seen in the table, as well.

Theorem 1 part (c) provides conditions for when using an incorrect origin of study

entry and treating the time-varying predictors as fixed results in consistent estimates of β. For the fixed deviation models (i.e., a_i), the discussion following Theorem 1 ensures that inference is valid if study entry is used as the time origin and the time-varying biomarker is treated as fixed (with or without covariate adjustment for study entry time), as in models (C) and (D), provided that the baseline hazard is constant. Similarly, if the baseline hazard follows an exponential form and both the value of the time-varying biomarker at study entry and study entry time are included in the model as covariates, as in (D), estimates of β are unbiased. For the time-varying deviations models (b_i), Theorem 1 (c) guarantees that estimates of β will be biased if study entry is used as the time origin and the time-varying biomarker is treated as a fixed covariate (with or without covariate adjustment for study entry time), as in (C) and (D). These theoretical findings are illustrated through our simulation results in columns four and five.

Table 1.1: Bias and analytical standard errors of Cox parameter estimates for the logit error model for specified covariate model and baseline hazard function using $n = 250$ and 5000 iterations

		Onset origin			Study entry origin			
Covariate model	β	A: $f(Z_i(u))$	A′: $f(\widehat{Z_i(u)})$	B: $f(Z_i(E_i))$	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $f(\widehat{Z_i(t+E_i)})$	F: $f(\widehat{Z_i(t+E_i)}) + E_i$
Fixed deviations, a_i								
Normal density based hazard	0	0.001 (0.131)	0.001 (0.132)	0 (0.034)	1.306 (0.108)	0.002 (0.169)	1.306 (0.108)	0.002 (0.169)
	0.25	0.003 (0.132)	0.003 (0.132)	-0.234 (0.033)	1.188 (0.104)	0.059 (0.168)	1.188 (0.104)	0.059 (0.168)
	0.5	0.004 (0.135)	0.004 (0.135)	-0.471 (0.032)	1.033 (0.117)	0.121 (0.174)	1.033 (0.117)	0.121 (0.174)
	1	0.006 (0.144)	0.002 (0.144)	-0.951 (0.031)	0.770 (0.158)	0.191 (0.174)	0.770 (0.158)	0.191 (0.174)
	1.5	0.008 (0.158)	-0.005 (0.157)	-1.438 (0.030)	0.436 (0.260)	0.162 (0.176)	0.436 (0.260)	0.162 (0.176)
Constant hazard	0	-0.002 (0.129)	-0.002 (0.129)	0 (0.031)	0 (0.030)	-0.002 (0.130)	0 (0.030)	-0.002 (0.130)
	0.25	0.002 (0.131)	0.002 (0.132)	-0.248 (0.014)	0.001 (0.019)	0.001 (0.132)	0.001 (0.019)	0.001 (0.132)
	0.5	0.005 (0.135)	0.003 (0.135)	-0.495 (0.014)	0.002 (0.030)	0.004 (0.136)	0.002 (0.030)	0.004 (0.136)
	1	0.008 (0.145)	0.003 (0.145)	-0.990 (0.015)	0.004 (0.055)	0.009 (0.146)	0.004 (0.055)	0.009 (0.146)
	1.5	0.009 (0.156)	-0.003 (0.157)	-1.484 (0.016)	0.006 (0.082)	0.010 (0.157)	0.006 (0.082)	0.010 (0.157)
Exponential-form hazard	0	0 (0.131)	0.001 (0.131)	0 (0.012)	0.729 (0.041)	0 (0.135)	0.729 (0.041)	0 (0.135)
	0.25	0.002 (0.131)	0.003 (0.131)	-0.248 (0.014)	0.708 (0.053)	0.002 (0.135)	0.708 (0.053)	0.002 (0.135)
	0.5	0.005 (0.135)	0.004 (0.135)	-0.495 (0.014)	0.688 (0.065)	0.004 (0.138)	0.688 (0.065)	0.004 (0.138)
	1	0.008 (0.145)	0.006 (0.145)	-0.989 (0.016)	0.652 (0.089)	0.009 (0.148)	0.652 (0.089)	0.009 (0.148)
	1.5	0.009 (0.155)	0.003 (0.156)	-1.483 (0.017)	0.615 (0.112)	0.011 (0.157)	0.615 (0.112)	0.011 (0.157)
Time-varying deviations, b_i								
Normal density based hazard	0	0 (0.021)	0 (0.021)	-0.001 (0.013)	0.556 (0.052)	0 (0.044)	0.289 (0.02)	0 (0.026)
	0.25	0.002 (0.025)	0.001 (0.026)	-0.206 (0.014)	0.360 (0.057)	0.122 (0.049)	0.063 (0.021)	0.014 (0.029)
	0.5	0.003 (0.034)	0 (0.035)	-0.434 (0.014)	0.156 (0.062)	0.133 (0.061)	-0.166 (0.022)	0.021 (0.040)
	1	0.005 (0.056)	-0.022 (0.078)	-0.917 (0.016)	-0.267 (0.071)	-0.118 (0.087)	-0.632 (0.023)	-0.011 (0.069)
	1.5	0.006 (0.083)	-0.065 (0.150)	-1.411 (0.016)	-0.696 (0.078)	-0.523 (0.108)	-1.102 (0.024)	-0.165 (0.133)
Constant hazard	0	0 (0.015)	0 (0.015)	0 (0.029)	0.001 (0.028)	-0.001 (0.081)	0 (0.013)	0 (0.015)
	0.25	0.001 (0.025)	0 (0.025)	-0.204 (0.013)	-0.014 (0.019)	0.155 (0.046)	0.001 (0.018)	0.001 (0.023)
	0.5	0.002 (0.035)	-0.002 (0.036)	-0.436 (0.014)	-0.113 (0.028)	0.244 (0.061)	-0.001 (0.029)	0.001 (0.032)
	1	0.004 (0.057)	-0.021 (0.078)	-0.921 (0.015)	-0.474 (0.044)	0.222 (0.083)	-0.010 (0.055)	-0.006 (0.057)
	1.5	0.007 (0.083)	-0.052 (0.137)	-1.410 (0.017)	-0.890 (0.055)	0.029 (0.104)	-0.026 (0.087)	-0.019 (0.087)
Exponential-form hazard	0	0 (0.018)	0 (0.018)	0 (0.012)	0.372 (0.031)	0 (0.032)	0.221 (0.017)	0 (0.019)
	0.25	0.001 (0.025)	0.001 (0.025)	-0.206 (0.013)	0.364 (0.047)	0.109 (0.042)	0.103 (0.022)	0.001 (0.024)
	0.5	0.002 (0.035)	-0.001 (0.036)	-0.433 (0.015)	0.348 (0.060)	0.152 (0.056)	-0.012 (0.027)	0.001 (0.034)
	1	0.006 (0.057)	-0.016 (0.073)	-0.913 (0.016)	0.181 (0.067)	0.088 (0.080)	-0.237 (0.040)	-0.001 (0.056)
	1.5	0.007 (0.080)	-0.049 (0.132)	-1.404 (0.018)	-0.195 (0.078)	-0.117 (0.099)	-0.464 (0.055)	-0.011 (0.085)

Theorem 1 part (a) corresponds to models (E) and (F). It asserts that for a constant baseline hazard, the model that uses study entry as the origin and includes $f(Z_i(t+E_i))$ as a time-varying covariate, with or without covariate adjustment for study entry time, will give consistent estimates of β for fixed or time-varying deviations. For a baseline hazard that is of an exponential form, fitting the model that uses study entry as the origin, including both $f(Z_i(t+E_i))$ and study entry time as covariates, will give consistent estimates of β. These results are supported through our simulations when $f(\widehat{Z_i(t+E_i)})$ is used to approximate $f(Z_i(t+E_i))$. When the baseline hazard does not follow an exponential form, Theorem 1 part (a) guarantees that estimates of β will be biased if a model is fit using study entry as the origin, including both $f(Z_i(t+E_i))$ and study entry time as covariates. For the case of the normal density kernel baseline hazard and fixed deviations, this is confirmed through our simulation results, when $f(\widehat{Z_i(t+E_i)})$ is used to approximate $f(Z_i(t+E_i))$, with greater bias corresponding to larger values of β. For the case of the normal density kernel baseline hazard and time-varying deviations, our simulation results indicate that the bias increases with β.

For the models with fixed-deviations (a_i), notice that fitting models (C) and (E) give identical results. The same is true for models (D) and (F). As noted at the end of Section 1.4.1, this is expected when the form of the covariate model is linear in u, as is the case with the sigmoid-based logit error model.

Table 4.2 shows that the type I error is within 0.06 for all analyses that use onset as the time origin. For the analyses that use study entry as the origin, the type I error is less than 0.06 for baseline hazard functions that are constant, or follow an exponential form, provided that study entry time is included as a covariate (models D, E); the type I error is inflated for all other analyses from study entry. This follows from Theorem 1 parts (a) and (c), which hold for $\beta = 0$; specifically, when $\beta = 0$ and the conditions of Theorem 1 parts (a) and (c) are not met, then the estimated regression coefficients are expected to be nonzero. Table 4.3 shows that the null hypothesis test is well powered, under the simulation parameters specified in Table 1.1, for all analyses except for when onset is used as the time origin and the time-varying covariate is treated as fixed (model B). In this case, the population-level covariate function is absorbed by the baseline hazard

function and the impact of the covariate is due to the individual-specific deviations, as in (1.9); it is likely that $\beta \cdot a_i$ is close to zero for all subjects so that the likelihood of rejecting the null is also small.

In additional simulations (not shown), variations in the truncation probability, π_{trunc}, had little effect on bias, type I error and power. Larger values of σ_a and σ_b, the standard deviations of a_i and b_i, were associated with smaller standard errors and greater power, but had little effect on type I error. Varying λ, the acceleration parameter of the sigmoid $S(u)$, gave comparable results with respect to bias, type I error and power, however greater values of β were required to see bias in the case of a normal density based baseline hazard and a covariate process with time-varying deviations (b_i). Reducing the sample size to $n = 100$ resulted in larger standard errors and less power, but had little effect on type I error and bias, except for the case of a normal density based baseline hazard and time-varying deviations (b_i) where greater values of β were needed to see bias.

The analytical standard errors from the Cox model do not account for the variability from estimating the sigmoidal parameters and may therefore be underestimated. Bootstrapped standard errors would account for the two sources of variation, however bootstrapping within this simulation study is not computationally feasible. Because of this, we present the analytical standard errors. In Section 1.6, where we apply our methods to Alzheimer's data, we compare the analytical and bootstrapped standard errors.

1.6 Application to Alzheimer's data

We applied our methodology to two Alzheimer's studies: one from the Rush Alzheimer's Disease Center (RADC) and the other from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We chose to study the impact of a single biomarker on disease progression. Recently, Capuano et al. (2016) at Rush University provided convincing evidence for a sigmoidal trajectory for a measure of global cognition in the RADC dataset. This complements the existing literature on sigmoidal biomarker models for AD proposed by Jack et al. (2010). Based on this theoretical framework for Alzheimer's biomarkers and empirical evidence, we assumed a sigmoidal trajectory (1.2) for the time-varying biomarker in

both analyses.

We fit models (A′), (B), (C), (D), (E), and (F). We also fit the model that adjusts the lateral shift based on an individual's disease severity in Section 1.4.3. In order to account for estimation of the covariate process in variance estimation, we employed a simple bootstrap with 5000 resamples. For each study, we describe below the analyses, study population, biomarker, error model, origin, endpoint and study entry definitions.

1.6.1 Global cognition

We obtained a limited data set from from the RADC that combines data from the Religious Orders Study and the Memory and Aging Project (ROS/MAP) (A Bennett et al., 2012a,b). The biomarker was an average of a battery of 19 cognitive test z-scores (Wilson et al., 2015), which was taken to be a measure of global cognitive function. The origin was defined as birth, study entry was defined as the time of baseline visit, and the endpoint was defined to be time of death or time at last cognitive testing based on the Mini-Mental State Examination (MMSE) (Folstein et al., 1975). Only subjects with at least one biomarker measurement and who carried a diagnosis of mild cognitive impairment (MCI) at the time of biomarker measurement were included in the study. The global measure of cognition was translated by two units to achieve positive values. Gender and body mass index (BMI) (Buchman et al., 2005) were included as covariates in the model that adjusts the lateral shift based on an individual's disease severity.

In choosing an error model, we used graphical diagnostics to determine goodness-of-fit of the absolute error and logit error models based on the sigmoid, $S(u)$, in (1.2). The absolute error model assumes that $Z(u) = S(u; \lambda, T_0) + a_i + b_i u$. We used least squares to estimate λ and T_0 based on data $(E_i, Z(E_i))$ and plotted the residuals, $Z(E_i) - S(E_i; \widehat{\lambda}, \widehat{T_0})$, against study entry time, E_i. We saw that the residuals were centered around zero with nearly constant variance, indicating that the sigmoid-based absolute error model with fixed deviations aligned well with the observed biomarker data. We also considered the sigmoid-based logit error model which assumes that the transformed biomarker, logit$(Z_i(u)/M)$, is linear in u. We plotted the transformed observed biomarker values collected at study entry, logit$(Z_i(E_i)/M)$, against time at study entry, E_i and saw a non-linear,

decreasing trend that accelerates with time. Based on these diagnostics, we assumed the global measure of cognition followed a sigmoid-based absolute error model with fixed deviations. As a sensitivity analysis, we also fit the sigmoid-based absolute error model with time-varying deviations, the sigmoid-based logit error model with both fixed and time-varying deviations, and the following linear absolute error model,

$$Z_i(u) = \lambda(u - T_0) + a_i + b_i u, \tag{1.14}$$

where λ and T_0 play similar roles as they do in the sigmoidal function. The corresponding severity adjusted model is given by

$$Z_i(u) = \lambda(u - \gamma' w_i) + a_i + b_i u, \tag{1.15}$$

where w_i is a vector of individual-specific covariates reflecting disease severity, which alter the lateral shift of the linear function.

Descriptive statistics are given in Table 1.2. A total of 737 subjects were included in the analysis, of whom 473 experienced the event. A plot of the cross-sectional global cognition score at time of baseline visit with a fitted loess curve is displayed in Figure 1.1, which highlights levels of (gender, overweight), where overweight corresponds to subjects with BMI greater than or equal to 25. Although based on cross-sectional, and not longitudinal data, a sigmoidal functional form is discernible, and a clear downward trend is present indicating a decline in global cognition with increasing age. There is no indication that gender and overweight status in tandem are associated with the global cognitive score from Figure 1.1.

The results of the analysis are displayed in Table 1.3. Note that models (B), (C) and (D), which include the time-varying biomarker value at study entry as a fixed covariate, do not involve estimation of the covariate function, and are thus listed with the same results for fixed (a_i) or time-varying deviations (b_i). Assuming the sigmoid-based absolute error model with fixed deviations, a_i, the hazard ratio comparing the biomarker level over a one unit time interval is 0.60 ($\exp(-0.51)$) based on our proposed method (A′). It is 0.49 ($\exp(-.72)$) based on (E) that assumes the origin is study entry and treats the biomarker as a fixed covariate, adjusting for time between the origin and study entry. This difference

Table 1.2: Descriptive statistics of subjects included in the RADC analyses

	n	(%)
Number of subjects	737	
Male	218	29.6
Education		
$\leq$12 years	149	20.2
13-16 years	296	40.2
17+ years	292	39.6
Converted to AD	473	64.2
	Mean	SD
Age at baseline visit	80.9	7.5
Global cognition score	0.42	0.46
Time from birth to last observation	87.8	7.2
Time from study entry to last visit	6.9	4.8

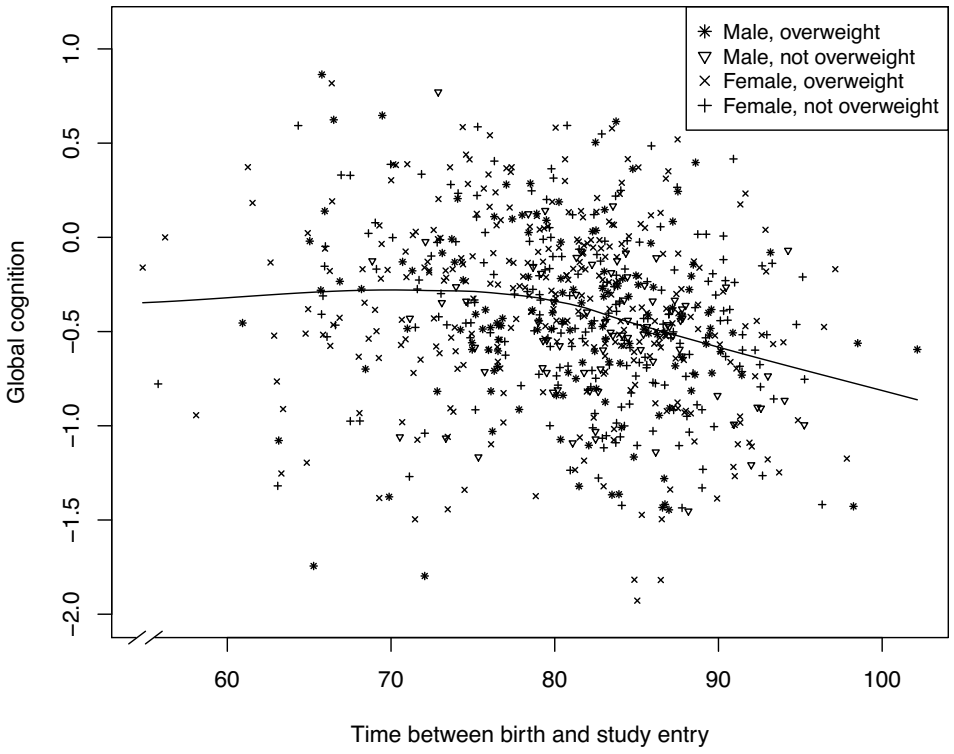

Figure 1.1: Plot of global cognition at study entry versus age at study entry for levels of (BMI, Gender) with fitted loess curve, where BMI$\geq$25 is defined as overweight.

highlights the importance of selection of the analytic approach for answering the clinical question. Given that there is support for the sigmoid-based absolute error model with fixed deviations based on our diagnostics, and it does not require any assumptions on the baseline hazard, we would report the hazard ratio of 0.60.

For the birth origin analyses, the estimates from the model that treats the biomarker as fixed (B) are smaller in magnitude by approximately 30% than those from the model that estimates the full covariate trajectory (A′), with the exception of the absolute error model with time-varying deviations (second row) where the effect is increased in magnitude by 30%. The estimates from the severity adjusted analyses using birth as the origin (column three) are also moderately different from the estimated time-varying model (A′), with more of a difference seen with the linear trend biomarker models, in which the change in magnitude of the regression estimates is more than 65%. This suggests that subject specific adjustments to the sigmoidal model are potentially important. This is seen as well with respect to the covariate function parameter estimates given in Table 4.5, which is given in the Supplemental Material section. When study entry is taken to be the origin, there are moderate differences across the four models that we fit (columns four through seven), and these estimates can vary dramatically from the estimates that are obtained when using birth as the origin (columns one through three).

In summary, for this data example the choice of time origin and treatment of time varying covariates do impact inference and thought must be given to the implications of these choices. Although the magnitude of the estimates for the models differ slightly, the direction of the effect among all models using both origins agree.

As mentioned in the Section 1.5, the analytical standard errors from the Cox model do not account for the variability in estimating the sigmoidal parameters and are likely to be underestimated. Table 4.4 compares the analytical standard errors with the bootstrapped standard errors. For the logit error models, the analytical standard errors are comparable to those that are obtained via bootstrapping. For the biomarker models based on the sigmoid with an absolute error model and based on the linear model, the analytical standard errors are greatly underestimated, suggesting that bootstrapped standard errors should be used.

1.6.2 Hippocampal volume

We also analyzed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Launched in 2003, the ADNI is a longitudinal study aimed at using biomarkers of AD, and clinical and neuropsychiatric assessments to measure the progression of MCI and AD (Mueller et al., 2005).

The origin was defined as birth, study entry was defined as the first biomarker measurement, and the endpoint was defined to be the time of the first diagnosis of AD in the ADNI study. Only subjects with at least one biomarker measurement and who carried a diagnosis of mild cognitive impairment (MCI) at the time of biomarker measurement were included in the study. Gender and APOE-4 allele status were included as covariates in the model that adjusts the lateral shift based on an individual's disease severity. After considering several potential markers in a preliminary analysis, we chose to focus on hippocampal volume, where a clear downward trend was seen in the cross-sectional plot. We standardized and shifted the values by three units so that all resulting biomarkers quantities were positive. Descriptive statistics for the study population are given in Table 4.6.

We used graphical diagnostics, as described for the Rush data, to chose an error model for these data. A plot of logit$(Z(E_i)/M)$ against E_i resulted in a clear linear trend with constant variability. In considering an absolute error model based on the sigmoid in (1.2) and linear trends in (1.14), neither of the residual plots had mean zero. Therefore, we assumed a sigmoid-based logit error model for this analysis. We fit the sigmoid- and linear trend-based absolute error models as sensitivity analyses.

The results of the analysis are reported in Table 4.7 in the Supplemental Material section. There were only small differences seen across the three birth origin models and the four study entry origin models that were fit, regardless of the biomarker model chosen. This was likely due to short follow-up and only small variations in the measured biomarker levels, but potentially a non-exponential functional form for the baseline hazard function.

Table 1.3: Parameter estimates and bootstrapped standard errors (5000 replications) of RADC analysis with global cognition biomarker

	Birth origin			Study entry origin			
Covariate model	A′: $\widehat{f(Z_i(u))}$	B: $f(Z_i(E_i))$	Sev. Adj.	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $\widehat{f(Z_i(t + E_i))}$	F: $\widehat{f(Z_i(t + E_i))} + E_i$
Absolute error, sigmoid, a_i	-0.51 (0.10)	-0.36 (0.10)	-0.35 (0.10)	-0.82 (0.10)	-0.72 (0.11)	-0.66 (0.09)	-0.72 (0.10)
Absolute error, sigmoid, b_i	-0.27 (0.09)	-0.36 (0.10)	-0.30 (0.10)	-0.82 (0.10)	-0.72 (0.11)	-0.61 (0.09)	-0.65 (0.10)
Logit error, sigmoid, a_i	-0.60 (0.07)	-0.40 (0.07)	-0.38 (0.07)	-0.93 (0.07)	-0.81 (0.08)	-0.67 (0.06)	-0.81 (0.07)
Logit error, sigmoid, b_i	-0.56 (0.07)	-0.40 (0.07)	-0.33 (0.07)	-0.93 (0.07)	-0.81 (0.08)	-0.58 (0.06)	-0.74 (0.07)
Absolute error, linear trend, a_i	-0.54 (0.10)	-0.36 (0.10)	-0.19 (0.09)	-0.82 (0.10)	-0.72 (0.11)	-0.59 (0.09)	-0.72 (0.10)
Absolute error, linear trend, b_i	-0.50 (0.09)	-0.36 (0.10)	-0.15 (0.09)	-0.82 (0.10)	-0.72 (0.11)	-0.52 (0.09)	-0.65 (0.10)

Table 1.4: Parameter estimates and bootstrapped standard errors (5000 replications) of ADNI analysis with hippocampal volume biomarker

	Birth origin			Study entry origin			
Covariate model	A′: $\widehat{f(Z_i(u))}$	B: $f(Z_i(E_i))$	Sev. Adj.	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $\widehat{f(Z_i(t + E_i))}$	F: $\widehat{f(Z_i(t + E_i))} + E_i$
Absolute error, sigmoid, a_i	-0.78 (0.07)	-0.75 (0.07)	-0.76 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.83 (0.08)	-0.83 (0.08)
Absolute error, sigmoid, b_i	-0.74 (0.07)	-0.75 (0.07)	-0.74 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.81 (0.07)	-0.81 (0.08)
Logit error, sigmoid, a_i	-0.83 (0.14)	-0.80 (0.13)	-0.81 (0.13)	-0.86 (0.09)	-0.93 (0.12)	-0.93 (0.12)	-0.93 (0.12)
Logit error, sigmoid, b_i	-0.83 (0.13)	-0.80 (0.13)	-0.80 (0.13)	-0.86 (0.09)	-0.93 (0.12)	-0.92 (0.12)	-0.92 (0.12)
Absolute error, linear trend, a_i	-0.79 (0.07)	-0.75 (0.07)	-0.73 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.82 (0.08)	-0.83 (0.08)
Absolute error, linear trend, b_i	-0.78 (0.07)	-0.75 (0.07)	-0.72 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.80 (0.07)	-0.81 (0.08)

1.7 Discussion

In settings such as Alzheimer's disease, using study entry as the time origin in a time-to-event analysis is convenient because time-varying biomarkers are often measured only at study entry and thus can be used as fixed baseline covariates. If study entry is not an appropriate time origin, the question of interest is "under what conditions is it valid to use an incorrect time origin of study entry in a Cox model with time-varying covariates?" The answer to this question for the case of fixed covariates was established by Korn et al. (1997). We have extended the work of Korn et al. to the case of a time-varying covariate; we have shown that using an incorrect origin of study entry is valid only for special cases of the baseline hazard function provided that the functional form of the covariate process is fully-observed. We further examined the implications of treating a continuous, time-varying covariate as fixed at its value at study entry for both the onset and study entry origins; we showed that this is only valid in certain cases of the baseline hazard and time-varying covariate. These theoretical findings underscore the necessity for full observation of time-varying covariates or good modeling of them. This is challenging unless good prior data is available.

For situations in which the full covariate function is not available, we developed methods for estimating regression coefficients when the time-varying covariate follows the general error model $f(Z_i(u)) = f(S(u)) + a_i + b_i u$, when only $f(Z_i(E_i))$ is observed; this is a generalization of the covariate model with fixed residuals given by Sperrin and Buchan (Sperrin and Buchan, 2013). While Sperrin and Buchan focused on predictive ability, we have examined the bias and standard errors of hazard model estimates.

A main limitation of our results in Theorem 1 parts (a) and (c) is that the time of onset must be known. In diseases with a long prodromal phase such as Alzheimer's disease, the time of onset is not easily determined; as a result, time from onset to study entry is not known and the analysis in Theorem 1 part (a) cannot be implemented only approximately. However, Theorem 1 part (c) can be applied if the baseline hazard function is constant; namely, it is only valid to use study entry as the origin and treat time-varying covariates as fixed when the baseline hazard measured from onset of disease is constant and the

continuous, time-varying covariate process satisfies the condition in Theorem 1 (c). In Alzheimer's disease, it is hard to imagine that the risk of the disease is constant over time, thus regression estimates are likely to be biased unless the biomarker does not vary over time.

Our analysis of the RADC data provides an example of where the choice of time origin and biomarker model can dramatically affect estimation in the presence of a time-varying covariate. Estimates from an analysis from study entry that treats the time-varying biomarker as fixed are quite different from the estimates of our proposed model, even when time between onset and study entry is adjusted for as a continuous covariate. On the other hand, our analysis of ADNI data provides an example of when incorrectly viewing a time-varying covariate as fixed had little effect on estimation which was likely due to covariate measurements that are nearly constant over a short follow-up period. However, the choice of time origin did have a noticeable effect, suggesting that the baseline hazard was not of the requisite exponential functional form to permit the use of study entry as the origin.

The strength of this work is that our analytical results and methods can be applied to any setting in which time-varying covariates follow continuous trajectories. It is important to note that our results are particular to the Cox model in which time-dependent population-level expressions are absorbed into the baseline hazard function, as in (1.9). For example, the methods we proposed cannot be directly applied to AFT regression, as the AFT model with time-varying covariates cannot be simplified in this way. However, Sperrin and Buchan (Sperrin and Buchan, 2013) showed through simulations that including individual-specific deviations in an AFT regression model is superior in its predictive ability to the model that includes the time-varying covariate treated as fixed.

As serial biomarker measurements are collected through studies such as the RADC studies and ADNI, the observation of the biomarker process at multiple timepoints would be useful for better parametric modeling and for identification of both components of individual level deviation, a_i and b_i. With two or more biomarker measurements per individual, our methods would allow for the estimation of the desired regression parameter whereas methods for Cox regression with missing data and joint modeling of survival

and longitudinal data are not likely to be robust with so few covariate measurements.

1.8 Appendix

Proof of Theorem 1 part (a) (Korn et al. extension)

Substituting the definition of time from study entry, $T_i = U_i - E_i$, into the definition of the hazard function

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) = \lim_{\delta \to 0} P(t < T_i \leq t + \delta | T_i \geq t, \overline{w_i}(\cdot), E_i)/\delta.$$

gives

$$\begin{aligned}
\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) &= \lim_{\delta \to 0} P(t + E_i < T_i + E_i \leq t + \delta + E_i | T_i + E_i \geq t + E_i, \overline{w_i}(\cdot), E_i)/\delta \\
&= \lim_{\delta \to 0} P(t + E_i < U_i \leq t + \delta + E_i | U_i \geq t + E_i, \overline{w_i}(\cdot), E_i)/\delta, \\
&= \lambda_{U_i}(t + E_i | \overline{w_i}(\cdot), E_i) \\
&\overset{(1.5)}{=} \lambda_{0U}(t + E_i) e^{\beta w_i(t+E_i)}. \qquad (1.16)
\end{aligned}$$

If $\lambda_{0U}(u) = ce^{\gamma u}$ for some $c > 0$, then (1.16) yields

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) = ce^{\gamma t} e^{\beta w_i(t+E_i) + \gamma E_i}.$$

Thus in comparison to model (1.6), it is clear that $\beta = \xi$ provided $\lambda_{0T}(t) = ce^{\gamma t}$.

If $\beta = \xi$, it follows that

$$\lambda_{0T}(t) = \lambda_{0U}(t + E_i) e^{-\gamma E_i},$$

because $\lambda_{T_i}(t) = \lambda_{0T}(t) e^{\xi w_i(t+E_i) + \gamma E_i}$ by (1.6) and $\lambda_{T_i}(t) = \lambda_{0U}(t + E_i) e^{\beta w_i(t+E_i)}$ by (1.16). In order for $\lambda_{0T}(t)$ to be independent of i, it must be that $\lambda_{0U}(u) = ce^{\gamma u}$.

Proof of Theorem 1 part (b)

First observe that the true survival model in (1.5) can be rewritten as

$$\lambda_{U_i}(u|\overline{w_i}(\cdot), E_i) = \lambda_{0U}(u) e^{\beta w_i(E_i)} e^{\beta[w_i(u) - w_i(E_i)]}. \qquad (1.17)$$

If β in (1.17) is equal to α in (1.7), then

$$\lambda_{1U_i}(u) = \lambda_{0U}(u) e^{\beta[w_i(u) - w_i(E_i)]} \qquad (1.18)$$

This implies that $w_i(u) - w_i(E_i) = g(u)$ for some function g that is independent of i; otherwise $\lambda_{1U}(u)$ would depend on i which is not the case. The condition $w_i(u) - w_i(E_i) = g(u)$ requires that the function g is zero on the support of E since it is zero at all study entry times, E_i. Note that $g(u)$ may be nonzero outside of the support of E.

If $w_i(u) = w_i(E_i) + g(u)$, then the true model given by (1.5) and fitted model given by (1.7) are equivalent and $\beta = \alpha$.

Proof of Theorem 1 part (c)

Note that the true survival model in (1.16) can be rewritten as

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) = \lambda_{0U}(t + E_i)e^{\beta[w_i(t+E_i)-w_i(E_i)]}e^{\beta w_i(E_i)}. \tag{1.19}$$

If β in (1.19) is equal to ζ in (1.8), then

$$\lambda_{2T}(t) = \lambda_{0U}(t + E_i)e^{\beta[w_i(t+E_i)-w_i(E_i)]-\eta E_i}. \tag{1.20}$$

It must be that $\lambda_{0U}(t + E_i)e^{-\eta E_i} = h(t)$ and $w_i(t + E_i) - w_i(E_i) = k(t)$ for some functions h and k independent of i; otherwise the right-hand side of (1.20) would depend on i and cannot be equal to $\lambda_{2T}(t)$, an expression independent of i. Since we are assuming that all hazard functions are continuous, $\lambda_{0U}(t + E_i)e^{-\eta E_i} = h(t)$ implies $\lambda_{0U}(u) = ce^{\eta u}$, for some $c > 0$.

If $w_i(t + E_i) - w_i(E_i) = k(t)$ and $\lambda_{0U}(u) = ce^{\eta u}$ for $c > 0$, then the true model in (1.19) is

$$\lambda_{T_i}(t|\overline{w_i}(\cdot), E_i) = ce^{\eta t+\beta k(t)}e^{\beta w_i(E_i)+\eta E_i}.$$

Comparing this to (1.8), it must be that $\beta = \zeta$ and $\lambda_{2T}(t) = ce^{\eta t+\beta k(t)}$.

Acknowledgement

The authors wish to thank Deborah Blacker for helpful feedback in preparation of this manuscript. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01

AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Time-to-event analysis when the event is defined on a finite time interval

Catherine Lee

Department of Biostatistics

Harvard T.H. Chan School of Public Health

Sebastien Haneuse

Department of Biostatistics

Harvard T.H. Chan School of Public Health

2.1 Introduction

Allogeneic hematopoietic stem cell transplantation (HCT) is the gold standard treatment for hematopoietic malignancies such as leukemia or multiple myeloma (Passweg et al., 2016). The procedure, however, carries a high risk of graft-versus-host disease (GVHD), a syndrome in which the donor cells attack those of the recipient and that is associated with increased morbidity and mortality (Choi et al., 2010). Clinically, current consensus diagnosis and staging criteria distinguish between acute GVHD and chronic GVHD on the basis of the clinical signs and symptoms that manifest post-transplant (Filipovich et al., 2005). For the latter they further distinguished between two sub-categories: *classic* acute GVHD, for which the clinical signs manifest within 100 days of the transplant, and *persistent, recurrent, or late-onset* acute GVHD, for which the signs manifest after 100 days.

For the most part clinical studies of classic acute GVHD (henceforth acute GVHD) have employed standard survival analysis techniques, such as the Kaplan-Meier estimate of the survivor function and the Cox model for the hazard function, or logistic regression (Remberger et al., 2002; Urbano-Ispizua et al., 2002; Weissinger et al., 2007; Paczesny et al., 2009; Levine et al., 2012; MacMillan et al., 2012; Holtan et al., 2015b). As applied to acute GVHD, however, these techniques fail to accommodate two important clinical issues. First, patients who undergo HCT are also at high risk for death in the short-term. In our motivating data from the Center for International Blood and Bone Marrow Transplant Research (CIBMTR), for example, 100-day mortality among 9,651 patients who underwent HCT between 1999-2011 was 15.4%. Performing statistical analyses that naïvely treat patients who die as censored (either in a standard survival analysis or in logistic regression analyses), however, is well-known to be problematic and may lead to erroneous conclusions. Towards resolving this, one could construct and model a composite endpoint of acute GVHD and death (Holtan et al., 2015a), although doing so changes the scientific question that is being addressed (Jazić et al., 2016). Another option is to perform a competing risks analysis (Pintilie, 2006), an approach adopted by a number of clinical papers (Baron et al., 2005; Kim et al., 2008; Levine et al., 2010; Sorror et al., 2014). One drawback of the competing risks framework, however, is that information on time from

acute GVHD to death is discarded; this is a loss of information that may affect estimation and inference and also limits the scope of scientific inquiry (Haneuse and Lee, 2016a). To make use of this information, a third option is to embed the study of acute GVHD within the semi-competing risks framework (Fine et al., 2001). Briefly, semi-competing risks refers to the setting where interest lies in a so-called *non-terminal* event (in our setting, acute GVHD) the occurrence of which is subject to a terminal event (in our setting, death). Broadly, methods for the analysis of semi-competing risks data can classified into three groups: those based on copulas (Fine et al., 2001; Peng and Fine, 2007; Hsieh et al., 2008; Lakhal et al., 2008); those grounded in causal inference (Egleston et al., 2007; Tchetgen Tchetgen, 2014); and those based on multi-state models, particularly the frailty illness-death model (Putter et al., 2007; Xu et al., 2010; Lee et al., 2015).

A second issue associated with standard survival analysis techniques for acute GVHD is that the support of the response variable (i.e. the time-to-event outcome of interest) is taken to be the positive part of the real line. Moreover, in the absence of censoring and competing risks, standard survival methods assume that all patients will experience the event of interest provided sufficient time has elapsed. This, however, is not the case for a diagnosis of classic acute GVHD since the support of the outcome is restricted to the finite interval (0, 100]. Survival models that ignore this defining feature therefore suffer from a form of misspecification which may also lead to erroneous conclusions. Towards accommodating the finite support of the non-terminal event, it is first important to acknowledge that only a subset of patients are expected to be diagnosed with acute GVHD. In a sense, as in cancer prognosis studies where there are often long-term survivors, the population of patients who undergo HCT can be viewed as a mixture of patients who are *susceptible* to acute GVHD and those who are *not susceptible*. In the statistical literature, cure models are often used to represent this phenomenon (Berkson and Gage, 1952; Farewell, 1982; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000).

Towards simultaneously accommodating both of the aforementioned clinical issues, in this paper we propose a novel multi-state model formulation for the study of time-to-event outcomes defined on a finite interval. As we elaborate upon, the proposed model builds on recent work by Conlon et al. (2014) by combining a logistic regression-based

cure model with an illness-death model. Crucially, the methods we develop extend the work of Conlon et al. (2014) in two important ways. First, the hazard function that dictates the rate at which patients transition experience the non-terminal event is specified so as to explicitly respect the finite time interval over which events can occur. Second, a subject-specific frailty term is introduced into the specification of each hazard function in the model specification. In doing so the proposed model is better-equipped to handle heterogeneity across and dependence within patients that is not captured by the (remaining) systematic components of the model (Wienke, 2010). The remainder of this paper is structured as follows. In Section 2.2 we provide a brief introduction to the CIBMTR data. Sections 2.3 and 2.4 presents the proposed multi-state model for survival analyses on a finite time interval and the framework for estimation/inference. Section 2.5 then provides a detailed simulation study aimed at evaluating small-sample operating characteristics of the proposed methods as well as investigating instances where competitor methods perform either poorly or well. In Section 3.6 we present an illustrative analysis of the CIBMTR data the goal of which is to investigate the impact of disease type, stage and HLA matching on the joint risk of acute GVHD and death. Finally, Section 2.7 concludes with a discussion and avenues for future work.

2.2 Data

The methods proposed in this paper are motivated by an on-going collaboration investigating risk factors for acute GVHD among patients who undergo HCT. The data were obtained from CIBMTR, a collaboration between the National Marrow Donor Program and the Medical College of Wisconsin representing a worldwide network of transplant centers that contribute detailed data on HCT. For the purposes of this paper we restrict attention to a sample of n=9,651 patients who underwent first HCT between 1999-2011 for treatment of acute myelogenous leukemia (AML), acute lymphoblastic leukemia (ALL), myelodysplastic syndrome (MDS) or chronic myelogenous leukemia (CML).

Of primary scientific interest is the time from transplantation to a diagnosis of grade III or IV acute GVHD, defined within 100 days post-transplant. We present Kaplan-

Meier estimated survivor functions for acute GVHD and death in Figure 2.1. Figure 2.1 illustrates two challenging features of acute GVHD: (1) the finite interval over which the outcome is defined and (2) the fraction of patients who are never diagnosed with acute GVHD. The estimated survivor function for acute GVHD is plotted for the first 100 days post-transplantation, the interval over which the outcome is defined, where deaths without acute GVHD within the first 100 days are treated as censored. At day 100 post-transplant, there were 1,701 patients diagnosed with acute GVHD grades III-IV with an estimated survival of 0.824 (s.e. 0.004). From the estimated survivor function for acute GVHD, we see that survival levels out long before 100 days post-transplant. At 30 and 60 days post-transplant, the survival estimates for acute GVHD grades III-IV are 0.883 (s.e. 0.003) and 0.839 (s.e. 0.004), respectively, indicating that grade III or IV acute GVHD is typically diagnosed within the first two months of transplantation. Since a diagnosis of acute GVHD is only given within the 100 day window post-transplant, the rate of acute GVHD after 100 days remains at 17.6%.

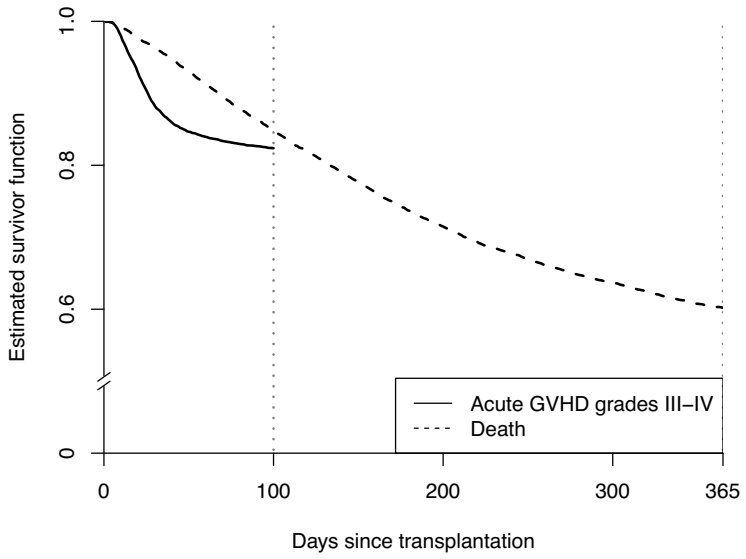

Figure 2.1: Kaplan-Meier estimates of the survivor functions corresponding to time to acute GVHD grades III-IV and death.

We administratively censor death at 365 days post-transplantation; as such, the estimated survivor function for death is displayed only until 365 days post-transplantation Figure 2.1.There were 5,822 observed deaths (40% censoring) with estimated one and five

year survival rates of 0.60 (s.e. 0.005) and 0.41 (s.e. 0.005), respectively.

2.3 The proposed model

Figure 2.2 provides a schematic of the proposed model for acute GVHD and death. Central to the model are four ingredients: (i) a generalized linear model for the latent binary susceptibility status, (ii) a hazard regression model for the terminal event among patients whose latent state is 'not susceptible', (iii) a finite-interval illness-death model for the non-terminal and terminal events, and (iv) a multiplicative subject-specific frailty that is taken to influence all time-to-event transitions.

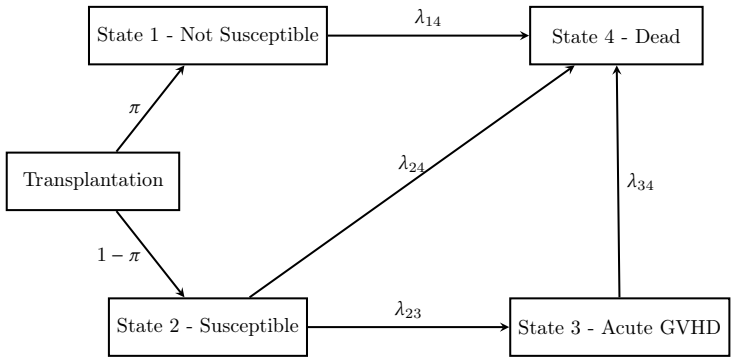

Figure 2.2: Schematic of proposed multi-state model for acute GVHD and death, incorporating a cure fraction for acute GVHD susceptibility.

2.3.1 Latent susceptibility state

Following convention in the cure fraction literature, we introduce a binary variable L to represent a latent cured state, such that $L = 1$ corresponds to a subject that will not experience acute GVHD and $L = 0$ corresponds to a subject that is susceptible to acute GVHD. Note that the cure indicator is partially observed; if a patient is alive at any time after day 100 post-transplant, it must be that $L = 1$. Let $\pi_i = P(L_i = 1|X_{s,i})$, which is the probability that a patient will not develop acute GVHD within τ days of transplant. We model the cure fraction, π, via a generalized linear model

$$\pi_i = P(L_i = 1|X_{s,i}) \; = \; g^{-1}(\beta_s X_{s,i}), \tag{2.1}$$

where X_s denotes a vector of covariates that may impact acute GVHD susceptibility.

2.3.2 Semi-competing risks with non-susceptibility fraction

Let T_1 and T_2 denote the nonterminal and terminal event times, respectively. For the stem cell transplant setting we consider, we assume that T_1 is defined on $(0, \tau)$, where $\tau = 101$. If a patient experiences the terminal event without experiencing the nonterminal event, we set $T_1 = \tau$. We define T_2 on the positive real line. We view patients as transitioning through a series of states post-transplantation (1. not susceptible to acute GVHD at transplantation; 2. susceptible to acute GVHD at transplantation; 3. acute GVHD; 4. death) in the following way: $(1 \to 4)$ transplantation (not susceptible to acute GVHD) to death; $(2 \to 4)$ transplantation (susceptible to acute GVHD) to death; $(2 \to 3)$ transplantation to acute GVHD; and $(3 \to 4)$ acute GVHD to death. This multi-state model combines a cure model and an illness-death model where a diagnosis of acute GVHD can only occur within finite window post-transplant, and is defined through the specification of a latent variable representing acute GVHD non-susceptibility status and the transition intensities or hazard functions defined by

$$\begin{aligned}
\lambda_{14}(t_2|L=1) &= \lim_{\Delta\to 0} P(T_2 \in [t_2, t_2+\Delta)|T_2 \geq t_2)/\Delta, \text{ for } t_2 > 0 \\
\lambda_{24}(t_2|L=0) &= \lim_{\Delta\to 0} P(T_2 \in [t_2, t_2+\Delta)|T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_2 < \tau \\
\lambda_{23}(t_1|L=0) &= \lim_{\Delta\to 0} P(T_1 \in [t_1, t_1+\Delta)|T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < \tau \\
\lambda_{34}(t_2|t_1, L=0) &= \lim_{\Delta\to 0} P(T_2 \in [t_2, t_2+\Delta)|T_1 = t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < t_2,
\end{aligned}$$

where the λ_{kj} denotes the hazard corresponding to the transition from state k to state j. Note that the 24-transition is also defined on $(0, \tau)$ since patients who have survived past τ days without a diagnosis of acute GVHD are necessarily cured, forcing deaths without acute GVHD among uncured patients to occur within τ days of transplantation.

For patients who are non-susceptible to acute GVHD at the time of transplantation, we model the hazard function corresponding to the time from transplantation to death by

$$\lambda_{14}(t_{2i}|X_{14,i}, L_i = 1) = \gamma_i \cdot \lambda_{0,14}(t_{2i}) \exp\{\beta_{14} X_{14,i}\}, \text{ for } t_{2i} > 0, \tag{2.2}$$

where γ is a subject-specific frailty that we choose from a Gamma distribution with mean 1 and variance θ, $\lambda_{0,14}$ is a baseline hazard function conditional on $\gamma = 1$, and X_{14} are covariates that may impact the hazard function that corresponds to time to death among patients who are non-susceptible. The shared frailty can be viewed as a patient-specific random effect that is multiplicative on each of the hazards in the same way, therefore inducing dependence between T_1 and T_2.

Patients who are susceptible to acute GVHD at the time of transplantation can subsequently be diagnosed with acute GVHD and/or can die. If a patient is diagnosed with acute GVHD, he or she could die thereafter. We define the hazard functions corresponding to the remaining three transitions by:

$$\lambda_{24}(t_{2i}|X_{24,i}, L_i = 0) = \gamma_i \cdot \lambda_{0,24}(t_{2i}) \exp\{\beta_{24} X_{24,i}\}, \text{ for } 0 < t_{2i} < \tau \tag{2.3}$$

$$\lambda_{23}(t_{1i}|X_{23,i}, L_i = 0) = \gamma_i \cdot \lambda_{0,23}(t_{1i}) \exp\{\beta_{23} X_{23,i}\}, \text{ for } 0 < t_{1i} < \tau \tag{2.4}$$

$$\lambda_{34}(t_{2i}|t_{1i}, X_{34,i}, L_i = 0) = \gamma_i \lambda_{0,34}(t_{2i}|t_{1i}) \exp\{\beta_{34} X_{34,i}\}, \text{ for } 0 < t_{1i} < t_{2i}, \tag{2.5}$$

where γ is the same subject-specific frailty that appears in (3.6) and X_{24}, X_{23} and X_{34} are covariates that are possibly associated with the hazard functions of the corresponding transitions. The hazard model that corresponds to the transition from state 2 (susceptible to acute GVHD at transplantation) to state 3 (acute GVHD) reflects the fact that T_1 is defined on $(0, \tau)$. In general, $\lambda_{0,34}(t_2|t_1)$ can be any function of t_1 and t_2. In a Markov model, $\lambda_{0,34}(t_2|t_1) = \lambda_{0,34}(t_2)$. We assume $\lambda_{0,34}(t_2|t_1) = \lambda_{0,34}(t_2 - t_1)$, a function of the sojourn time between states 3 and 4; this is referred to as the semi-Markov property. These three transition hazard models correspond to the standard illness-death with shared frailty model (Xu et al., 2010; Lee et al., 2015), without a cure fraction. Note that models (3.6)-(3.9) represent a modification of the multi-state model that is presented in Conlon et al. (2014). We define the 23- and 24-transitions over a finite time window to better reflect the structure of the data, and we include a patient-specific shared frailty γ to the transition hazard regression models to better handle heterogeneity that is not captured by covariates; additionally, the shared frailty induces additional dependence between the two outcomes.

The baseline hazard functions in (3.6), (3.7) and (3.9) can be defined using any survival distribution that is defined on the positive real line; similarly, the baseline hazard

function corresponding to the nonterminal event in (3.8) can be defined by any survival distribution defined on $(0, \tau)$. For example, the hazard corresponding to a truncated density f defined on $(0, \tau)$ is given by

$$\lambda(t) = \frac{f(t)}{F(\tau) - F(0)},$$

where F is the cumulative distribution that corresponds to f. We assume that the baseline hazard functions in (3.6)-(3.9) correspond to Weibull distributions (Lai, 2014):

$$\lambda_{0,\omega}(t) = \begin{cases} \alpha_\omega \kappa_\omega t^{\alpha_\omega - 1}, & \text{if } \omega \in \{14, 34\}, \text{ for } t > 0 \\ \dfrac{\alpha_\omega \kappa_\omega t^{\alpha_\omega - 1} e^{-\kappa_\omega t^{\alpha_\omega}}}{e^{-\kappa_\omega t^{\alpha_\omega}} - e^{-\kappa_\omega \tau^{\alpha_\omega}}}, & \text{if } \omega \in \{24, 23\}, \text{ for } t \in (0, \tau). \end{cases}$$

2.4 Estimation and inference

2.4.1 The observed data likelihood

Let C_i denote the censoring time; we assume the pair, (T_{i1}, T_{i2}), is independent of C_i conditional on X_i, where X_i is a matrix of covariates. For patient i, the observed data are $D_i = \{Y_{i1}, \delta_{i1}, Y_{i2}, \delta_{i2}, X_i\}$, where $Y_{i1} = \min(T_{i1}, T_{i2}, C_i, \tau)$ with censoring indicator $\delta_{i1} = \mathbb{1}\{T_{i1} \leq \min(T_{i2}, C_i, \tau)\}$, and $Y_{i2} = \min(T_{i2}, C_i)$ with censoring indicator $\delta_{i2} = \mathbb{1}\{T_{i2} \leq C_i\}$. Note that if $Y_{i1} = \tau$, then it is necessarily the case that $\delta_{i1} = 0$.

Towards developing the form of the likelihood, we consider the eight distinct post-transplant trajectories presented in Table 2.1. These trajectories vary across combinations of δ_{i1}, δ_{i2} and $\mathbb{1}(Y_{i1} < \tau)$ and correspond to distinct likelihood contributions. When death without acute GVHD is observed within τ days of transplantation, we are unable to distinguish between patients that are susceptible to acute GVHD at time of transplantation and those who are not based on the observed data; in both cases $(\delta_{i1}, \delta_{i2}, \mathbb{1}(Y_{i1} < \tau)) = (0, 1, 1)$. Consequently, the likelihood contribution of these subjects is the sum of the two likelihood contributions. Similarly, when neither acute GVHD nor death are not observed within τ days post-transplant, $(\delta_{i1}, \delta_{i2}, \mathbb{1}(Y_{i1} < \tau)) = (0, 0, 1)$, the likelihood contribution of patients who pass through either the susceptible or non-susceptible state is the sum of the two respective likelihood contributions. Let ϕ denote the collection of unknown parameters, $(\alpha_{14}, \kappa_{14}, \beta_{14}, \alpha_{24}, \kappa_{24}, \beta_{24}, \alpha_{23}, \kappa_{23}, \beta_{23}, \alpha_{34}, \kappa_{34}, \beta_{34}, \theta)$. The marginal likelihood

with respect to the distribution of γ of the proposed model is given by

$$\begin{aligned} & L(\phi) \\ & = \prod_{i=1}^{n} f_1(Y_{i1}, Y_{i2})^{\delta_{i1}\delta_{i2}} \cdot f_2(Y_{i1}, Y_{i2})^{\delta_{i1}(1-\delta_{i2})} \cdot f_3(Y_{i1}, Y_{i2})^{(1-\delta_{i1})\delta_{i2}\mathbb{1}(Y_{i1}=\tau)} \cdot \\ & \qquad f_4(Y_{i1}, Y_{i2})^{(1-\delta_{i1})\delta_{i2}\mathbb{1}(Y_{i1}<\tau)} \cdot f_5(Y_{i1}, Y_{i2})^{(1-\delta_{i1})(1-\delta_{i2})\mathbb{1}(Y_{i1}=\tau)} \cdot \\ & \qquad f_6(Y_{i1}, Y_{i2})^{(1-\delta_{i1})(1-\delta_{i2})\mathbb{1}(Y_{i1}<\tau)} \end{aligned} \tag{2.6}$$

where

$$f_k(Y_{i1}, Y_{i2}) = \int f_k(Y_{i1}, Y_{i2}|\gamma_i) f(\gamma_i)\, \mathrm{d}\gamma_i,$$

for $k = 1, \ldots, 6$, and

$$\begin{aligned} f_1(Y_{i1}, Y_{i2}|\gamma_i) &= (1-\pi_i) \cdot \lambda_{23}(Y_{i1}|\gamma_i) \cdot S_2(Y_{i1}|\gamma_i) \lambda_{34}(Y_{i2} - Y_{i1}|\gamma_i) \cdot S_3(Y_{i2}|Y_{i1}, \gamma_i) \\ f_2(Y_{i1}, Y_{i2}|\gamma_i) &= (1-\pi_i) \cdot \lambda_{23}(Y_{i1}|\gamma_i) \cdot S_2(Y_{i1}|\gamma_i) \cdot S_3(Y_{i2}|Y_{i1}, \gamma_i) \\ f_3(Y_{i1}, Y_{i2}|\gamma_i) &= \pi_i \cdot \lambda_{14}(Y_{i2}|\gamma_i) \cdot S_1(Y_{i2}|\gamma_i) \\ f_4(Y_{i1}, Y_{i2}|\gamma_i) &= \pi_i \cdot \lambda_{14}(Y_{i2}|\gamma_i) \cdot S_1(Y_{i2}|\gamma_i) + (1-\pi_i) \cdot \lambda_{24}(Y_{i2}|\gamma_i) \cdot S_2(Y_{i2}|\gamma_i) \\ f_5(Y_{i1}, Y_{i2}|\gamma_i) &= \pi_i \cdot S_1(Y_{i2}|\gamma_i) \\ f_6(Y_{i1}, Y_{i2}|\gamma_i) &= \pi_i \cdot S_1(Y_{i2}|\gamma_i) + (1-\pi_i) \cdot S_2(Y_{i2}|\gamma_i). \end{aligned}$$

These likelihood expressions depend on the transition survival functions for remaining in states 1, 2 and 3 which are defined as follows:

$$\begin{aligned} S_1(t|\gamma) &= \exp\left(-\int_0^t \lambda_{14}(u|\gamma)\, du\right) \\ S_2(t|\gamma) &= \exp\left(-\int_0^t \lambda_{23}(u|\gamma)\, du - \int_0^t \lambda_{24}(u|\gamma)\, du\right) \\ S_3(t_2|t_1, \gamma) &= \exp\left(-\int_0^{t_2-t_1} \lambda_{34}(u|\gamma)\, du\right), \text{ for } t_2 > t_1. \end{aligned}$$

The expressions for $f_k(Y_{i1}, Y_{i2})$ based on Weibull distributions, for $k = 1, \ldots, 6$ are included in the Supplementary Materials section.

2.4.2 Maximum likelihood estimation and inference

We use maximum likelihood estimation to obtain the parameter estimates of the model. Under certain regularity conditions and a correctly specified model, the maximum likelihood estimator

$$\hat{\phi} = \arg\max_{\phi} \log L(\phi)$$

Table 2.1: Likelihood contributions of eight possible post-transplant trajectories.

Trajectory	Observed data $(\delta_{i1}, \delta_{i2}, y_{i1} < \tau)$	Likelihood Expression
Acute GVHD, death	$(1, 1, -)$	f_1
Acute GVHD, cens. before death	$(1, 0, -)$	f_2
Death $\geq \tau$ days without acute GVHD	$(0, 1, 0)$	f_3
Not susceptible, death $< \tau$ days without acute GVHD	$(0, 1, 1)$	f_4
Susceptible, death $< \tau$ days without acute GVHD	$(0, 1, 1)$	f_4
Alive without acute GVHD at $\geq \tau$ days	$(0, 0, 0)$	f_5
Not susceptible, alive without acute GVHD at $< \tau$ days	$(0, 0, 1)$	f_6
Susceptible, alive without acute GVHD at $< \tau$ days	$(0, 0, 1)$	f_6

is consistent for the truth ϕ_0 and

$$\sqrt{n}(\hat{\phi} - \phi_0) \xrightarrow{\mathcal{L}} MVN\left(0, \Sigma\right),$$

where

$$\Sigma = I(\phi_0)^{-1},$$

the inverse of the expected information matrix (Ferguson, 1996),

$$I(\phi) = -E\left[\frac{\partial^2}{\partial \phi_0^2} \log L(\phi_0)\right].$$

We use a pre-existing non-linear optimization function to estimate the maximum likelihood estimator (described below). We use the inverse of the observed information (inverse Hessian matrix) to estimate the variance of the maximum likelihood estimator as the observed information is a consistent estimator for the expected information. This matrix is included as output by the numerical optimizers used in the estimation process. We use the results regarding the asymptotic distribution of $\hat{\phi}$ to construct 95% confidence intervals for our estimates.

2.4.3 Practical considerations

We maximize the log-likelihood using a quasi-Newton non-linear numerical optimization algorithm (Wright and Nocedal, 1999). Initial attempts at estimation revealed that estimates were sensitive to the starting values used in the numerical optimization algorithm; this is likely due to the complex nature of the log-likelihood function. To boost the efficacy

of the numerical optimization at finding the maximum of the log-likelihood function, we perform five separate attempts at maximum likelihood estimation, retaining the estimates that maximize the log-likelihood function among the five attempts. If more than one set of estimates correspond to the same maximum log-likelihood value, we retain the set of estimates that are closest, with respect to the L^2-norm, to the mean of the set of estimates. Starting values must be provided for the numerical optimization procedure; we generate a set of base starting values for the five maximum likelihood estimation attempts as follows. We first create an indicator variable representing cured status, where subjects who have died without acute GVHD or are censored after 100 days are considered cured, and subjects who have a diagnosis of acute GVHD (alive or censored for death) are not cured; the remaining subjects (who are censored before 100 days), are randomly assigned cured status. We then fit an illness-death model with shared frailty among the subjects who are not cured to obtain parameter estimates for the 24-, 23-, and 34-transitions and θ, the variance of the frailty term; we fit a univariate Weibull regression model on the 14-transition time to obtain parameter estimates for the 14-transition; and we fit a logistic regression model to the cure variable to obtain the non-susceptibility regression coefficients. For each maximum likelihood estimation attempt, we perturb the base starting values with randomly chosen additive or multiplicative noise generated from a Uniform($-0.1i$,$0.1i$) or Uniform($1-0.1i$,$1+0.1i$) distribution, respectively, where $i = 1, \ldots, 5$ is the maximum likelihood attempt.

2.5 Simulation study

We conducted a series of simulation studies with the overarching goal being to evaluate the performance of the proposed model and framework of Section 2.3, and to better understand the performance of comparable methods that fail to take the finite nature of the nonterminal event into consideration.

2.5.1 Data generation

We based our simulated data on a data analysis using a subset of the population consisting of patients 50 years or older with intermediate or advanced stage disease. We use the estimates from this analysis to initially set the parameters of our simulation study. In order to highlight the effect of defining acute GVHD using a truncated Weibull distribution, we update the truncated Weibull parameters of the 23-transition so that the distributions between the truncated and standard Weibull distributions are very distinct. We refer to data generated with these parameters as the *base scenario*.

We generated a total of $1,000$ simulated data sets of size $5,000$ under our proposed model defined by by (3.6)-(3.9) with $\tau = 101$ days using the following procedure. For each patient i, we first generate a binary covariate X_i from a Bernoulli distribution with mean 0.5. We then generate an individual-specific frailty γ_i using a Gamma distribution with mean 1 and variance θ, where $\theta = 0.18$. We generate a nonsusceptibility (or cured) indicator, L_i, using a Bernoulli distribution with nonsusceptibility rate $\pi_i = \text{expit}(\beta_{0s} + \beta_{1s}X_i)$, where $\beta_{0s} = -0.41$ and $\beta_{1s} = 0.50$, so that the cure rate is 46%. If $L_i = 1$ (cured of acute GVHD), we set $\alpha_{14} = 1.4$, $\kappa_{14} = 0.0002$ and $\beta_{14} = 0.25$, and generate time to death T_{2i} using the model in (3.6) which corresponds to a standard Weibull distribution. We then define T_{1i} to be the minimum of τ and T_{2i}. If $L_i = 0$ (susceptible to acute GVHD), time from transplantation to a diagnosis of acute GVHD is generated using the model in (3.8) via the inverse cumulative distribution transform method (Bender et al., 2005) with $\alpha_{23} = 0.5$, $\kappa_{23} = 0.1$ and $\beta_{23} = 0.25$. We set $\alpha_{24} = 1.4$, $\kappa_{24} = 0.005$, and $\beta_{24} = 0.5$ and generate time from transplantation to death using the model in (3.7), which corresponds to a truncated Weibull distribution. If $T_{1i} \geq T_{2i}$, we re-define T_{1i} to be the minimum of τ and T_{2i}. If $T_{1i} < T_{2i}$, we re-define T_{2i} to be the sum of T_{1i} and time from a diagnosis of acute GVHD to death generated using model (3.9), with $\alpha_{34} = 1.3$, $\kappa_{34} = 0.002$, and $\beta_{34} = 0.15$. We do not introduce censoring in our base simulations and set $C_i = \infty$. We determine the observed data $(Y_{1i}, \delta_{1i}, Y_{2i}, \delta_{2i})$ based on C_i, T_{1i}, T_{2i} and τ.

2.5.2 Additional data scenarios

We also examine the effect of varying each of the following separately on estimation and inference: (i) the rate of censoring within the 100 day window by introducing a censoring variable drawn from an exponential distribution with mean 100; (ii) administrative censoring of death at 365 days; (iii) increasing the the non-susceptibility fraction from 46% to 76%; and (iv) varying the Weibull parameters of the $23-$transition so that the truncated Weibull and standard Weibull distributions are comparable on the interval $(0, \tau)$.

2.5.3 Analyses

We compare our proposed model defined by (3.6)-(3.9), which accommodates the finite window over which acute GVHD is defined, to the following models: (1) our proposed model without a shared frailty; (2) the model presented in Conlon et al. combining a cure fraction and illness-death model (Conlon et al., 2014) with an added shared frailty; (3) the model presented in Conlon et al.; (4) illness-death model with shared frailty (Xu et al., 2010; Lee et al., 2015); (5) semiparametric cure model for a univariate outcome (Sy and Taylor, 2000; Peng and Dear, 2000; Cai et al., 2012); and (6) univariate Cox regression. Note that in our proposed model as well as models (1)-(4), we assume the baseline hazard functions correspond to Weibull distributions.

Analyses (2), (3) and (4) ignore the finite window over which acute GVHD is defined. Fitting these models will allow us to better understand the effect of mis-specifying time-to-acute GVHD. Analyses (5) and (6) leave the baseline hazard function unspecified, however, fail to account for death as a competing risk. Examining the estimates from these models will allow us to asses the impact of failing to account for death on the estimated regression coefficients corresponding to acute GVHD.

We estimate parameters and corresponding standard errors using the mean and standard deviation of the sampling distribution of each parameter, respectively. We estimate the coverage probability by the average number of times the true parameter value falls into a Wald-based 95% confidence interval based on the estimated parameters and standard errors.

All statistical analyses were conducted in the R environment for statistical computing, version 3.2.2 (R Core Team, 2014). In our implementation of our proposed method and models (1) and (2), we use: the `optim` function in the `stats` package for quasi-Newton non-linear optimization, respectively; and the `weibreg` function in the `eha` package for univariate Weibull regression (Brostrm, 2016). We use the `FreqID_HReg` function in the `SemiCompRisks` package (Lee and Haneuse, 2016) to fit an illness-death model with shared frailities in model (3). We use the `smcure` function in the `smcure` package (Cai et al., 2012) for semiparametric cure model estimation. Note that the latent susceptibility indicator in `smcure` assigns value one to patients that are susceptible. Recall that we define the susceptibility status with $L = 0$, so that estimates from the logistic regression model for susceptibility from `smcure` are negated as to be comparable to the estimates from our proposed model.

2.5.4 Results

We present the parameter estimates from the simulations corresponding to the base scenario in Table 2.2. The estimates from fitting our proposed model show negligible bias. Failing to account for the shared frailty in the model fitting (see 'Proposed, no frailty') resulted in bias across half of the parameters estimates, notably the intercept of the cure model and the regression parameters corresponding to the 14- and 24-transitions with approximately 20% bias. Not accounting for the finite support of acute GVHD (see 'Conlon, frailty') resulted in biased estimation for many of the parameters, with over 20% bias in the regression estimates corresponding to the intercept of the cure model, and the 14- and 23-transitions. Fitting the model proposed by Conlon et al., which fails to account for the finite interval and omits the frailty term, results in bias across most of the Weibull parameters, however, several of the regression parameter estimates are close to the truth. Fitting a standard illness-death model to the data resulted in bias across all parameter values with regression coefficients corresponding to the 23- and 24-transitions that are in the opposite direction of the true values. Ignoring the competing risk death via a univariate analysis (standard cure and Cox models) resulted in biased estimation, with the regression coefficient in the Cox model in the opposite direction as the truth.

Table 2.2: Simulation results corresponding to the 'base scenario', with 1000 replicates of size 5000. Parameter estimates calculated as mean of sampling distributions.

	Truth	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon, no frailty	Illness-death	Univ. Cure	Univ. Cox
Intercept	-0.41	-0.41	-0.31	-0.55	-0.40		0.25	
X_s	0.50	0.50	0.51	0.53	0.54		0.42	
X_{14}	0.25	0.25	0.20	0.33	0.23			
X_{24}	0.50	0.50	0.41	0.52	0.45	-0.08		
X_{23}	0.25	0.25	0.25	0.30	0.25	-0.31	0.32	-0.26
X_{34}	0.15	0.15	0.14	0.16	0.14	0.11		
$\log(\kappa_{14})$	-8.52	-8.51	-7.06	-9.06	-7.41			
$\log(\alpha_{14})$	0.34	0.33	0.14	0.38	0.18			
$\log(\kappa_{24})$	-5.30	-5.30	-5.17	-5.27	-5.09	-5.35		
$\log(\alpha_{24})$	0.34	0.34	0.23	0.30	0.25	0.08		
$\log(\kappa_{23})$	-2.16	-2.17	-2.31	-1.81	-1.77	-1.98		
$\log(\alpha_{23})$	-0.69	-0.69	-0.75	-0.63	-0.64	-1.10		
$\log(\kappa_{34})$	-6.21	-6.22	-5.66	-6.20	-5.66	-7.01		
$\log(\alpha_{34})$	0.26	0.26	0.15	0.26	0.15	0.37		
$\log(\theta)$	-1.71	-1.73		-2.00		-0.49		

Estimates of the standard errors (empirical and analytical) and coverage probabilities for the base scenario simulations are included in the Supplementary Material. For our proposed method and all other analyses with one exception, the empirical standard errors align with the means of the analytical standard errors. However, for Analysis (2), the model proposed by Conlon and colleagues with an added shared frailty, the empirical standard errors are generally much larger than the mean of the analytical standard errors; this is due to the existence of a small number of outliers that skew the estimates of the standard deviations of the sampling distributions. When the empirical standard errors are calculated using the median absolute deviation, the empirical standard deviations align with the medians of the sampling distributions of the parameter estimates. For our proposed method, all parameters achieve nearly 95% coverage. For the other analyses considered, coverage is generally poor due to biased estimates discussed earlier.

The results from the four other simulation scenarios described in Section 2.5.2 are included in the Supplementary Materials. The results from these simulations generally align with the finding from the base scenario setting, with respect to bias and coverage. However, when censoring is added (either via an exponentially distributed variable with mean 100 or administratively at 365 days), the standard errors of the parameter estimates are generally larger than those corresponding to the base scenario for all analyses, which

is to be expected. Increasing the cure fraction from 46% in the base scenario to 75% redistributed the data from the 24-, 23- and 34-transitions to the 14-transition, thereby decreasing the efficiency of the 24-, 23- and 34-transitions parameters while increasing the precision of 14-transition parameter estimates. Because the main difference between our proposed model and the model proposed by Conlon and colleagues with an added frailty, Analysis (2), is the finite interval over which the 23-transition is defined, for the last simulation scenario, we chose Weibull parameters for the 23-transition so that the truncated Weibull and standard Weibull distributions are similar. The regression estimates corresponding to Analysis (2) have negligible bias, however, the estimates for the intercept in the cure model, α_{14} and the gamma frailty parameter, θ, all have more than 10% bias, with corresponding coverage probabilities that are below the nominal level.

Overall, our simulations illustrate that for data generated under our proposed model, failing to account for the finite time interval over which acute GVHD is defined, include a shared frailty term, and/or incorporate death can lead to bias estimation and suboptimal coverage probabilities. The main distinction between our model and the model proposed by Conlon et al. with the added frailty is that the baseline hazard function for the 24- and 23-transitions are defined on $(0, \tau)$. Thus, differences in estimation and inference between the two models are due to the differences in baseline hazard specification.

2.6 Analysis of data from CIBMTR

We illustrate our proposed methodology through an analysis of stem cell transplant data presented in Section 2.2. We fit the data to our proposed model, including dummy variables indicating disease type, disease stage, and HLA compatibility group as covariates in all transitions. We compare the results of our proposed model to the following competing models: the model proposed by Conlon et al. with an added frailty; an illness-death model with shared gamma frailty; a semiparametric cure model for acute GVHD; and a Cox proportional hazards model. The first two are multi-state models with a shared gamma frailty, both of which fail to account for the finite interval over which acute GVHD is defined and the latter of which also fails to model the fraction of subjects who are

never diagnosed with acute GVHD. The latter two models are semiparametric univariate models; the cure model accounts for the subset of patients that are diagnosed with the condition, whereas the Cox model does not.

We present the results from our analyses in Table 2.3. For our proposed model, the estimates for Advanced stage are associated with greater risk for both outcomes than the estimates for intermediate stage, for all transitions. Specifically, for the 14-transition, the estimated regression coefficients corresponding to Intermediate and Advanced stage disease are 0.30 and 1.13, respectively. In examining the effects of HLA compatibility, the regression estimates corresponding to a 7/8 donor match are generally associated with higher risk than an 8/8 match. Of note, for the 34-transition, the estimated regression coefficients corresponding to the HLA 7/8 and 8/8 match indicator variables are 0.29 and 0.08, respectively. The parameter estimates and standard errors corresponding to our proposed model are comparable to the results from assuming the model proposed by Conlon et al. with added frailty for the cure fraction, 14- , 23- and 34-transitions. However, for the 24-transition, the effects are several folds smaller for our proposed model with the regression parameter corresponding to the dummy variable Intermediate Stage in the opposite direction as in the Conlon model with frailty. There are also sizable differences in Weibull parameter estimates between the two models, with a 43% decrease in magnitude of the estimate of θ comparing our proposed model with the one proposed by Conlon et al. The standard errors for the Conlon model are much larger than our proposed model for the 24-transition. The comparison of log-likelihood values between our proposed model and the model proposed by Conlon et al. with a shared frailty shows that our model has a better fit to the data, with a log-likelihood value of -39542.03 compared to -39611.93 for the Conlon model with frailty.

The standard illness-death model with shared frailty fails to account for the cure fraction in the joint analysis of acute GVHD and death. In doing so, the model assumes that all subjects will receive a diagnosis of acute GVHD over time, which does not align with the real data. The corresponding estimates are dramatically different from those of our proposed model for nearly all parameters. The estimate for θ is 75% smaller in magnitude for the illness-death model than our proposed model. Using a univariate approach to

modeling acute GVHD (cure and Cox models) results in estimates that also differ greatly from the estimates from fitting the proposed model.

Overall, this analysis illustrates that estimation can vary widely across the models considered. As a specific example, consider the regression parameter corresponding to CML used in the modeling of the 23-transition. The estimated value for the five models considered (our proposed model, Conlon with frailty, illness-death, cure, Cox) are 0.12, 0.09, 0.63, -0.54, 0.35. We advocate for the use of our proposed model in for this data set as it specifically addresses both the finite interval over which acute GVHD is defined and the fraction of subjects who never receive a diagnosis of acute GVHD.

2.7 Discussion

We proposed a novel multi-state model that accounts for mortality through joint modeling of acute GVHD and death, and explicitly acknowledges the finite time scale in which acute GVHD occurs. Our simulations show that fitting competing models that do not account for these two factors can lead to bias and compromised inference. This is supported by our data example where wildly different parameter estimates were observed across the different models considered.

Although the focus of this paper is acute GVHD, we emphasize that the key phenomenon that the event of interest is only well-defined on a finite interval is more general. For example, brief psychotic disorder (APA et al., 2013) is defined within four weeks after exposure to a traumatic event, and acute respiratory distress syndrome is characterized by lung injury within one week of lung injury (Ferguson et al., 2012). A key component of hospital quality of care assessments are rates of hospital readmission within 30 days of discharge (Kim et al., 2012).

Note that an attractive application of our fully-parametric proposed model is that the joint density can be determined on the full support of (T_1, T_2), so that likelihood based inference can be performed. Moreover, the joint density can be used to calculate patient-specific absolute risk profiles for acute GVHD and death simultaneously. Specifically, note that at any time t following transplantation, a patient falls into one of the following

Table 2.3: Parameter estimates and standard errors from analysis of stem cell transplantation data fitting our proposed model and four other competing models.

	Proposed		Conlon, with frailty		Illness-death		Cure		Cox	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
π: Cure fraction										
Intercept	1.94	0.06	2.07	0.06			-4.68	20.13		
Intermediate stage	-0.14	0.07	-0.15	0.07			0.59	15.48		
Advanced stage	-0.33	0.09	-0.35	0.08			1.05	9.57		
HLA: 7/8	-0.75	0.08	-0.77	0.09			1.15	17.77		
HLA: 8/8	-0.42	0.07	-0.47	0.07			1.15	14.28		
Disease type: ALL	-0.17	0.08	-0.20	0.08			0.59	13.81		
Disease type: CML	-0.44	0.08	-0.49	0.08			-15.31	11.41		
Disease type: MDS	-0.25	0.10	-0.29	0.10			0.85	13.03		
14: Cured, HCT − Death										
Intermediate stage	0.30	0.06	0.27	0.06						
Advanced stage	1.13	0.06	1.05	0.06						
HLA: 7/8	0.32	0.07	0.31	0.07						
HLA: 8/8	0.14	0.05	0.11	0.05						
Disease type: ALL	-0.08	0.06	-0.09	0.06						
Disease type: CML	-0.53	0.09	-0.52	0.08						
Disease type: MDS	-0.52	0.08	-0.50	0.08						
14: Not cured, HCT − Death										
Intermediate stage	-0.04	0.20	0.68	0.36	0.43	0.13				
Advanced stage	0.19	0.23	1.24	0.37	1.63	0.13				
HLA: 7/8	0.68	0.22	1.28	0.49	1.55	0.15				
HLA: 8/8	0.37	0.19	1.19	0.46	1.02	0.12				
Disease type: ALL	0.63	0.20	1.37	0.43	0.49	0.14				
Disease type: CML	0.07	0.23	1.09	0.48	0.08	0.16				
Disease type: MDS	0.11	0.28	0.75	0.47	-0.03	0.17				
23: Not cured, HCT − Acute GVHD										
Intermediate stage	-0.30	0.09	-0.30	0.08	0.01	0.11	0.34	0.31	0.05	0.06
Advanced stage	-0.24	0.10	-0.22	0.08	0.51	0.11	-0.13	0.25	0.21	0.06
HLA: 7/8	0.44	0.11	0.42	0.10	1.24	0.13	0.24	0.33	0.66	0.07
HLA: 8/8	0.44	0.09	0.37	0.08	0.87	0.10	-0.15	0.24	0.40	0.06
Disease type: ALL	0.13	0.10	0.09	0.09	0.33	0.12	0.24	0.24	0.07	0.07
Disease type: CML	0.12	0.10	0.09	0.09	0.63	0.13	-0.54	0.38	0.35	0.07
Disease type: MDS	0.04	0.12	0.02	0.10	0.33	0.14	-0.19	0.67	0.19	0.08
34: Not cured, Acute GVHD − Death										
Intermediate stage	0.15	0.09	0.15	0.09	0.33	0.13				
Advanced stage	0.59	0.09	0.58	0.08	1.22	0.13				
HLA: 8/8	0.29	0.10	0.29	0.09	0.85	0.14				
HLA: 8/8	0.08	0.08	0.09	0.08	0.37	0.12				
Disease type: ALL	0.01	0.10	0.01	0.09	0.12	0.14				
Disease type: CML	0.15	0.10	0.15	0.10	0.42	0.14				
Disease type: MDS	0.11	0.11	0.09	0.10	0.30	0.16				
Weibull baseline hazard parameters										
$\log(\kappa_{14})$	-7.98	0.19	-7.56	0.17						
$\log(\alpha_{14})$	0.14	0.03	0.08	0.03						
$\log(\kappa_{24})$	-13.52	0.11	-11.77	0.95	-7.43	0.17				
$\log(\alpha_{24})$	0.80	0.05	0.73	0.08	0.32	0.03				
$\log(\kappa_{23})$	-6.83	0.18	-6.58	0.18	-14.14	0.27				
$\log(\alpha_{23})$	0.66	0.02	0.62	0.03	1.05	0.02				
$\log(\kappa_{34})$	-5.43	0.16	-5.33	0.15	-7.91	0.19				
$\log(\alpha_{34})$	-0.10	0.03	-0.13	0.03	0.07	0.03				
$\log(\theta)$	-0.90	0.08	-1.47	0.15	2.31	0.04				
Log-likelihood	-39542.03		-39611.93							

four disjoint categories: (1) alive with a diagnosis of acute GVHD; (2) dead, carrying a diagnosis of acute GVHD; (3) dead without acute GVHD; or (4) alive without acute GVHD. The probability of each of the four categories can be estimated through integration of the joint density. Paired with a decision rule, the set of joint absolute risks could be used to classify patients into one of the four joint outcome categories.

If the main estimates of interest are those corresponding to the 23-transition (HCT$-$acute GVHD), our simulation and data example results provide examples where the regression estimates from fitting the model proposed by Conlon with an added shared frailty align with those of our proposed model. As our results illustrate, the model that does not acknowledge the finite time interval in which acute GVHD occurs may in some cases be adequate for estimation, however, correctly defining the nonterminal event on $(0, \tau)$ is crucial for risk prediction. If the nonterminal event is instead defined on the entire real line, then the joint density would contain mass over the entire upper wedge $(0 < T_1 < T_2)$. It is not immediately clear how the probability mass in the region $(T_1 > \tau) \cap (T_2 > \tau)$ should be assigned when calculating four absolute risks described in the previous paragraph. A model that fails to define the nonterminal event on (0,τ) is clearly mis-specified in the setting of risk prediction.

Joint risk prediction in the semicompeting risks setting

Catherine Lee

Department of Biostatistics

Harvard T.H. Chan School of Public Health

Sebastien Haneuse

Department of Biostatistics

Harvard T.H. Chan School of Public Health

3.1 Introduction

Semi-competing risks refers to the setting where interest lies the the time-to-event for some so-called *non-terminal* event, the observation of which is subject to some *terminal* event (Fine et al., 2001). In contrast to standard competing risks, where each of the outcomes under consideration is typically terminal (e.g. death due to some cause or another), in the semi-competing risks setting it is possible to observe both events on the same study unit, so that there is at least partial information on their joint distribution (Fine et al., 2001; Xu et al., 2010). As an example, patients who have undergone hematopoetic stem cell transplantation (HCT) for the treatment of leukemia may experience graft-versus-host disease (GVHD), a debilitating condition associated with significant morbidity and compromised quality of life (Lee et al., 2006; Shlomchik, 2007; Joseph et al., 2008; Ferrara et al., 2009). Unfortunately, mortality is also high among patients undergoing HCT. As such, studies seeking to investigate risk factors for GVHD must contend with death as a competing risk. Other examples are the study of readmission following a diagnosis of heart failure (Haneuse and Lee, 2016b) and the study of Alzheimer's disease among the elderly (Jacqmin-Gadda et al., 2014).

Towards the analysis of semi-competing risks data, the statistical literature has focused on three broad frameworks that seek to exploit the joint information T_1, the time to the non-terminal event, and T_2, the time to the terminal event (Varadhan et al., 2014). First, building on the original paper by Fine et al. (2001), a number of authors have pursued modeling the joint distribution of (T_1, T_2) by first developing models for the marginal distributions of T_1 and T_2 and then inducing dependence via a copula (Wang, 2003; Lakhal et al., 2008; Peng and Fine, 2007; Hsieh et al., 2008). Recent developments include the incorporation of covariates (Peng and Fine, 2007; Hsieh et al., 2008) and methodology to account for left truncation (Peng and Fine, 2006). The philosophy underpinning this approach is that patients begin in some initial state at time zero and may transition into the 'non-terminal' and/or 'terminal' state. Analyses typically proceed through the development of models for transition-specific hazard functions (which dictate the rate at which patients experience the events), often with the use of subject-specific frailties as a means

to capture heterogeneity across subjects (and thus within-subject dependence between T_1 and T_2). Recent work in this area has sought to generalize illness-death model specifications to accommodate a broad range of dependence structures, including clustering of study participants (Han et al., 2014; Lee et al., 2016). Finally, a more limited literature has sought to embed the analysis of semi-competing risks data within the causal inference paradigm (Egleston et al., 2007; Tchetgen Tchetgen, 2014). Sometimes referred to as accounting for *truncation by death*, the approach proceeds by first defining counterfactual outcomes for both the non-terminal and terminal events, the defining an causal contrast of interest (e.g. the so-called survivor adjusted causal effect) together with identifying conditions, and then developing infrastructure for valid estimation and inference.

While the analysis of semi-competing risks data has received considerable recent attention, the vast majority of this work has focused on estimation and inference for risk factor associations and in dealing within non-standard data settings (e.g. left truncation and/or interval censoring). Moreover, very little attention has been paid to risk prediction in the semi-competing risks context. Jacqmin-Gadda et al. (2014) used an illness-death model to develop a prognostic score for predicting risk of dementia while accounting for the competing risk of death, while Han et al. (2014) used an illness-death model with random effects to calculate the marginal survival probability of the terminal event at any given time. Common to both of these papers is that, while the underlying modeling structure considered both events, the output (in terms of a numerical prediction) focused on just one of the events. That is, neither paper sought to calculate risk for the two events simultaneously. Thus information on the joint risk of the two events that could, in principle, be useful in clinical contexts is not reported or even calculated. As an example, consider again patients who undergo HCT for the treatment of leukemia. The quantification of the joint risk of GVHD and death could have a number of potentially important uses, particularly towards enabling individualized patient-centered decisions. First, estimating the absolute risk of the two events as a function of the interplay between the characteristics of the patient and potential unrelated donors could help inform decisions about whether to pursue transplantation, and which donor to select. Second, this information could be useful making decisions regarding post-transplant immunosuppressive strategies, partic-

ularly for patients at low risk of GVHD for whom aggressive immunosuppression may not be necessary. Finally, from a research perspective, the quantification of joint risk could also be helpful in identifying patients for recruitment in clinical trials.

Moving forward, we propose to consider prediction in the semi-competing risks setting through the calculation and evaluation of patient-specific *absolute risk profiles* for the non-terminal and terminal events simultaneously. In particular, we note that at any given point in time after the initiating event/state, a patient will have: (1) experienced both events; (2) experienced the non-terminal event but not the terminal event; (3) experienced the terminal event without having experienced the non-terminal event; or (4) experienced neither event. Thus, in contrast to much of the prediction literature, we consider the task of prediction as being one where we seek to classify patients at any given point in time into one of four categories based on a vector of probabilities that add to 1.0. To the best of our knowledge, only one paper by Putter et al. (2007) has sought to calculate and report such a profile. We build on that work in two important ways. First, while the illness-death model considered by Putter and colleagues permitted the inclusion of covariates into each of the transition-specific hazard functions, it did not permit the inclusion of a subject-specific frailty. Not including a frailty simplifies estimation of the model components, since each transition is treated independently, and thus also risk prediction. While, as mentioned, the inclusion of frailties can be useful in explaining residual heterogeneity, the calculation of risk prediction is more complex because of the dependence that is induced. Our first contribution, therefore, is to develop a framework for calculation of risk profiles based on an illness-death model that incorporates a subject-specific frailty. Second, although Putter and colleagues report estimated risk profiles, they do not consider the evaluation of the predictive performance of the profiles. While the evaluation of risk prediction tools is well-established when the outcome is binary (Steyerberg et al., 2010), less work has been published on methods for multi-category outcomes (Scurfield, 1996; Mossman, 1999; Dreiseitl et al., 2000; Nakas and Yiannoutsos, 2004; Li and Fine, 2008) and none specifically in the semi-competing risk setting. The second contribution of this paper, therefore, is to propose a framework for evaluation of predictive performance for risk profiles based on the *hypervolume under the manifold* (HUM) statistic, an extension

of the well-known area-under-the-curve (AUC) statistic for univariate binary outcomes. As part of this, we propose a method for estimating the HUM statistic in the presence of potential verification bias which arises when the true outcome category is unknown (Alonzo and Pepe, 2005; Chi and Zhou, 2008; Duc et al., 2016; He et al., 2009; Rotnitzky et al., 2006; Zhang and Alonzo, 2016). Throughout, we illustrate the proposed methods using a motivating dataset of N=1,416 patients who underwent HCT between 1999 and 2011.

3.2 Acute GVHD following HCT

Data were abstracted from databases maintained by the Center for International Blood and Marrow Transplant Research, an international network of stem cell transplant registries on 1,416 patients who underwent first HCT between the ages of 20 and 59, with tissue from a donor that is an identical sibling or an 8/8 HLA match for treatment of intermediate or advanced stage acute myelogenous leukemia (AML) between 1999 and 2011. Across these patients, after administrative censoring at 365 days, the median (minimum, maximum) observed follow-up was 332 (2, 365) days. Although patients undergoing HCT may be diagnosed with either acute or chronic GVHD (Filipovich et al., 2005), for the purposes of this paper we focus on the former. From Table 3.1 we see that approximately 17.9% of the 1,416 patients experienced an acute GVHD event during the observed follow-up period. We also see that mortality is very high among these patients, with 62.3% being observed to die during the observed follow-up.

In practice, a patient may be diagnosed with either *classic* acute GVHD or *persistent, recurrent, or late-onset* acute GVHD, depending on the specific symptoms that manifest as well as the timing of the diagnosis (Filipovich et al., 2005). When interest lies in classic acute GVHD, an interesting feature is such a diagnosis is only given within 100 days of transplantation. That is, according to current consensus criteria, a patient should not be diagnosed with classic acute GVHD beyond 100 days post-transplantation. As such, in contrast to standard survival models where it is assumed that the time-to-event is defined on the entire real line, time to classic acute GVHD is only well-defined (at least clinically)

Table 3.1: Description of patients with AML between the ages of 20 and 60 years of age who underwent stem cell transplantation with an HLA identical sibling or 8/8 donor.

			Observed outcome category, %			
	N	%	Both acute GVHD & death	Acute GVHD & censored for death	Death without acute GVHD	Censored for both
Total	1,416		14.0	3.9	48.3	33.8
Gender						
Male	757	53.5	14.8	4.1	47.8	33.3
Female	659	46.5	13.1	3.6	48.9	34.4
Age, years						
20-29	266	18.8	15.0	3.8	44.8	36.5
30-39	280	19.8	16.5	4.3	38.9	40.4
40-49	423	29.9	10.9	3.5	47.8	37.8
50-59	447	31.6	14.8	4.0	46.9	24.4
Disease status						
Intermediate	603	42.6	9.9	4.6	36.7	48.8
Advanced	813	57.4	17.0	3.3	46.9	22.8
Karnofsky score						
90-100	804	56.8	11.8	5.1	44.1	38.9
<90	612	43.2	16.8	2.3	53.8	27.1
HLA compatibility						
Identical sibling	598	42.2	11.5	4.3	49.8	34.3
8/8	818	57.8	15.8	3.5	47.2	33.5
Conditioning intensity						
Myeloblative	1,091	77.0	14.1	4.3	47.1	34.5
Reduced/non-myeloblative	325	23.0	13.5	2.5	52.3	31.7
In vivo T-cell depletion						
No	1,133	80.0	14.9	4.1	46.7	34.2
Yes	283	20.0	10.2	2.8	54.8	32.2

on the finite interval of (0, 100] days.

In Chapter 2, we proposed a novel multi-state model that simultaneously: (i) accounts for mortality through joint modeling of acute GVHD and death, and (ii) explicitly acknowledges the finite time scale in which the event of interest can take place. We provided methodology for estimation and inference via maximum likelihood estimation, for which employ in the development of joint risk prediction of acute GVHD and deaths simultaneously.

3.3 Joint risk prediction

We present novel methodology for joint risk prediction in the semicompeting risk framework for the shared frailty illness-death model proposed by Xu et al. (2010). We focus on predictions of both events simultaneously at time $T_1 = T_2 = t$, where predictions

are generated at the time of the initiating event/origin. In the stem cell transplant example, this corresponds to predicting the joint absolute risk of acute GVHD and death at t-days post-transplantation, with predictions made at the time of transplantation. We provide parallel developments for both the shared frailty illness-death model and the finite interval cure model.

3.3.1 Standard shared frailty illness-death model

Model

The illness-death model with shared gamma frailty is a multi-state model where subjects are viewed as making one or more of the following transitions: (1) initiating event to nonterminal event; (2) initiating event to terminal event; (3) nonterminal event to terminal event. A schematic of the model is presented in Figure 3.1. Let T_1 and T_2 denote the time between the initiating event and the nonterminal and terminal event times, respectively. The hazard functions corresponding to the three transitions are defined by

$$\begin{aligned}
\lambda_1(t_1) &= \lim_{\Delta \to 0} P(T_1 \in [t_1, t_1 + \Delta) | T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } t_1 > 0, \\
\lambda_2(t_2) &= \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta) | T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } t_2 > 0, \\
\lambda_3(t_2|t_1) &= \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta) | T_1 = t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < t_2.
\end{aligned}$$

One way in which the transition hazards can be modeled is through a function of subject-specific covariates and a shared-frailty

$$\lambda_1(t_{1i}; X_{1,i}, \gamma_i) = \gamma_i \cdot \lambda_{0,1}(t_{1i}) \exp\{\beta_1 X_{1,i}\}, \text{ for } t_{1i} > 0 \tag{3.1}$$

$$\lambda_2(t_{2i}; X_{2,i}, \gamma_i) = \gamma_i \cdot \lambda_{0,2}(t_{2i}) \exp\{\beta_2 X_{2,i}\}, \text{ for } t_{2i} > 0 \tag{3.2}$$

$$\lambda_3(t_{2i}; t_{1i}, X_{2,i}, \gamma_i) = \gamma_i \cdot \lambda_{0,3}(t_{2i}|t_{1i}) \exp\{\beta_3 X_{3,i}\}, \text{ for } 0 < t_{1i} < t_{2i}, \tag{3.3}$$

where $X_{g,i}$ are sets of covariates that may impact the transition hazards λ_g, for $g = 1, 2, 3$, and γ_i is a shared Gamma frailty with mean 1 and variance θ. The role of the shared frailty term is two-fold; it induces an additional level of dependence between T_1 and T_2, and accounts for additional heterogeneity across subjects analogous to random effects in a mixed model. In general, $\lambda_{0,3}(t_2|t_1)$ can be defined as any function of t_1 and

t_2. A Markov model assumes that $\lambda_{0,3}(t_2|t_1) = \lambda_{0,3}(t_2)$. In this paper, we assume that $\lambda_{0,3}(t_2|t_1) = \lambda_{0,3}(t_2 - t_1)$, a function of the sojourn time, which is referred to as the semi-Markov property.

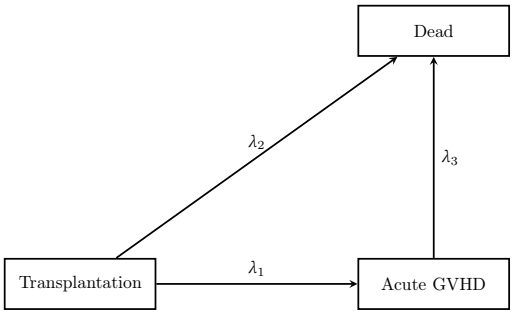

Figure 3.1: Schematic of illness-death model.

The transition hazards in (3.1), (3.2) and (3.3) determine the joint distribution of (T_1, T_2). We denote the joint density on the observable region $T_1 < T_2$, termed the upper wedge, by $f_U(t_1, t_2)$, for $t_1 > t_2$. We use the convention introduced by Xu et al. (2010) and assign the remaining probability mass along the line $T_1 = \infty$, with density denoted $f_{T_1=\infty,T_2}(t_2)$, representing subjects who never experience the nonterminal event T_1.

Several methods have been proposed for fitting the model in (3.1)-(3.3). Xu et al. (2010) left the baseline hazard functions unspecified and obtained estimates through maximum likelihood estimation. Lee et al. (2015) and Han et al. (2014) assumed flexible, piecewise exponential baseline hazard functions and used Bayesian methods for estimation. For the purpose of risk prediction, the form of the baseline hazard must be known on the entire domain of the transition, $t_1 > 0$ and $t_2 > 0$, as will be evident in Section 3.3.1.

Joint density

The joint density for the illness-death model with shared gamma frailty was derived by Lee et al. (2015) and is presented below.

$$f_U(t_1, t_2|\gamma) = \lambda_1(t_1|\gamma)\lambda_3(t_2|t_1, \gamma) \exp\left\{-\int_0^{t_1} [\lambda_1(v|\gamma) + \lambda_2(v|\gamma)]\, dv - \int_{t_1}^{t_2} \lambda_3(v|t_1, \gamma)\, dv\right\}$$

$$f_{T_1=\infty,T_2}(t_2|\gamma) = \lambda_2(t_2|\gamma) \exp\left\{-\int_0^{t_1} [\lambda_1(v|\gamma) + \lambda_2(v|\gamma)]\, dv\right\}$$

Consolidating the previous two expressions, we denote the joint density by the following expression

$$f(t_1, t_2|\gamma) = f_U(t_1, t_2|\gamma) \cdot \mathbb{1}(t_1 < t_2) + f_{T_1=\infty,T_2}(t_2|\gamma) \cdot \mathbb{1}(T_1 = \infty). \tag{3.4}$$

Patient-specific risk profiles

We present the methods for calculating the absolute risk profile for both the nonterminal and terminal events at any time t following the initiating event. This is done via integration of the joint density in Section 3.3.1. We define the four components of the joint absolute risk, conditional on γ, as follows:

$$\begin{aligned}
p_{(1)}(t|\gamma) &= P(\text{experienced nonterminal event, but not the terminal event at time } t|\gamma) \\
&= \int_t^\infty \int_0^t f_U(u, v|\gamma)\, du\, dv \\
p_{(2)}(t|\gamma) &= P(\text{experienced both events by time } t|\gamma) \\
&= \int_0^t \int_u^t f_U(u, v|\gamma)\, dv\, du \\
p_{(3)}(t|\gamma) &= P(\text{experienced terminal event without nonterminal event by time } t|\gamma) \\
&= \int_0^t f_{T_1=\infty,T_2}(v|\gamma)\, dv \\
p_{(4)}(t|\gamma) &= P(\text{neither events occur by time } t|\gamma) \\
&= \int_t^\infty f_{T_1=\infty,T_2}(v|\gamma)\, dv + \int_t^\infty \int_u^\infty f_U(u, v|\gamma)\, dv\, du.
\end{aligned}$$

A plot describing the bounds of integration is included in Figure 4.1 in the Supplementary Materials. Note that the regions of integration are disjoint and exhaustive so that $p_{(1)}(t|\gamma)+p_{(2)}(t|\gamma)+p_{(3)}(t|\gamma)+p_{(4)}(t|\gamma)$=1. The vector, $(p_{(1)}(t|\gamma), p_{(2)}(t|\gamma), p_{(3)}(t|\gamma), p_{(4)}(t|\gamma))$, is conditional on the patient-specific frailties and covariates, whereby we refer to as patient-specific risk profiles.

We also consider the marginal subject-specific risk profiles defined by

$$p_{(j)}(t) = \int_\gamma p_{(j)}(t|\gamma) \cdot f_\gamma(\gamma)\, d\gamma,$$

which averages the risk profiles, $(p_{(1)}(t|\gamma), p_{(2)}(t|\gamma), p_{(3)}(t|\gamma), p_{(4)}(t|\gamma))$, over all values of γ.

3.3.2 Finite interval shared frailty illness-death model

Model

As discussed in Section 3.2, an analysis of acute GVHD requires careful treatment of the competing force of death and special features of the outcome. We addressed these issues in Chapter 2. A representation of the model is presented in Figure 3.2. We briefly review the model below.

Let a binary variable L_i represent a latent cure state, such that $L_i = 1$ indicates that a subject that will not experience acute GVHD and $L_i = 0$ indicates that a subject that is susceptible to acute GVHD. Note that if $L_i = 0$, it is assumed that the subject will receive a diagnosis of acute GVHD, in the absence of censoring or death. The cure state accommodates a mixture distribution of subjects who experience the event of interest and those who never do, which is observed in acute GVHD among transplant recipients. Note that the cure indicator is partially observed; if a patient is alive at any time after day 100 post-transplant, it must be that $L_i = 1$. The cure fraction, π_i, is modeled using logistic regression, where

$$\pi_i = P(L_i = 1|X_{s,i}) = g^{-1}(\beta_s X_{s,i}), \tag{3.5}$$

which is the probability that a patient will not develop acute GVHD within τ days of transplant, and X_s denotes a vector of covariates that may impact acute GVHD susceptibility. Subjects who are not cured of acute GVHD are at risk of both acute GVHD and death, whereas, subjects who are cured of acute GVHD are only at risk for death.

As in Section 3.3.1, time to acute GVHD and death are jointly modeled via a multistate model depicted in Figure 3.2 with the following states: 1. not susceptible to acute GVHD at transplantation; 2. susceptible to acute GVHD at transplantation; 3. acute GVHD; 4. death. As before, let T_1 and T_2 denote the time to acute GVHD and death, respectively, where now T_1 is defined on $(0, \tau)$, $\tau = 101$. If a patient experiences the terminal event without experiencing the nonterminal event, we set $T_1 = \tau$. Note that in the standard illness-death model $\tau = \infty$. Let X_i denote a vector of patient-specific baseline covariate values. The transition hazards are defined by

$$\lambda_{14}(t_2) = \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta)|T_2 \geq t_2)/\Delta, \text{ for } t_2 > 0$$

$$\lambda_{24}(t_2) = \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta) | T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_2 < \tau$$
$$\lambda_{23}(t_1) = \lim_{\Delta \to 0} P(T_1 \in [t_1, t_1 + \Delta) | T_1 \geq t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < \tau$$
$$\lambda_{34}(t_2|t_1) = \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta) | T_1 = t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < t_2,$$

where the λ_{kj} denotes the hazard corresponding to the transition from state k to state j . Note that λ_{23}, λ_{24}, and λ_{34} correspond to λ_1, λ_2 and λ_3 in Section 3.3.1.

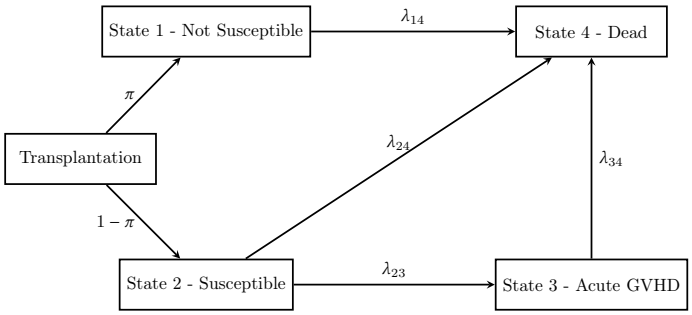

Figure 3.2: Schematic of illness-death model.

For patients who are cured of (nonsusceptible) acute GVHD at the time of transplantation, the hazard function corresponding to the time from transplantation to death is modeled by

$$\lambda_{14}(t_{2i}; X_{14,i}, L_i = 1) = \gamma_i \cdot \lambda_{0,14}(t_{2i}) \exp\{\beta_{14} X_{14,i}\}, \text{ for } t_{2i} > 0, \tag{3.6}$$

where γ is a subject-specific frailty following a Gamma distribution with mean 1 and variance θ and $X_{14,i}$ are vector of covariates impacting susceptibility status.

For patients who are not cured (susceptible) of acute GVHD at the time of transplantation, the hazard functions corresponding to the remaining three transitions are defined by:

$$\lambda_{24}(t_{2i}; X_{24,i}, L_i = 0) = \gamma_i \cdot \lambda_{0,24}(t_{2i}) \exp\{\beta_{24} X_{24,i}\}, \text{ for } 0 < t_{2i} < \tau \tag{3.7}$$
$$\lambda_{23}(t_{1i}; X_{23,i}, L_i = 0) = \gamma_i \cdot \lambda_{0,23}(t_{1i}) \exp\{\beta_{23} X_{23,i}\}, \text{ for } 0 < t_{1i} < \tau \tag{3.8}$$
$$\lambda_{34}(t_{2i}; t_{1i}, X_{34,i}, L_i = 0) = \gamma_i \lambda_{0,34}(t_{2i} - t_{1i}) \exp\{\beta_{34} X_{34,i}\}, \text{ for } 0 < t_{1i} < t_{2i}, \tag{3.9}$$

where γ is the same subject-specific frailty that appears in (3.6) and X_{24}, X_{23} and X_{34} are covariates that are possibly associated with the hazard functions of the corresponding

transitions. These equations correspond to (3.2), (3.1) and (3.3), respectively, in Section 3.3.1.

The model is fit using maximum likelihood estimation, the details of which can be found in Chapter 2, Section 2.4.2.

Joint density

We derive the joint density corresponding to the model in Section 3.3.2. First let us introduce the notation $\Lambda_\omega(t) = \int_0^t \lambda_\omega(s)\, ds$. The joint density for (T_1, T_2) is given by:

$$\begin{aligned}
g(t_1, t_2|\gamma) &= g(T_1 = t_1, T_2 = t_2|\gamma) \\
&= P(L=1) \cdot g(t_1, t_2|L=1, \gamma) + P(L=0) \cdot g(t_1, t_2|L=0, \gamma) \\
&= \pi \cdot g(t_1, t_2|L=1, \gamma) + (1-\pi) \cdot g(t_1, t_2|L=0, \gamma) \\
&= \pi \cdot \lambda_{14}(t_2|\gamma) \cdot \exp\left\{-\int_0^{t_2} \lambda_{14}(v)\, dv\right\} \cdot \mathbb{1}(T_1 = \tau) + (1-\pi) \cdot f(t_1, t_2|\gamma) \\
&= \pi \cdot \lambda_{14}(t_2|\gamma) \cdot \exp\left\{-\int_0^{t_2} \lambda_{14}(v)\, dv\right\} \cdot \mathbb{1}(T_1 = \tau) + \\
&\quad (1-\pi) \cdot [\lambda_{23}(t_1|\gamma)\lambda_{34}(t_2 - t_1|t_1, \gamma) \cdot \\
&\qquad \exp\left\{-\Lambda_{23}(t_1|\gamma) - \Lambda_{24}(t_1|\gamma) - \int_0^{t_2-t_1} \lambda_{34}(v|t_1, \gamma)\, dv\right\} \cdot \mathbb{1}(t_1 < t_2) + \\
&\qquad \lambda_{24}(t_2|\gamma) \exp\left\{-\Lambda_{23}(t_2|\gamma) - \Lambda_{34}(t_2|\gamma)\right\} \cdot \mathbb{1}(T_1 = \tau)],
\end{aligned}$$

where $f(t_1, t_2|\gamma)$ represents the joint distribution corresponding to the standard illness-death model without the cure fraction in (3.4). We denote the joint density over the region $(0 < T_1 < T_2) \cap (0 < T_1 < \tau)$ by $g_U(t_1, t_2|\gamma)$. We assign the remaining probability mass over the line $T_1 = \tau$ with probability density denoted, $g_{T_1=\tau, T_2}(t_2|\gamma)$. The line $T_1 = \tau$ corresponds to patients who never receive a diagnosis of acute GVHD. Collecting like terms in the derivation of $g(t_1, t_2|\gamma)$ above, we see that the probability density along the line $T_1 = \tau$ is given by

$$\begin{aligned}
g_{T_1=\tau, T_2}(t_2|\gamma) = \pi \cdot \lambda_{14}(t_2|\gamma) \cdot \exp\left\{-\int_0^{t_2} \lambda_{14}(v)\, dv\right\} + \\
(1-\pi) \cdot \lambda_{24}(t_2|\gamma) \exp\left\{-\Lambda_{23}(t_2|\gamma) - \Lambda_{34}(t_2|\gamma)\right\}
\end{aligned}$$

and the probability density over the region $(0 < T_1 < T_2) \cap (0 < T_1 < \tau)$ is given by

$$g_U(t_1, t_2|\gamma) = (1-\pi) \cdot \lambda_{23}(t_1|\gamma)\lambda_{34}(t_2 - t_1|t_1, \gamma) \cdot$$

$$\exp\left\{-\Lambda_{23}(t_1|\gamma) - \Lambda_{24}(t_1|\gamma) - \int_0^{t_2-t_1} \lambda_{34}(v|t_1,\gamma)\, dv\right\}.$$

Patient-specific risk profiles

As in Section 3.3.1, we define the subject-specific risk profiles as follows.

$$\begin{aligned}
p_{(1)}(t|\gamma) &= P(\text{experienced nonterminal event, but not the terminal event at time } t|\gamma) \\
&= \int_t^\infty \int_0^t g_U(u,v|\gamma)\, du\, dv \\
p_{(2)}(t|\gamma) &= P(\text{experienced both events by time } t|\gamma) \\
&= \int_0^t \int_u^t g_U(u,v|\gamma)\, dv\, du \\
p_{(3)}(t|\gamma) &= P(\text{experienced terminal event without nonterminal event by time } t|\gamma) \\
&= \int_0^t g_{T_1=\tau,T_2}(v|\gamma)\, dv \\
p_{(4)}(t|\gamma) &= P(\text{neither events occur by time } t|\gamma) \\
&= \int_t^\infty g_{T_1=\tau,T_2}(v|\gamma)\, dv + \int_t^\infty \int_u^\tau g_U(u,v|\gamma)\, dv\, du.
\end{aligned}$$

A plot describing the bounds of integration is included in Figure 4.2 in the Supplementary Materials. In contrast to Section 3.3.1, the component $p_{(4)}(t|\gamma)$ is calculated over the finite support of T_1. As in Section 3.3.1, we define the marginal subject-specific risk profiles by

$$p_{(j)}(t) = \int_\gamma p_{(j)}(t|\gamma) \cdot f_\gamma(\gamma)\, d\gamma.$$

3.4 Decision rules

Beyond quantifying joint absolute risk, the subject-specific risk profiles outlined in Section 3.3 can be used to guide decisions if paired with a decision rule. In the setting of this paper, a decision rule would assign each subject to one of the four joint outcome categories based on his or her estimated risk profile. We present two examples of existing three-category decision rules, which we refer to in later sections. Note that in practice, a decision rule should be devised incorporating subject matter understanding under the advisement of substantive experts.

In the first example, for subject i, a continuous marker, T_i, is used to assign patients to one of three ordinal disease categories, $D_i = 1, 2, 3$, as follows:

$$\begin{aligned} &\text{if } T_i \leq c_1, \text{ then } \hat{D}_i = 1, \\ &\text{if } c_1 < T_2 < c_2, \text{ then } \hat{D}_i = 2, \\ &\text{if } T_i \geq c_2, \text{ then } \hat{D}_i = 3, \end{aligned} \tag{3.10}$$

where $c_1 < c_2$ are constants and $\hat{D}_i$ denotes the predicted disease category. This decision rule appeared in work by Scurfield (1996) and Nakas and Yiannoutsos (2004), which we will revisit in Section 3.5.

In the second example, for subject i, an absolute risk profile, $p^{(i)} = (p_{(1i)}, p_{(2i)}, p_{(3i)})$, is used to assign patients to one of three un-ordered disease categories, $D_i = 1, 2, 3$, where $p_{(ki)} = P(D_i = k)$ for $k = 1, 2, 3$, such that the components sum to 1. Let e_j be the vector of length three with 1 in the j-th position and 0 elsewhere. Using the intuition that a risk profile equal or close to e_j places most confidence in assigning the subject to disease category $D = j$, the decision rule is defined as follows: select one subject from each of the three disease categories, $D = 1, 2, 3$, with corresponding risk profiles $p^{(1)}$, $p^{(2)}$ and $p^{(3)}$, and assign the three subjects to disease groups k_1, k_2 and k_3 such that the expression,

$$||p^{(1)} - e_{k_1}|| + ||p^{(2)} - e_{k_2}|| + ||p^{(3)} - e_{k_3}||,$$

is minimized, where $(e_{k_1}, e_{k_2}, e_{k_3})$ is a permutation of (e_1, e_2, e_3). In essence, this minimizes the distance between the three risk profiles, $(p^{(1)}, p^{(2)}, p^{(3)})$, and $(e_{k_1}, e_{k_2}, e_{k_3})$. This decision rule was introduced by Mossman (1999), which he denoted R_{III}, and was later extended to $M > 3$ categories by Li and Fine (2008).

3.5 Evaluating predictive peformance

Having developed methodology for estimating the absolute risk for both nonterminal and terminal events jointly at time t, we seek to assess the predictive performance of our methodology given a decision rule. We focus on the capacity of our predictions to discriminate among the four disease categories using an extension of the area under the ROC curve (AUC) to more than two outcomes, for which there is extensive literature

(Scurfield, 1996; Mossman, 1999; Dreiseitl et al., 2000; Nakas and Yiannoutsos, 2004; Li and Fine, 2008). We consider the multicategory AUC measure in two settings: 1) all outcome categories are known; and 2) some outcome categories are missing due to censoring and a verification bias adjusted measure is needed.

3.5.1 Measures of discrimination for multicategory outcomes

For a dichotomous outcome, such as disease status, the area under the curve (AUC) of the ROC curve is a global measure of discrimination of a set of predictions. The standard ROC curve plots the true positive rate for diseased subjects against the false positive rate for non-diseased subjects based on a series of dichotomous decision rules; it can be equivalently plotted as a correct (true) classification rate for diseased subjects against the correct (true) classification rate for non-diseased subjects. The area under this ROC curve is equivalent to the c-index, which represents the probability of correctly assigning higher risk to patients that are diseased compared to patients who are not diseased (Bamber, 1975). There are several extensions of the AUC to the case of a multicategory outcome with M categories. In general, these extensions either consider the concordance of two groups at a time or all M classes simultaneously.

The AUC of predictions corresponding to a dichotomous outcome is well understood and is relatively intuitive since an AUC of 0.5 corresponds to random discrimination and a value of 1 corresponds to perfect discrimination. Thus, it may be natural to reduce an M-category classification problem to the comparison of two groups. Hand and Till (2001) extended the definition to the case of more than two classes by considering the discriminatory performance of each of $M(M-1)/2$ pairs of categories separately, and averaging over all pairs of classes. In the competing risks literature, AUC has also been re-defined as a comparison between two groups as AUC_k, comparing individuals with outcome k against individuals without outcome k. For example, if there are two competing risks (referred to as diseases 1 and 2), then AUC_1 compares individuals with disease 1 versus individuals with disease 2 or who are disease-free (Saha and Heagerty, 2010; Zheng et al., 2012; Gerds et al., 2014). In the setting of post-transplantation trajectories of acute GVHD and death, the focus does not lie in any one of the four categories so an approach that

compares one disease category against the rest is not necessarily of interest. Instead, we consider a measure of discrimination that accounts for all M outcomes simultaneously.

A direct extension of the ROC and the AUC to the case of three categories was introduced independently by Scurfield (1996) and Mossman (1999). Scurfield assumed ordinal outcome categories, $D = 1, 2, 3$, where predictions are made using a continuous marker T_i and the decision rule in (3.10), as was introduced in the first example of Section 3.4. For each pair of cutoff points, (c_1, c_2), the decision rule in (3.10) is used to predict disease status, $\hat{D}_i$, and correct classification rates, $P(\hat{D}_i = k|D_i = k)$ for $k = 0, 1, 2$, are calculated for each of the three disease categories. Plotting the three-tuple of correct classification rates in three-dimensions and varying the pair of cutoff points creates a two-dimensional ROC surface. Note that this process is exactly analogous to the construction of the ROC curve for a dichotomous outcome, where the axes are taken to be the sensitivity and specificity. The analogue of the AUC for a dichotomous outcome is the *volume under the ROC surface* (VUS), which is interpreted as the probability that a set of markers corresponding to each of the three categories will lead to correct classification. In this setting, the VUS is defined as

$$VUS = P(T_1 < T_2 < T_3), \tag{3.11}$$

where T_1, T_2 and T_3 are marker values from subjects from disease categories 1, 2, and 3 respectively.

Mossman worked under the assumption of three, un-ordered categories and a marker which is an absolute risk profile, denoted $p^{(i)} = (p_{(1i)}, p_{(2i)}, p_{(3i)})$, where $p_{(1i)} + p_{(2i)} + p_{(3i)} = 1$. He proposed several decision rules for a subject-specific risk profile; we focus on R_{III} presented in Section 3.4. In this setting, the VUS is defined by

$$\text{VUS} = \Pr\{CR(p^{(1)}, p^{(2)}, p^{(3)}) = 1\}, \tag{3.12}$$

where $p^{(k)}$ are absolute risk profiles corresponding to subjects from disease group $D = k$, for $k = 1, 2, 3$, and the function CR takes on value 1 if the three risk profiles are correctly classified based on decision rule R_{III} and 0 otherwise (Dreiseitl et al., 2000). Note that the expression for the VUS in (3.11) is essentially the expression in (3.12) when the decision

rule is taken to be (3.10). It is clear from from the definition of the VUS in (3.11) and (3.12) that the noninformative value of the VUS corresponds to the random assignment of a set of subjects, one from each disease group, which has probability $1/(3!) = 0.17$.

Nonparametric estimators were later proposed for the VUS. For classification of ordinal outcome categories based on a continuous marker and the decision rule in (3.10), Nakas and Yiannoutsos (2004) proposed an estimator for the VUS in (3.11) as follows

$$\widehat{VUS} = \frac{1}{n_1 \cdot n_2 \cdot n_3} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} I(T_{1i_1} < T_{2i_2} < T_{3i_3}), \tag{3.13}$$

where T_k and n_k, for $k = 1, 2, 3$, are continuous marker values and sample sizes corresponding to subjects from $D = 1, 2, 3$, respectively.

For classification of un-ordered outcome categories based on absolute risk profiles and the decision rule, R_{III}, Dreiseitl et al. (2000) proposed an estimator for the VUS in (3.12) as follows:

$$\widehat{VUS} = \frac{1}{n_1 \cdot n_2 \cdot n_3} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} CR(p^{(i_1)}, p^{(i_2)}, p^{(i_3)}), \tag{3.14}$$

where $p^{(i_1)}$, $p^{(i_2)}$ and $p^{(i_3)}$ are absolute risk profiles corresponding to subjects in outcome categories 1, 2, and 3, and n_D are the sample sizes of categories $D = 1, 2, 3$. Li and Fine (2008) extended the estimator in (3.14) to the case of $M > 3$ categories. The expression in (3.14) extends naturally to $M > 3$ dimensions as the empirical mean of $CR(p^{(1)}, p^{(2)}, \ldots p^{(M)})$, where $p^{(1)}, p^{(2)}, \ldots p^{(M)}$ are absolute risk profiles corresponding to subjects in outcome category 1 through M. For $M > 3$, the ROC surface has dimension $M - 1$, and the corresponding volume under this $(M-1)-$dimensional ROC surface was coined the *hypervolume under the manifold*, abbreviated HUM (Li and Fine, 2008). In the case of M categories, the noninformative HUM is $1/(M!)$.

In predicting the nonterminal and terminal events jointly with semicompeting risks data, absolute risk profiles of length four are used to predict the four un-ordered outcome categories. If there were no censoring of the terminal event before time t, then the disease outcome is known for all subjects. To assess the ability of our predictions to discriminate among the four groups, we can directly apply the estimator of the HUM in (3.14) and assess the variability of our estimates through bootstrapping.

Multicategory AUC accounting for verification bias

In the setting of our problem, for subjects who are censored for the terminal event before time t, the disease outcome cannot be determined. In this case, the estimators for the multicategory AUC in (3.13) and (3.14) cannot be used because they require knowledge of the disease category for every patient. A complete case analysis can be performed, however, this is only justified when the categories are missing completely at random. For the $M = 3$, Zhang and Alonzo (2016) proposed an unbiased estimator of the VUS in (3.11), when the outcome categories are assumed to be missing at random using an inverse probability weighting (IPW) approach. We present their proposed verification bias adjusted multicategory AUC measure and provide an extension to the setting of four un-ordered disease categories and classification based on absolute risk profiles.

Zhang and Alonzo (2016) re-expressed the non-parametric estimator of the VUS in (3.13) by

$$\widehat{VUS} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I(T_i < T_j < T_k)I(D_i < D_j < D_k)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I(D_i < D_j < D_k)}.$$

Now suppose the true disease category is unknown for some subjects. Let V denote a verification status indicator, where $V = 1$ if the true disease status is known, and 0 otherwise, and assume that the verification status, V, is independent of disease category, D, conditional on the continuous marker, T and covariates, X (missing at random). They proposed the following estimator of the VUS that accounts for verification bias:

$$\widehat{VUS}_{IPW} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} V_iV_jV_kI(T_i < T_j < T_k)I(D_i < D_j < D_k)/(\pi_i\pi_j\pi_k)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} V_iV_jV_kI(D_i < D_j < D_k)/(\pi_i\pi_j\pi_k)}, \quad (3.15)$$

where $\pi_i = P(V_i|T_i, X_i)$ for covariates X_i. Applying iterated expectations and the Weak Law of Large Numbers shows that the estimator in (3.15) is unbiased and consistent. Furthermore, the estimator in (3.15) can be re-expressed as a U-estimator for which standard asymptotic results for U-statistics follow.

We extend the verification-bias corrected VUS in (3.15) to our setting as follows. Let $D_i = d$ denote the true disease category, for $d = 1, 2, 3, 4$, let V_i denote a verification indicator that takes on value 1 if the true disease status is observed and 0 otherwise, and let X_i represent a vector of covariates related to V_i and D_i. We assume V_i and D_i are

independent conditional on X_i (missing at random). We first re-express the estimator in (3.12) by the following expression:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} I(D_i < D_j < D_k < D_l) \cdot CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)})}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} I(D_i < D_j < D_k < D_l)}. \tag{3.16}$$

Note that the product of the indicator functions in the numerator corresponds to correct classification of individuals from each of the four possible categories and the denominator is equivalent to $n_1 \cdot n_2 \cdot n_3 \cdot n_4$. We define the verification-bias corrected HUM using a parallel argument as in (3.15) by the following:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} V_i V_j V_k V_l I(D_i < D_j < D_k < D_l) CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)})/(\pi_i \pi_j \pi_k \pi_l)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} V_i V_j V_k V_l I(D_i < D_j < D_k < D_l)/(\pi_i \pi_j \pi_k \pi_l)}. \tag{3.17}$$

In the numerator:

$$\begin{aligned}
&E[{\pi_i}^{-1}{\pi_j}^{-1}{\pi_k}^{-1}{\pi_l}^{-1} \cdot V_i V_j V_k V_l \cdot I(D_i < D_j < D_k < D_l) \cdot CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)})] \\
&\quad = E\left\{E[{\pi_i}^{-1}{\pi_j}^{-1}{\pi_k}^{-1}{\pi_l}^{-1} \cdot V_i V_j V_k V_l \cdot I(D_i < D_j < D_k < D_l) \cdot \right. \\
&\qquad \left. CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)}) | X_i, X_j, X_k, X_l]\right\} \\
&\quad = E\left\{CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)}) E[I(D_i < D_j < D_k < D_l) | X_i, X_j, X_k, X_l]\right\} \\
&\quad = E\left\{CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)}) I(D_i < D_j < D_k < D_l)\right\} \\
&\quad = E\left\{CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)}) | D_i < D_j < D_k < D_l\right\} \cdot P(D_i < D_j < D_k < D_l) \\
&\quad = E\left\{CR(p^{(i)}, p^{(j)}, p^{(k)}, p^{(l)})\right\} \cdot P(D_i = 1) \cdot P(D_j = 2) \cdot P(D_k = 3) \cdot P(D_l = 4) \\
&\quad = \text{HUM} \cdot P(D_i = 1) \cdot P(D_j = 2) \cdot P(D_k = 3) \cdot P(D_l = 4)
\end{aligned}$$

Similar arguments show that the $E[{\pi_i}^{-1}{\pi_j}^{-1}{\pi_k}^{-1}{\pi_l}^{-1} V_i V_j V_k V_l I(D_i < D_j < D_k < D_l)] = P(D_i = 1) \cdot P(D_j = 2) \cdot P(D_k = 3) \cdot P(D_l = 4)$. Using the Weak Law of Large Numbers and Slutky's Theorem (Ferguson, 1996), it follows that the estimator in (3.17) is consistent for the true HUM.

3.6 Data application

3.6.1 Data

We analyze the stem cell transplant data presented in Section 3.2. Towards building a risk prediction instrument, we consider two nested models that include the following indicator variables corresponding to: HLA 8/8 status (poor donor-patient matching compared

to a match from an identical sibling), advanced stage disease, use of reduced myeloablative conditioning, use of in vivo T-cell depletion, Karnofsky score < 90 (indicating a loss of function).

3.6.2 Models and estimating subject-specific risk profiles

We fit the data to the model presented in Section 3.3.2 that accommodates the finite support of acute GVHD using the methods developed in Chapter 2, where the baseline hazard functions in (3.6)-(3.9) are assumed to follow truncated Weibull distributions:

$$\lambda_{0,\omega}(t) = \begin{cases} \alpha_\omega \kappa_\omega t^{\alpha_\omega - 1}, & \text{if } \omega \in \{14, 34\} \text{ for } t > 0 \\ \dfrac{\alpha_\omega \kappa_\omega t^{\alpha_\omega - 1} e^{-\kappa_\omega t^{\alpha_\omega}}}{e^{-\kappa_\omega t^{\alpha_\omega}} - e^{-\kappa_\omega \tau^{\alpha_\omega}}}, & \text{if } \omega \in \{24, 23\}, \text{ where } t \in (0, \tau), \end{cases}$$

with $\tau = 101$. We generate predicted subject-specific risk profiles presented in Section 3.3.2 at day $t = 5, 10, \ldots, 150$ post-transplantation. To assess the impact of the shared frailty on joint risk prediction, we also calculate predicted subject-specific risk profiles at day 100 post-transplantation for the following values of the shared frailty, $\gamma = 0.5, 0.6, \ldots, 1.5$. We compare these estimated absolute risk profiles to the estimated marginal subjected-specific risk profile that averages over the distribution of the shared frailty, defined in Section 3.3.2.

3.6.3 Measures of discrimination

We then assess predictive performance using the four category HUM based on predicted absolute risk profiles presented in Section 3.5. There is no censoring of death within 100 days of transplantation, so that four possible categories of acute GVHD and death status is known. This allows us to calculate the HUM with the full dataset using the estimator in (3.14). We use a simple bootstrap with 1000 resamples to estimate the standard errors of our HUM estimates.

To test the efficacy of our verification bias adjusted HUM estimator in (3.17), we artificially create missingness following a missing at random scheme, as was done by Zhang and Alonzo (2016) in their analysis of an Alzheimer's data set. Specifically, we use a logistic regression model to model the non-missingness indicator, V, as a function of covariates. The logistic regression model is used to calculate the probability of nonmissingness,

$P(V_i = 1)$, for all subjects, which are then used to generate Bernoulli random variables with probability of success of $P(V_i = 1)$. We then calculate the verification bias adjusted HUM in Section 3.5.1 and compare the estimated value to the true HUM for the full data set and to the HUM calculated for the complete cases.

In our calculation of the HUM, we divide the dataset into training and test sets of equal size. We fit the training data to the semicompeting risks model. Based on the estimated parameters from the trained model, we calculate predictions based on both the training and test data. We expect the measure of predictive performance (HUM) on the training set to be biased upward; we refer to the HUM in this setting as *optimistic*. We expect the HUM calculated on the independent test set to be a more accurate measure of predictive performance, and refer to this measure as *realistic*.

Note that our data set includes only categorial covariates. This results in predicted risk profiles that may be the same across several subjects. Again following the methodology introduced by Zhang and Alonzo (2016) to account for ties in marker values, we add a small amount of noise to each of the covariate values to ensure that the subject-specific risk profiles do not coincide among subjects. Specifically, we generate noise from a Uniform$(-\epsilon, \epsilon)$, where $\epsilon = 0.00001$.

Finally, we estimate the optimistic and realistic HUM values at times $t = 5, 10, \ldots, 150$ to investigate possible temporal trends in the HUM.

3.6.4 Results

From Table 3.1, we see that approximately 57% of the population have Advanced stage disease, 43% have a Karnofsky score less than 90, and 58% received transplanted cells from HLA 8/8 donor. These patients are more likely to die with or without acute GVHD, compared to patients in the corresponding reference group. Twenty-three percent of patients received a reduced myeloablative conditioning regimen and 20% received in vivo T-cell depletion; these patients appeared to have a lowered risk of acute GVHD.

The estimated model parameters and the log-likeihood value (given the data) for the two models are presented in Table 3.2. For Model 1, the effect of lower functionality (Karnofsky score less than 90) is associated with an increased risk of death in the $14-$,

$24-$ and $34-$transitions, and a decrease in risk of being cured of acute GVHD. For the 23-transition, the results suggest that being female, reduced conditioning, and in vivo T-cell depletion are protective for acute GVHD. An HLA: 8/8 match is associated with an increased risk for acute GVHD, and a decrease in risk of being cured of acute GVHD. The effects of these variables are comparable in Model 2. In addition, the effect of advanced stage disease is associated with an increased risk of death in the 14-, 24- and 34-transitions, and a decrease in risk of being cured of acute GVHD. Note that Model 1 is nested within Model 2; Model 2 differs from Model 1 by the addition of an Advance disease status indicator to the 14-, 24- and 34- transitions. A likelihood ratio test comparing the two nested models indicates that Model 2 is a better fit to the data ($p < 0.001$).

We plotted the stacked estimated risk profiles at $t = 1, 2, \ldots, 150$ for three patients (labeled A, B, and C) under both Models 1 and 2 in Figure 3.3 so that at any time t the vertical distance between the: solid and dashed line is $p_{(2)}$ (the probability of dying carrying a diagnosis of acute GVHD by time t); dashed and dotted line is $p_{(3)}$ (the probability of death without acute GVHD); and dotted line and the value 1 is $p_{(4)}$ (the probability of being alive without acute GVHD). At any time t, Patient A has the lowest risk of death, and Patient B has the highest risk of acute GVHD. Both Patients B and C have a large risk of death, however, the rate of increase differs between the two patients. For Patients A and B, we see that the risk of acute GVHD (with and without death) increases until around day 60 post-transplantation at which point it stabilizes; for Patient C, we see that the risk of acute GVHD increases through day 80 post-transplantation. The risk of death steadily increases through the 150 days post-transplantation for all three patients.

We present a graph of the stacked estimated risk profile (conditional on the shared frailty term, γ) of Patient C at $t = 100$ for $\gamma = 0.5, 0.6, \ldots, 1.5$ in Figure 3.3. The marginal stacked risk profile for Patient C at $t = 100$ is plotted to the right. We see a dramatic increase in the absolute risk of death for increasing values of the shared frailty term, particularly for death without acute GVHD, thereby providing evidence that the shared frailty term accounts for heterogeneity in the data set that is not accounted for by covariates. Comparing the conditional and marginal stacked risk profiles, we see that the marginal risk profile components align with values of γ between 0.8 and 1.25.

Table 3.2: Estimated parameters for our proposed multi-state model for acute GVHD and death considering two sets of covariates.

	Model 1		Model 2	
	Est	SE	Est	SE
Cure model				
Intercept	1.65	0.14	1.85	0.16
HLA donor match: 8/8	-0.30	0.15	-0.32	0.15
Disease status: Advance			-0.41	0.16
Karnofsky score < 90	-0.21	0.17	-0.15	0.16
Cured: HCT to Death				
Disease status: Advance			0.89	0.11
Karnofsky score < 90	0.54	0.10	0.38	0.11
Not cured: HCT to Death				
Disease status: Advance			0.05	0.44
Karnofsky score < 90	0.14	0.55	0.28	0.43
Not cured: HCT to Acute GVHD				
Female			-0.33	0.18
HLA donor match: 8/8	0.13	0.19	0.23	0.19
Conditioning: Red. intensity/Non-myel.	-0.91	0.23	-0.88	0.23
In vivo T-cell depletion			-0.63	0.28
Not cured: Acute GVHD to Death				
Disease status: Advance			0.33	0.18
Karnofsky score < 90	0.51	0.17	0.48	0.17
Weibull baseline hazard				
log(kappa14)	-6.95	0.42	-7.68	0.37
log(alpha14)	0.07	0.06	0.11	0.05
log(kappa24)	-14.73	0.69	-17.65	0.08
log(alpha24)	0.86	0.35	0.84	0.07
log(kappa23)	-6.19	0.42	-6.08	0.36
log(alpha23)	0.59	0.07	0.59	0.06
log(kappa34)	-5.17	0.34	-5.41	0.37
log(alpha34)	-0.10	0.07	-0.09	0.07
log(theta)	-1.32	0.26	-1.28	0.27
Log-likelihood	-6838.97		-6790.66	
Num. of parameters	17		23	

The estimated values of the HUM are presented Table 3.3. Observe that all HUM values for Model 2 and larger than the corresponding values for Model 1. The optimistic and realistic HUM for Model 2 using the full data set are 0.102 and 0.087, respectively. Recall that the noninformative HUM for $M = 4$ categories is $1/(M!) = 0.04$; the HUM values observed in this analysis are many times larger than this noninformative value. When we

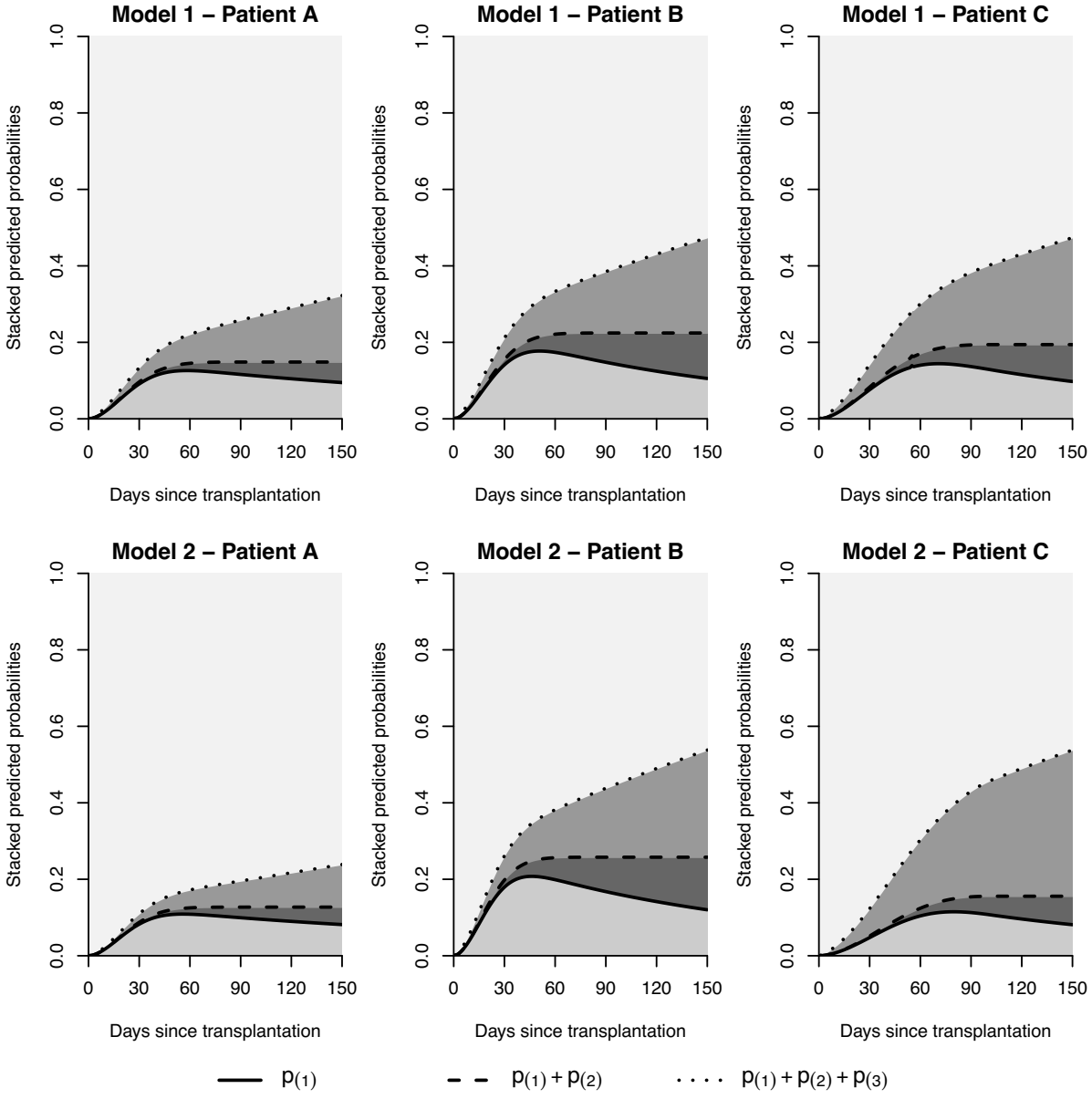

Figure 3.3: Estimated absolute risk profiles calculated for the first 150 post-transplantation corresponding to three patients (A, B, C).

artificially added missingness to the outcome categories directly, the IPW estimator of the HUM is comparable to the HUM of the full data. For both models the estimated HUM corresponding to a complete case analysis are smaller than the verification bias adjusted HUM values. The distribution of outcomes for each analysis are summarized in Table 4.24

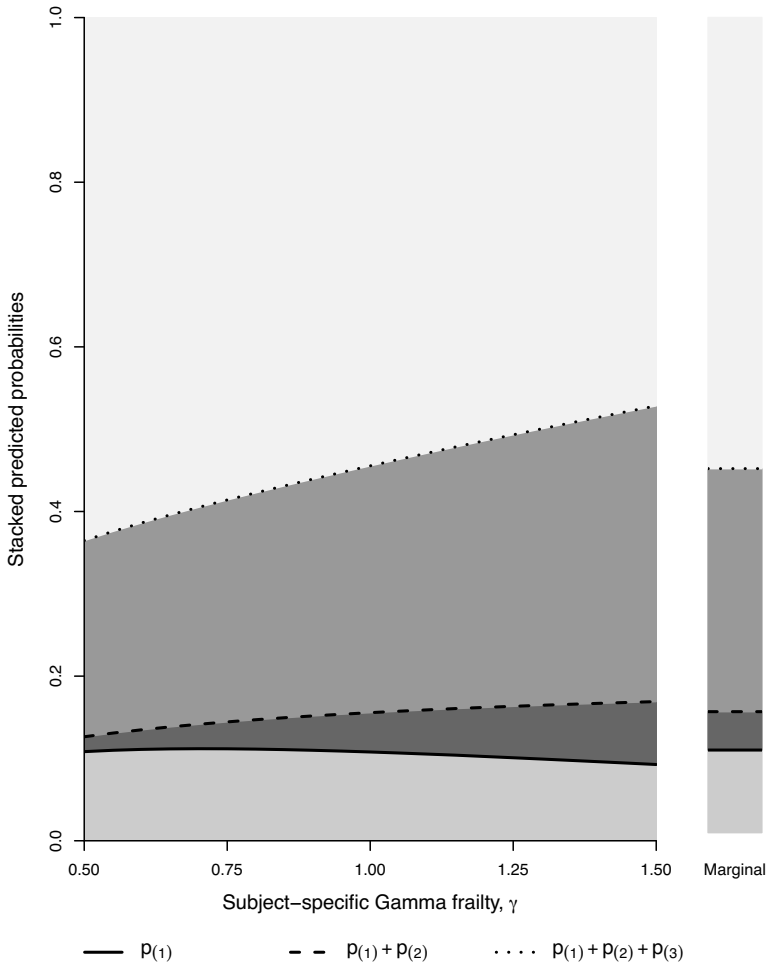

Figure 3.4: Left panel: Estimated absolute risk profiles (conditional on frailty, γ) for Patient C at 100 days post-transplantation for varying values of the gamma frailty, γ. Right panel: Estimated marginal absolute risk profiles for Patient C at 100 days post-transplantation.

We present a plot of the HUM values (optimistic and realistic) for both Models 1 and 2 corresponding to time $t = 5, 10, \ldots, 150$ in Figure 3.5. At each time point considered, the HUM value (optimistic or realistic) for Model 2 is greater than the corresponding HUM value for Model 1.The realistic HUM values appear to increase over time for both models, with a faster rate of increase corresponding to Model 2.

In summary, we have illustrated our methods for risk prediction using semicompeting risks data outlined in this paper. For the two nested models considered, we see that

Table 3.3: HUM estimates for the two models considered. HUM IPW denotes the verification bias adjusted HUM presented in Section 3.5.1.

Noninformative HUM: 0.042				
	Model 1		Model 2	
	Est	SE_b	Est	SE_b
Full data set (no censoring)				
Optimistic	0.074	0.015	0.102	0.018
Realistic	0.069	0.014	0.087	0.017
Added missingness to disease status				
HUM IPW				
Optimistic	0.078	0.017	0.100	0.020
Realistic	0.069	0.017	0.086	0.019
HUM CC				
Optimistic	0.074	0.016	0.094	0.020
Realistic	0.063	0.015	0.078	0.018

an improvement in the fit between Model 1 and 2 corresponds to an improvement in the HUM of the predictions. When we artificially generate missingness in our outcome categories in our data set, our extension of the verification bias adjusted HUM estimator, defined by (3.17), shows little bias when compared with the HUM that is calculated from the full data set.

3.7 Discussion

We have presented novel methodology for calculating the joint absolute risk of the nonterminal and terminal events at time t in semicompeting risks data via integration over the joint density. In essence, we have turned a survival problem into a multicategory outcome problem. A natural question is: "Why not just use standard multicategory modeling methods, such as multinomial logistic regression, to estimate the absolute risk profiles?" When there is no censoring of the terminal event before time t, the four outcome categories can be determined for all subjects. Then a standard multicategory regression or classification model can be applied to the data. However, when there is censoring of the terminal event before time t for some subjects, the outcome categories cannot be determined for those subjects. As a result, those observations with missing outcome categories

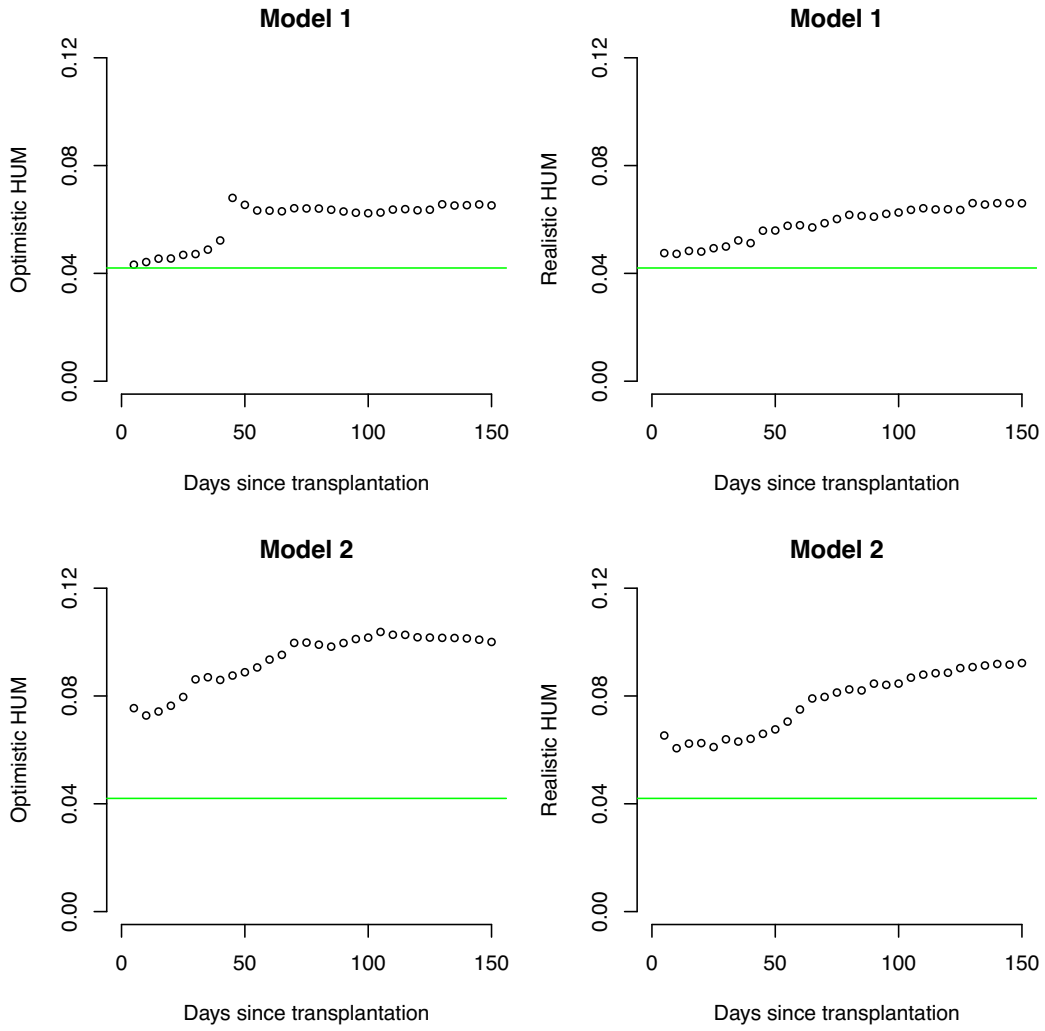

Figure 3.5: HUM over time for Models 1 and 2. Non-informative HUM is plotted in green.

cannot be used in a standard multicategory regression model. This is not a problem in our proposed methodology since the semicompeting risks model accommodates censoring.

We used the multicategory extension of the AUC for a dichotomous outcome, the HUM, to assess capacity of our predictions to discriminate among the four groups simultaneously. Although the interpretation of the measure is appealing and intuitive, the actual HUM value for M categories is non-intuitive since the noninformative HUM corresponding to random classification has value $1/(M!)$. For four categories, the noninformative HUM is 0.042. As with the AUC for a dichotomous outcome, the HUM can be used to compare the predictive performance of competing models. In our data example,

we used the HUM to compare two models with respect to discrimination. However, in isolation, the optimistic HUM value of 0.102 for Model 2 is hard to make sense of; it is several folds larger than the non-informative HUM but seems very far from the HUM for perfect discrimination which is a value of one.

As discussed above, the outcome category cannot be determined when there is censoring of death before time t. When the censoring mechanism does not depend on the true outcome category and covariates, the missing outcome categories are missing completely at random and a complete case estimate of the HUM is expected to be unbiased for the true HUM. However, unbiased estimates of the HUM are not guaranteed for other missing data mechanisms. Zhang and Alonzo (2016) addressed the problem of the HUM for ordinal disease outcomes that are missing at random through an IPW estimator of the multicategory AUC for three dimensions. We translated and extended their measure to the setting of our problem (four, non-ordered outcome categories). Although the dataset that we considered did not have any subjects who were censored before time $t = 100$ days, our methods can be applied when such censoring arises.

In this paper, we chose to assess predictive performance of our predicted risk profiles using a measure of discrimination, the multicategory AUC, however, other measures that target calibration and risk stratification capacity may be informative and complementary. We assumed fully-parametric baseline hazard functions in order to integrate the joint density at any time t following transplantation, however, more flexible parametric baseline hazards using splines may be more appealing and are the focus of future work.

Supplementary Materials

4.1 Time-to-event data with time-varying biomarkers measured only at study entry, with applications to Alzheimer's disease

Table 4.1: Values of R_{trunc} used in data generating procedure.

		R_{trunc}					
	β	$n = 250$ $\lambda = 0.25$ $\pi_{\text{trunc}} = 0.2$ $\sigma_a = 0.5$ $\sigma_b = 0.05$	$n = 250$ $\lambda = 0.25$ $\pi_{\text{trunc}} = 0.4$ $\sigma_a = 0.5$ $\sigma_b = 0.05$	$n = 100$ $\lambda = 0.25$ $\pi_{\text{trunc}} = 0.2$ $\sigma_a = 0.5$ $\sigma_b = 0.05$	$n = 250$ $\lambda = 0.1$ $\pi_{\text{trunc}} = 0.2$ $\sigma_a = 0.5$ $\sigma_b = 0.05$	$n = 250$ $\lambda = 0.4$ $\pi_{\text{trunc}} = 0.2$ $\sigma_a = 0.5$ $\sigma_b = 0.05$	$n = 250$ $\lambda = 0.25$ $\pi_{\text{trunc}} = 0.2$ $\sigma_a = 1$ $\sigma_b = 0.01$
Fixed deviations, a_i							
Normal density kernel based hazard	0	32	40	32	32	32	32
	0.25	32	44	33	32	34	32
	0.5	33	44	34	32	34	33
	1	32	46	35	32	36	32
	1.5	32	48	35	32	38	32
Constant hazard	0	31	50	31	30	30	31
	0.25	74	102	74	72	72	74
	0.5	79	104	79	98	72	79
	1	75	100	75	85	68	75
	1.5	67	92	67	75	62	67
Exponential-form hazard	0	89	118	89	88	90	89
	0.25	81	108	81	86	77	81
	0.5	76	100	76	82	72	76
	1	70	92	70	78	66	70
	1.5	66	90	66	76	69	66
Time-varying deviations, b_i							
Normal density kernel based hazard	0	79	106	79	78	78	79
	0.25	78	104	78	80	78	78
	0.5	76	102	75	78	74	76
	1	76	102	75	78	72	76
	1.5	74	100	75	76	72	74
Constant hazard	0	79	74	31	30	90	79
	0.25	78	102	78	70	77	78
	0.5	78	108	78	70	72	78
	1	76	104	76	78	66	76
	1.5	75	100	75	80	69	75
Exponential-form hazard	0	90	118	90	90	90	90
	0.25	82	108	82	88	78	82
	0.5	75	102	75	84	73	75
	1	70	93	70	80	67	70
	1.5	66	90	66	76	63	66

Table 4.2: Type I error for the logit error model for specified covariate model and baseline hazard function using $n = 250$ and 5000 iterations

	Onset origin			Study entry origin			
Covariate model	A: $f(Z_i(u))$	A′: $f(\widehat{Z_i(u)})$	B: $f(Z_i(E_i))$	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $f(\widehat{Z_i(t + E_i)})$	F: $f(\widehat{Z_i(t + E_i)}) + E_i$
Fixed deviations, a_i							
Normal density based hazard	0.051	0.052	0.054	1	0.124	1	0.124
Constant hazard	0.048	0.048	0.052	0.053	0.048	0.053	0.048
Exponential-form hazard	0.054	0.054	0.048	1	0.055	1	0.055
Time-varying deviations, b_i							
Normal density based hazard	0.061	0.060	0.050	1	0.060	1	0.070
Constant hazard	0.056	0.056	0.044	0.049	0.053	0.050	0.058
Exponential-form hazard	0.050	0.052	0.052	1	0.052	1	0.051

Table 4.3: Power corresponding to $\beta = 0.5$ for the logit error model for specified covariate model and baseline hazard function using $n = 250$ and 5000 iterations

	Onset origin			Study entry origin			
Covariate model	A: $f(Z_i(u))$	A′: $f(\widehat{Z_i(u)})$	B: $f(Z_i(E_i))$	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $f(\widehat{Z_i(t + E_i)})$	F: $f(\widehat{Z_i(t + E_i)}) + E_i$
Fixed deviations, a_i							
Normal density based hazard	0.97	0.97	0.15	1	0.98	1	0.98
Constant hazard	0.97	0.97	0.07	1	0.97	1	0.97
Exponential-form hazard	0.97	0.97	0.07	1	0.96	1	0.96
Time-varying deviations, b_i							
Normal density based hazard	1	1	0.99	1	1	1	1
Constant hazard	1	1	0.99	1	1	1	1
Exponential-form hazard	1	1	0.99	1	1	1	1

Table 4.4: Analytical standard errors versus bootstrapped standard errors (5000 replications) for RADC analysis

	Birth origin						Study entry origin							
	$\widehat{f(Z_i(u))}$		$f(Z_i(E_i))$		Sev. Adj.		$f(Z_i(E_i))$		$f(Z_i(E_i)) + E_i$		$\widehat{f(Z_i(t+E_i))}$		$\widehat{f(Z_i(t+E_i))} + E_i$	
	Wald	Boot	Wald	Boot	Wald	Boot	Wald	Boot	Wald	Boot	Wald	Boot	Wald	Boot
Absolute error, sigmoid, a_i	0.06	0.10	0.06	0.10	0.06	0.10	0.06	0.10	0.06	0.11	0.06	0.09	0.06	0.10
Absolute error, sigmoid, b_i	0.05	0.09	0.06	0.10	0.06	0.10	0.06	0.10	0.06	0.11	0.05	0.09	0.06	0.10
Logit error, sigmoid, a_i	0.06	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.08	0.07	0.06	0.07	0.07
Logit error, sigmoid, b_i	0.06	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.08	0.06	0.06	0.06	0.07
Absolute error, linear trend, a_i	0.06	0.10	0.06	0.10	0.06	0.09	0.06	0.10	0.06	0.11	0.06	0.09	0.06	0.10
Absolute error, linear trend b_i	0.05	0.09	0.06	0.10	0.06	0.09	0.06	0.10	0.06	0.11	0.05	0.09	0.06	0.10

Table 4.5: Estimates of the sigmoidal and linear parameters for RADC analysis

			Severity adjusted				
	$\widehat{T}_0$	$\widehat{\lambda}$	$\widehat{\lambda}_{\text{sev}}$	$\widehat{\gamma}_{\text{BMI}}$	$\widehat{\gamma}_{\text{MALE}}$	$min(\widehat{T_{0,i}})$	$max(\widehat{T_{0,i}})$
Absolute error, sigmoid, a_i	106	-0.03	-0.001	37	-212	449	1702
Absolute error, sigmoid, b_i	105	-0.03	-0.001	32	-157	401	1450
Logit error, sigmoid, a_i	104	-0.03	-0.001	33	-172	411	1499
Logit error, sigmoid, b_i	103	-0.03	-0.001	31	-141	397	1435
Absolute error, linear trend, a_i	188	-0.04	0.028	-2	7	-97	-27
Absolute error, linear trend, b_i	186	-0.04	0.028	-2	6	-100	-28

Table 4.6: Descriptive statistics of subjects included in the ADNI analyses

	n	(%)
Number of subjects	786	
Male	465	59
Education		
$\leq$12 years	136	17
13-16 years	332	42
17-20 years	318	40
One or more APOE4 alleles	408	52
	Mean	SD
Age at first Hippocampal measurement	73	7
Hippocampal volume	6735	1158
Time from birth to last visit	76	8
Time from study entry to last visit	3	2
Converted to AD	280	

Table 4.7: Parameter estimates and bootstrapped standard errors (5000 replications) of ADNI analysis with hippocampal volume biomarker

	Birth origin			Study entry origin			
Covariate model	A′: $\widehat{f(Z_i(u))}$	B: $f(Z_i(E_i))$	Sev. Adj.	C: $f(Z_i(E_i))$	D: $f(Z_i(E_i)) + E_i$	E: $\widehat{f(Z_i(t+E_i))}$	F: $\widehat{f(Z_i(t+E_i))} + E_i$
Absolute error, sigmoid, a_i	-0.78 (0.07)	-0.75 (0.07)	-0.76 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.83 (0.08)	-0.83 (0.08)
Absolute error, sigmoid, b_i	-0.74 (0.07)	-0.75 (0.07)	-0.74 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.81 (0.07)	-0.81 (0.08)
Logit error, sigmoid, a_i	-0.83 (0.14)	-0.80 (0.13)	-0.81 (0.13)	-0.86 (0.09)	-0.93 (0.12)	-0.93 (0.12)	-0.93 (0.12)
Logit error, sigmoid, b_i	-0.83 (0.13)	-0.80 (0.13)	-0.80 (0.13)	-0.86 (0.09)	-0.93 (0.12)	-0.92 (0.12)	-0.92 (0.12)
Absolute error, linear trend, a_i	-0.79 (0.07)	-0.75 (0.07)	-0.73 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.82 (0.08)	-0.83 (0.08)
Absolute error, linear trend, b_i	-0.78 (0.07)	-0.75 (0.07)	-0.72 (0.07)	-0.72 (0.07)	-0.83 (0.08)	-0.80 (0.07)	-0.81 (0.08)

4.2 Time-to-event analysis when the event is defined on a finite time interval

Table 4.8: Description of study population

			Observed outcome category, %			
	N	%	Both acute GVHD & death	Acute GVHD & censored for death	Death without acute GVHD	Censored for both
Total subjects	9651		11.0	6.7	28.7	53.7
Gender						
Male	5366	55.6	11.7	6.9	28.3	53.1
Female	4285	44.4	10.1	6.4	29.2	54.4
Age, years						
<10	653	6.8	5.7	7.0	21.6	65.7
10-19	1162	12.0	8.5	8.3	21.8	61.4
20-29	1572	16.3	11.3	8.1	25.9	54.7
30-39	1581	16.4	11.6	6.3	26.5	55.6
40-49	2095	21.7	11.4	6.8	29.6	52.2
50-59	2008	20.8	12.6	5.6	35.2	46.5
60+	580	6.0	11.9	3.1	38.1	46.9
Disease type						
AML	4919	51.0	10.1	6.1	31.8	52.0
ALL	2071	21.5	10.4	6.6	29.3	53.7
CML	1525	15.8	12.9	8.3	18.0	60.9
MDS	1136	11.8	13.1	7.1	28.5	51.2
Disease status						
Early	4873	50.5	9.0	7.3	21.1	62.5
Intermediate	2316	24.0	11.2	7.0	28.4	53.5
Advanced	2462	25.5	14.6	5.1	43.9	36.4
HLA compatibility						
Identical sibling	3941	40.8	7.9	5.7	26.0	60.4
8/8	4100	42.5	11.8	7.3	29.7	51.1
7/8	1610	16.7	16.3	7.5	32.7	43.5

Table 4.9: Observed outcomes corresponding to 100 simulated data sets each of size 5000 for each of the five simulations considered in Section 5.

	Base Scenario		Censoring via exponential		Administrative censoring at 365 days		Increased cure fraction		Conlon with frailty sufficient	
	n	%	n	%	n	%	n	%	n	%
Acute GVHD and death within 100 days	933	18.7	563	11.3	491	9.8	439	8.8	491	9.8
Acute GVHD and death after 100 days	875	17.5	161	3.2	633	12.7	417	8.3	695	13.9
Acute GVHD and censored before 100 days	0	0.0	747	14.9	0	0.0	0	0.0	0	0.0
Acute GVHD and censored after 100 days	0	0.0	162	3.2	62	1.2	0	0.0	0	0.0
Death within 100 days without acute GVHD	1202	24.0	895	17.9	1823	36.5	924	18.5	1823	36.5
Death after 100 days without acute GVHD	1991	39.8	139	2.8	983	19.7	3220	64.4	1991	39.8
Censored after 100 days without either	0	0.0	581	11.6	1008	20.2	0	0.0	0	0.0
Censored within 100 days without either	0	0.0	1751	35.0	0	0.0	0	0.0	0	0.0

Table 4.10: Simulation results corresponding to the 'base scenario', with 1000 replicates of size 5000. Empirical and analytical standard errors calculated using the standard deviations of the sampling distributions and means of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty	Illness-death	Univ. cure	Univ. Cox
	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$
Intercept	0.05 0.05	0.05 0.05	0.75 0.05	0.05 0.05		0.04 0.03	
X.Cure	0.06 0.06	0.06 0.06	0.56 0.06	0.06 0.07		0.06 0.06	
X14	0.05 0.05	0.05 0.04	1.32 0.06	0.05 0.05			
X24	0.08 0.08	0.09 0.08	0.60 0.08	0.08 0.08	0.06 0.05		
X23	0.06 0.06	0.06 0.06	2.90 0.06	0.06 0.06	0.06 0.05	0.05 0.05	0.05 0.05
X34	0.06 0.05	0.05 0.05	1.23 0.05	0.05 0.05	0.06 0.06		
log(kappa14)	0.29 0.27	0.20 0.18	1.56 0.37	0.22 0.20			
log(alpha14)	0.03 0.03	0.03 0.02	0.13 0.04	0.03 0.02			
log(kappa24)	0.15 0.14	0.14 0.14	0.74 0.13	0.13 0.14	0.10 0.10		
log(alpha24)	0.03 0.03	0.03 0.03	0.43 0.03	0.03 0.03	0.04 0.03		
log(kappa23)	0.12 0.11	0.15 0.14	0.76 0.05	0.05 0.05	0.04 0.04		
log(alpha23)	0.03 0.03	0.03 0.03	0.64 0.02	0.02 0.02	0.04 0.03		
log(kappa34)	0.14 0.14	0.12 0.11	0.78 0.15	0.12 0.11	0.18 0.17		
log(alpha34)	0.02 0.02	0.02 0.02	0.24 0.02	0.02 0.02	0.03 0.02		
log(theta)	0.12 0.12		4.35 0.17		0.16 0.12		

Table 4.11: Base scenario. Coverage probabilities. 95% confidence intervals via estimates and standard errors obtained using the means of the sampling distributions and standard deviations of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty
Intercept	0.94	0.44	0.46	0.95
X.Cure	0.95	0.95	0.91	0.90
X14	0.95	0.75	0.87	0.92
X24	0.94	0.74	0.95	0.90
X23	0.95	0.94	0.95	0.96
X34	0.94	0.92	0.94	0.92
log(kappa14)	0.95	0.00	0.72	0.00
log(alpha14)	0.94	0.00	0.76	0.00
log(kappa24)	0.94	0.85	0.93	0.67
log(alpha24)	0.95	0.09	0.86	0.08
log(kappa23)	0.95	0.92	0.00	0.00
log(alpha23)	0.95	0.54	0.03	0.30
log(kappa34)	0.95	0.00	0.95	0.00
log(alpha34)	0.95	0.00	0.96	0.00
log(theta)	0.95		0.96	

Table 4.12: Simulation results for the case of 'moderate censoring', with 1000 replicates of size 5000. Parameter estimates calculated as mean of sampling distribution

	Truth	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon, no frailty	Illness-death	Univ. Cure	Univ. Cox
Intercept	-0.41	-0.40	-0.31	-0.60	-0.52		0.25	
X.Cure	0.50	0.50	0.48	0.57	0.56		0.42	
X14	0.25	0.26	0.17	0.47	0.40			
X24	0.50	0.50	0.43	0.52	0.49	-0.05		
X23	0.25	0.25	0.23	0.24	0.24	-0.18	0.32	-0.21
X34	0.15	0.15	0.14	0.15	0.14	0.03		
log(kappa14)	-8.52	-8.52	-7.26	-10.46	-9.33			
log(alpha14)	0.34	0.33	0.17	0.52	0.40			
log(kappa24)	-5.30	-5.32	-5.25	-5.23	-5.14	-5.20		
log(alpha24)	0.34	0.34	0.26	0.29	0.25	0.26		
log(kappa23)	-2.16	-2.17	-2.27	-1.83	-1.81	-2.08		
log(alpha23)	-0.69	-0.69	-0.74	-0.64	-0.67	-0.64		
log(kappa34)	-6.21	-6.21	-5.95	-6.19	-5.95	-7.38		
log(alpha34)	0.26	0.26	0.21	0.25	0.21	0.40		
log(theta)	-1.71	-1.75		-2.02		0.81		

Table 4.13: Simulation results for the case of 'moderate censoring', with 1000 replicates of size 5000. Empirical and analytical standard errors calculated using the standard deviations of the sampling distributions and means of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty	Illness-death	Univ. cure	Univ. Cox
	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$
Intercept	0.07 0.06	0.06 0.06	0.15 0.08	0.07 0.07		0.05 0.04	
X.Cure	0.08 0.08	0.08 0.08	0.12 0.08	0.08 0.08		0.07 0.07	
X14	0.16 0.16	0.14 0.14	0.53 0.21	0.18 0.18			
X24	0.10 0.09	0.10 0.09	0.10 0.10	0.09 0.09	0.09 0.09		
X23	0.07 0.07	0.07 0.07	0.07 0.07	0.07 0.07	0.07 0.07	0.06 0.06	0.05 0.05
X34	0.08 0.08	0.08 0.07	0.08 0.08	0.08 0.07	0.10 0.10		
log(kappa14)	0.91 0.87	0.67 0.69	1.17 1.22	0.98 0.99			
log(alpha14)	0.11 0.11	0.10 0.10	0.12 0.12	0.11 0.11			
log(kappa24)	0.15 0.15	0.15 0.16	0.14 0.14	0.14 0.14	0.11 0.12		
log(alpha24)	0.04 0.03	0.04 0.04	0.03 0.03	0.03 0.03	0.03 0.03		
log(kappa23)	0.14 0.13	0.18 0.17	0.06 0.05	0.05 0.05	0.05 0.05		
log(alpha23)	0.03 0.03	0.03 0.03	0.03 0.03	0.02 0.02	0.03 0.03		
log(kappa34)	0.18 0.18	0.16 0.17	0.19 0.19	0.16 0.17	0.21 0.20		
log(alpha34)	0.03 0.03	0.03 0.03	0.03 0.03	0.03 0.03	0.03 0.03		
log(theta)	0.23 0.23		0.67 0.67		0.07 0.07		

Table 4.14: Coverage probabilities for the case of 'moderate censoring', with 1000 replicates of size 5000. 95% confidence intervals via estimates and standard errors obtained using the means of the sampling distributions and standard deviations of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty
Intercept	0.95	0.63	0.34	0.66
X.Cure	0.95	0.94	0.86	0.89
X14	0.95	0.90	0.82	0.89
X24	0.95	0.87	0.94	0.95
X23	0.94	0.93	0.95	0.94
X34	0.95	0.94	0.95	0.94
log(kappa14)	0.93	0.54	0.67	0.89
log(alpha14)	0.94	0.62	0.66	0.89
log(kappa24)	0.95	0.92	0.92	0.78
log(alpha24)	0.95	0.45	0.72	0.17
log(kappa23)	0.95	0.97	0.00	0.00
log(alpha23)	0.95	0.71	0.50	0.83
log(kappa34)	0.96	0.63	0.95	0.63
log(alpha34)	0.96	0.56	0.95	0.56
log(theta)	0.97		0.99	

Table 4.15: Simulation results for the case of 'administrative censoring at 365 days', with 1000 replicates of size 5000. Parameter estimates calculated as mean of sampling distribution

	Truth	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon, no frailty	Illness-death	Univ. Cure	Univ. Cox
Intercept	-0.41	-0.41	-0.34	-0.54	-0.48		0.25	
X.Cure	0.50	0.50	0.50	0.54	0.53		0.42	
X14	0.25	0.25	0.21	0.33	0.29			
X24	0.50	0.50	0.41	0.49	0.46	-0.03		
X23	0.25	0.25	0.23	0.22	0.22	-0.23	0.32	-0.26
X34	0.15	0.15	0.14	0.15	0.14	0.01		
log(kappa14)	-8.52	-8.51	-7.53	-9.66	-8.78			
log(alpha14)	0.34	0.33	0.21	0.45	0.35			
log(kappa24)	-5.30	-5.30	-5.16	-5.24	-5.10	-5.50		
log(alpha24)	0.34	0.34	0.23	0.31	0.25	0.32		
log(kappa23)	-2.16	-2.17	-2.38	-1.80	-1.79	-2.02		
log(alpha23)	-0.69	-0.69	-0.76	-0.62	-0.66	-0.71		
log(kappa34)	-6.21	-6.22	-5.79	-6.18	-5.79	-7.87		
log(alpha34)	0.26	0.26	0.18	0.25	0.18	0.48		
log(theta)	-1.71	-1.73		-1.89		0.80		

Table 4.16: Simulation results for the case of 'administrative censoring at 365 days', with 1000 replicates of size 5000. Empirical and analytical standard errors calculated using the standard deviations of the sampling distributions and means of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty	Illness-death	Univ. cure	Univ. Cox
	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$
Intercept	0.05 0.05	0.05 0.05	0.05 0.05	0.05 0.05		0.04 0.03	
X.Cure	0.06 0.06	0.06 0.06	0.06 0.06	0.06 0.06		0.06 0.06	
X14	0.07 0.07	0.06 0.06	0.07 0.07	0.07 0.07			
X24	0.08 0.08	0.09 0.08	0.08 0.08	0.08 0.08	0.08 0.08		
X23	0.06 0.06	0.06 0.06	0.06 0.06	0.05 0.06	0.07 0.07	0.05 0.05	0.05 0.05
X34	0.06 0.05	0.05 0.05	0.06 0.05	0.05 0.05	0.08 0.08		
log(kappa14)	0.43 0.42	0.37 0.37	0.52 0.55	0.44 0.47			
log(alpha14)	0.05 0.05	0.05 0.05	0.05 0.06	0.05 0.05			
log(kappa24)	0.15 0.14	0.13 0.13	0.13 0.13	0.13 0.13	0.11 0.11		
log(alpha24)	0.03 0.03	0.03 0.03	0.03 0.03	0.02 0.03	0.02 0.02		
log(kappa23)	0.12 0.11	0.17 0.16	0.05 0.05	0.05 0.05	0.05 0.05		
log(alpha23)	0.03 0.03	0.03 0.03	0.02 0.02	0.02 0.02	0.03 0.03		
log(kappa34)	0.14 0.14	0.12 0.12	0.14 0.15	0.12 0.12	0.17 0.16		
log(alpha34)	0.02 0.02	0.02 0.02	0.02 0.02	0.02 0.02	0.02 0.02		
log(theta)	0.14 0.14		0.23 0.24		0.05 0.05		

Table 4.17: Coverage probabilities for the case of 'administrative censoring at 365 days', with 1000 replicates of size 5000. 95% confidence intervals via estimates and standard errors obtained using the means of the sampling distributions and standard deviations of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty
Intercept	0.95	0.69	0.32	0.74
X.Cure	0.95	0.95	0.91	0.91
X14	0.95	0.87	0.80	0.92
X24	0.95	0.76	0.94	0.91
X23	0.95	0.93	0.93	0.91
X34	0.94	0.93	0.94	0.93
log(kappa14)	0.95	0.24	0.46	0.93
log(alpha14)	0.94	0.23	0.49	0.95
log(kappa24)	0.94	0.81	0.92	0.63
log(alpha24)	0.95	0.08	0.82	0.06
log(kappa23)	0.95	0.88	0.00	0.00
log(alpha23)	0.95	0.46	0.12	0.62
log(kappa34)	0.95	0.07	0.95	0.07
log(alpha34)	0.95	0.01	0.94	0.01
log(theta)	0.96		0.96	

Table 4.18: Simulation results for the case of 'increased cure fraction', with 1000 replicates of size 5000. Parameter estimates calculated as mean of sampling distribution

	Truth	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon, no frailty	Illness-death	Univ. Cure	Univ. Cox
Intercept	0.84	0.84	0.96	0.74	0.91		1.32	
X.Cure	0.50	0.50	0.54	0.53	0.56		0.47	
X14	0.25	0.25	0.21	0.28	0.22			
X24	0.50	0.50	0.33	0.52	0.41	0.14		
X23	0.25	0.25	0.30	0.23	0.30	-0.39	0.30	-0.40
X34	0.15	0.15	0.14	0.15	0.14	0.14		
log(kappa14)	-8.52	-8.50	-7.23	-8.86	-7.34			
log(alpha14)	0.34	0.33	0.16	0.37	0.18			
log(kappa24)	-5.30	-5.31	-5.31	-5.33	-5.19	-6.44		
log(alpha24)	0.34	0.34	0.21	0.32	0.25	0.08		
log(kappa23)	-2.16	-2.18	-2.09	-1.83	-1.73	-2.76		
log(alpha23)	-0.69	-0.69	-0.72	-0.62	-0.63	-1.45		
log(kappa34)	-6.21	-6.22	-5.67	-6.26	-5.67	-5.91		
log(alpha34)	0.26	0.26	0.15	0.27	0.15	0.20		
log(theta)	-1.71	-1.74		-1.68		-2.97		

Table 4.19: Simulation results for the case of 'increased cure fraction', with 1000 replicates of size 5000. Empirical and analytical standard errors calculated using the standard deviations of the sampling distributions and means of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty	Illness-death	Univ. cure	Univ. Cox
	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$
Intercept	0.05 0.05	0.05 0.05	0.06 0.06	0.06 0.06		0.05 0.03	
X.Cure	0.08 0.08	0.08 0.08	0.08 0.08	0.08 0.08		0.08 0.08	
X14	0.04 0.04	0.04 0.03	0.04 0.27	0.04 0.03			
X24	0.12 0.12	0.19 0.16	0.14 0.31	0.15 0.15	0.03 0.03		
X23	0.10 0.10	0.10 0.10	0.10 0.75	0.10 0.10	0.07 0.07	0.07 0.07	0.07 0.07
X34	0.08 0.08	0.08 0.07	0.09 0.88	0.08 0.07	0.08 0.07		
log(kappa14)	0.24 0.24	0.16 0.14	0.33 0.33	0.17 0.15			
log(alpha14)	0.03 0.03	0.02 0.02	0.04 0.04	0.02 0.02			
log(kappa24)	0.22 0.22	0.25 0.26	0.26 0.80	0.24 0.24	0.11 0.10		
log(alpha24)	0.05 0.05	0.06 0.06	0.13 0.63	0.04 0.05	0.02 0.02		
log(kappa23)	0.18 0.17	0.18 0.17	0.08 0.73	0.08 0.07	0.05 0.06		
log(alpha23)	0.04 0.05	0.04 0.05	0.04 0.03	0.03 0.03	0.02 0.03		
log(kappa34)	0.19 0.19	0.17 0.17	0.20 0.86	0.17 0.17	0.19 0.19		
log(alpha34)	0.03 0.03	0.03 0.03	0.03 0.03	0.03 0.03	0.03 0.03		
log(theta)	0.16 0.16		0.92 0.19		1.02 0.54		

Table 4.20: Coverage probabilities for the case of 'increased cure fraction', with 1000 replicates of size 5000. 95% confidence intervals via estimates and standard errors obtained using the means of the sampling distributions and standard deviations of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty
Intercept	0.95	0.37	0.63	0.81
X.Cure	0.95	0.93	0.93	0.90
X14	0.95	0.76	0.92	0.83
X24	0.95	0.79	0.95	0.90
X23	0.95	0.91	0.95	0.94
X34	0.95	0.93	0.95	0.93
log(kappa14)	0.95	0.00	0.85	0.00
log(alpha14)	0.95	0.00	0.85	0.00
log(kappa24)	0.94	0.96	0.95	0.92
log(alpha24)	0.95	0.47	0.95	0.54
log(kappa23)	0.95	0.87	0.01	0.00
log(alpha23)	0.95	0.93	0.43	0.49
log(kappa34)	0.95	0.12	0.94	0.12
log(alpha34)	0.95	0.02	0.94	0.02
log(theta)	0.96		0.91	

Table 4.21: Simulation results for the case where 'Conlon with frailty is adequate', with 1000 replicates of size 5000. Parameter estimates calculated as mean of sampling distribution

	Truth	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon, no frailty	Illness-death	Univ. Cure	Univ. Cox
Intercept	-0.41	-0.41	-0.35	-0.38	-0.34		0.52	
X.Cure	0.50	0.50	0.51	0.50	0.51		0.65	
X14	0.25	0.25	0.21	0.24	0.20			
X24	0.50	0.50	0.45	0.49	0.46	0.02		
X23	0.25	0.25	0.22	0.25	0.23	-0.45	0.36	-0.46
X34	0.15	0.15	0.13	0.15	0.13	-0.01		
log(kappa14)	-8.52	-8.51	-7.28	-8.22	-7.21			
log(alpha14)	0.34	0.34	0.17	0.30	0.16			
log(kappa24)	-5.30	-5.30	-5.08	-5.29	-5.04	-6.05		
log(alpha24)	0.34	0.34	0.24	0.35	0.27	0.40		
log(kappa23)	-7.13	-7.14	-6.70	-7.17	-6.78	-4.49		
log(alpha23)	0.64	0.64	0.55	0.65	0.57	-0.05		
log(kappa34)	-6.21	-6.23	-5.70	-6.16	-5.70	-7.73		
log(alpha34)	0.26	0.26	0.15	0.25	0.15	0.48		
log(theta)	-1.71	-1.73		-1.90		0.58		

Table 4.22: Simulation results for the case where 'Conlon with frailty is adequate', with 1000 replicates of size 5000. Empirical and analytical standard errors calculated using the standard deviations of the sampling distributions and means of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty	Illness-death	Univ. cure	Univ. Cox
	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$	SD $\widehat{\text{SE}}$
Intercept	0.05 0.05	0.05 0.05	0.05 0.05	0.05 0.05		0.05 0.03	
X.Cure	0.06 0.06	0.06 0.06	0.06 0.06	0.06 0.06		0.07 0.07	
X14	0.05 0.05	0.05 0.04	0.05 0.05	0.05 0.04			
X24	0.06 0.06	0.06 0.06	0.06 0.06	0.06 0.06	0.08 0.07		
X23	0.07 0.08	0.07 0.07	0.07 0.07	0.07 0.07	0.08 0.08	0.07 0.07	0.06 0.06
X34	0.07 0.07	0.07 0.06	0.07 0.07	0.07 0.06	0.10 0.09		
log(kappa14)	0.27 0.27	0.20 0.19	0.26 0.26	0.20 0.19			
log(alpha14)	0.03 0.03	0.03 0.02	0.03 0.03	0.03 0.02			
log(kappa24)	0.13 0.12	0.10 0.11	0.12 0.12	0.11 0.11	0.25 0.16		
log(alpha24)	0.03 0.03	0.03 0.03	0.02 0.02	0.02 0.02	0.06 0.04		
log(kappa23)	0.18 0.17	0.15 0.16	0.17 0.17	0.15 0.15	0.17 0.13		
log(alpha23)	0.03 0.03	0.03 0.03	0.03 0.03	0.02 0.02	0.07 0.04		
log(kappa34)	0.16 0.16	0.14 0.14	0.16 0.16	0.14 0.14	0.21 0.19		
log(alpha34)	0.02 0.02	0.02 0.02	0.02 0.03	0.02 0.02	0.03 0.02		
log(theta)	0.13 0.13		0.15 0.16		0.11 0.07		

Table 4.23: Coverage probabilities for the case where 'Conlon with frailty is adequate', with 1000 replicates of size 5000. 95% confidence intervals via estimates and standard errors obtained using the means of the sampling distributions and standard deviations of the estimated analytical standard errors

	Proposed model	Proposed, no frailty	Conlon, frailty	Conlon no frailty
Intercept	0.95	0.78	0.91	0.70
X.Cure	0.95	0.95	0.95	0.95
X14	0.95	0.82	0.94	0.80
X24	0.95	0.81	0.95	0.88
X23	0.96	0.92	0.96	0.94
X34	0.95	0.91	0.95	0.91
log(kappa14)	0.95	0.00	0.79	0.00
log(alpha14)	0.95	0.00	0.79	0.00
log(kappa24)	0.94	0.46	0.95	0.35
log(alpha24)	0.94	0.02	0.92	0.12
log(kappa23)	0.95	0.22	0.94	0.37
log(alpha23)	0.95	0.04	0.94	0.15
log(kappa34)	0.95	0.06	0.94	0.06
log(alpha34)	0.95	0.00	0.93	0.00
log(theta)	0.95		0.84	

Marginal likelihood contributions

Let $\eta_\omega = X_\omega \beta_\omega$ and $\Lambda_{0,\omega}(y) = \int_0^y \lambda_{0,\omega}(s)\, ds$, for $\omega \in \{14, 24, 23, 34\}$.

$$
\begin{aligned}
f_1(y_1, y_2; X) &= \int_0^\infty f_1(y_1, y_2; X, \gamma) f_\gamma(\gamma)\, d\gamma \\
&= \int_0^\infty \left[(1-\pi) \cdot \gamma \cdot \lambda_{0,23}(y_1) e^{\eta_{23}} \cdot \exp\left\{ -\gamma \left(\Lambda_{0,23}(y_1) e^{\eta_{23}} + \Lambda_{0,24}(y_1) e^{\eta_{24}} \right) \right\} \cdot \right. \\
&\quad \gamma \cdot \lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \cdot \exp\left\{ -\gamma \Lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \right\} \cdot \\
&\quad \left. \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \right] d\gamma \\
&= \frac{(1-\pi) \cdot \lambda_{0,23}(y_1) e^{\eta_{23}} \cdot \lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \cdot (1+\theta)}{\left\{ 1 + \theta \left[\Lambda_{0,23}(y_1) e^{\eta_{23}} + \Lambda_{0,24}(y_1) e^{\eta_{24}} + \Lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \right] \right\}^{1/\theta + 2}}
\end{aligned}
$$

$$
\begin{aligned}
f_2(y_1, y_2; X) &= \int_0^\infty f_2(y_1, y_2; X, \gamma) f_\gamma(\gamma)\, d\gamma \\
&= \int_0^\infty \left[(1-\pi) \cdot \gamma \cdot \lambda_{0,23}(y_1) e^{\eta_{23}} \cdot \exp\left\{ -\gamma \left(\Lambda_{0,23}(y_1) e^{\eta_{23}} + \Lambda_{0,24}(y_1) e^{\eta_{24}} \right) \right\} \cdot \right. \\
&\quad \left. \exp\left\{ -\gamma \Lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \right\} \cdot \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \right] d\gamma \\
&= (1-\pi) \cdot \lambda_{0,23}(y_1) e^{\eta_{23}} \cdot \\
&\quad \left\{ 1 + \theta \left[\Lambda_{0,23}(y_1) e^{\eta_{23}} + \Lambda_{0,24}(y_1) e^{\eta_{24}} + \Lambda_{0,34}(y_2 - y_1) e^{\eta_{34}} \right] \right\}^{-1/\theta - 1}
\end{aligned}
$$

$$
\begin{aligned}
f_3(y_1, y_2; X) &= \int_0^\infty f_3(y_1, y_2; X, \gamma) f_\gamma(\gamma)\, d\gamma \\
&= \int_0^\infty \left[\pi \cdot \gamma \cdot \lambda_{0,14}(y_2) e^{\eta_{14}} \cdot \exp\left\{ -\gamma \Lambda_{0,14}(y_2) e^{\eta_{14}} \right\} \cdot \right. \\
&\quad \left. \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \right] d\gamma \\
&= \pi \cdot \lambda_{0,14}(y_2) e^{\eta_{14}} \cdot \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \cdot \\
&\quad \int_0^\infty \gamma^{1/\theta} \cdot \exp\left\{ -\gamma \left[\frac{1}{\theta} + \Lambda_{0,14}(y_2) e^{\eta_{14}} \right] \right\} d\gamma \\
&= \pi \cdot \lambda_{0,14}(y_2) e^{\eta_{14}} \left[1 + \theta \cdot \Lambda_{0,14}(y_2) e^{\eta_{14}} \right]^{-1/\theta - 1}
\end{aligned}
$$

$$
\begin{aligned}
f_4(y_1, y_2; X) &= \int_0^\infty f_4(y_1, y_2; X, \gamma) f_\gamma(\gamma)\, d\gamma \\
&= \int_0^\infty \left[f_3(y_1, y_2; X, \gamma) + \right. \\
&\quad \left. (1-\pi) \cdot \gamma \cdot \lambda_{0,24}(y_2) e^{\eta_{24}} \cdot \exp\left\{ -\gamma \left(\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}} \right) \right\} \right] \cdot f_\gamma(\gamma)\, d\gamma
\end{aligned}
$$

$$
\begin{aligned}
&= f_3(y_1, y_2; X) + \\
&\quad \int_0^\infty (1-\pi) \cdot \gamma \cdot \lambda_{0,24}(y_2) e^{\eta_{24}} \cdot \exp\left\{-\gamma \left(\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}}\right)\right\} \cdot \\
&\quad \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \, d\gamma \\
&= \pi \cdot \lambda_{0,14}(y_2) e^{\eta_{14}} \left[1 + \theta \cdot \Lambda_{0,14}(y_2) e^{\eta_{14}}\right]^{-1/\theta - 1} + \\
&\quad (1-\pi) \cdot \lambda_{0,24}(y_2) e^{\eta_{24}} \left\{1 + \theta \left[\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}}\right]\right\}^{-1/\theta - 1}
\end{aligned}
$$

$$
\begin{aligned}
f_5(y_1, y_2; X) &= \int_0^\infty f_5(y_1, y_2; X, \gamma) \cdot f_\gamma(\gamma) \, d\gamma \\
&= \int_0^\infty \pi \exp\left\{-\gamma \cdot \Lambda_{0,14}(y_2) e^{\eta_{14}}\right\} \cdot \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \, d\gamma \\
&= \pi \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \int_0^\infty \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma \left[\frac{1}{\theta} + \Lambda_{0,14}(y_2) e^{\eta_{14}}(t)\right]\right\} \, d\gamma \\
&= \pi \left[1 + \theta \cdot \Lambda_{0,14}(y_2) e^{\eta_{14}}\right]^{-1/\theta}
\end{aligned}
$$

$$
\begin{aligned}
f_6(y_1, y_2; X) &= \int_0^\infty f_6(y_1, y_2; X, \gamma) f_\gamma(\gamma) \, d\gamma \\
&= \int_0^\infty \left[f_5(y_1, y_2; X, \gamma) + \right. \\
&\quad \left. (1-\pi) \cdot \exp\left\{-\gamma \left(\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}}\right)\right\}\right] \cdot f_\gamma(\gamma) \, d\gamma \\
&= f_5(y_1, y_2; X) + \\
&\quad \int_0^\infty (1-\pi) \cdot \exp\left\{-\gamma \left(\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}}\right)\right\} \cdot \\
&\quad \left[\Gamma\left(1/\theta\right)\right]^{-1} \theta^{-\frac{1}{\theta}} \gamma^{\frac{1}{\theta}-1} \exp\left\{-\gamma/\theta\right\} \, d\gamma \\
&= \pi \left[1 + \theta \cdot \Lambda_{0,14}(y_2) e^{\eta_{14}}\right]^{-1/\theta} + (1-\pi) \left\{1 + \theta \left[\Lambda_{0,23}(y_2) e^{\eta_{23}} + \Lambda_{0,24}(y_2) e^{\eta_{24}}\right]\right\}^{-1/\theta}
\end{aligned}
$$

4.3 Joint risk prediction in the semicompeting risks setting

Table 4.24: Distribution of outcomes

	Acute GVHD, without death	Death with acute GVHD	Death without acute GVHD	Alive without acute GVHD	Missing
Full data	174	79	202	961	0
(%)	(12.3)	(5.6)	(14.3)	(67.9)	–
Adding missing category	123	61	164	686	382
(%)	(8.7)	(4.3)	(11.6)	(48.4)	(27.0)

Equivalence of absolute risk profiles in Section 3.3 and the transition probabilities in Putter et al. (2007)

We show the equivalence of the transition probabilities discussed in Putter et al. and the absolute risk profiles corresponding to the standard illness-death model using joint density in (3.4), for the Markov model, $\lambda_{0,3}(t_2|t_1) = \lambda_{0,3}(t_2)$. Let $\Lambda_\omega(t) = \int_0^t \lambda_\omega(s)\, ds$.

$$
\begin{aligned}
p_{\text{aGVHD only}}(t|\gamma) &= \int_t^\infty \int_0^t f_U(u,v|\gamma)\, du\, dv \\
&= \int_t^\infty \int_0^t \lambda_1(u|\gamma)\cdot\lambda_3(v|\gamma)\cdot \\
&\quad \exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)-\Lambda_3(v|\gamma)+\Lambda_3(u|\gamma)\}\, du\, dv \\
&= \left[\int_0^t \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)+\Lambda_3(u|\gamma)\}\, du\right] \\
&\quad \left[\int_t^\infty \lambda_3(v|\gamma)\exp\{-\Lambda_3(v|\gamma)\}\, dv\right] \\
&= \int_0^t \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)+\Lambda_3(u|\gamma)\}\cdot \\
&\quad \exp\{-\Lambda_3(t|\gamma)\}\, du \\
&= P_{12}(0,t|\gamma)
\end{aligned}
$$

$$
\begin{aligned}
p_{\text{aGVHD and death}}(t|\gamma) &= \int_0^t \int_u^t f_U(u,v|\gamma)\, dv\, du \\
&= \int_0^t \int_u^t \lambda_1(u|\gamma)\cdot\lambda_3(v|\gamma)\cdot \\
&\quad \exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)-\Lambda_3(v|\gamma)+\Lambda_3(u|\gamma)\}\, dv\, du \\
&= \int_0^t \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)\}\cdot
\end{aligned}
$$

$$
\begin{aligned}
&\left[\int_u^t \lambda_3(v|\gamma)\cdot\exp\{-\Lambda_3(v|\gamma)+\Lambda_3(u|\gamma)\}\,dv\right]\,du\\
&= P_{13}^2(0,t|\gamma)
\end{aligned}
$$

$$
\begin{aligned}
p_{\text{death only}}(t|\gamma) &= \int_0^t f_{T_1=\tau,T_2}(v|\gamma)\,dv\\
&= \int_0^t \lambda_2(v|\gamma)\cdot\exp\{-\Lambda_1(v|\gamma)-\Lambda_2(v|\gamma)\}\,dv\\
&= P_{13}^1(0,t|\gamma)
\end{aligned}
$$

$$
\begin{aligned}
p_{\text{alive without aGVHD}}(t|\gamma) &= \int_t^\infty\int_u^\infty f_U(u,v|\gamma)\,dv\,du + \int_t^\infty f_{T_1=\infty,T_2}(v|\gamma)\,dv\\
&= \int_t^\infty\int_u^\infty \lambda_1(u|\gamma)\cdot\lambda_3(v|\gamma)\cdot\\
&\qquad \exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)-\Lambda_3(v|\gamma)+\Lambda_3(u|\gamma)\}\,dv\,du+\\
&\qquad \int_t^\infty \lambda_2(v|\gamma)\cdot\exp\{-\Lambda_1(v|\gamma)-\Lambda_2(v|\gamma)\}\,dv\\
&= \int_t^\infty \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)+\Lambda_3(u|\gamma)\}\cdot\\
&\qquad \left[\int_u^\infty \lambda_3(v|\gamma)\exp\{-\Lambda_3(v|\gamma)\}\,dv\right]\,du+\\
&\qquad \int_t^\infty \lambda_2(v|\gamma)\cdot\exp\{-\Lambda_1(v|\gamma)-\Lambda_2(v|\gamma)\}\,dv\\
&= \int_t^\infty \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)+\Lambda_3(u|\gamma)\}\cdot\\
&\qquad \exp\{-\Lambda_3(u|\gamma)\}\,du+\\
&\qquad \int_t^\infty \lambda_2(v|\gamma)\cdot\exp\{-\Lambda_1(v|\gamma)-\Lambda_2(v|\gamma)\}\,dv\\
&= \int_t^\infty \lambda_1(u|\gamma)\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)\}\,du+\\
&\qquad \int_t^\infty \lambda_2(v|\gamma)\cdot\exp\{-\Lambda_1(v|\gamma)-\Lambda_2(v|\gamma)\}\,dv\\
&= \int_t^\infty (\lambda_1(u|\gamma)+\lambda_2(u|\gamma))\cdot\exp\{-\Lambda_1(u|\gamma)-\Lambda_2(u|\gamma)\}\,du\\
&= \exp\{-\Lambda_1(t|\gamma)-\Lambda_2(t|\gamma)\}\\
&= P_{11}(0,t|\gamma)
\end{aligned}
$$

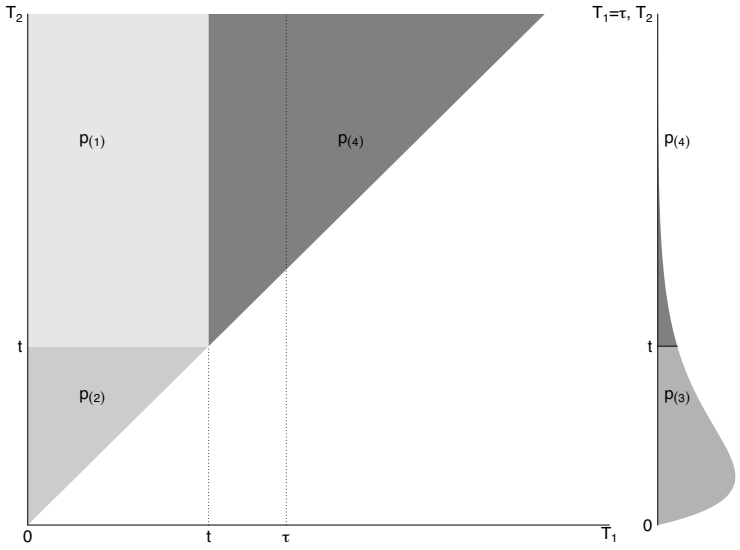

Figure 4.1: Diagram indicating the bounds of integration for calculating subject-specific risk profiles for the standard shared frailty illness-death model.

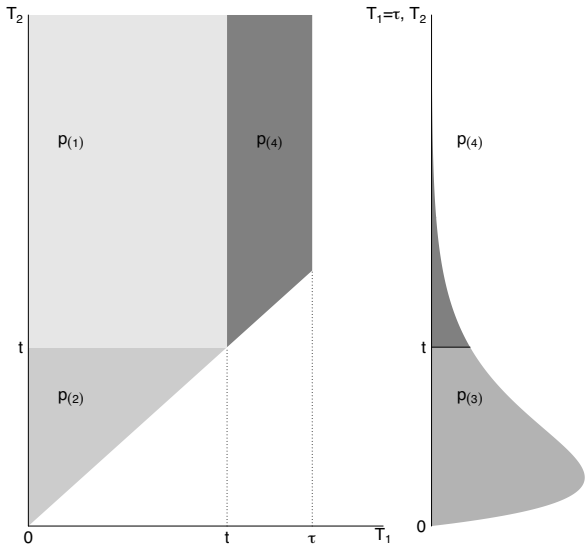

Figure 4.2: Diagram indicating the bounds of integration for calculating subject-specific risk profiles for the finite interval shared frailty illness-death model.

References

A BENNETT, D., A SCHNEIDER, J., ARVANITAKIS, Z. and S WILSON, R. (2012a). Overview and findings from the religious orders study. *Current Alzheimer Research* **9** 628–645.

A BENNETT, D., A SCHNEIDER, J., S BUCHMAN, A., L BARNES, L., A BOYLE, P. and S WILSON, R. (2012b). Overview and findings from the Rush Memory and Aging Project. *Current Alzheimer Research* **9** 646–663.

ALONZO, T. A. and PEPE, M. S. (2005). Assessing accuracy of a continuous screening test in the presence of verification bias. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54** 173–190.

ALZHEIMER'S ASSOCIATION, . (2017). 2017 Alzheimer's Disease Facts and Figures. *Alzheimer's and Dementia* **13** 325–373.

APA ET AL. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

BAMBER, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12** 387–415.

BARON, F., MARIS, M. B., SANDMAIER, B. M., STORER, B. E., SORROR, M., DIACONESCU, R., WOOLFREY, A. E., CHAUNCEY, T. R., FLOWERS, M. E., MIELCAREK, M. ET AL. (2005). Graft-versus-tumor effects after allogeneic hematopoietic cell transplantation with nonmyeloablative conditioning. *Journal of Clinical Oncology* **23** 1993–2003.

BENDER, R., AUGUSTIN, T. and BLETTNER, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine* **24** 1713–1723.

BERKSON, J. and GAGE, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47** 501–515.

BRAZZALE, A. R. (2005). hoa: An R package bundle for higher order likelihood inference. *Rnews* **5/1 May 2005** 20–27.

BROSTRM, G. (2016). *eha: Event History Analysis*. R package version 2.4-4.
URL `http://CRAN.R-project.org/package=eha`

BUCHMAN, A. S., WILSON, R. S., BIENIAS, J. L., SHAH, R. C., EVANS, D. A. and BENNETT, D. A. (2005). Change in body mass index and risk of incident alzheimer disease. *Neurology* **65** 892–897.

CAI, C., ZOU, Y., PENG, Y. and ZHANG, J. (2012). smcure: An r-package for estimating semiparametric mixture cure models. *Computer methods and programs in biomedicine* **108** 1255–1260.

CAPUANO, A. W., WILSON, R. S., LEURGANS, S. E., DAWSON, J. D., BENNETT, D. A. and HEDEKER, D. (2016). Sigmoidal mixed models for longitudinal data. *Statistical Methods in Medical Research* .

CHEN, H. Y. and LITTLE, R. J. A. (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Association* **94** 896–908.

CHI, Y.-Y. and ZHOU, X.-H. (2008). Receiver operating characteristic surfaces in the presence of verification bias. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **57** 1–23.

CHOI, S. W., LEVINE, J. E. and FERRARA, J. L. (2010). Pathogenesis and management of graft-versus-host disease. *Immunology and allergy clinics of North America* **30** 75–101.

CNAAN, A. and RYAN, L. (1989). Survival analysis in natural history studies of disease. *Statistics in Medicine* **8** 1255–68.

CONLON, A., TAYLOR, J. and SARGENT, D. J. (2014). Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in medicine* **33** 1750–1766.

COX, D. R. and OAKES, D. (1984). *Analysis of survival data*. Chapman and Hall, New York.

DREISEITL, S., OHNO-MACHADO, L. and BINDER, M. (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making* **20** 323–331.

DUC, K. T., CHIOGNA, M. and ADIMARI, G. (2016). Estimation of the volume under the ROC surface in presence of nonignorable verification bias. *arXiv preprint arXiv:1610.08805* .

EGLESTON, B. L., SCHARFSTEIN, D. O., FREEMAN, E. E. and WEST, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8** 526–545.

FAREWELL, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1041–1046.

FERGUSON, N. D., FAN, E., CAMPOROTA, L., ANTONELLI, M., ANZUETO, A., BEALE, R., BROCHARD, L., BROWER, R., ESTEBAN, A., GATTINONI, L. ET AL. (2012). The berlin definition of ards: an expanded rationale, justification, and supplementary material. *Intensive care medicine* **38** 1573–1582.

FERGUSON, T. S. (1996). *A course in large sample theory*, vol. 49. Chapman & Hall London.

FERRARA, J. L., LEVINE, J. E., REDDY, P. and HOLLER, E. (2009). Graft-versus-host disease. *The Lancet* **373** 1550–1561.

FIEBERG, J. and DELGIUDICE, G. D. (2009). What time is it? Choice of time origin and scale in extended proportional hazards models. *Ecology* **90** 1687–1697.

FILIPOVICH, A. H., WEISDORF, D., PAVLETIC, S., SOCIE, G., WINGARD, J. R., LEE, S. J., MARTIN, P., CHIEN, J., PRZEPIORKA, D., COURIEL, D. ET AL. (2005). National institutes of health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. diagnosis and staging working group report. *Biology of blood and marrow transplantation* **11** 945–956.

FINE, J. P., JIANG, H. and CHAPPELL, R. (2001). On semi-competing risks data. *Biometrika* 907–919.

FISHER, L. D. and LIN, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health* **20** 145–57.

FOLSTEIN, M. F., FOLSTEIN, S. E. and MCHUGH, P. R. (1975). Mini-mental state?: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12** 189–198.

GERDS, T. A., ANDERSEN, P. K. and KATTAN, M. W. (2014). Calibration plots for risk prediction models in the presence of competing risks. *Statistics in Medicine* **33** 3191–3203.

GRIFFIN, B. A., ANDERSON, G. L., SHIH, R. A. and WHITSEL, E. A. (2012). Use of alternative time scales in Cox proportional hazard models: implications for time-varying environmental exposures. *Statistics in Medicine* **31** 3320–3327.

HAN, B., YU, M., DIGNAM, J. J. and RATHOUZ, P. J. (2014). Bayesian approach for flexible modeling of semicompeting risks data. *Statistics in Medicine* **33** 5111–5125.

HAND, D. J. and TILL, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* **45** 171–186.

HANEUSE, S. and LEE, K. H. (2016a). Semi-competing risks data analysis. *Circulation: Cardiovascular Quality and Outcomes* **9** 322–331.

HANEUSE, S. and LEE, K. H. (2016b). Semi-competing risks data analysis. *Circulation: Cardiovascular Quality and Outcomes* **9** 322–331.

HE, H., LYNESS, J. M. and MCDERMOTT, M. P. (2009). Direct estimation of the area under the receiver operating characteristic curve in the presence of verification bias. *Statistics in Medicine* **28** 361–376.

HOLTAN, S. G., DEFOR, T. E., LAZARYAN, A., BEJANYAN, N., ARORA, M., BRUNSTEIN, C. G., BLAZAR, B. R., MACMILLAN, M. L. and WEISDORF, D. J. (2015a). Composite endpoint of graft-versus-host disease-free, relapse-free survival after allogeneic hematopoietic cell transplantation. *Blood* blood–2014.

HOLTAN, S. G., VERNERIS, M. R., SCHULTZ, K. R., NEWELL, L. F., MEYERS, G., HE, F., DEFOR, T. E., VERCELLOTTI, G. M., SLUNGAARD, A., MACMILLAN, M. L. ET AL. (2015b). Circulating angiogenic factors associated with response and survival in patients with acute graft-versus-host disease: results from blood and marrow transplant clinical trials network 0302 and 0802. *Biology of Blood and Marrow Transplantation* **21** 1029–1036.

HSIEH, J.-J., WANG, W. and ADAM DING, A. (2008). Regression analysis based on semi-competing risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 3–20.

INGRAM, D. D., MAKUC, D. M. and FELDMAN, J. J. (1997). Re: 'Time-to-event analysis of longitudinal follow-up of a survey: choice of the time scale'. *American Journal of Epidemiology* **146** 528–529.

JACK, C. R., KNOPMAN, D. S., JAGUST, W. J., PETERSEN, R. C., WEINER, M. W., AISEN, P. S., SHAW, L. M., VEMURI, P., WISTE, H. J., WEIGAND, S. D. ET AL. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* **12** 207–216.

JACK, C. R., KNOPMAN, D. S., JAGUST, W. J., SHAW, L. M., AISEN, P. S., WEINER, M. W., PETERSEN, R. C. and TROJANOWSKI, J. Q. (2010). Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology* **9** 119–128.

JACQMIN-GADDA, H., BLANCHE, P., CHARY, E., LOUBÈRE, L., AMIEVA, H. and DARTIGUES, J.-F. (2014). Prognostic score for predicting risk of dementia over 10 years while accounting for competing risk of death. *American Journal of Epidemiology* **180** 790–798.

JAZIĆ, I., SCHRAG, D., SARGENT, D. J. and HANEUSE, S. (2016). Beyond composite endpoints analysis: semi-competing risks as an underutilized framework for cancer research. *Journal of the National Cancer Institute* **108** djw154.

JOSEPH, R. W., COURIEL, D. R. and KOMANDURI, K. V. (2008). Chronic graft-versus-host disease after allogeneic stem cell transplantation: challenges in prevention, science, and supportive care. *J Support Oncol* **6** 361–372.

KEIDING, N. and KNUIMAN, M. (1990). Comments on 'Survival analysis in natural history studies of disease' by Cnaan and Ryan Stat Med. 1989 Oct;8(10):1255-68. *Statistics in Medicine* **9** 1221–2.

KIM, D. H., LEE, N. Y., LEE, M.-H. and SOHN, S. K. (2008). Vascular endothelial growth factor gene polymorphisms may predict the risk of acute graft-versus-host disease following allogeneic transplantation: preventive effect of vascular endothelial growth factor gene on acute graft-versus-host disease. *Biology of Blood and Marrow Transplantation* **14** 1408–1416.

KIM, N., OTT, L. S., SPIVACK, S., XU, X., LIU, A., BHAT, K. R. and OLADELE, C. (2012). Hospital-level, risk-standardized payment associated with a 30-day episode of care for AMI (Version 1.0) .

KORN, E. L., GRAUBARD, B. I. and MIDTHUNE, D. (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American Journal of Epidemiology* **145** 72–80.

KUK, A. Y. and CHEN, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 531–541.

LAI, C.-D. (2014). Generalized weibull distributions. In *Generalized Weibull Distributions*. Springer, 23–75.

LAKHAL, L., RIVEST, L.-P. and ABDOUS, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics* **64** 180–188.

LAMARCA, R., ALONSO, J., GOMEZ, G. and MUNOZ, A. (1998). Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **53** M337–43.

LEE, K. H., DOMINICI, F., SCHRAG, D. and HANEUSE, S. (2016). Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. *Journal of the American Statistical Association* **111** 1075–1095.

LEE, K. H. and HANEUSE, S. (2016). *SemiCompRisks: Hierarchical Models for Parametric and Semi-Parametric Analyses of Semi-Competing Risks Data*. R package version 2.3.2.

LEE, K. H., HANEUSE, S., SCHRAG, D. and DOMINICI, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64** 253–273.

LEE, S., KIM, H., HO, V., CUTLER, C., ALYEA, E., SOIFFER, R. and ANTIN, J. (2006). Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplantation* **38** 305–310.

LEVINE, J. E., LOGAN, B., WU, J., ALOUSI, A. M., HO, V., BOLAÑOS-MEADE, J., WEISDORF, D. ET AL. (2010). Graft-versus-host disease treatment: predictors of survival. *Biology of Blood and Marrow Transplantation* **16** 1693–1699.

LEVINE, J. E., LOGAN, B. R., WU, J., ALOUSI, A. M., BOLAÑOS-MEADE, J., FERRARA, J. L., HO, V. T., WEISDORF, D. J. and PACZESNY, S. (2012). Acute graft-versus-host disease biomarkers measured during therapy can predict treatment outcomes: a blood and marrow transplant clinical trials network study. *Blood* **119** 3854–3860.

LI, J. and FINE, J. P. (2008). ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics* **9** 566–576.

LIESTOL, K. and ANDERSEN, P. K. (2002). Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Statistics in Medicine* **21** 3701–3714.

LIN, D. Y. and YING, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88** pp. 1341–1349.

MACMILLAN, M. L., DEFOR, T. E. and WEISDORF, D. J. (2012). What predicts high risk acute graft-versus-host disease (gvhd) at onset?: identification of those at highest risk by a novel acute gvhd risk score. *British journal of haematology* **157** 732–741.

MOSSMAN, D. (1999). Three-way ROCs. *Medical Decision Making* **19** 78–89.

MUELLER, S. G., WEINER, M. W., THAL, L. J., PETERSEN, R. C., JACK, C. R., JAGUST, W., TROJANOWSKI, J. Q., TOGA, A. W. and BECKET, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's Dementia* **1** 55–66.

NAKAS, C. T. and YIANNOUTSOS, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine* **23** 3437–3449.

PACZESNY, S., KRIJANOVSKI, O. I., BRAUN, T. M., CHOI, S. W., CLOUTHIER, S. G., KUICK, R., MISEK, D. E., COOKE, K. R., KITKO, C. L., WEYAND, A. ET AL. (2009). A biomarker panel for acute graft-versus-host disease. *Blood* **113** 273–278.

PASSWEG, J., BALDOMERO, H., BADER, P., BONINI, C., CESARO, S., DREGER, P., DUARTE, R., DUFOUR, C., KUBALL, J., FARGE-BANCEL, D. ET AL. (2016). Hematopoietic stem cell transplantation in europe 2014: more than 40 000 transplants annually. *Bone marrow transplantation* **51** 786–792.

PENCINA, M. J., LARSON, M. G. and D'AGOSTINO, R. B. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* **26** 1343–1359.

PENG, L. and FINE, J. P. (2006). Nonparametric estimation with left-truncated semicompeting risks data. *Biometrika* **93** 367–383.

PENG, L. and FINE, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics* **63** 96–108.

PENG, Y. and DEAR, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56** 237–243.

PINTILIE, M. (2006). *Competing risks: a practical perspective*, vol. 58. John Wiley & Sons.

PUTTER, H., FIOCCO, M. and GESKUS, R. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* **26** 2389–2430.

R CORE TEAM (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

REMBERGER, M., PERSSON, U., HAUZENBERGER, D. and RINGDÉN, O. (2002). An association between human leucocyte antigen alleles and acute and chronic graft-versus-host disease after allogeneic haematopoietic stem cell transplantation. *British journal of haematology* **119** 751–759.

ROTNITZKY, A., FARAGGI, D. and SCHISTERMAN, E. (2006). Doubly robust estimation of the area under the receiver-operating characteristic curve in the presence of verification bias. *Journal of the American Statistical Association* **101** 1276–1288.

SAHA, P. and HEAGERTY, P. (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics* **66** 999–1011.

SCURFIELD, B. K. (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology* **40** 253–269.

SHLOMCHIK, W. D. (2007). Graft-versus-host disease. *Nature Reviews Immunology* **7** 340–352.

SORROR, M. L., MARTIN, P. J., STORB, R. F., BHATIA, S., MAZIARZ, R. T., PULSIPHER, M. A., MARIS, M. B., DAVIS, C., DEEG, H. J., LEE, S. J. ET AL. (2014). Pretransplant comorbidities predict severity of acute graft-versus-host disease and subsequent mortality. *Blood* **124** 287–295.

SPERRIN, M. and BUCHAN, I. (2013). Modelling time to event with observations made at arbitrary times. *Statistics in Medicine* **32** 99–109.

STEYERBERG, E. W., VICKERS, A. J., COOK, N. R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M. J. and KATTAN, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)* **21** 128.

SY, J. P. and TAYLOR, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics* **56** 227–236.

TCHETGEN TCHETGEN, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in medicine* **33** 3601–3628.

THERNEAU, T. M. (2015). *A Package for Survival Analysis in S*. Version 2.38.

THIÉBAUT, A. and BÉNICHOU, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in medicine* **23** 3803–3820.

TSIATIS, A. A., DEGRUTTOLA, V. and WULFSOHN, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association* **90** pp. 27–37.

URBANO-ISPIZUA, A., ROZMAN, C., PIMENTEL, P., SOLANO, C., DE LA RUBIA, J., BRUNET, S., PÉREZ-OTEYZA, J., FERRÁ, C., ZUAZU, J., CABALLERO, D. ET AL. (2002). Risk factors for acute graft-versus-host disease in patients undergoing transplantation with cd34+ selected blood cells from hla-identical siblings. *Blood* **100** 724–727.

VARADHAN, R., XUE, Q.-L. and BANDEEN-ROCHE, K. (2014). Semicompeting risks in aging research: methods, issues and needs. *Lifetime Data Analysis* **20** 538–562.

WANG, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 257–273.

WANG, Y. and TAYLOR, J. M. G. (2001). Jointly modeling longitudinal and event time

data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96** pp. 895–905.

WEISSINGER, E. M., SCHIFFER, E., HERTENSTEIN, B., FERRARA, J. L., HOLLER, E., STADLER, M., KOLB, H.-J., ZANDER, A., ZÜRBIG, P., KELLMANN, M. ET AL. (2007). Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood* **109** 5511–5519.

WIENKE, A. (2010). *Frailty models in survival analysis*. CRC Press.

WILSON, R. S., BOYLE, P. A., YU, L., BARNES, L. L., SYTSMA, J., BUCHMAN, A. S., BENNETT, D. A. and SCHNEIDER, J. A. (2015). Temporal course and pathologic basis of unawareness of memory loss in dementia. *Neurology* **85** 984–991.

WRIGHT, S. J. and NOCEDAL, J. (1999). Numerical optimization. *Springer Science* **35** 7.

WU, L., LIU, W., YI, G. Y. and HUANG, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics* **2012**.

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** pp. 330–339.

XU, J., KALBFLEISCH, J. D. and TAI, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics* **66** 716–725.

ZHANG, Y. and ALONZO, T. A. (2016). Inverse probability weighting estimation of the volume under the ROC surface in the presence of verification bias. *Biometrical Journal* **58** 1338–1356.

ZHENG, Y., CAI, T., JIN, Y. and FENG, Z. (2012). Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics* **68** 388–396.

ZHOU, H. and PEPE, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **82** pp. 139–149.