Sequential stratification methods for estimating effects of time-dependent treatments on multivariate survival outcomes

by

Abigail R. Smith

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2016

Doctoral Committee:

Professor Douglas Schaubel, Chair
Research Assistant Professor Sehee Kim
Professor Emeritus Robert Merion
Research Professor Ananda Sen

To Xavier and Zoe.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Doug Schaubel, for all of his support and encouragement over the past four years. I learned many valuable skills from him regarding both methodological research and applied clinical research, and our meetings were always very productive and stimulating, and never failed to move projects forward. I would also like to thank my committee members, Drs. Bob Merion, Sehee Kim, and Ananda Sen. Their thoughtful comments were very helpful in clarifying ideas and producing a polished final product. I am especially grateful to Dr. Merion for supporting my graduate studies financially for the past six years.

I am also very grateful for my parents, Liz Hilder and Randy Smith, for always believing in me. They instilled in me the value of education at a very young age, and throughout their lives have modeled the importance of finding a career you are passionate about.

Finally, I want to thank my husband, Taylor Blachley, and our children, Xavier and Zoe, for being my beacon of sanity during this process. I would not have completed this degree without the support of my husband, and I am so lucky to have found not only my best friend, but someone who I can talk about work with and is an amazing father to our kids. He truly is a SuperDad and I am thankful for him every day. I will also always be grateful for the unconditional love of my kids, and the balance they provided during this process; this dissertation is dedicated to them. I hope they one day find their passion as I have.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

Recurrent events are a common outcome of interest in many clinical applications, and are often used to evaluate new treatments. Many important clinical outcomes related not only to morbidity but also healthcare costs can occur repeatedly, and analysis of recurrent event rates allows for the estimation of treatment effects on these types of outcomes. Common examples of recurrent events include repeat infections, myocardial infarctions, and complications related to medical procedures as well as resource utilization outcomes like hospitalizations and medical costs. However, when treatment initiation occurs after the start of follow-up existing methods for the analysis of recurrent events are generally inapplicable or yield treatment effect parameters with unsatisfactory interpretations. In this dissertation we propose methodology that evaluates the effect of time-dependent treatments on multivariate survival outcomes.

In Chapter II we consider a time-dependent treatment that is relatively rare, and develop a two-stage method of estimating the effect of this treatment on the recurrent event rate. Since treatment is initiated after the start of follow-up the ideal comparison is between a subject that receives treatment at time s and the same subject under the scenario where the treatment does not exist. Since the

counterfactual absence-of-treatment experience is not observable in practice, our goal is to use other "similar" subjects to mimic this counterfactual experience. We identify these subjects by using a conditional prognostic model (i.e., conditional on previous events and any other relevant history) to find subjects with similar pre-treatment trajectories. Subjects that remain untreated at time s are then matched to the subject treated at time s if their prognostic score is within a given distance from the score of the treated subject. We then use the sequential stratification method (Schaubel et al, 2009) to estimate the effect of treatment on the recurrent event rate with each set of treated subjects and matched controls serving as a stratum in the analysis. The method conditions on the history up until the time of treatment (s), but is marginal beyond s. Note that matched subjects who are subsequently treated are censored from stratum where they serve as controls. While this generally results in dependent censoring, with a rare treatment, estimates remain unbiased. We seek to identify the threshold above which the dependent censoring results in bias.

In Chapter III we extend this method in two important directions. First, treatment is no longer assumed to be rare. Second, multiple treatments are available. In particular, we consider the case where there is both a "standard" and an "experimental" treatment. In this scenario all subjects begin follow-up untreated, some go on to receive standard treatment, others go on to receive the experimental treatment, and still others remain untreated. Our objective is to compare the recurrent event rates under experimental treatment and "conventional therapy"; i.e., beginning follow-up untreated and remaining in that state or subsequently receiving standard treatment. In this setting, subjects serving as matched controls are censored if they receive experimental treatment, but remain in the comparison group if they go on to receive standard treatment. We also assume that the experimental treatment is

more common, i.e. it exceeds the threshold determined in Chapter II that results in bias due to dependent censoring. To account for this we use a variant of Inverse Probability of Censoring Weighting (IPCW, Robins and Rotnitzky, 1992, Robins and Finkelstein, 2000). We model the hazard of experimental treatment using traditional proportional hazards methods adjusting for relevant covariates and history, and use this to construct weights representing the hazard of experimental treatment between time s and $s + t$. Sequential stratification methods are then used to determine the effect of experimental treatment on the recurrent event rate.

In Chapter IV we return to the scenario where there is only one treatment but remove the assumption that the treatment is rare. In this chapter we develop a method for estimating the effect of a time-dependent treatment on correlated recurrent event and survival outcomes. Here we assume that there is a latent process (unobserved random effect) related to both the recurrent and terminal events, and that the terminal event stops all subsequent observations of the recurrent event. Under this scenario we propose to jointly model the pre-treatment recurrent and terminal events using a frailty model in order to estimate the treatment-free trajectories for both event types as well as the subject-specific frailties. The matching of treated subjects then proceeds using the estimated intensities and hazards as well as the frailties. The final estimates of the treatment effect on the recurrent event rate and terminal event hazard are estimated separately using sequential stratification and the weighting method described in Chapter III.

The methods described above are applied to data from the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL). A2ALL is a multicenter NIH-funded cohort studying morbidity and mortality related to living donor liver transplant (LDLT) for both donors and recipients. The study collected both ret-

rospective and prospective data on potential LDLT recipients transplanted between 1998 and 2014. Data collection included demographic and clinical information related to recipient's liver disease and health status at donor evaluation, intraoperative information, vital status and laboratory values at specified time points post-donation, as well as information on complications and hospitalizations. Data collection began at the time of donor evaluation; some potential recipients went on to receive an LDLT, some received a deceased donor transplant (DDLT), and others remained on the waitlist.

CHAPTER II

Estimating the effect of a rare time-dependent treatment on the recurrent event rate

2.1 Introduction

Recurrent events often serve as the basis for measuring treatment effects in observational studies. A reduction in outcomes, such as repeated myocardial infarction or opportunistic infections, indicates that a treatment has a positive effect on morbidity. Reductions in hospital admission rates among the treatment group would imply that lower morbidity as well as reduced health care costs are associated with (or caused by) treatment.

Methods for analyzing recurrent events have been well described in the literature. Models have been developed that condition on the event history (Anderson and Gill, 1982) or previous number of events (Pepe and Cai, 1993). Marginal models, such as those of Lawless and Nadeau (1995) or Lin et al. (2000), allow for an interpretation of covariate effects on the recurrent event rate that does not require patients to have similar event histories. Few papers to date have explored methods for using recurrent events as an evaluation of an experimental treatment, with exceptions being Cook et al. (2009) and Schaubel and Zhang (2010).

Treatment can be initiated after the beginning of follow-up, which occurs frequently in studies without randomization. While some existing recurrent event

methods can incorporate time-dependent covariates (Chen et al, 2013), these traditional methods often do not give interpretations that satisfy the research question of interest. In the settings often of interest, treatment initiation depends on internal processes such as disease progression or the event history itself, violating the assumption of most time-dependent recurrent event methods that time-dependent covariates be external (Kalbfleisch and Prentice, 2002). Ideally, we would begin follow-up of an untreated patient, and after treatment initiation we would compare the recurrent event rate to that of the same patient had they remained untreated. This counterfactual experience is unobservable in practice, however.

In an attempt to compare each treated subject with their unobservable counterfactual treatment-free experience, this chapter will extend the sequential stratification method described by Schaubel et al. (2009) to the recurrent event setting. For every subject treated at time s, subjects that are eligible to receive treatment at time s but do not are matched to the treated subject. Each treated subjects's post-treatment recurrent event rate is compared to the averaged matched recurrent event rate in what can be conceptualized as a subject-level experiment. Matched subjects that subsequently receive treatment are censored from experiments for which they serve as controls, and begin their own experiment as the treated subject. Note that,in every experiment, the comparison of interest begins at time s, such that recurrent events that occur in $[0, s)$ are not considered.

Schaubel et al (2009) proposed combining hard covariate matching and adjustment to ensure that matched subjects were 'similar' to the treated subject in addition to the requirement that they remain untreated at time s. This method was proposed in the univariate survival setting where failure times prior to treatment are not observed for treated subjects. However, information regarding pre-treatment recurrent event

trajectories are available on all subjects in the setting described above. Given that event history is a strong predictor of the recurrent event rate, we propose to leverage this information using a two-stage modeling approach. In the first stage, we use a conditional rate model to describe pre-treatment event trajectories for all subjects. We then use the linear predictor from this first stage model to caliper-match as yet untreated patients to those receiving treatment at time s. The goal is to create a control group with an event trajectory similar to that which the treated patient would have experienced had treatment not been available. The final model for the recurrent event rate includes only the treatment effect and a measure of distance between the prognostic score of the treated subject and that of the matched controls.

The method proposed is not restricted to "treatment"' in the classical sense, and is in fact applicable to any state change. Often this state change is in the form of treatment such as initiation of new medication or performance of a procedure, but this is not always the case. Diagnosis of disease or experience of a medical event such as injury could constitute a state change for which comparing the recurrent event rate in the presence and absence of the state change is of clinical or policy interest. This will be discussed further in relation to the application of the method to liver transplantation.

The remainder of this chapter proceeds as follows. In Section 2.2 we introduce the notation and proposed models and describe the parameter estimation. Section 2.3 presents results of simulation studies to demonstrate the performance of the treatment effect estimator in moderate sized samples. An application to living donor liver transplant is described in Section 2.4 using data from the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL). Some concluding remarks are offered in Section 2.5.

2.2 Methods

2.2.1 Notation

In the following, i represents subject ($i = 1, \ldots, n$), T_i is treatment time, with $T_i \geq 0$, and $\boldsymbol{Z}_i^*(t)$ represents the time-dependent covariate for subject i. We assume for the purposes of this chapter that subjects treated at time T_i remain treated for the duration of follow-up. The true number of events for subject i in $[0, t]$ is defined as $N_i^*(t) = \int_0^t dN_i^*(u)$. Event and treatment times are subject to independent right censoring by C_i, assumed to be administrative in this setting without loss of generality. The number of observed events is given by $N_i(t) = \int_0^t I(C_i > u)\, dN_i^*(u)$.

The number of pre-treatment events in $(0, t]$ is given by the counting process

$$N_i^0(t) = \int_0^t I(T_i > u)\, dN_i^*(u). \tag{2.1}$$

If patient i receives treatment at time s; i.e., $T_i = s$, then the post-treatment event counter is defined as

$$N_i^\star(t; s) = I(T_i = s) \int_s^{s+t} dN_i^*(u). \tag{2.2}$$

Note that it will be our convention that $N(t; s)$ refers to the interval of length t, but starting at time s; a single time index, as in the previously-defined $N_i^0(t)$, pertains to the $(0, t]$ time interval. Correspondingly, we define an event counter representing the events that would have been experienced in the absence of treatment, also beginning at time s,

$$N_i^0(t; s) = \int_s^{s+t} I(T_i > u) dN_i^*(u). \tag{2.3}$$

Note that (2.3) is the pre-treatment event counter described in (2.1) but instead of $(0, t]$ the counter $N_i^0(t; s)$ tracks the patient on $(s, s + t]$. For a subject eligible to receive the treatment at time s, (i.e. $I(T_i \geq s)$), if $T_i = s$, the counting process (2.2)

takes effect; if the treatment had not been available, process (2.3) takes effect. The subject is untreated on $(0, s)$ under either scenario to which (2.2) and (2.3) pertain.

Finally, we define a 0/1 process for being observed to receive treatment,

$$N_i^T(t) = \int_0^t I(C_i > u) dI(T_i \leq u). \tag{2.4}$$

2.2.2 Proposed Models

As described above, the goal of this method is to compare the post-treatment recurrent event mean to the corresponding event mean under no treatment. We denote the mean of (2.2) by

$$\mu_i^\star(t; s) \;=\; E\left[\int_0^t N_i^\star(du; s) | T_i = s, \boldsymbol{H}_i(s)\right], \tag{2.5}$$

where $\boldsymbol{H}_i(s) = \{\boldsymbol{Z}_i^*(u), N_i(u), I(T_i > u), I(C_i > u);\ 0 \leq u < s\}$ represents the observed pre-treatment history for subject i on $[0, s)$.

Similarly, in the absence of treatment, the mean of (2.3) can be written as

$$\mu_i^0(t; s) \;=\; E\left[\int_0^t N_i^0(du; s) | \boldsymbol{H}_i(s), T_i > u\right]. \tag{2.6}$$

Note, both models are partly conditional (Pepe and Couper, 1997, Zheng and Heagerty 2005, Gong and Schaubel 2013) in the sense that they condition on the history up until time s as opposed to $s + t$. We do not model either (2.5) or (2.6) directly, instead, our model of interest is given by

$$\mu_i^\star(t; s) = \mu_i^0(t; s) \exp\{\beta_\star\}, \tag{2.7}$$

which can equivalently be expressed in terms of a rate function by

$$\mu_i^\star(dt; s) = \mu_i^0(dt; s) \exp\{\beta_\star\}. \tag{2.8}$$

In this model $\mu_i^0(t; s)$, the treatment-free mean number of events is scaled up or down by $\exp\{\beta_\star\}$ if subject i received the experimental treatment at time s. The mean

number of post-treatment events is then compared to the mean number of treatment-free events after time s. It is conceivable that the treatment effect could depend on time since treatment, t, or time of treatment, s, and this model can be extended to accommodate a time-dependent $\beta_\star$ in the form $\beta_\star(t;\cdot)$, $\beta_\star(\cdot;s)$ or $\beta_\star(t;s)$. Note that these different time-dependent forms of $\beta_\star$ could be any parametric function of time such as linear or log-linear; time could also be categorized to examine the functional form of the time-dependent effect.

Since we cannot observe a patients' pre-treatment experience once treatment is initiated, a patient treated at time s will be compared to similar patients who did not start treatment at follow-up time s but were eligible to do so. Similar to Schaubel et al. (2009), we use the concept that each treatment time initiates an "experiment", in which the recipient of the treatment is compared to 'similar' treatment-eligible candidates. Note that 'similar', in this context, refers to current status (i.e., at time s) and history on $[0, s)$. Eligibility for the comparison is defined as

$$e_i(s) = I(T_i = s) + I(T_i > s),$$

i.e., at time s, patient i either received the treatment or remained untreated.

Our method of estimating $\beta_\star$ from (2.8) involves a stratified analysis. Each treated patient generates a stratum, which will include the index patient as well as similar treatment-eligible patients. Here we define similar as both treatment eligible at s, $e_i(s)$, and similar with regard to accumulated covariate and recurrent event history on $(0, s]$, $\boldsymbol{H}_i(s)$. In order to quantify each subject's history, we use a prognostic score (Hansen, 2008) based on the pre-treatment event rate, modeled using a time-dependent proportional rates model,

$$d\mu_i^0(t) = E\left[dN_i^*(t)|\boldsymbol{H}_i(t), T_i > t\right] = \exp\{\boldsymbol{\alpha}_0^T \boldsymbol{Z}_i(t)\} d\mu_0(t), \tag{2.9}$$

where the covariate $\boldsymbol{Z}_i(t)$ is chosen to capture the pertinent components of the history, $E[dN_i^*(t)|\boldsymbol{H}_i(t), T_i > t] = E[dN_i^*(t)|\boldsymbol{Z}_i(t), T_i > t]$. Model (2.9) resembles the marginal Lin et al. (2000) model, but is more accurately interpreted as the conditional Andersen-Gill (1982) model, due to the explicit dependence on the prior event history, a property avoided by Lin et al. (2000). The regression parameter $\boldsymbol{\alpha}_0$ from (2.9) can be computed by solving the unweighted Cox (1975) score equation. Due to the dependence on internal covariates (Kalbfleisch and Prentice, 2002), elements of $\boldsymbol{\alpha}_0$ are difficult to interpret. However, the purpose of this model is matching similar subjects on $[0, s)$, not interpretation.

The purpose of the prognostic score is to match patients that have similar pre-treatment event rates, the rationale being that previous event rate is the most important predictor of the current event rate. Unlike a propensity score, which uses the treatment event rate to match subjects with similar probabilities of being treated, the prognostic score aims to compare the effect of treatment on the event rate among subjects that were on the same trajectory with respect to their pre-treatment event rate. The use of prognostic scores in conjunction with, or as an alternative to, propensity scores has been considered in several reports (Rubin and Thomas, 2000, Stuart, Lee, and Leacy, 2013, Leacy and Stuart, 2014, Li, Schaubel, and He, 2014) and will be discussed later. Once the prognostic scores have been estimated, caliper matching is used to assign untreated control subjects to a subject receiving treatment at time s. Caliper matching requires that the prognostic scores of matched subjects be within a certain radius of the prognostic score of the index subject. Appropriate selection of the caliper involves balancing the need for homogeneity within-stratum

with the need to have an adequate number of matches for each index subject. The discrepancy between prognostic scores for experimental subject j and control subject i can be quantified through the subject-pair specific rate ratio,

$$\psi_{i,j}(s) = \frac{d\mu_i^0(s)}{d\mu_j^0(s)} = \exp\{\boldsymbol{\alpha}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)]\}.$$

Subject i is 'similar' on $[0, s)$ to subject j if $|\log \psi_{ij}(s)| \leq \epsilon$, where $\epsilon > 0$ is a pre-determined constant.

Combining the eligibility indicators and prognostic scores, patient i is included in the stratum generated by patient j if $m_{ij}(s) = 1$, where

$$m_{ij}(s) \;\; = \;\; e_i(s)I(T_i > s)e_j(s)I(T_j = s)I(|\log \widehat{\psi}_{ij}(s)| \leq \epsilon),$$

with $\widehat{\psi}_{ij}(s) = \exp\{\widehat{\boldsymbol{\alpha}}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)]\}$. In order to account for the residual difference between patients i and j, we propose to adjust for $\log \widehat{\psi}_{ij}(s)$ in the final model. Incorporating the eligibility indicator and the prognostic score distance, the final fitted model for the event mean for stratum j is then

$$\mu_{ij}^\star(t; s) = m_{ij}(s)\mu_i^0(t; s) \exp\{\beta_\star I(T_i = s) + \beta_\psi \log \widehat{\psi}_{ij}(s)\}. \tag{2.10}$$

In (2.10), j is the stratum (generated by patient j through $T_j = s$) and i is the patient within stratum. The model governs the treated patient through the indicator $I(T_i = s)$, which equals 1 if $i = j$. The vector of parameters to be estimated and the corresponding covariates are given by

$$\boldsymbol{\beta}_{\star\psi} = \begin{bmatrix} \beta_\star \\ \beta_\psi \end{bmatrix} \quad \boldsymbol{Z}_i^\star(s) = \begin{bmatrix} I(T_i = s) \\ \log \widehat{\psi}_{ij}(s) \end{bmatrix}, \tag{2.11}$$

such that model (2.10) can be re-written as $\mu_i^\star(t; s) = m_{ij}(s)\mu_i^0(t; s) \exp\{\beta_{\star\psi}^T \boldsymbol{Z}_i^\star(s)\}$.

Subjects matched to the treated subject enter the experiment without receiving any treatment, but could subsequently receive treatment. If a matched subject receives treatment after time s they are censored from all experiments in which they

serve as controls and begin their own experiment as the index subject. This generally results in dependent censoring since, although treatment can be considered random given $\boldsymbol{H}_i(s+t)$, the model for $\mu^{\star}_{ij}(t;s)$ from (2.10) only conditions on $\boldsymbol{H}_i(s)$, the pre-treatment history up to time s. While this could be addressed though Inverse Probability of Censoring Weighting (IPCW, Robins and Finkelstein, 2000, Miloslavsky et al, 2004, Smith and Schaubel, 2015), in this chapter we consider treatments that are relatively rare, with rates small enough such that bias due to dependent censoring is negligible. Section 3 will investigate through simulation treatment rates at which dependent censoring needs to be addressed.

2.2.3 Parameter Estimation

In order to estimate $\beta_{\star}$ we define the pertinent risk set indicator for stratum j,

$$Y_{ij}(t;s) \;=\; m_{ij}(s)I(C_i > s+t)\{I(T_i = s) + I(T_i > s+t)\}.$$

If, given $\boldsymbol{H}_i(s)$ matched subjects are randomly assigned to treatment after time s, the process

$$m_{ij}(s)\int_0^{\tau-s} M_{ij}(du;s), \tag{2.12}$$

where $m_{ij}(s)$ is the matching indicator described above, $M_{ij}(du;s) = Y_{ij}(u;s)\{N_i(du;s) - \mu_{ij}(du;s)\}$, and τ is chosen to satisfy $P(C_i \geq \tau) > 0$ and often set to $\max\{C_1,\ldots,C_n\}$, would have mean zero. As mentioned above, bias due to censoring of subsequently treated controls is expected to be minimal in the setting of rare treatment, so we assume the condition above holds.

Aggregating across subjects for the experiment occurring at time s produces the set of zero mean processes,

$$\textstyle\sum_{i=1}^{n} m_{ij}(s)\int_0^t M_{ij}(du;s) \tag{2.13}$$

and

$$\sum_{i=1}^{n} m_{ij}(s) \int_0^t \boldsymbol{Z}_i^\star(s) M_{ij}(du; s). \tag{2.14}$$

We reorganize this system to solve implicitly for the baseline mean, $\mu_0^0(u; s)$ in (2.13), then substitute into (2.14). Then, aggregating across all experiments yields the final estimating function for $\boldsymbol{\beta}_{\star\psi}$,

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{n} \sum_{i=1}^{n} \int_0^\tau m_{ij}(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^\star(s) - \overline{\boldsymbol{Z}}_\star(u; s)\} N_i(du; s) dN_j^T(s), \tag{2.15}$$

where

$$\overline{\boldsymbol{Z}}_\star(u; s) = \frac{\sum_{\ell=1}^{n} Y_{\ell j}(u; s) \boldsymbol{Z}_\ell^\star(s) \exp\{\boldsymbol{\beta}_{\star\psi}^T \boldsymbol{Z}_\ell^\star(s)\}}{\sum_{\ell=1}^{n} Y_{\ell j}(u; s) \exp\{\boldsymbol{\beta}_{\star\psi}^T \boldsymbol{Z}_\ell^\star(s)\}}. \tag{2.16}$$

Since $U(\boldsymbol{\beta})$ from (2.15) behaves asymptotically like a zero-mean estimating function, the solution to $U(\boldsymbol{\beta}) = \mathbf{0}$, denoted by $\widehat{\boldsymbol{\beta}}_{\star\psi}$, should yield a consistent estimator of $\boldsymbol{\beta}_{\star\psi}$.

2.2.4 Asymptotic Properties

To proceed with inference on $\widehat{\boldsymbol{\beta}}$ we need to estimate the variance of $\widehat{\boldsymbol{\beta}}$. To do this we first explore the distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ as $n \to \infty$. Using results initially derived by Lin et al (2000) it can be shown that

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = A^{-1}(\boldsymbol{\beta}) n^{-1/2} \sum_{i=1}^{n} U_i(\boldsymbol{\beta}) + o_p(1), \tag{2.17}$$

where

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{n} \sum_{i=1}^{n} \int_0^\tau m_{ij}(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^\star(s) - \overline{\boldsymbol{Z}}_\star(u; s)\} N_i(du; s) dN_j^T(s)$$

as above and $A(\boldsymbol{\beta})$ is the limiting value of the second derivative matrix, given by

$$\widehat{A}(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^{n} \sum_{i=1}^{n} \int_0^\tau m_{ij}(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^\star(s) - \overline{\boldsymbol{Z}}_\star(u; s)\}^{\otimes 2} N_i(du; s)\, dN_j^T(s).$$

Note that in order to show (2.17) above it must be shown that (2.15) has mean zero. To do this we first rewrite the interior integral as from (2.15) as

$$\sum_{i=1}^{n} \int_0^{\tau-s} \{\boldsymbol{Z}_i^{\star}(s) - \overline{\boldsymbol{Z}}_{\star}(u;s)\} M_i(du;s), \tag{2.18}$$

where $M_i(t;s)$ is defined as above. Since (2.10) is not the intensity model of Anderson and Gill (1982) but a proportional rates model similar to Lin et al (2000), the $M_i(t;s)$ are not martingales and therefore the martingale central limit theorem does not apply. However, $E\{dM_i(t;s)|\boldsymbol{Z}_{\star}(u;s)\} = 0$ in the setting described above given

$$M_i(dt;s) = Y_i(t;s)[N_i(dt;s) - \mu_i^0(dt;s)\exp\{\beta_{\star\psi}^T \boldsymbol{Z}_i^{\star}(s)\}]$$

and therefore (2.18) has mean zero under the assumed model. It follows that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} \int_0^{\tau} \int_0^{\tau-s} \{\boldsymbol{Z}_i^{\star}(s) - \overline{\mathbf{z}}_{\star}(u;s)\} M_i(du;s) dN_j^T(u)$$

also has mean zero because $\{\boldsymbol{Z}_i^{\star}(s) - \overline{\mathbf{z}}_{\star}(u;s)\} M_i(du;s)$ and $dN_j^T(t)$ are independent when $i \neq j$ and when $i = j$ $dN_j^T(t) = 1$ is embedded in the covariate $\boldsymbol{Z}_i^{\star}(s)$.

It can be shown through empirical process theory (Shorack and Wellner, 1986, Karatzas and Shreve, 1988, Pollard, 1990, Bilias et al 1997) that $n^{-1/2}U(\boldsymbol{\beta})$ converges weakly to a zero mean Gaussian process under certain regularity conditions such as those listed in Lin et al (2000). Further, by the Weak Law of Large Numbers, the matrix $\widehat{A}(\boldsymbol{\beta})$ converges in probability to $A(\boldsymbol{\beta})$. Then, by applying Slutsky's Theorem, (2.17) converges to a normal distribution with mean zero and variance $A^{-1}(\boldsymbol{\beta})B(\boldsymbol{\beta})A^{-1}(\boldsymbol{\beta})$, where $B(\boldsymbol{\beta}) = E[U_i(\boldsymbol{\beta})U_i^T(\boldsymbol{\beta})]$. The form of the variance of (2.17) suggests the robust sandwich estimator with $\widehat{A}(\boldsymbol{\beta})$ as given above and $B(\boldsymbol{\beta})$ estimated by $\widehat{B}(\boldsymbol{\beta})$ where

$$\widehat{B}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i(\boldsymbol{\beta})^{\otimes 2}$$

with

$$\widehat{U}_i(\boldsymbol{\beta}) = \sum_{j=1}^{n}\sum_{i=1}^{n}\int_0^{\tau} m_{ij}(s)\int_0^{\tau-s}\{\boldsymbol{Z}_i^{\star}(s) - \overline{\boldsymbol{Z}}_{\star}(u;s)\}\widehat{M}_i(du;s)dN_j^T(s).$$

2.2.5 Variance Estimation

The variance of $\boldsymbol{\beta}$ can be estimated via the robust sandwich estimator. This was used for several reasons. First, recurrent events are clustered within subject, and therefore observed events are not independent. Second, subjects can serve as controls in multiple strata, i.e., the data set for the final model may include repeated instances of a subject's recurrent event experience. The performance of this robust sandwich estimator will be tested through simulation.

2.3 Simulation Study

2.3.1 Simulations of Proposed Method

We conducted simulations to demonstrate the properties of the proposed estimator in moderate sized samples. For each scenario we simulated 1000 subjects 500 times. In addition to the observed experience, the counterfactual, treatment-free experience was generated for each subject in order to determine target values for $\beta_{\star}$, which, given the complex data structure, were difficult to pre-specify. Independent adjustment covariates Z_{i1} and Z_{i2} were generated to follow a Bernoulli(0.5) distribution. Correlation between recurrent events for each subject was induced through a frailty variate, Q_i, distributed Gamma with mean 1 and variance 0.5. The frailty was capped at 2, the 90^{th} percentile. Pre-treatment recurrent event experience was then generated though a frailty model with rate parameter $Q_i d\mu_0 \exp\{\alpha_1 Z_{i1} + \alpha_2 Z_{i2}\}$. An additional unobserved event process related to treatment was also generated using a similar model to ensure dependent censoring. Treatment times, T_i^S, were then generated to

follow the hazard $\lambda_0^T \exp\{\delta_1 Z_{i1} + \delta_2 Z_{i2} + \delta_3 \log(N_{i1}(t^-) + 1) + \delta_4 \log(N_{i2}(t^-) + 1)\}$, where $N_{i1}(t)$ is the outcome of interest and $N_{i2}(t)$ is the unobserved event process also related to treatment. The recurrent event times post-treatment were generated from rate parameter $Q_i d\mu_0^T \exp\{\phi_1 Z_{i1} + \phi_2 Z_{i2} + \phi_3 \log(N_i(T_i) + 1)\}$.

Once the data were generated, prognostic scores representing pre-treatment event trajectories were obtained from the model $d\mu_i^0(t) = \exp\{\alpha_{01} Z_{i1} + \alpha_{02} Z_{i2} + \alpha_{03} N_i^0(t^-)\} d\mu_0^0(t)$. Subjects were matched if $|\log \widehat{\psi}_{ij}| \leq 0.025$.

Parameters used in the simulation studies are as follows. For the pre-treatment event rates, we set $d\mu_0 = 3$, $\alpha_1 = 0.3$, and $\alpha_2 = -0.1$ for the observed process and $d\mu_0 = 6$, $\alpha_1 = 0.3$, and $\alpha_2 = -0.1$ for the unobserved process. For the treatment hazard, $\lambda_0^T = 0.01$, $\delta_1 = -0.2$, $\delta_2 = 0.1$, $\delta_3 = 0.2$, and $\delta_4 = 0.5$. Finally, for the post-treatment event rate, $d\mu_0^T$ was given values of 1, 1.5, 2, 3, 4, and 5, $\phi_1 = 0.3$, $\phi_2 = -0.1$, and $\phi_3 = 0.2$. This resulted in values of $\beta_\star$ of -0.265, -0.002, 0.206, 0.514, 0.751, and 0.945. In the simulated data 9.13% of the sample received treatment, and the mean number of events was 12.4 (sd = 9.8).

Results from the simulations are shown in Table 2.1. Absolute bias ranged from 0.005 to 0.016, and coverage probabilities ranged from 0.91 to 0.94, close to the target level of 0.95. Histograms of the difference between the estimated and target values of $\beta_\star$ are shown in Figure 2.1. They show a relatively normal distribution centered at zero, supporting the claim that $\beta_\star$ is unbiased and asymptotically normal. The robust variance estimator performed well in this setting, with asymptotic standard error estimates similar to their empirical counterparts. This will be discussed further in Section 2.5.

Table 2.1: Simulation results for proposed method estimating rare time-dependent treatment effects on the recurrent event rate

Scenario	$d\mu_0^1$	$\beta_\star$	Estimate	Bias	ESE	ASE	CP
1	1.0	−0.265	−0.271	−0.005	0.084	0.078	0.91
2	1.5	−0.002	−0.013	−0.010	0.078	0.074	0.94
3	2.0	0.206	0.189	−0.016	0.077	0.071	0.93
4	3.0	0.514	0.506	−0.008	0.072	0.068	0.93
5	4.0	0.751	0.742	−0.009	0.071	0.066	0.92
6	5.0	0.945	0.930	−0.015	0.072	0.065	0.92

ESE=empirical standard error; ASE=asymptotic standard error; CP=coverage probability

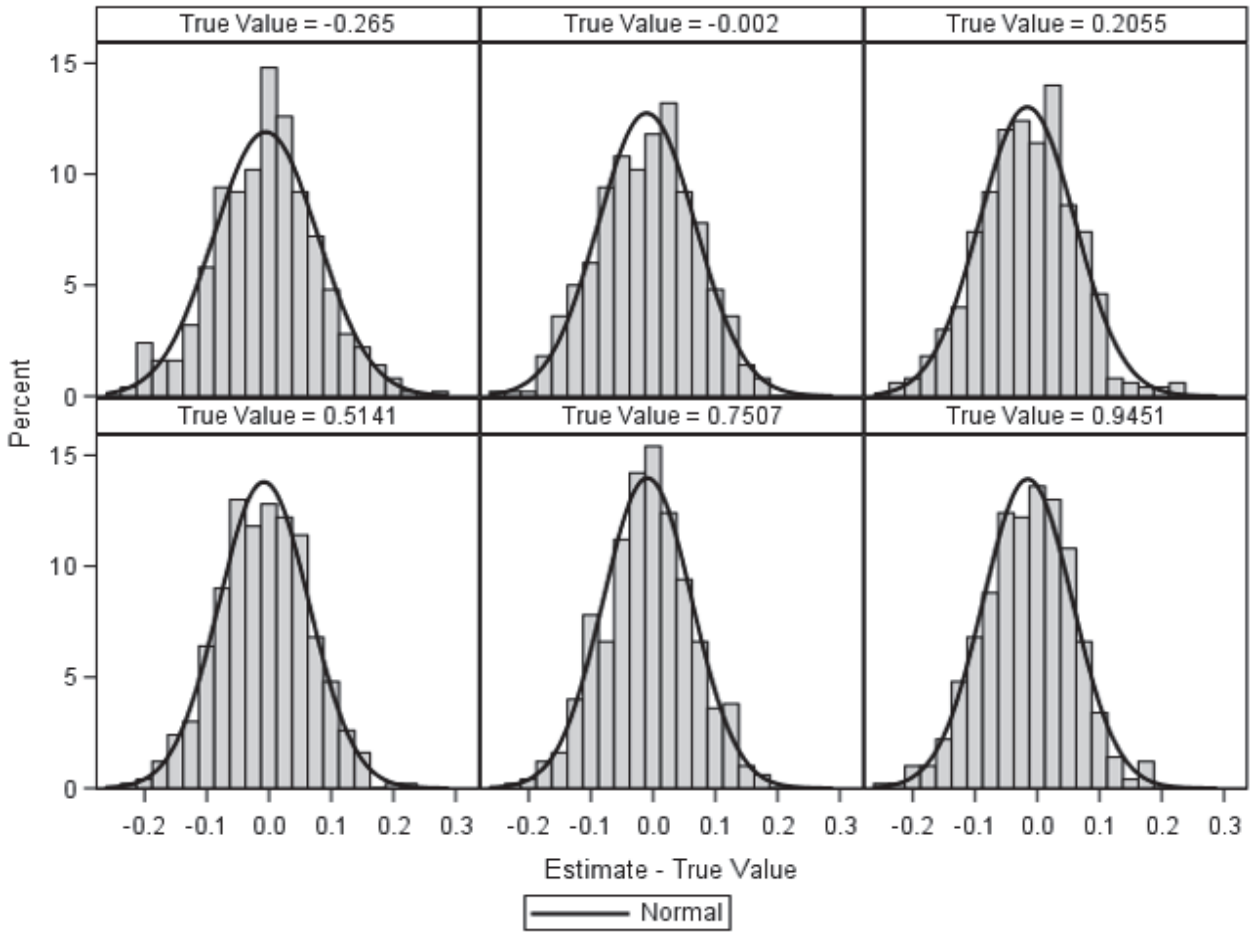

Figure 2.1: Histogram of parameter estimates from proposed method estimating rare time-dependent treatment effects on the recurrent event rate with normal density

Table 2.2: Simulation results for proposed method with increasing percent treated demonstrating bias due to dependent censoring

Scenario	% Treated	$\beta_\star$	Estimate	Bias	ESE	ASE	CP
1	9.44	−0.305	−0.262	0.043	0.118	0.108	0.89
2	24.82	−0.314	−0.276	0.038	0.073	0.071	0.91
3	36.69	−0.330	−0.280	0.050	0.066	0.062	0.87
4	56.66	−0.361	−0.292	0.069	0.062	0.057	0.74
5	68.71	−0.387	−0.300	0.086	0.066	0.057	0.62
6	85.62	−0.443	−0.305	0.138	0.075	0.062	0.42

ESE=empirical standard error; ASE=asymptotic standard error;
CP=coverage probability

2.3.2 Investigation of dependent censoring

Recall that subjects are censored from strata in which they serve as controls if they subsequently receive treatment. Since treatment depends on, among other things, the event history, this will result in dependent censoring in cases where treatment is not rare. We used simulation to explore the point at which more common treatments result in substantial bias. To do this we simulated 500 patients using a similar set up to that of the previous section except that $d\mu_0^T$ was set at 1 and λ_0^T took on values of 0.01, 0.05, 0.15, and 0.3. This resulted in 9.44%, 24.82% 36.69%, 56.66%, 68.71%, and 85.62% of subjects receiving treatment, respectively. Results of these simulations are shown in Table 2.2.

As shown in Figure 2.2, increasing the proportion of subjects treated increases bias and decreases coverage probability. At approximately one-third of subjects treated bias is at 0.05, and this almost triples to 0.14 when 86% of subjects are treated, with increased bias resulting in lower coverage. Bias and coverage are similar from 10-25% treated, likely due in this case to the smaller absolute number of treated subjects.

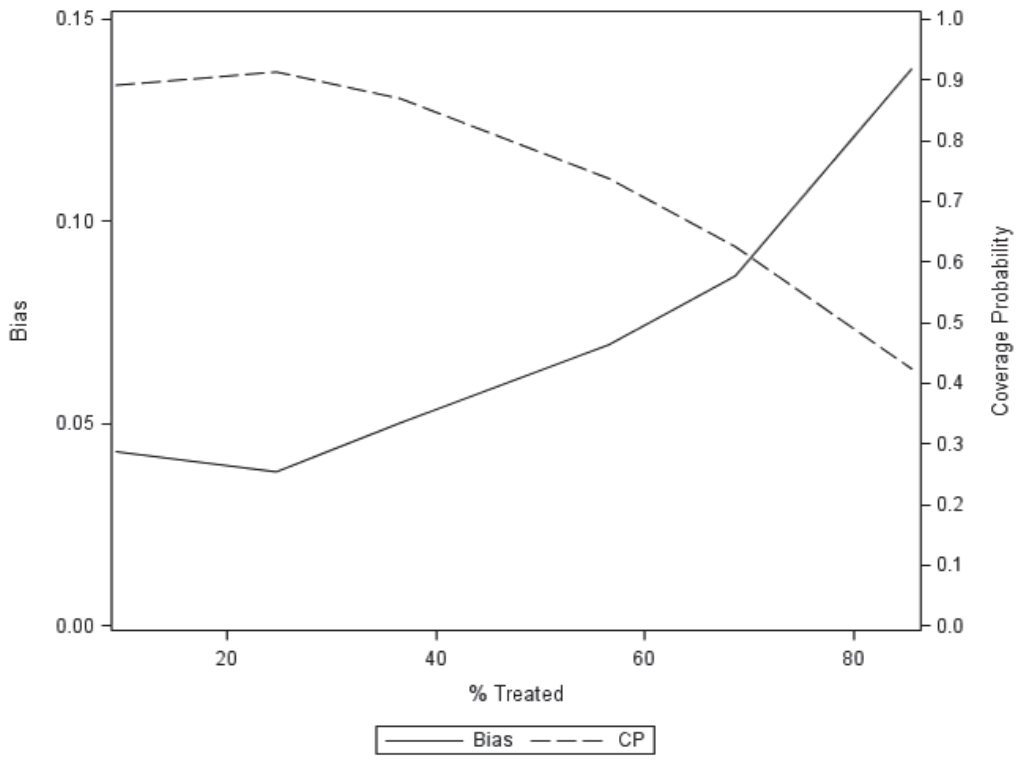

Figure 2.2: Bias and coverage probability with increasing percentage treated

Given this trajectory we recommend that if the proportion of treated patients is less than 20% methods such as IPCW aimed at correcting dependent censoring are not necessary, however, once the proportion of treated subjects exceeds 20% weighting is necessary to correct bias.

2.4 Application to Liver Transplantation

Development of End Stage Renal Disease (ESRD) post-liver transplant leads to increased patient morbidity and mortality, and places increased burden on heath care resources. We will use the proposed method to evaluate effect of ESRD development post-liver transplant on the number of days hospitalized in the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL). In this setting the "treatment" of interest is development of ESRD, defined as initiation of dialysis or kidney trans-

plant post-liver transplant. As mentioned previously, the proposed method is generalizable to time-dependent state changes such as development of post-LT ESRD. In this setting we utilize time-dependent markers of kidney function such as creatinine to estimate the hopitalization trajectory from the time of transplant to the development of ESRD, and use these to match with patients on similar trajectories that do not develop ESRD. Comparing the rate of days hospitalized for a patient that develops post-LT ESRD compared to the rate that would have been observed had the patient not developed ESRD is a critical component to the estimation of the costs of post-LT care.

A2ALL is a multi-center NIH-funded consortium composed of 12 North American transplant centers. Potential living donor liver transplant (LDLT) recipients transplanted between January 1, 1998 and January 31, 2014 were enrolled. Retrospective and prospective data collection included post-transplant vital status and laboratory information as well as hospitalization admission and discharge information. Data were supplemented from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States; these data are submitted by the members of OPTN and have been described elsewhere. The Health Resources and Services Administration (US Department of Health and Human Services) provides oversight for the activities of the OPTN and SRTR contractors.

There were 55 ESRD events out of 1447 transplanted patients in A2ALL. Median post-transplant follow-up time was 5 years, and the average number of days hospitalized per patient year was 14.9 for non-ESRD patients and 37.2 for ESRD patients (median days 2.3 and 5.1, respectively). Hospitalization admissions that occurred after discharge from the transplant hospitalization but before onset of ESRD were

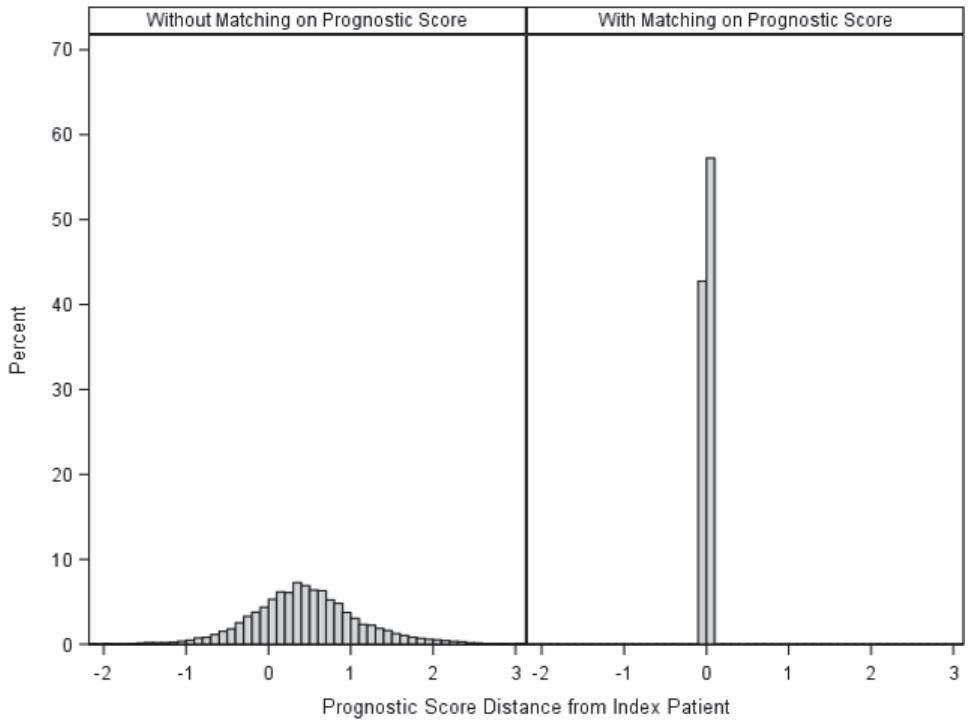

Figure 2.3: Distribution of prognostic score distance from index ESRD patient with and without matching

used to build the prognostic model, which was adjusted for the event history as well as other transplant and post-transplant time-dependent predictors. Results from the prognostic model are shown in Table 2.3. Each additional day of hospitalization history was associated with a 2% increase in the rate of future hospitalization days ($p < 0.001$).

Using prognostic scores derived from the model in Table 2.3, the distribution of prognostic score distance from index patient with and without matching is shown in Figure 2.3. Prior to matching on prognostic score the range of distance between the index subject and matched controls spans from -7.2 to 5.4, with 98% of matched controls within the interval $[-1.5, 3.5]$ from the index subject. When the 55 patients that developed ESRD post-transplant were matched to patients that had not yet developed ESRD based on prognostic score, with all control subjects within ± 0.02, the

Table 2.3: A2ALL analysis: Prognostic model of pre-ESRD rate of days hospitalized

Parameter	Rate Ratio	95% Confidence Interval	p-value
Recipient age at Transplant (ref=65+)			
18−40	1.00	(0.95, 1.06)	0.873
40−50	0.77	(0.73, 0.81)	<.001
50−55	0.79	(0.75, 0.83)	<.001
55−60	0.78	(0.74, 0.82)	<.001
60−65	0.87	(0.82, 0.92)	<.001
Recipient diagnosis: HCV	1.07	(1.04, 1.10)	<.001
African−American (ref=all others)	0.75	(0.71, 0.79)	<.001
Diabetes	0.89	(0.86, 0.92)	<.001
Ln(creatinine) (time−dependent)	1.31	(1.27, 1.35)	<.001
Ln(bilirubin) (time−dependent)	1.09	(1.07, 1.10)	<.001
Ln(albumin) (time−dependent)	0.36	(0.34, 0.38)	<.001
Donor age (ref=70+)			
<18	0.64	(0.55, 0.74)	<.001
18−40	0.76	(0.67, 0.87)	<.001
40−50	0.78	(0.69, 0.89)	<.001
50−60	1.02	(0.90, 1.16)	0.722
60−70	0.72	(0.63, 0.83)	<.001
DCD (ref=non−DCD)	1.33	(1.21, 1.47)	<.001
Regional (ref=Local)	1.32	(1.25, 1.40)	<.001
National (ref=Local)	1.41	(1.32, 1.51)	<.001
Split Liver	1.18	(1.08, 1.29)	<.001
Living Donor (ref=Deceased Donor)	0.96	(0.88, 1.05)	0.397
Hospitalization History (per day)	1.02	(1.02, 1.02)	<.001

distribution of score distance is much tighter around zero. The matching resulted in a median of 14 matches, with 6 (11%) patients that developed ESRD being excluded due to lack of matches.

The proposed method was then used to fit a stratified model to determine the effect of ESRD on the rate of days hospitalized (Model I) using model (2.10). The following traditional time-dependent proportional rates models were also fitted where ESRD status was treated as a time-dependent predictor adjusted for the same predictors in the prognostic model (Table 2.3):

$$\mu_i(t) = \mu_i^0(t) \exp\{\theta_{\text{II}} I(T_i \leq t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\} \tag{2.19}$$

and

$$\mu_i(t) = \mu_i^0(t) \exp\{\theta_{\text{III}} I(T_i \leq t) + \boldsymbol{\beta}^T \boldsymbol{Z}_i\}. \tag{2.20}$$

In the first model, model (2.19), additional time-dependent predictors thought to be associated with the progression to ESRD were included, such as lab values and hospitalization history, while in second model, (2.20), only baseline, i.e. at transplant, values of these predictors were used. The results from all three models are shown in Table 2.4. In Model I, which uses the proposed method, patients that develop ESRD have a rate of days hospitalized that is 2.9 times higher than patients that have not yet developed ESRD. By contrast, results from Model II give a rate that is only 1.4 times higher for patients that have developed ESRD, but this comparison is to patients without ESRD that have the same lab values and hospitalization history at time t. By contrast, adjusting only for baseline values of factors associated with the development of ESRD (Model III) estimates that patients that develop ESRD have a days hospitalized rate 3.2 times higher than patients that do not have ESRD at time t and were similar at transplant. This comparison demonstrates how use of

Table 2.4: A2ALL analysis of the effect of post-LT ESRD development on the rate of days hospitalized: Comparison of proposed method with traditional approaches

Model	Equation	Parameter	RR	95% CI	p-value
I: Proposed Method	(10)	$\beta_\star$	2.90	(1.69, 4.97)	<0.001
II: Time-dependent Adjustment Covariates	(16)	θ_{II}	1.44	(1.35, 1.53)	<0.001
III: Baseline Adjustment Covariates	(17)	θ_{III}	3.17	(3.01, 3.35)	<0.001

RR=Rate Ratio; CI = Confidence Interval

the proposed method balances the opposing biases of over- and under-adjustment. As a sensitivity analysis we also tested interactions with time since development of ESRD and time of development of ESRD, but no significant variations in the effect of ESRD development were found (both rate ratios ≈ 1, $p = 0.23$ and $p = 0.68$, respectively).

2.5 Discussion

In this chapter we lay out a two-stage method for estimating the effect of time-dependent treatments on recurrent events using an extension of the method of sequential stratification. The method proposed is partly conditional in the sense that information up until treatment time, s, is used in the prognostic model, but the final model for $\mu^\star_{ij}(t; s)$ does not condition on covariates after s. A purely conditional model (e.g. Anderson and Gill, 1982), which would include covariate information on $[s, s + t)$, would tend to dampen the effect of treatment because it would require comparison subjects to have the same history at the time of treatment. A marginal analysis such as that of Lin et al (2000), on the other hand, would exaggerate the effect of treatment because since treatment depends on the history, subjects that receive treatment at time s may differ from those that do not. Our method ensures subjects are similar up to s through conditioning, and is marginal thereafter, so that

the post-treatment comparison averages over the treatment-free experience of the matched controls.

The opposing biases of baseline and time-dependent adjustment in traditional time-dependent models is demonstrated in the application to development of post-liver transplant ESRD in the A2ALL study. While all three methods produced significant results, the fully time-dependent model, which included a time-dependent indicator for development of ESRD as well as time-dependent lab values and previous number of days hospitalized, all factors associated with progression of renal failure, underestimated the effect of ESRD on the rate of days hospitalized by almost half, while the similar model with only baseline factors associated with the development of ESRD (i.e. at transplant values), overestimated the effect. In addition, the comparison groups for these models are not constructed in a way that gives the desired interpretation, i.e. a comparison of the event rate in the time period following treatment in the presence and absence of treatment.

Note that the outcome chosen in the application was days hospitalized instead of hospital admissions. Analyses of hospital admissions often ignore the fact that patients are not at risk for hospitalization during the period in which they are in the hospital. This can be accounted for by removing the duration of hospitalization from the risk set, however, this step is often ignored. Modeling days hospitalized instead of hospital admissions automatically removes this potential for error, however, in some situations, hospital admissions may be a more relevant outcome.

As mentioned previously, it follows from the dependence of treatment initiation on the event history that censoring of matched controls due to treatment would constitute dependent censoring. We have shown that when the treatment is rare, bias is not substantial, and therefore IPCW to correct for dependent censoring is not

necessary. However, for more common treatments, bias will be induced and therefore some sort of weighting must be done to preserve the unbiased properties of $\beta_\star$.

It should also be noted that while the sandwich estimator performed well in the setting of rare treatment, issues related to the limiting value of $\overline{\boldsymbol{Z}}_\star(u;s)$ may affect variance estimates. In particular, the condition that $\overline{\boldsymbol{Z}}_\star(u;s)$ converges to a constant $\overline{\mathbf{z}}_\star(u;s)$ is difficult to justify in this setting because the component of $\boldsymbol{Z}_i^\star(s)$ corresponding to treatment will always equal 1 within stratum j due to the continuous time scale, and as $n \to \infty$ this will not change. The performance of the asymptotic variance estimator in this setting does not imply equivalent performance in more complex settings, such as those described in subsequent chapters. One potential solution to this issue is to use a discrete time scale, as is often done in practice. If patients receive treatment within a given unit of discrete time (e.g. day, week) and match on prognostic score, the group of treated patients could then be matched to appropriate controls. In this setting, as $n \to \infty$, a limiting value of $\overline{\boldsymbol{Z}}_\star(u;s)$, $\overline{\mathbf{z}}_\star(u;s)$, is a more defensible assumption. In finite samples, however, where increasing the number of treated patients per stratum is not feasible, and the data structure is more complex than that described in this chapter, this problem may persist. Potential solutions to this issue, including additional variance estimation procedures, are described in Chapters III and IV.

The proposed method makes use of the prognostic score in order to match as yet untreated patients into strata. Another viable alternative would be propensity matching (Rosenbaum and Rubin, 1983), i.e., matching on the probability of receiving treatment. A time-dependent propensity score has been proposed by Lu (2005), and could be used in this setting. Our goal, however, was to create a comparison group that mirrored the treatment-free experience of a subject treated at time s. It

was therefore necessary to ensure that the event trajectories up until s were the same between treated and control subjects, a property that the propensity score does not preserve. Some combination of prognostic and propensity scores could also be used.

One limitation of the proposed method is that it does not account for terminating events that halt all further occurrences of the recurrent event. The terminal event is often correlated with the recurrent event, and when treated as a censoring mechanism, can result in dependent censoring. Numerous methods for the simultaneous modeling of recurrent and terminal events are currently available. Marginal models of recurrent events while subjects are alive have been described by Cook and Lawless (1997) and Ghosh and Lin (2000). Joint modeling methods where the recurrent and terminal event processes are linked through a subject-level random effect have been proposed by Liu et al. (2004), Ye et al. (2007), Kalbfleisch et al. (2013), and others. Development of these methods in the setting described above is explored in Chapter IV.

CHAPTER III

Time-dependent prognostic score matching for recurrent event analysis with a common multi-state treatment

3.1 Introduction

Events that can occur repeatedly for the same subject are often of interest in clinical settings. Examples include repeated hospitalizations, post-surgery complications, viral infections, or myocardial infarctions. Many methods for modeling recurrent events have been described in the literature, including conditional models (e.g., Andersen and Gill, 1982; Pepe and Cai, 1993) and marginal approaches (Lawless and Nadeau, 1995; Lin et al., 2000; Pena, Strawderman and Hollander, 2001). A comprehensive survey of methodology for recurrent events is provided by Cook and Lawless (2007).

It may be of interest to study recurrent events as a basis for evaluating a treatment, and a limited number of methods are available for this purpose (e.g., Cook et al., 2009; Schaubel and Zhang, 2010). For chronic conditions, it is often the case that treatment begins some time after the start of follow-up, such that treatment is time-dependent. Although methods exist for analyzing recurrent events in the presence of time-varying covariates (Chen et al., 2013), such methods were not designed for evaluating treatments. Most recurrent event methods assume that time-dependent elements are external (Kalbfleisch and Prentice, 2002), which essentially rules out

their application to evaluate time-dependent treatments that are not randomized.

In the setting we consider in this report, treatment is time-dependent, in that subjects typically begin follow-up untreated, with some going on to receive treatment. Two forms of treatment are available, a new 'experimental' treatment and the standard treatment. The conventional course for a subject is to start follow-up untreated, then possibly receive the standard treatment. Our objective is to compare the recurrent event rate for subjects receiving the experimental treatment to that which they would have experienced under conventional therapy (which would be a combination of 'untreated' and 'standard treatment'). In Figure 3.1, we depict the recurrent event experience of the same subject i under two scenarios. For the top time line, subject i receives the experimental treatment at follow-up time s; after time s, the event course changes in relation to what it would have been on conventional therapy. The bottom time line pertains to the same subject, but under a scenario where experimental treatment does not exist. The comparison of interest is the ratio of the recurrent event rates t time units following time s. Note that the comparison of interest begins at the time of experimental treatment initiation, s, since the experimental treatment would clearly have no effect on subject i during the $[0, s)$ time interval. Ideally, we would like to compare each experimentally treated patient with their 'ghost', who remains on the conventional course; naturally, we never observe the latter if we observe the former.

Additional features of the data structure we consider are as follows. There are longitudinal measures available on each subject; i.e., time-dependent covariates. Moreover, the event history itself (e.g., up to the time of treatment) naturally provides information on the anticipated event rate (e.g., post-treatment). Treatment is not randomized, and generally depends on the time-varying covariates and the observed

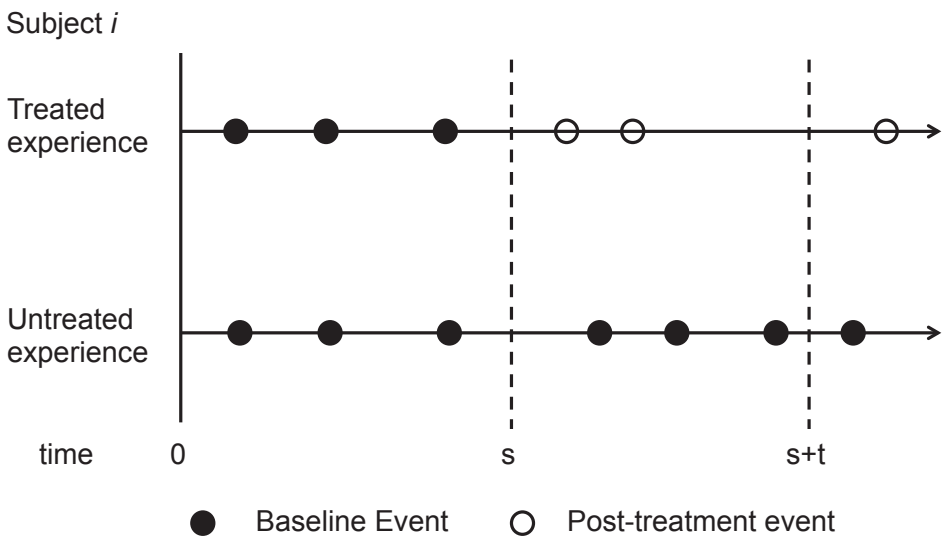

Figure 3.1: Recurrent event process for subject i under in the presence and absence of experimental treatment.

recurrent event experience. Finally, the experimental treatment effect may not be homogeneous and could be modified by the time-dependent entities.

The above-described data structure arises in the study of hospitalization rates after living donor liver transplantion (LDLT). The standard treatment for end-stage liver disease (ESLD) is deceased donor liver transplantation (DDLT). However, LDLT is a relatively new treatment option for ESLD that involves using a partial graft from a living donor in place of the traditional whole graft from a deceased donor. Our goal is to compare hospitalization rates for a patient who receives an LDLT relative to what that patient's hospitalization rate would have been had they remained on the waiting list and potentially received a DDLT.

Note that a traditional analysis (e.g., proportional rates model) featuring a time-dependent experimental treatment indicator 0/1 would generally produce a biased estimate of the contrast of interest (described in the preceding paragraph). This is largely due to the comparison groups having not been constructed appropriately (i.e., in a manner which respects the timing of the treatment assignments across subjects).

We illustrate this phenomenon in Section 3.4 through the data set which motivated our work.

Since it is not possible to observe the treatment-free event process for subjects that undergo treatment, a potential alternative is to compare those who receive a certain time-dependent treatment at time s to other similar subjects who were eligible to receive the treatment at time s but do not. The averaged event rates after s of matched subjects are then compared to the post-treatment event rate of the index subject. Methodology for this type of analysis currently exists for the univariate survival setting (Schaubel et al., 2009). In the current report, we extend this method to the recurrent event setting.

There are several key differences in the methods we propose in this report and those of Schaubel et al. (2009). First, the outcome is a recurrent event and, hence, more complex than a survival time. Second, we use a two-stage modeling approach. The first stage features a conditional rate model intended, not for interpretation, but to track each patient's recurrent event history and projected prognosis. The linear predictor from the Stage 1 model is used at Stage 2 to caliper-match each patient receiving the experimental treatment to yet-untreated patients as opposed to the hard matching and covariate adjustment used by Schaubel et al. (2009). The final model for the recurrent event rate then consists only of the treatment effect and a distance measure which accounts for any small differences in the pre-treatment event trajectories between the experimental patient and the matched controls. The matching we propose in this report aims to create counterfactual 'ghosts' to recover the conventional therapy experience of the experimentally treated patients.

The remainder of this report is organized as follows. In Section 3.2, we introduce the notation and describe the proposed methods, including parameter estimation

and inference. A simulation study is presented in Section 3.3. In Section 3.4, the proposed methods are applied to data from the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL) to estimate the impact of LDLT on hospital admission rates relative to conventional therapy (remaining on the wait list and potentially receiving a DDLT). Some concluding remarks are offered in Section 3.5, including areas of future work and work currently in progress.

3.2 Methods

3.2.1 Notation and Data Set-up

In the development that follows, i represents subject ($i = 1, \ldots, n$). The time-dependent covariate for subject i is denoted $\boldsymbol{Z}_i^*(t)$. Subjects are generally untreated at time 0, with treatment time given by $T_i = T_i^E \wedge T_i^S$, where T_i^E represents the time of 'experimental' treatment initiation, T_i^S denotes time of 'standard' treatment initiation and $a \wedge b = \min(a, b)$. Note that at most one of T_i^E and T_i^S occurs. Treatment (be it experimental or standard) is assumed to be non-reversible, in the sense that patients are considered 'treated' from the time of treatment initiation forward. The true number of events for subject i in $[0, t]$ is defined as $N_i^*(t) = \int_0^t dN_i^*(u)$. Event and treatment times are subject to independent right censoring by C_i. The number of observed events is given by $N_i(t) = \int_0^t I(C_i > u)\, dN_i^*(u)$. Note that right censoring is administrative in this setting.

To fix ideas, we make the distinction between 'treatment' and 'therapy'. In particular:

- There are two forms of *treatment*: *standard* and *experimental*
- *Conventional therapy* involves the treatment course that all patients would need to follow if the experimental version of the treatment did not exist. Under

conventional therapy, patients begin follow-up *untreated*, with some going on to receive the *standard treatment*.

Our definition of *therapy* as being a sequence of *treatments* is not standard, but allows for convenient labeling. While untreated, a patient is actually following conventional therapy, with experimental treatment being initiated at time T_i^E. An untreated patient who later receives standard treatment is also following conventional therapy.

We define several event counting processes. The number of pre-treatment events in $(0, t]$ is given by

$$N_i^0(t) = \int_0^t I(T_i > u)\, dN_i^*(u).$$

Consider patient i at follow-up time s under two scenarios. In the first scenario, patient i receives the experimental treatment at time s; i.e., $T_i^E = s$. The pertinent post-experimental treatment event counter is defined as

$$N_i^\star(t; s) = I(T_i^E = s) \int_s^{s+t} dN_i^*(u). \tag{3.1}$$

Note that it will be our convention that $N(t; s)$ refers to the interval of length t, but starting at time s; a single time index, as in the previously-defined $N_i^0(t)$, pertains to the $(0, t]$ time interval. Correspondingly, we define an event counter representing the events that would have been experienced in the absence of experimental treatment, also beginning at time s,

$$N_i^{CT}(t; s) = I(T_i \geq s) \int_s^{s+t} I(T_i^E > u) dN_i^*(u), \tag{3.2}$$

where the superscript CT denotes 'conventional therapy'. Parallelling $N_i^\star(t; s)$, the counter $N_i^{CT}(t; s)$ tracks the patient on $(s, s + t]$. To emphasize the correspondence between (3.1) and (3.2), consider a subject eligible to receive the experimental treatment at time s, such that $I(T_i \geq s)$; note that $I(T_i \geq s) = I(T_i^E = s) \cup (T_i^E >$

$s, T_i^S > s$). Under the setting where $T_i^E = s$, the counting process (3.1) takes effect; had the experimental therapy not been an option, process (3.2) takes effect. An important point is that the patient is untreated on $(0, s)$ under either scenario to which (3.1) and (3.2) pertain.

Finally, we define a 0/1 process for being observed to receive the experimental treatment,

$$N_i^E(t) = \int_0^t I(C_i > u) dI(T_i^E \leq u). \tag{3.3}$$

As shown in Section 2.2, $N_i^E(t)$ is used to fit a model needed for inverse weighting.

3.2.2 Proposed Models and Estimation Methods

As described above, the goal of this method is to compare the post-experimental treatment recurrent event mean to the corresponding event mean under conventional therapy (CT). For a subject with $T_i^E = s$, the comparison would be on $(s, s+t]$ for $t > 0$, such that the pertinent counting processes are given in (3.1) and (3.2). We denote the mean of (3.1) by

$$\mu_i^\star(t; s) \;=\; E\left[\int_0^t N_i^\star(du; s) | T_i^E = s, \boldsymbol{H}_i(s)\right], \tag{3.4}$$

where $\boldsymbol{H}_i(s) = \{\boldsymbol{Z}_i^*(u), N_i(u), I(T_i > u), I(C_i > u);\ 0 \leq u < s\}$ represents the observed pre-treatment history for subject i on $[0, s)$. As we describe shortly, we avoid explicitly modeling (3.4). Note that (3.4) represents a partly conditional model (Pepe and Couper, 1997; Zheng and Heagerty, 2005; Gong and Schaubel, 2013) since it conditions on the history only up to time s, not $s+t$. We make this choice for the sake of interpretation, but it is not without consequences with respect to parameter estimation, as detailed shortly. This idea of conditioning on only part of the history is connected to landmark analysis (van Houwelingen, 2007; van Houwelingen and Putter, 2008; 2012).

Having defined the post-experimental recurrent event mean in (3.4), the corresponding quantity (in the absence of experimental treatment) is given by

$$\mu_i^{CT}(t;s) \;=\; E\left[\int_0^t N_i^{CT}(du;s)|\boldsymbol{H}_i(s), T_i > s, T_i^E > u\right], \tag{3.5}$$

which, like (3.4), we avoid modeling directly. The model of chief interest is then given by

$$\mu_i^{\star}(t;s) = \mu_i^{CT}(t;s)\exp\{\beta_\star\}, \tag{3.6}$$

which can equivalently be expressed in terms of a rate function by

$$\mu_i^{\star}(dt;s) = \mu_i^{CT}(dt;s)\exp\{\beta_\star\}. \tag{3.7}$$

In this model $\mu_i^{CT}(t;s)$, the experimental treatment-free mean number of events is scaled up or down by $\exp\{\beta_\star\}$ if subject i received the experimental treatment at time s. The mean number of post-experimental treatment events is then compared to the mean number of experimental treatment-free events after time s.

More general versions of model (3.7) are possible, including

$$\mu_i^{\star}(dt;s) = \mu_i^{CT}(dt;s)\exp\{\beta_\star(t;s)\}, \tag{3.8}$$

where $\beta_\star(t;s)$ is a parametric function of both time of experimental treatment initiation (given by s) and time since $T_i^E = s$, given by t. Model (3.8) is more flexible, in the sense that the effect of receiving the experimental treatment, on the event rate, can depend on the time until and time since treatment was received.

In reality, once a subject receives the experimental treatment, we can no longer observe the subject's experience receiving conventional therapy. Therefore, a patient experimentally treated at time s will be compared to similar patients who did not initiate experimental treatment at follow-up time s but were eligible to do so. Similar

to Schaubel et al. (2009), we use the concept that each experimental treatment time initiates an "experiment", in which the recipient of the experimental treatment is compared to 'similar' treatment-eligible candidates. Note that 'similar', in this context, refers to current status (i.e., at time s) and history on $[0, s)$. Eligibility for the comparison is defined as

$$e_i(s) = I(T_i^E = s) + I(T_i > s),$$

i.e., at time s, patient i either received the experimental treatment or remained untreated.

Our method of estimating $\beta_\star$ from (3.7) involves a stratified analysis, as will be formalized shortly. Each experimental treatment patient generates a stratum, which will include the index patient as well as similar treatment-eligible patients. The most relevant definition of similar, here, is with respect to treatment-eligibility at s, denoted by $e_i(s)$, and accumulated covariate and recurrent event history on $(0, s]$, represented by $\boldsymbol{H}_i(s)$. In order to quantify each subject's history, we use a prognostic score (Hansen, 2008) based on the pre-treatment event rate, modeled using a time-dependent proportional rates model,

$$d\mu_i^0(t) = E\left[dN_i^*(t)|\boldsymbol{H}_i(t), T_i > t\right] = \exp\{\boldsymbol{\alpha}_0^T \boldsymbol{Z}_i(t)\} d\mu_0(t), \tag{3.9}$$

where the covariate $\boldsymbol{Z}_i(t)$ is chosen to capture the pertinent components of the history, $E[dN_i^*(t)|\boldsymbol{H}_i(t), T_i > t] = E[dN_i^*(t)|\boldsymbol{Z}_i(t), T_i > t]$. Model (3.9) resembles the Lin et al. (2000) model, but is more accurately interpreted as an Andersen-Gill (1982) model, due to the explicit dependence on the prior event history, a property avoided by Lin et al. (2000). The regression parameter $\boldsymbol{\alpha}_0$ from (3.9) can be computed by solving the unweighted Cox (1972) score equation. Due to the dependence on internal covariates (Kalbfleisch and Prentice, 2002), elements of $\boldsymbol{\alpha}_0$ are difficult

to interpret. This is not a problem, in our case, since we do not care to interpret, let alone carry out inference on $\boldsymbol{\alpha}_0$. Note that we focus on untreated experience since, at any time s, the patient generating the stratum (by initiating the experimental treatment) and the potential matches are all necessarily untreated with either experimental or standard treatment on $[0, s)$.

The purpose of the prognostic score is to match patients that have similar pre-treatment event rates, the rationale being that previous event rate is the most important predictor of the current event rate. Unlike a propensity score, which uses the treatment event rate to match subjects with similar probabilities of being treated, the prognostic score aims to compare the effect of treatment on the event rate among subjects that were on the same trajectory with respect to their pre-treatment event rate. The use of prognostic scores in conjunction with, or as a alternative to, propensity scores has been considered in several reports (e.g., Rubin and Thomas, 2000; Stuart, Lee and Leacy, 2013; Leacy and Stuart, 2014; Li, Schaubel, and He, 2014) and will be addressed further in Section 5. Once the prognostic scores have been estimated, caliper matching is used to assign untreated control subjects to a subject receiving the experimental treatment at time s. Caliper matching requires that the prognostic scores of matched subjects be within a certain radius of the prognostic score of the index subject. Appropriate selection of the caliper involves balancing the need for homogeneity within-stratum with the need to have an adequate number of matches for each index subject. We propose that prior to data analysis the caliper size be chosen such that the median number of matches is between 10 and 20. It is important to verify that some level of balance in the event rate at time s is achieved with the chosen caliper. Methods for checking balance in propensity matching have been discussed by Hansen (2008) and Harder, Stuart, and Anthony (2010). This will

be explored further in the application to living donor liver transplant described in Section 4. The discrepancy between prognostic scores for experimental subject j and control subject i can be quantified through the subject-pair specific rate ratio,

$$\psi_{i,j}(s) = \frac{d\mu_i^0(s)}{d\mu_j^0(s)} = \exp\{\boldsymbol{\alpha}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)]\}.$$

Subject i is considered to have a path on $[0, s)$ sufficiently similar to subject j if $|\log \psi_{ij}(s)| \leq \epsilon$, where $\epsilon > 0$ is a pre-determined constant.

Combining the eligibility indicators and prognostic scores, patient i is included in the stratum generated by patient j if $m_{ij}(s) = 1$, where

$$m_{ij}(s) \;\; = \;\; e_i(s)I(T_i > s)e_j(s)I(T_j^E = s)I(|\log \widehat{\psi}_{ij}(s)| \leq \epsilon),$$

with $\widehat{\psi}_{ij}(s) = \exp\{\widehat{\boldsymbol{\alpha}}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)]\}$. In order to account for the residual difference between patients i and j, we propose to adjust for $\log \widehat{\psi}_{ij}(s)$ in the final model. Incorporating the eligibility indicator and the prognostic score distance, the final fitted model for the event mean for stratum j is then

$$\mu_{ij}^\star(t; s) = m_{ij}(s)\mu_i^{CT}(t; s) \exp\{\beta_\star I(T_i^E = s) + \beta_\psi \log \widehat{\psi}_{ij}(s)\}. \tag{3.10}$$

To clarify the representation in model (3.10), j is the stratum (generated by patient j through $T_j^E = s$) and i is the patient within stratum. The model governs the experimental patient through the indicator $I(T_i^E = s)$, which equals 1 if $i = j$. The vector of parameters to be estimated and the corresponding covariates are given by

$$\boldsymbol{\beta}_{\star\psi} = \begin{bmatrix} \beta_\star \\ \beta_\psi \end{bmatrix} \quad \boldsymbol{Z}_i^\star(s) = \begin{bmatrix} I(T_i^E = s) \\ \log \widehat{\psi}_{ij}(s) \end{bmatrix}, \tag{3.11}$$

such that model (3.10) can be re-written as $\mu_i^\star(t; s) = m_{ij}(s)\mu_i^{CT}(t; s) \exp\{\beta_{\star\psi}^T \boldsymbol{Z}_i^\star(s)\}$.

Subjects that are matched to the experimental subject enter the experiment without receiving any treatment, but could subsequently receive either the experimental

treatment or standard treatment. If a patient receives the standard treatment after time s, follow-up in the experiment continues since the goal is to compare experimental treatment to conventional therapy (i.e., beginning untreated and potentially later receiving standard treatment). However, if a matched subject receives the experimental treatment after time s they are censored from all experiments in which they serve as controls and begin their own experiment as the index subject. This generally results in dependent censoring since, although $dI(T_i^E \leq s+t)$ can be considered random given $\boldsymbol{H}_i(s+t)$, the model for $\mu_{ij}^{\star}(t;s)$ from (3.10) only conditions on $\boldsymbol{H}_i(s)$, the pre-treatment history up to time s. Since $\{\boldsymbol{H}_i(s+t);\ t>0\}$ can affect $dI(T_i^E \leq s+t)$ and, by definition, affects $N_i(t;s)$, the censoring of matched controls via experimental treatment initiation constitutes dependent censoring.

We apply a variant of Inverse Probability of Censoring Weighting (IPCW; Robins and Rotnitzky, 1992; Miloslavsky et al., 2004) to account for this dependent censoring by modeling the censoring mechanism (experimental treatment) using the following model,

$$\lambda_i^E(t) = \lambda_0^E(t)\exp[\boldsymbol{\beta}_E^T \boldsymbol{Z}_i(t)],$$

fitted through standard partial likelihood (Cox, 1975).

3.2.3 Parameter Estimation

We now describe the estimation of the parameter of interest, $\beta_{\star}$, through the estimation of $\boldsymbol{\beta}_{\star\psi}$. To begin, we define the pertinent risk set indicator for stratum j,

$$Y_{ij}(t;s) \;=\; m_{ij}(s)I(C_i > s+t)\{I(T_i^E = s) + I(T_i^E > s+t)\}.$$

For now, suppose that, subsequent to being matched to an index experimental treatment patient j at time s, matched controls were randomly selected to receive exper-

imental treatment given $\boldsymbol{H}_i(s)$. In this restrictive case, the following process,

$$m_{ij}(s)\int_s^{s+t} Y_{ij}(u;s)\{N_i(du;s)-\mu_i(du;s)\} \tag{3.12}$$

would have mean zero.

As mentioned above, we will use IPCW to account for censoring of matched patients who receive the experimental treatment in each experiment. IPCW involves weighting using the inverse probability of being censored or, in this case, receiving the experimental treatment. Since the experiment begins at time s, the weight only accounts for the probability of receiving the experimental treatment during $[s, s+t)$. Correspondingly, we define the quantity

$$G_i(t;s)=\frac{G_i(s+t)}{G_i(s)}=\exp\left\{-\int_s^{s+t}\lambda_i^E(u)\,du\right\},$$

where

$$G_i(t)=\exp\left\{-\int_0^{t}\lambda_i^E(u)\,du\right\}.$$

The weight function can then be defined as

$$W_{ij}(t;s)=Y_{ij}(t;s)G_i(t;s)^{-I(T_i^E>s)},$$

such that the index experimental treatment patient is appropriately self-weighting; censoring after T_i^E can only occur independently via C_i.

Applying the weight function to (3.12) produces the zero-mean process,

$$m_{ij}(s)\int_0^{\tau-s} W_{ij}(u;s)M_{ij}(du;s), \tag{3.13}$$

where we define $M_{ij}(du;s)=Y_{ij}(u;s)\{N_i(du;s)-\mu_{ij}(du;s)\}$, with τ chosen to satisfy $P(C_i\geq\tau)>0$ and often set to $\max\{C_1,\ldots,C_n\}$. Aggregating across subjects for the experiment occurring at time s produces the set of zero mean processes,

$$\textstyle\sum_{i=1}^n m_{ij}(s)\int_0^t W_{ij}(u;s)M_{ij}(du;s) \tag{3.14}$$

and

$$\sum_{i=1}^{n} m_{ij}(s) \int_0^t \boldsymbol{Z}_i^{\star}(s) W_{ij}(u;s) M_{ij}(du;s). \tag{3.15}$$

We reorganize this system to solve implicitly for the baseline mean, $\mu_0^{CT}(u;s)$ in (3.14), then substitute into (3.15). Then, aggregating across all experiments yields the final estimating function for $\boldsymbol{\beta}_{\star\psi}$,

(3.16)

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{n} \sum_{i=1}^{n} \int_0^{\tau} m_{ij}(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^{\star}(s) - \overline{\boldsymbol{Z}}_{\star}(u;s)\} \widehat{W}_{ij}(u;s) N_i(du;s) dN_j^E(s),$$

where $\widehat{W}_{ij}(t;s) = Y_{ij}(t;s) \exp\{[\widehat{\Lambda}_i^E(s+t) - \widehat{\Lambda}_i^E(s)] I(T_i^E > s)\}$, and with

$$\overline{\boldsymbol{Z}}_{\star}(u;s) = \frac{\sum_{\ell=1}^{n} m_{\ell j}(s) \widehat{W}_{\ell}(u;s) \boldsymbol{Z}_{\ell}^{\star}(s) \exp\{\boldsymbol{\beta}_{\star\psi}^T \boldsymbol{Z}_{\ell}^{\star}(s)\}}{\sum_{\ell=1}^{n} m_{\ell j}(s) \widehat{W}_{\ell}(u;s) \exp\{\boldsymbol{\beta}_{\star\psi}^T \boldsymbol{Z}_{\ell}^{\star}(s)\}}.$$

Since $U(\boldsymbol{\beta})$ from (3.16) behaves asymptotically like a zero-mean estimating function, the solution to $U(\boldsymbol{\beta}) = \mathbf{0}$, denoted by $\widehat{\boldsymbol{\beta}}_{\star\psi}$, should yield a consistent estimator of $\boldsymbol{\beta}_{\star\psi}$.

3.2.4 Variance Estimation, Asymptotic Properties, and Inference

Due to the complex nature of the data structure in this setting we investigated the use of both the robust sandwich estimator and the bootstrap estimator for variance estimation and propose the use of the bootstrap estimator (Efron, 1979). In moderate sized samples approximately 100 bootstrap samples should be used. If the normal approximation is reasonable, 100 bootstrap samples should be adequate for hypothesis testing and interval estimation. If the normal approximation is in question, then upwards of 10,000 may be necessary to estimate the $2.5th$ and $97.5th$ percentiles of the distribution for hypothesis testing. A straightforward pre-test method for determining the number of necessary bootstrap samples is described by Davidson and MacKinnon (2000).

3.3 Simulation Study

Simulation studies were conducted to assess the performance of the method described above. For each subject who received the experimental treatment, the subsequent (counterfactual) experience (i.e., in the absence of experimental treatment receipt) was also generated. This counterfactual experience was not used in the modeling, but was used to determine target values of $\beta_\star$, which are difficult to pre-specify.

First, for each subject i, a frailty variate was generated from a Gamma (1, 0.5) distribution, bounded at 2, the 90^{th} percentile. Next, two independent Bernoulli (0.5) variates, Z_{i1} and Z_{i2} were created. Pre-treatment recurrent event experience was then generated though a frailty model with rate parameter $\gamma_i d\mu_0 \exp\{\alpha_1 Z_{i1} + \alpha_2 Z_{i2}\}$. The time of standard treatment receipt, T_i^S, was then generated to follow the hazard $\lambda_0^S \exp\{\beta_{S1} Z_{i1} + \beta_{S2} Z_{i2} + \beta_{S3} \log(N_i(t^-) + 1)\}$. The recurrent event times post-standard treatment were generated from rate parameter $\gamma_i d\mu_0^S \exp\{\delta_1 Z_{i1} + \delta_2 Z_{i2}\}$. Thus, as mentioned in the first paragraph of this subsection, we initially generated the conventional therapy (pre-treatment plus post-standard treatment) recurrent event experience for all subjects. We then generated T_i^E from hazard function, $\lambda_0^E \exp\{\beta_{E1} Z_{i1} + \beta_{E2} Z_{i2} + \beta_{E3} \log(N_i(t^-) + 1)\}$. The experimental treatment was only considered to actually be received if $T_i^E < T_i^S$ and, in such cases, the post-experimental recurrent event experience then followed rate model $\gamma_i d\mu_0^E \exp\{\phi_1 Z_{i1} + \phi_2 Z_{i2}\}$.

For each data set generated, we modeled the hazard for T_i^E using the model, $\lambda_i^E(t) = \lambda_0^E(t) \exp\{\beta_{E1} Z_{i1} + \beta_{E2} Z_{i2} + \beta_{E3} \log(N_i(t^-) + 1)\}$, in order to obtain the estimated weights for IPCW. We then modeled the pre-treatment event rate using the model $d\mu_i^0(t) = \exp\{\alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \alpha_3 N_i^0(t^-)\} d\mu_0^0(t)$, in order to obtain prognostic

scores. Subjects were matched if $|\log \widehat{\psi}_{ij}| \leq 0.1$, resulting in approximately 8 matches per stratum.

Parameters used in the simulation studies are as follows. For the pre-treatment event rate, we set $d\mu_0 = 5$, $\alpha_1 = -0.3$, and $\alpha_2 = 0.7$. For the post-standard treatment event rate, $d\mu_0^S = 1$, $\delta_1 = -0.6$, and $\delta_2 = 0.2$. For the experimental treatment hazard, $\lambda_0^E = 0.5$, $\beta_{E1} = -0.7$, $\beta_{E2} = -0.5$, and $\beta_{E3} = 0.3$. For the standard treatment hazard, $\lambda_0^S = 0.5$, $\beta_{S1} = -0.8$, $\beta_{S2} = -0.7$, and $\beta_{S3} = 0.1$. Finally, for the post-experimental treatment rate, $d\mu_0^E$ was given values of 1, 1.5, 2.5, and 3, $\phi_1 = -0.7$, and $\phi_2 = 0.7$. This resulted in values of $\beta_\star$ of -0.901, -0.457, 0.082, and 0.261. In the simulated data 63% of the sample received experimental treatment, 36% received standard treatment, and 1% received no treatment. The mean number of events was 9.46 (sd = 0.72) in the no treatment state, 18.92 (sd = 1.49) in the post-conventional treatment state, and ranged from 16.41 to 24.96 in the post-experimental treatment state. These numbers describe the simulated data but are not directly related to the different treatment effects tested, due to the timing of experimental treatment. A total of $n = 200$ subjects were simulated 250 times, with 50 bootstrap samples used for variance estimation.

The results are shown in Table 3.1 for the estimate of the parameter for the experimental treatment effect. Direct matching and prognostic matching gave similar bias (range 0.022-0.029 and 0.015-0.030, respectively). However, prognostic score matching resulted in empirical standard errors that were 9-10% lower. Adjusting for $\log \psi_{ij}$ further reduces bias in the estimate of the experimental treatment effect (range 0.004-0.017 across the 4 scenarios). Figure 3.2 shows histograms of the parameter estimates centered on their target value with normal density curves. All estimates appear to follow a normal density centered at zero, supporting the theory that $\hat{\beta}_\star$ is

Table 3.1: Results of simulation study using proposed method to estimate the effect of a time-dependent treatment on the recurrent event rate using IPCW

Scenario	$\beta_\star$	Matching	Bias	ESE	ASE	CP
1	-0.901	Prognostic	0.030	0.130	0.158	0.980
2	-0.457		0.029	0.127	0.151	0.972
3	0.082		0.024	0.121	0.142	0.972
4	0.261		0.015	0.119	0.140	0.976
1	-0.901	Prognostic	0.017	0.129	0.158	0.976
2	-0.457	w/adjustment	0.017	0.125	0.153	0.980
3	0.082		0.012	0.120	0.144	0.984
4	0.261		0.004	0.118	0.141	0.976
1	-0.901	# of Events	0.022	0.142	0.188	0.988
2	-0.457	with Covariate	0.027	0.140	0.181	0.976
3	0.082	Adjustment	0.029	0.135	0.172	0.976
4	0.261		0.024	0.133	0.170	0.980

ESE=empirical standard error; ASE=asymptotic standard error;
CP=coverage probability

asymptotically normal with mean $\beta_\star$.

3.4 Application to Living Donor Liver Transplant

The Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL) is a multi-center NIH-sponsored study investigating the post-transplant morbidity and mortality of LDLT recipients and their donors. Potential living donor transplant recipients who had a donor evaluated between January 1, 1998 and August 31, 2009 were recruited at each of the 9 A2ALL centers beginning in the third quarter of 2004 and followed through August 31, 2010. These potential recipients may have ultimately received an LDLT, a DDLT, or neither. Clinical data, including laboratory

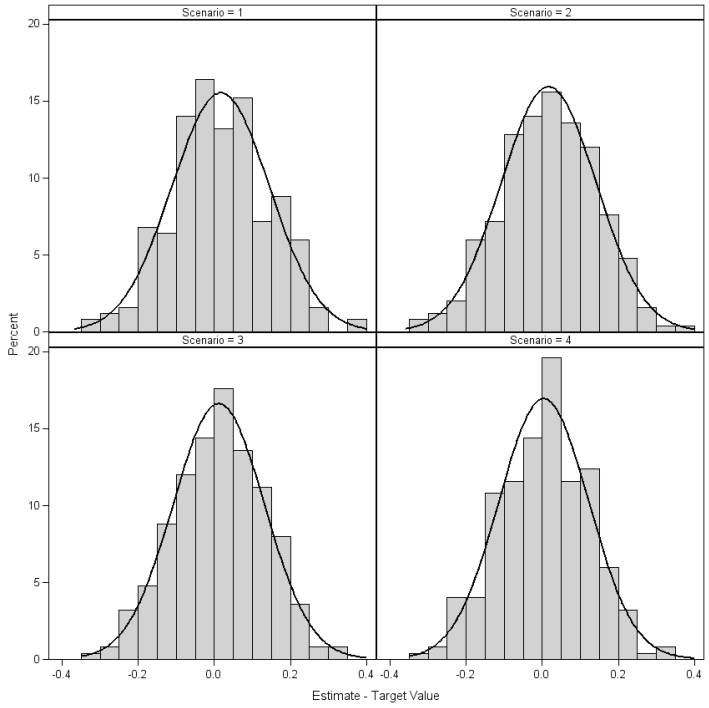

Figure 3.2: Histogram of parameter estimates from proposed method to estimate the effect of a time-dependent treatment on the recurrent event rate using IPCW with normal density

information, hospitalizations, and complications were collected based on a common protocol, and supplemented with data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States; these data are submitted by the members of OPTN and have been described elsewhere. The Health Resources and Services Administration (US Department of Health and Human Services) provides oversight for the activities of the OPTN and SRTR contractors. Data collection for pre- and post-transplant hospitalizations included date of admission and discharge, reason for hospitalization, and discharge diagnosis. Subjects could enroll either before or after transplant, and information prior to enrollment was collected via chart review.

LDLT is a technically complicated surgery. Since we essentially wish to eval-

uate whether LDLT results in a higher post-transplant hospitalization rate than conventional therapy, our objective is consistent with a non-inferiority test. The comparison of interest is between LDLT recipients and their counterparts who remain on the waiting list and potentially receive a DDLT. The time origin was date of first donor evaluation. This analysis includes 1467 liver transplant candidates, of which 714 went on to receive an LDLT, 455 received a DDLT, and 298 remained untransplanted.

The pre-treatment model for the event rate was adjusted for age, race, recipient diagnosis, Model for End Stage Liver Disease (MELD), event history (previous number of hospital admissions), and the following characteristics measured at the time of donor evaluation: hospitalization status, ascites, mechanical ventilation, dialysis status, and transplant center. Subjects with $|\log \psi_{ij}(s)| \leq 0.025$ who were untransplanted at the time of LDLT were matched. This resulted in a median of 20 matches (range 1-71, IQR 11-34). Of the 714 LDLTs, 692 (97%) were matched to at least one "control" subject. The mean $\log \widehat{\psi}_{ij}(s)$ was 0.000033 (SD = 0.014), indicating that the index patients and their matched counterparts were very similar at the time of matching, with respect to prognosis.

Both traditional time-dependent models and models using the proposed method were fitted. The time-dependent model was adjusted for age, ethnicity, diagnosis, MELD, transplant center, diabetes and ascites at the time of donor evaluation, and event history defined as above. In the model using the proposed method 100 bootstrap samples were generated to estimate the variance of parameter estimates. Table 3.2 shows the results of two models, one using only a main effect for LDLT and another using an interaction between LDLT and MELD score, each fit for both methods mentioned above (a total of four models). In the time-dependent version

Table 3.2: A2ALL results: Effect of LDLT on hospital admission rate

		Proposed Model			Time-dependent Model	
Model		Rate Ratio	95% CI	p-value	Rate Ratio	p-value
1	LDLT	0.986	0.857-1.134	0.8425	1.217	<0.0001
2	LDLT MELD 6-11	1.152	0.849-1.562	0.3628	1.606	<0.0001
	LDLT MELD 12-15	0.962	0.752-1.230	0.7589	1.351	<0.0001
	LDLT MELD 16-19	0.915	0.696-1.202	0.5226	1.009	0.8906
	LDLT MELD 20-29	0.918	0.702-1.202	0.5344	1.040	0.5483
	LDLT MELD 30-40	1.150	0.136-9.702	0.8978	1.143	0.3148

CI=confidence interval

of Model 1, LDLT recipients have a significantly higher rate of hospitalizations compared to the combination of waiting list patients and DDLT recipients (rate ratio [RR] = 1.22, p <0.0001), consistent with Merion et al, 2010. However, using the proposed methods, there is no difference in hospitalization rates between subjects who receive LDLT and those receiving conventional therapy (RR = 0.986, p = 0.8245). We also fit models with interactions between LDLT and MELD (Model 2). In the time-dependent model lower MELD subjects receiving LDLT had significantly higher hospitalization rates, with a rate ratio of 1.606 for subjects with MELD 6-11 and 1.351 for subjects with MELD 12-15 (p <0.0001 in both cases). Subjects receiving LDLT in higher MELD categories did not have significantly different hospitalization rates relative to subjects waiting for DDLT. Similar to the main effect model, the higher hospitalization rates in low MELD subjects receiving LDLT are not seen when the proposed method is used. Rate ratios range from 0.915-1.152 with p-values ranging from 0.3628-0.8978.

One potential reason for the differing results between the two methods is that in

the proposed model patients are matched on pre-transplant event rate on the day of the index transplant, so the method compares the LDLT patient to patients with a similar pre-transplant event history. By contrast, in the traditional time-dependent analysis, pre- and post-transplant event rates are averaged over in both the LDLT and non-LDLT groups; patients in the comparison group may have differing event rates at the time a given subject receives an LDLT. To explore this further we investigated interactions with time and LDLT using the proposed method to explore potential differences in the rate ratios over time. Figure 3.3 shows that there is a sharp increase in the hospitalization rate immediately after LDLT. However, the RR declines quickly, to the point where it is significantly below 1 for a period, then continues to rise for the remainder of follow-up. The estimated RR actually rises above 1 (although non-significantly) towards the end of the follow-up period. The pronounced and steady rise in the RR (after the drop to significantly below 1) is due to the nature of the comparison groups. Specifically, as time progresses, a greater percentage of subjects in the conventional therapy groups are DDLT (as opposed to pre-transplant) patients. If the comparison were between experimental treatment and pre-transplant, then it is highly likely that the RR may have stabilized.

3.5 Discussion

Despite the wide array of methods available for modeling recurrent event data, there are relatively few such methods devoted specifically to estimating the effect of treatment. A limited number of methods have been developed to accommodate time-dependent covariates and informative censoring (e.g., Huang, Qin and Wang, 2010; Zhao et al, 2012). However, such methods apply under different data structures and assumptions. Our goal in this report was to develop methods to estimate treatment

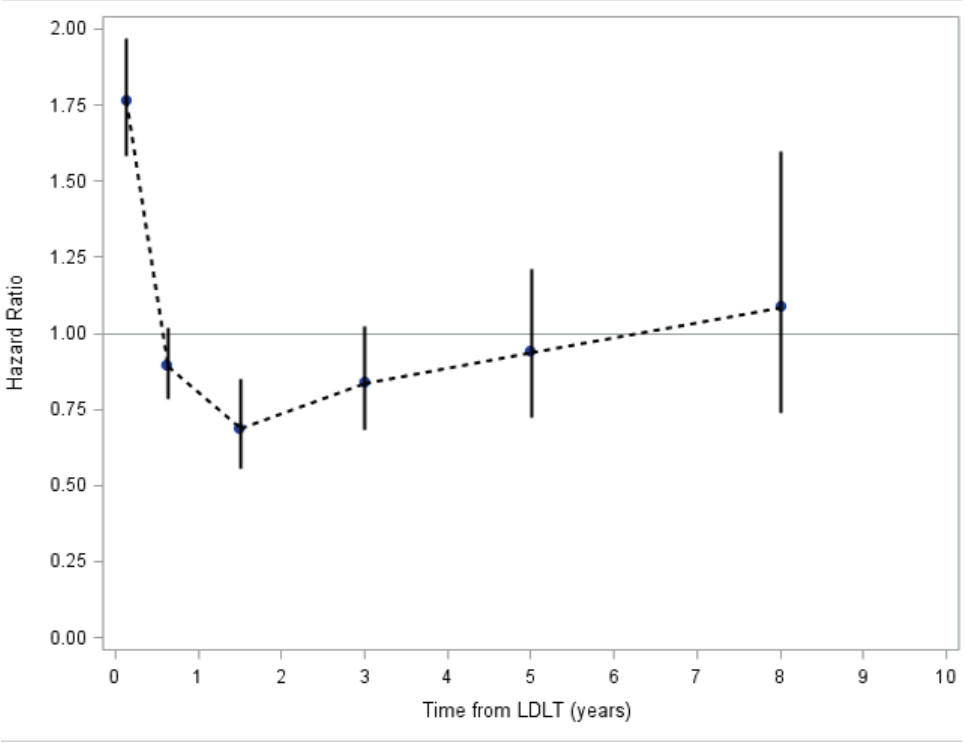

Figure 3.3: Time-dependent rate ratio (RR) for LDLT and 95% confidence intervals

effects in a fairly complex setting: treatment is time-dependent, different forms of treatment are available, and the event rate in the absence of the treatment of interest is dependently censored.

One key aspect to this type of analysis is appropriate matching of control subjects to experimental subjects. In the method proposed prognostic scores were used to match index patients to control patients within a certain radius. It should be noted that this is slightly different than the usual prognostic score setting because all patients are observed in the untreated state for some period of time, not only the control patients, and thus all patients contribute to the event rate model used to generate the prognostic scores. An alternative would be to use k-nearest neighbor matching. However, this method appears to be less appropriate for two reasons. First, while some experimental patients may have many appropriate matches, requiring a pre-specified number of matches per experimental patient could result in large

heterogeneity in some strata. Second, Abadie and Imbens (2008) have demonstrated that the bootstrap estimator is invalid in the setting of nearest neighbor matching. It could also be argued that propensity matching would be an appropriate matching method for this data structure. In that scenario the goal would be to simulate a randomized trial by matching patients with similar probabilities of receiving experimental treatment. A time-dependent propensity score similar to that proposed by Lu (2005), may be appropriate.

A limitation to the method as described is that it does not take into account terminal events such as death. Many methods for recurrent/terminal event data have been proposed (e.g., Ghosh and Lin, 2002; Liu, Wolfe and Huang, 2004; Huang and Wang, 2004; Ye, Kalbfleisch and Schaubel, 2007; Kim et al., 2012, Kalbfleisch et al., 2013) and could be applied to the method described above. Extension of the proposed methods to incorporate terminal events is described in the following chapter.

CHAPTER IV

Estimating time-dependent treatment effects for correlated recurrent and terminal events

4.1 Introduction

In most clinical settings a recurrent event process can be stopped by a terminal event such as death. Events such as hospitalizations, infections, or myocardial infarctions cease to occur once a patient dies. If the recurrent event process is independent of the terminal event then parameter estimates measuring the effects of covariates of interest on the recurrent event rate remain unbiased. The assumption of independence is a strong one and is not often justifiable in practice. It is more often the case that an increase in the recurrent event rate increases the probability of the terminal event occurring, which means that methods which treat the terminating event as independent censoring generally lead to biased estimation.

Under the framework we consider in this chapter, recurrent events are stopped, not censored by, the terminal event. This is in contrast to some proposed methods which view the recurrent event process as a latent process that continues, unobserved, after the terminal event occurs. Methods using this framework have been proposed by Ghosh and Lin (2003) and Miloslavsky (2004), and adapt methodology for dependent censoring to this setting. If the terminal event is something other than death, such as study termination for medical reasons, then this framework and associated estimation

methods are applicable and have reasonable interpretation. However, if death is the terminating event, the recurrent event rate while the patient is alive is often of more clinical interest, therefore we will use the conditional framework for the remainder of the chapter.

Under the assumption that the terminating event stops future occurrences of the recurrent event, several methods for modeling recurrent events have been proposed. Ghosh and Lin (2000, 2002), Strawderman (2000), and Schaubel and Zhang (2010) proposed models of the mean number of recurrent events, i.e., averaging over the terminal event and pre-terminal event experience. Another method models the recurrent event rate conditional on "surviving" free of the terminal event (Liu et al, 2004, Ye et al, 2007). Cook and Lawless (1997) proposed several mean and rate functions for recurrent events, conditional on the terminating event time. A detailed report of recurrent event analysis methods generally is available from the same authors (Cook and Lawless, 2007).

A popular approach under this framework has been to condition on a subject-specific random effect associated with both the recurrent event process and the terminal event, such as methods proposed by Lancaster and Intrator (1998), Wang, et al. (2001), and Huang and Wang (2004). These frailty models have become a common method of evaluating recurrent event rates in the presence of terminal events, and methods that explicitly model the latent frailty have been proposed by Liu et al. (2004), Ye et al. (2007), and others. Multiple estimation methods for these frailty models have been proposed. Given that the latent frailty is unobserved, Liu et al. (2004) and Huang and Liu (2007) proposed to use the missing data problem approach and use Expectation-Maximization (EM). Other estimation proposals include penalized partial likelihood (Rondeau et al., 2007)), non-parametric maximum

likelihood (Zeng and Lin, 2009), and estimating equations (Kalbfleisch et al., 2013). Many of these methods can estimate parameters for time-varying covariates, and methods for estimating time-varying coefficients have also been proposed by Yu et al. (2014).

Current methods for evaluating treatment effects in the setting of correlated recurrent and terminal events are limited to baseline treatments (Chen and Cook, 2004, Pan and Schaubel, 2009, Schaubel and Zhang, 2010). As previously described, estimating treatment effects through traditional time-dependent methods yields inappropriate interpretations. Therefore, we propose the use of sequential stratification methods (Schaubel et al, 2009) to estimate the effect of treatment assigned during follow-up on both recurrent and terminal events.

In this chapter the methods of Chapter III are extended to the setting in which recurrent events are stopped by correlated terminating events. As in Chapter II there is only one treatment. However, the treatment is no longer assumed to be relatively rare. The objective is to estimate the treatment effect on both the recurrent event rate and the terminal event hazard in a way that respects the timing of treatment and yields an appropriate interpretation.

The remainder of the chapter proceeds as follows. In Section 4.2 the notation and proposed models are introduced along with the parameter estimation and asymptotic properties. Section 4.3 provides results from simulation studies, and Section 4.4 describes an application to the A2ALL study. Finally, some concluding remarks are offered in Section 4.5.

4.2 Methods

4.2.1 Notation

Let D_i and C_i denote terminal event and censoring times, respectively, where i represents subject ($i = 1, \ldots, n$) with time-dependent covariate $\boldsymbol{Z}_i(t)$, which does not include any parametric form of the event history. The observed time is given by $X_i = D_i \wedge C_i$ where $a \wedge b = \min(a, b)$. Censoring in this setting is administrative and depends, at most, on $\boldsymbol{Z}_i(t)$. The counting process notation for the terminal event time is given by $N_i^{D*}(t) = \int_0^t dN_i^{D*}(u)$. The terminal event is observed if $D_i < C_i$, i.e., $N_i^D(t) = \int_0^t I(C_i > u)\, dN_i^{D*}(u)$. Subjects begin follow-up untreated, and after some time some subjects may be treated at time T_i. Some subjects will experience the terminal event prior to treatment ($D_i < T_i$), and some may never experience treatment at all ($T_i = \infty$). We also define a subject-specific random effect, $\gamma_i \sim N(0, \theta)$, representing underlying processes that effects both the terminal event hazard and recurrent event intensity.

We use counting process notation to set up the recurrent events as well. For all subjects, the true number of recurrent events is given by $N_i^*(t) = \int_0^t dN_i^*(u)$. The number of observed events is given by $N_i(t) = \int_0^t I(C_i > u)\, dN_i^*(u)$, where $dN_i^*(t) = I(D_i > t)\, dN_i^*(t)$, as censoring and the terminal event preclude further recurrent events. In the following, it will be necessary to distinguish between pre- and post-treatment recurrent and terminal events, therefore we will use

$$N_i^{D0}(t) = \int_0^t I(T_i > u)\, dN_i^{D*}(u) \tag{4.1}$$

to represent the pre-treatment terminal event for subject i and

$$N_i^0(t) = \int_0^t I(T_i > u)\, dN_i^*(u) \tag{4.2}$$

to represent pre-treatment recurrent events for subject i. After receiving treatment at $T_i = s$, the respective quantities are

$$N_i^{D1}(t;s) = I(T_i = s) \int_s^{s+t} dN_i^{D*}(u) \tag{4.3}$$

and

$$N_i^{1}(t;s) = I(T_i = s) \int_s^{s+t} dN_i^{*}(u) \tag{4.4}$$

for terminal and recurrent events, respectively. Note here that we employ the notation $N_i^{D1}(t;s)$ and $N_i^{1}(t;s)$ to denote a time interval beginning at s with length t. $N_i^{D0}(t)$ and $N_i^{0}(t)$ are similarly defined with $s = 0$. Since the event processes $N_i^{D1}(t;T_i)$ and $N_i^{1}(t;T_i)$ can only begin at treatment initiation, information in $[0, T_i)$ is not useful for assessment of the impact of treatment on subject i.

Lastly, we define a counting process for receiving treatment, $N_i^T(t) = \int_0^t I(X_i > u)\, dN_i^T(u)$, which will be used for inverse weighting as described below.

4.2.2 Proposed Models

As previously noted, the comparison of interest is between the treatment and treatment-free experience for a given subject. If this were observable in practice, the quantities of interest would be given by the terminal event hazard

$$d\Lambda_i^1(t;s) = E\left[\, dN_i^{D1}(t;s) | \boldsymbol{Z}_i(s), \gamma_i \right] \tag{4.5}$$

and instantaneous rate function

$$dR_i^1(t;s) = E\left[\, dN_i^{1}(t;s) | \boldsymbol{Z}_i(s), \gamma_i \right], \tag{4.6}$$

where, as noted above, γ_i is a zero mean normal variate with variance θ and $\boldsymbol{Z}_i(s)$ is free of functions of the event history. The associated counterfactual (i.e., treatment

free) experience can be defined as

$$d\Lambda_i^0(t;s) = E\left[\, dN_i^{D0}(t;s) | \boldsymbol{Z}_i(s), \gamma_i \right] \tag{4.7}$$

and

$$dR_i^0(t;s) = E\left[\, dN_i^1(t;s) | \boldsymbol{Z}_i(s), \gamma_i \right] \tag{4.8}$$

for the hazard and rate functions, respectively. Note that we do not model these quantities directly, but instead define

$$d\Lambda_i^1(t;s) = d\Lambda_i^0(t;s) \exp\{\alpha_\star\} \tag{4.9}$$

and

$$dR_i^1(t;s) = dR_i^0(t;s) \exp\{\beta_\star\}. \tag{4.10}$$

Here, $d\Lambda_i^0(t;s)$ and $R_i^0(t;s)$ are scaled up or down by $\exp\{\alpha_\star\}$ and $\exp\{\beta_\star\}$, respectively, as a result of treatment at time s.

If subject i is treated at time s we will never observe their treatment-free experience on $[s, s+t)$, therefore we propose to use prognostic matching similar to those described in Chapters II and III to generate a set of "similar" patients untreated at time s. Unlike previous chapters, however, we now have both recurrent and terminal events, and must account for each, as well as their correlation, in the prognostic model. We propose to do so using frailty models similar to those proposed by Liu et al (2004). In particular, we assume the following frailty models for the pre-treatment hazard and intensity functions for terminal and recurrent events, respectively,

$$d\Lambda_i^0(t) = d\Lambda_0(t) \exp\{\boldsymbol{\delta}_0^T \boldsymbol{Z}_i(t) + \gamma_i\} \tag{4.11}$$

and

$$dR_i^0(t) = dR_0(t) \exp\{\boldsymbol{\xi}_0^T \boldsymbol{Z}_i(t) + \gamma_i\}, \tag{4.12}$$

where we assume $\gamma_i \sim N(0, \theta)$ as noted above. In order to make the method feasible in larger datasets, we use a two-stage procedure to fit models (4.11) and (4.12) in order to speed up computation time. In the first stage we estimate $\widehat{\boldsymbol{\xi}}$ and $(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_n)$ using the recurrent event data only by fitting the following piecewise constant rate model,

$$r_{ik}^0(t) = r_{0k}^0 \exp\{\boldsymbol{\xi}_0^T \boldsymbol{Z}_i(t_k) + \gamma_i\}, \tag{4.13}$$

where the baseline rate parameters, $r_{01}^0, \ldots, r_{0K}^0$, are intended to closely approximate dR_0 from (4.12) and $\boldsymbol{Z}_i(t_k)$ is the potentially time-dependent covariate (excluding any functions of the event history) taking the value associated with the beginning of the interval $(t_k, t_{k+1}]$. This is essentially a piecewise Poisson regression model, fitted by restricted pseudo-likelihood methods. This type of linearization method uses a first order Taylor series with expansion around estimates of the best linear unbiased predictors (BLUPS) of the subject-specific random effects. At the second stage, model (4.11) is fitted using a Cox proportional hazards model with the $\widehat{\gamma}_i$ from the first stage as an offset.

Following the model fitting procedure described above, each subject will then have two prognostic scores, one for their pre-treatment terminal event hazard and one for their pre-treatment recurrent event intensity. Caliper matching will be used to match treated patients to as-yet-untreated patients based on their pre-treatment trajectories. The distance between the treated patient i and a matched patient j will be calculated as

$$\psi_{i,j}^D(s) = \frac{\lambda_i^0(s)}{\lambda_j^0(s)} = \exp\{\boldsymbol{\delta}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)] + \hat{\gamma}_i - \hat{\gamma}_j\}$$

for the terminal event hazard and

$$\psi_{i,j}^R(s) = \frac{dR_i^0(s)}{dR_j^0(s)} = \exp\{\boldsymbol{\xi}_0^T[\boldsymbol{Z}_i(s) - \boldsymbol{Z}_j(s)] + \hat{\gamma}_i - \hat{\gamma}_j\}$$

for the recurrent event intensity.

As described in more detail below, while the prognostic model used a joint frailty set up, the final models will estimate the treatment effect on the terminal event hazard and the recurrent event intensity separately. As a result matching could proceed in several ways. Successful matching could require that both $|\log \psi_{ij}^D(s)| \leq \epsilon^D$ and $|\log \psi_{ij}^R(s)| \leq \epsilon^R$ for both models, where ϵ^D and ϵ^R are predetermined constants not necessarily equal to each other. This would imply that the matched sets are the same in both models. Alternatively, we could match separately for terminal and recurrent event models, i.e. requiring $|\log \psi_{ij}^D(s)| \leq \epsilon^D$ for the former and $|\log \psi_{ij}^R(s)| \leq \epsilon^R$ for the latter, resulting in potentially different matched sets for the two models. Some combination of these two matching scenarios could also be applied, and different combinations will be explored in simulation.

We can now create an indicator function that determines whether patient j is matched to patient i, a subject treated at time s. If we use the first prognostic score matching method described above which results in the same matched sets being used for both models, then $m_{ij}(s) = 1$ indicates a successful match where

$$m_{ij}(s) = I(T_j = s)I(T_i > s)I(|\log \psi_{ij}^D(s)| \leq \epsilon^D)I(|\log \psi_{ij}^R(s)| \leq \epsilon^R).$$

If we want to match separately for each model, then for the terminal and recurrent event models

$$m_{ij}^D(s) = I(T_j = s)I(T_i > s)I(|\log \psi_{ij}^D(s)| \leq \epsilon^D) = 1$$

and

$$m_{ij}^R(s) = I(T_j = s)I(T_i > s)I(|\log \psi_{ij}^R(s)| \leq \epsilon^R) = 1$$

would indicate successful matches in the respective models. As in Chapters II and III we will again adjust for the appropriate distance measure in each model to account

for residual differences between treated patients and matched controls. We also adjust for the individual frailties estimated from the pre-treatment recurrent event experience. Our final fitted models for the terminal event hazard and recurrent event mean can then be written as

$$(4.14)\ d\Lambda^1_{ij}(t;s) = m^D_{ij}(s) d\Lambda^0_i(t;s) \exp\{\alpha_\star I(T_i = s) + \alpha_\psi \log \psi^D_{ij}(s) + \alpha_\gamma \hat{\gamma}_i\}$$

and

$$(4.15)\, dR^1_{ij}(t;s) = m^R_{ij}(s)\, dR^0_i(t;s) \exp\{\beta_\star\ I(T_i = s) + \beta_\psi \log \psi^R_{ij}(s) + \beta_\gamma \hat{\gamma}_i\},$$

with associated parameter and covariate vectors

$$\boldsymbol{\alpha}_{\star\psi\gamma} = \begin{bmatrix} \alpha_\star \\ \alpha_\psi \\ \alpha_\gamma \end{bmatrix} \quad \boldsymbol{Z}^{\star D}_i(s) = \begin{bmatrix} I(T_i = s) \\ \log \widehat{\psi}^D_{ij}(s) \\ \hat{\gamma}_i \end{bmatrix},$$

$$\boldsymbol{\beta}_{\star\psi\gamma} = \begin{bmatrix} \beta_\star \\ \beta_\psi \\ \beta_\gamma \end{bmatrix} \quad \boldsymbol{Z}^{\star R}_i(s) = \begin{bmatrix} I(T_i = s) \\ \log \widehat{\psi}^R_{ij}(s) \\ \hat{\gamma}_i \end{bmatrix}.$$

Models (4.14) and (4.15) can be expressed as $d\Lambda^1_{ij}(t;s) = m^D_{ij}(s) d\Lambda^0_i(t;s) \exp\{\boldsymbol{\alpha}'_{\star\psi\gamma} \boldsymbol{Z}^{\star D}_i(s)\}$ and $dR^1_{ij}(t;s) = m^R_{ij}(s)\, dR^0_i(t;s) \exp\{\boldsymbol{\beta}'_{\star\psi\gamma} \boldsymbol{Z}^{\star R}_i(s)\}$.

4.2.3 Parameter Estimation

In order to proceed with estimation of $\alpha_\star$ and $\beta_\star$ we first define risk set indicators for the two models. Let

$$Y^D_{ij}(t;s) = m^D_{ij}(s) I(X_j \wedge T_i > s + t)$$

and

$$Y^R_{ij}(t;s) = m^R_{ij}(s) I(X_j \wedge T_i > s + t)$$

indicate whether patient j is included in the risk set for stratum i. Recall that both $m_{ij}^D(s)$ and $m_{ij}^R(s)$ require that patient j be untreated at time s. In addition, if patient j is treated at $s+t$, they are censored from the strata in which they serve as a control and begin their own stratum as the treated patient. Since treatment likely depends on the recurrent event process, terminal event hazard, and the latent frailty, this constitutes dependent censoring. We propose to use a variant of Inverse Probability of Censoring Weighting (IPCW, Robins and Rotnitzky, 1992, Robins and Finkelstein, 2000) to correct this problem. We will weight control subjects by the inverse probability of treatment over the period $[s, s+t)$ as follows. We first model the probability of treatment using a traditional proportional hazards model fitted through standard partial likelihood (Cox, 1975) as

$$\lambda_i^T(t) = \lambda_0^T(t) \exp\{\boldsymbol{\beta}_T' \boldsymbol{Z}_i^H(t)\}.$$

Note that this model holds under the assumption that $\lambda_i^T(t|\boldsymbol{H}_i(t)) = \lambda_i^T(t|\boldsymbol{Z}_i^H(t))$, where $\boldsymbol{H}_i(t) = \{\boldsymbol{Z}_i(u), N_i^{D0}(u), N_i^0(u), I(T_i > u), I(X_i > u); 0 \leq u < t\}$, the observed history for subject i on $[0,t)$. In addition, $\boldsymbol{Z}_i^H(t)$ is a different covariate vector than $\boldsymbol{Z}_i(t)$ defined above, and can include the event history.

The probability of treatment on $[0,t)$ can then be defined as

$$G_i(t) = \exp\left\{-\int_0^t \lambda_i^T(u)\, du\right\}.$$

Since we are only interested in the probability of treatment on $[s, s+t)$ for a given subject we further define

$$G_i(t;s) = \frac{G_i(s+t)}{G_i(s)} = \exp\left\{-\int_s^{s+t} \lambda_i^T(u)\, du\right\}$$

and the associated weight functions

$$W_{ij}^D(t;s) = Y_{ij}^D(t;s) G_i(t;s)^{-I(T_i>s)}$$

and

$$W_{ij}^R(t;s) = Y_{ij}^R(t;s)G_i(t;s)^{-I(T_i>s)}$$

for the terminal and recurrent events, respectively.

Using these weight functions it can be shown that:

$$E\left[m_{ij}^D(s)W_{ij}^D(t;s)\,dM_{ij}^D(t;s)|\boldsymbol{Z}_i^{\star D}(s),\gamma_i\right] = 0$$

and that

$$E\left[m_{ij}^R(s)W_{ij}^R(t;s)\,dM_{ij}^R(t;s)|\boldsymbol{Z}_i^{\star R}(s),\gamma_i\right] = 0$$

where

$$dM_{ij}^D(t;s) = Y_{ij}^D(s)\{\,dN_i^D(s) - \lambda_{ij}(s)\}$$

and

$$dM_{ij}^R(t;s) = Y_{ij}^R(s)\{\,dN_i(s) - \,dR_{ij}(s)\}.$$

These zero mean processes can be used to set up the following sets of estimating equations:

$$\sum_{i=1}^{n} m_{ij}^D(s)\int_0^t W_{ij}^D(u;s)\,dM_{ij}^D(u;s), \tag{4.16}$$

$$\sum_{i=1}^{n} m_{ij}^D(s)\int_0^\tau \boldsymbol{Z}_i^{\star D}(s)W_{ij}^D(u;s)\,dM_{ij}^D(u;s), \tag{4.17}$$

$$\sum_{i=1}^{n} m_{ij}^R(s)\int_0^t W_{ij}^R(u;s)\,dM_{ij}^R(u;s), \tag{4.18}$$

and

$$\sum_{i=1}^{n} m_{ij}^R(s)\int_0^\tau \boldsymbol{Z}_i^{\star R}(s)W_{ij}^R(u;s)\,dM_{ij}^R(u;s), \tag{4.19}$$

where $\tau = \max\{X_1,\ldots,X_n\}$.

The estimation proceeds by solving for $d\Lambda_i^0(u;s)$ and $dR_i^0(u;s)$ in (4.16) and (4.18) respectively and substituting into (4.17) and (4.19). Summing across all experiments gives final estimating equations for $\boldsymbol{\alpha}_{\star\psi\gamma}$ and $\boldsymbol{\beta}_{\star\psi\gamma}$,

(4.20)
$$U^D(\boldsymbol{\alpha}) = \sum_{j=1}^{n}\sum_{i=1}^{n}\int_0^{\tau} m_{ij}^D(s)\int_0^{\tau-s}\{\boldsymbol{Z}_i^{\star D}(s) - \overline{\boldsymbol{Z}}_{\star}^D(u;s)\}\widehat{W}_{ij}^D(u;s)\,dN_i^D(u;s)\,dN_j^T(s)$$

and

(4.21)
$$U^R(\boldsymbol{\beta}) = \sum_{j=1}^{n}\sum_{i=1}^{n}\int_0^{\tau} m_{ij}^R(s)\int_0^{\tau-s}\{\boldsymbol{Z}_i^{\star R}(s) - \overline{\boldsymbol{Z}}_{\star}^R(u;s)\}\widehat{W}_{ij}^R(u;s)\,dN_i(u;s)\,dN_j^T(s)$$

where

$$\widehat{W}_{ij}^D(t;s) = Y_{ij}^D(t;s)\exp\{[\widehat{\Lambda}_i^T(s+t) - \widehat{\Lambda}_i^T(s)]I(T_i > s)\},$$

$$\widehat{W}_{ij}^R(t;s) = Y_{ij}^R(t;s)\exp\{[\widehat{\Lambda}_i^T(s+t) - \widehat{\Lambda}_i^T(s)]I(T_i > s)\},$$

$$\overline{\boldsymbol{Z}}_{\star}^D(u;s) = \frac{\sum_{\ell=1}^n Y_{\ell j}^D(s)\widehat{W}_{\ell}^D(u;s)\boldsymbol{Z}_{\ell}^{\star D}(s)\exp\{\boldsymbol{\alpha}_{\star\psi}'\boldsymbol{Z}_{\ell}^{\star D}(s)\}}{\sum_{\ell=1}^n Y_{\ell j}^D(s)\widehat{W}_{\ell}^D(u;s)\exp\{\boldsymbol{\alpha}_{\star\psi}'\boldsymbol{Z}_{\ell}^{\star D}(s)\}},$$

and

$$\overline{\boldsymbol{Z}}_{\star}^R(u;s) = \frac{\sum_{\ell=1}^n Y_{\ell j}^R(s)\widehat{W}_{\ell}^R(u;s)\boldsymbol{Z}_{\ell}^{\star R}(s)\exp\{\boldsymbol{\beta}_{\star\psi}'\boldsymbol{Z}_{\ell}^{\star R}(s)\}}{\sum_{\ell=1}^n Y_{\ell j}^R(s)\widehat{W}_{\ell}^R(u;s)\exp\{\boldsymbol{\beta}_{\star\psi}'\boldsymbol{Z}_{\ell}^{\star R}(s)\}}.$$

$\boldsymbol{\alpha}_{\star\psi\gamma}$ and $\boldsymbol{\beta}_{\star\psi\gamma}$ can then be estimated consistently by the solutions to $U^D(\boldsymbol{\alpha}) = \mathbf{0}$ and $U^R(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{0}$ denotes a vector with all elements equal to zero.

4.2.4 Asymptotic Properties

As described in Chapter II variance estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are challenging to derive in this complex setting, and asymptotic estimates may underestimate the true variance of these parameters. One option would be to use the bootstrap, similar to

the method described in Chapter III, however, this can be computationally intensive, especially in the setting where there are two outcomes of interest.

We propose instead a variation of a method proposed by Lin et al (2000) to construct confidence bands for the mean function of the proportional means model. We first impose the following regularity conditions:

(a) $\left[N_i(t), C_i(t), N_i^T(t), N_i^D(t), \boldsymbol{Z}_i(t), \gamma_i\right]$ are independent and identically distributed.

(b) $P(X_i \geq \tau) > 0$ for all i.

(c) $N_i(\tau) < \infty$ for all i.

(d) $E[I(T_i \leq \tau)] > 0$ for all i.

(e) $\boldsymbol{Z}_i(t)$ is of bounded variation.

Similar to Chapter II, under these conditions it can be shown for the terminal event that

$$n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = A_D^{-1}(\boldsymbol{\alpha}) n^{-1/2} \sum_{i=1}^{n} U_i^D(\boldsymbol{\alpha}) + o_p(1), \tag{4.22}$$

where

$$U_i^D(\boldsymbol{\alpha}) = \int_0^\tau m_i^D(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^{\star D}(s) - \overline{\mathbf{z}}_\star^D(u;s)\} dM_i^D(s) dF^T(s),$$

$$A_D(\boldsymbol{\alpha}) = E\left[\int_0^\tau m_i^D(s) \int_0^{\tau - s} \{\boldsymbol{Z}_i^{\star D}(s) - \overline{\mathbf{z}}_\star^D(u;s)\}^{\otimes 2} \exp\{\boldsymbol{\alpha}^T \boldsymbol{Z}_i^{\star D} d\Lambda_i^0(u;s) dF^T(s)\right],$$

$m_i^D(s)$ is a 0/1 indicator for patient i being matched on terminal event hazard to the index patient treated at time s and $F^T(s) = E[N^T(s)]$. Similarly, for the recurrent event, we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = A_R^{-1}(\boldsymbol{\beta}) n^{-1/2} \sum_{i=1}^{n} U_i^R(\boldsymbol{\beta}) + o_p(1), \tag{4.23}$$

where

$$U_i^R(\boldsymbol{\beta}) = \int_0^\tau m_i^R(s) \int_0^{\tau-s} \{\boldsymbol{Z}_i^{\star R}(s) - \overline{\mathbf{z}}_\star^R(u;s)\} dM_i^R(s) dF^T(s),$$

$$A_R(\boldsymbol{\beta}) = E\left[\int_0^\tau m_i^R(s) \int_0^{\tau-s} \{\boldsymbol{Z}_i^{\star R}(s) - \overline{\mathbf{z}}_\star^R(u;s)\}^{\otimes 2} \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i^{\star R}\} dR_i^0(u;s) dFT(s)\right],$$

$m_i^R(s)$ is an indicator taking value 1 when patient i is matched in terms of recurrent event rate to the index patient treated at time s and $F^T(s)$ is defined as above.

Using this, we can estimate the distribution of (4.22) and (4.23) by repeatedly sampling from $\widehat{A}_D^{-1}(\widehat{\boldsymbol{\alpha}}) n^{-1/2} \sum_{i=1}^n \widehat{U}_i^D(\widehat{\boldsymbol{\alpha}}) P_i$ for the terminal event and $\widehat{A}_R^{-1}(\widehat{\boldsymbol{\beta}}) n^{-1/2} \sum_{i=1}^n \widehat{U}_i^R(\widehat{\boldsymbol{\beta}}) P_i$ for the recurrent event, where $P_i \sim Exp(1)$. Reasons and implications for choosing the exponential distribution are discussed further in Section 4.5.

4.3 Simulation Study

Simulation was used to assess the performance of the proposed method in moderate sized samples. True values of $\alpha_\star$ and $\beta_\star$ were determined by generating both observed and counterfactual data for each subject. A frailty for each subject was generated from a Normal (0, 0.5) distribution, and baseline covariates Z_{i1} and Z_{i2} were generated from a Uniform [-1,1] distribution. Pre-treatment recurrent and terminal events were generated from the following models:

$$dR_i^0(t) = dR_0(t) \exp\{\beta_{01} Z_{i1} + \beta_{02} Z_{i2} + \gamma_i\}$$

and

$$d\Lambda_i^0(t) = d\Lambda_0(t) \exp\{\alpha_{01} Z_{i1} + \alpha_{02} Z_{i2} + \gamma_i\}.$$

Treatment times were then generated from

$$\lambda_i^T(t) = \lambda_0^T(t) \exp\{\delta_1 Z_{i1} + \delta_2 Z_{i2} + \delta_3 \log(N_i^0(t^-) + 1)\}.$$

For subjects with $T_i < D_i$, post-treatment recurrent and terminal events were generated from the following intensity and hazard:

$$dR_i^1(t) = dR_0^1(t)\exp\{\beta_{11}Z_{i1} + \beta_{12}Z_{i2} + \gamma_i\}$$

and

$$d\Lambda_i^1(t) = d\Lambda_0^1(t)\exp\{\alpha_{11}Z_{i1} + \alpha_{12}Z_{i2} + \gamma_i\}.$$

Note that for each subject with $T_i < D_i$, we generated both treatment-free and post-treatment recurrent and terminal event experience. However, only one of these is observed. The generated data are combined such that subject i has a recurrent event intensity of $dR_i^0(t)$ on $[0, T_i)$ and a recurrent event intensity and terminal event hazard of $dR_i^1(t)$ and $d\Lambda_i^1(t)$, respectively, on $[T_i, T_i + t)$. For subjects with $T_i > D_i$, $dR_i^1(t)$ and $d\Lambda_i^1(t)$ are never observed, and only $dR_i^0(t)$ and $d\Lambda_i^0(t)$ are generated and modeled for these subjects.

Once the data were generated, the hazard for treatment was modeled as $\lambda_i^T(t) = \lambda_0^T(t)\exp\{\beta_{T1}Z_{i1} + \beta_{T2}Z_{i2} + \beta_{T3}\log(N_i(t^-) + 1)\}$ in order to calculate weights for IPCW. At this stage we also fit the prognostic model using the frailty model

$$dR_i^0(t) = dR_0(t)\exp\{\xi_{01}Z_{i1} + \xi_{02}Z_{i2} + \gamma_i\}$$

and

$$d\Lambda_i^0(t) = d\Lambda_0(t)\exp\{\delta_{01}Z_{i1} + \delta_{02}Z_{i2} + \hat{\gamma}_i\}.$$

For the terminating event model, subjects were matched if $|\log\psi_{ij}^D(s)| \leq 0.1$, and for the recurrent event model, matching was successful if $|\log\psi_{ij}^R(s)| \leq 0.1$. This resulted in a median of 16 matches per stratum for the recurrent event model and a median of 17 matches per stratum for the terminal event model. Parameters used to generate are as follows: $dR_0 = 6, \beta_{01} = 0.3, \beta_{02} = 0.1, d\Lambda_0 = 0.3, \alpha_{01} = 0.3, \alpha_{02} =$

Table 4.1: Descriptive statistics from simulations or correlated recurrent and terminal events

Scenario	1	2	3	4	5	6
% Tx	66%	66%	66%	66%	66%	66%
% Died	60%	69%	65%	53%	54%	54%
% Died Pre-Tx	87%	87%	87%	87%	87%	87%
% Died Post-Tx	46%	59%	54%	36%	39%	37%
Event Mean	11.81	10.88	12.81	10.44	11.01	11.60

$0.1, \lambda_0^T = 0.25, \delta_1 = 0.4, \delta_2 = 0.2, \delta_3 = 0.5, \beta_{11} = 0.3, \beta_{12} = 0.1, \alpha_{11} = 0.3, \alpha_{12} = 0.1$. In the different scenarios dR_0^1 took on values of 4, 4.5, 5, 6, 6.5, and 8, and $d\Lambda_0^1$ took on values of 0.2, 0.3, 0.4, and 0.5. Each scenario generated 400 subjects and was simulated 500 times. Standard error estimates were generated from 100 perturbations of the weighted score residuals, and these estimates were compared to standard errors estimated from 50 bootstrapped samples.

Descriptive information regarding proportion treated and died and average number of events is given in Table 4.1, and simulation results are given in Table 4.2. Bias for the frailty variance, θ, was 0.020, with empirical and asymptotic standard errors approximately equal and coverage probability close to the desired level of 0.95, indicating that the prognostic model was capturing the variance in the random effects correctly. Most scenarios showed reasonably small bias for $\alpha_\star$, ranging in absolute value from 0.005 to 0.028, with bias exceeding 0.02 in only one scenario. Standard error estimates based on the proposed method were similar to empirical standard errors, resulting in coverage probabilities ranging from 0.966-0.982. Finally, estimates of $\beta_\star$ showed relatively small amounts of bias, ranging from 0.003 to 0.017, with coverage probabilities ranging from 0.948 to 0.972. Similar to $\alpha_\star$, the proposed

Table 4.2: Simulation results for correlated recurrent and terminal events using a frailty prognostic model

Scenario	Outcome	$\alpha_\star$ or $\beta_\star$	Est	Bias	ESE	ASE	CP
1	Recurrent Event	−0.019	−0.016	0.003	0.055	0.059	0.964
1	Survival	0.005	−0.013	−0.018	0.212	0.235	0.980
2	Recurrent Event	0.058	0.041	−0.017	0.053	0.060	0.966
2	Survival	0.510	0.483	−0.028	0.186	0.222	0.982
3	Recurrent Event	0.267	0.262	−0.005	0.054	0.057	0.960
3	Survival	0.289	0.264	−0.025	0.194	0.228	0.982
4	Recurrent Event	−0.423	−0.413	0.010	0.059	0.062	0.966
4	Survival	−0.404	−0.396	0.008	0.230	0.248	0.976
5	Recurrent Event	−0.305	−0.293	0.012	0.056	0.061	0.972
5	Survival	−0.402	−0.388	0.014	0.225	0.248	0.966
6	Recurrent Event	−0.199	−0.185	0.014	0.057	0.060	0.948
6	Survival	−0.406	−0.400	0.005	0.223	0.248	0.982

ESE=empirical standard error; ASE=asymptotic standard error;
CP=coverage probability

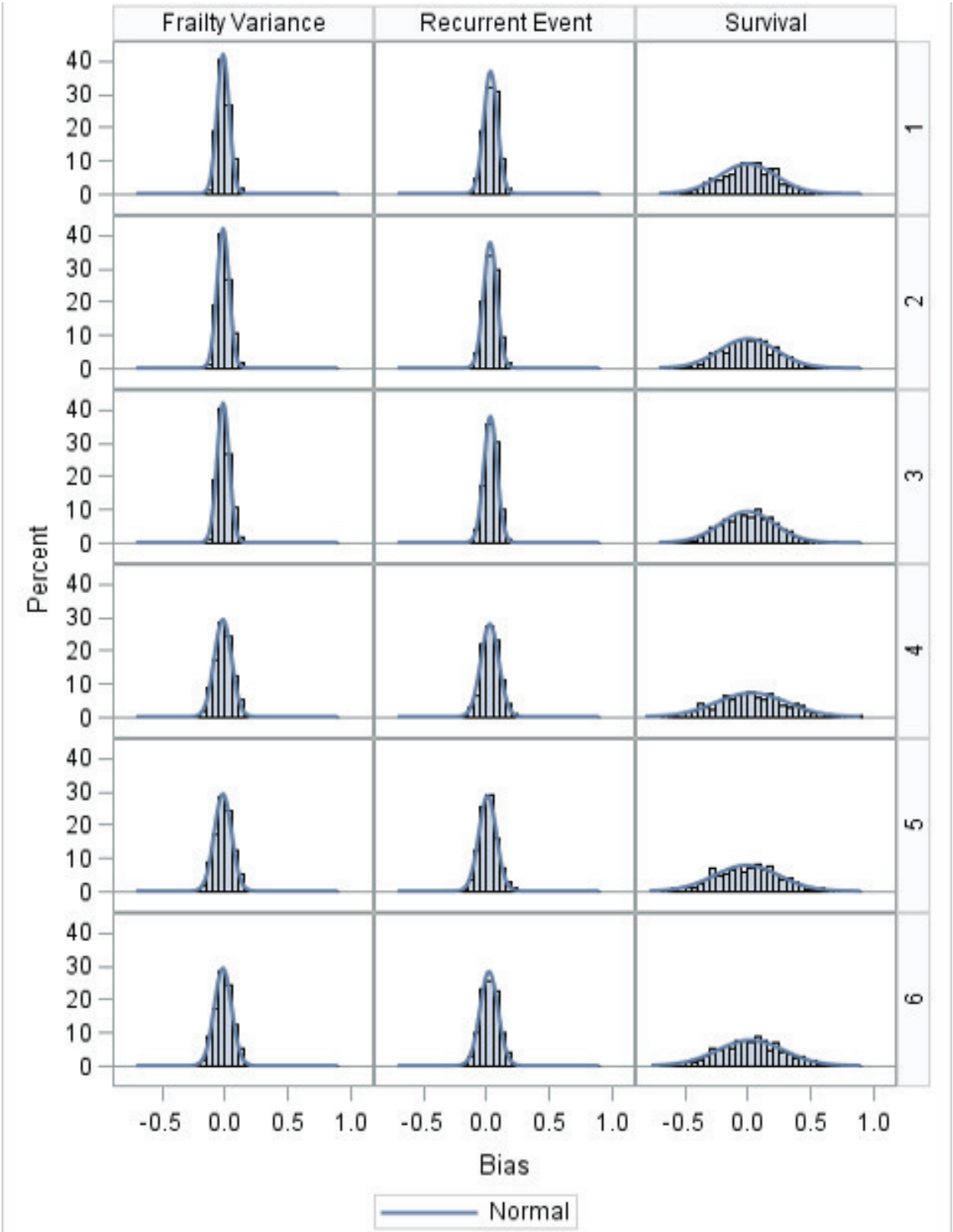

Figure 4.1: Histogram of parameter estimates from proposed models of correlated recurrent and terminal events using a frailty prognostic model with normal density

Table 4.3: Simulation results for correlated recurrent and terminal events with bootstrapped standard errors

Scenario	Outcome	$\alpha_\star$ or $\beta_\star$	Est	Bias	ESE	ASE	CP
1	Recurrent Event	−0.019	−0.017	0.002	0.055	0.050	0.905
1	Survival	0.005	−0.015	−0.019	0.204	0.208	0.944
2	Recurrent Event	0.058	0.042	−0.016	0.056	0.051	0.909
2	Survival	0.510	0.474	−0.036	0.188	0.191	0.942
3	Recurrent Event	0.267	0.260	−0.007	0.052	0.049	0.922
3	Survival	0.289	0.270	−0.019	0.197	0.199	0.960
4	Recurrent Event	−0.423	−0.412	0.011	0.058	0.053	0.899
4	Survival	−0.404	−0.401	0.002	0.231	0.224	0.940
5	Recurrent Event	−0.305	−0.294	0.011	0.053	0.052	0.922
5	Survival	−0.402	−0.402	0.001	0.233	0.222	0.940
6	Recurrent Event	−0.199	−0.187	0.013	0.055	0.051	0.924
6	Survival	−0.406	−0.397	0.008	0.233	0.225	0.950

ESE=empirical standard error; ASE=asymptotic standard error;
CP=coverage probability

method for variance estimation produced standard error estimates of $\widehat{\beta_\star}$ which were similar to empirical standard errors. Histograms of the estimates of $\alpha_\star$, $\beta_\star$, and θ are shown in Figure 4.1 with normal density curves, demonstrating the normality of these estimates centered at their true value.

Table 4.3 shows simulation results when the bootstrap method is used to obtain standard errors instead of the proposed method. This method is more computationally intensive, and gives some underestimation of the standard errors for the recurrent event parameter, with coverage ranging from 0.899 to 0.924. The proposed method of variance estimation, by contrast, is slightly more conservative, in some

cases overestimating the standard errors and giving coverage probabilities slightly over the nominal level of 0.95.

4.4 Application to Liver Transplantation

The method described above was applied to data from the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL), supplemented with data from the Scientific Registry of Transplant Recipients, both of which are described in Chapters II and III. Patients with ESRD have increased morbidity, as shown in Chapter II as well as increased mortality. We will use the method proposed in this chapter to evaluate the difference in the rate of days hospitalized and hazard of mortality between liver transplant recipients that develop ESRD post-transplant compared to the rate and hazard that would have occurred had they not developed ESRD.

Post-LT ESRD occurred in 55 of 1447 LT recipients in A2ALL. The post-ESRD rate of days hospitalized was 10.6 per patient year at risk compared to 4.5 days hospitalized per patient year at risk pre-ESRD. Of the 55 patients that developed post-LT ESRD 23 died, while 261 deaths occurred in the 1392 patients that did not develop post-LT ESRD. A frailty prognostic model was fitted and included donor and recipient age, recipient race, diabetes, hepatitis C diagnosis, at transplant values of creatinine, bilirubin, and albumin, donation after cardiac death, local, regional, or national share, and indicators for split liver and living donor transplant. The estimated variance of the subject-specific frailty was 3.6 (SE=0.16). Matching for the final analysis of days hospitalized rate was successful if $|\log \widehat{\psi}_{ij}^{R}(s)|<0.05$, resulting in a median of 13 matches per stratum, with 2 (3.6%) strata excluded due to inability to match. Similarly, for the mortality model, matching was successful if $|\log \widehat{\psi}_{ij}^{D}(s)|<0.15$, resulting in a median of 17.5 matches per stratum and 3 (5.5%) strata excluded due

Table 4.4: Morbidity and mortality related to ESRD development post-LT: Results from proposed method compared to traditional models

Outcome	Model	HR	CI	p-value
Recurrent Event	Proposed Model	2.45	1.55−3.87	<0.001
	Traditional Baseline Model	3.17	3.01−3.35	<0.001
	Traditional Time-Dependent Model	1.44	1.35−1.52	<0.001
Survival	Proposed Model	1.88	1.03−3.45	0.04
	Traditional Baseline Model	3.52	2.23−5.56	<0.001
	Traditional Time-Dependent Model	1.65	1.00−2.72	0.05

to lack of eligible matches. Two sets of traditional time-dependent ("naive") models were also fitted, one adjusted for baseline covariates only and one adjusting for time-dependent covariates including the event history.

Results from the three sets of models are shown in Table 4.4. Using the proposed method, patients developing post-LT ESRD had a days hospitalized rate 2.45 times higher than similar patients that had not developed ESRD at the time of the index patient. This result is similar to that given in Section 2.4, although slightly lower, indicating a slight overestimation of the effect estimate as a result of ignoring the correlated terminal event in Chapter II. Mortality was also higher in patients developing ESRD, with an 88% higher risk of death in these patients using the proposed method. Similar to Chapter II, we demonstrate that the traditional baseline Cox model overestimates the effect of ESRD development on both the rate of days hospitalized and the hazard of death, while the traditional time-dependent Cox model including the event history underestimates this effect.

4.5 Discussion

In the above chapter we propose a method for estimating effects of a time-dependent treatment on correlated recurrent and terminal event outcomes. The proposed method uses a frailty prognostic model to match patients that receive treatment at time s to those that are untreated at s with similar trajectories in their recurrent event intensity and terminal event hazards based both on covariate effects as well as underlying frailty. The method of sequential stratification is then used to compare the recurrent event rate and terminal event hazard between treated patients and a matched control group representing the treated patients' experience had they not received treatment at time s.

The proposed method incorporates two important aspects that reduce bias when estimating treatment effects on the recurrent event rate in the presence of a correlated terminal event. First, it corrects bias related to the correlated terminal event by incorporating a latent frailty into the prognostic model. We use a model similar to that proposed by Liu et al (2004), and include the estimated frailty in the final recurrent and terminal event models. We fit the frailty prognostic model using estimating equations in a similar vein to methods proposed by Kalbfleisch et al (2013). Second, we use a partly conditional model in order to correct bias related to over and under adjustment that often occurs in traditional recurrent event models such as those of Anderson and Gill (1982) and Lin et al (2000). In the setting where treatment assignments are made during the course of follow-up the proposed method conditions on the history prior to treatment and marginalizes over the history after treatment in order to make appropriate comparisons.

In the proposed method inference depends on estimation of the asymptotic distri-

bution of (4.22) and (4.23) via an approximation involving repeated sampling from an exponential distribution with unit mean and variance. In the original method proposed by Lin et al (2000) the authors simulate from a standard normal distribution to approximate the asymptotic distribution of the mean function, however, given the underestimation of standard errors by the sandwich estimator, the heavier tails of the exponential distribution are able to correct the underestimation more effectively than the standard normal. With the exponential distribution we end up with slight overestimation of standard errors, therefore an in between distribution, such as a t-distribution with 15 degrees of freedom, could be explored. The over-coverage could also be caused by treating the estimated frailty, γ_i, and the weight function as known quantities, as is done in the proposed model, however, treating the weight function especially as a known is commonly done in the literature.

Limitations to this method include parametric assumptions on both the prognostic model's baseline hazard and frailty. We propose the use of a log-normal frailty, although other distributions could by used, such as the gamma distribution which is a common choice. Since the underlying frailty is unobserved, it may be beneficial in some settings to assume a non-parametric form of the frailty. An additional limitation is the proportional hazards assumption on the parameter estimates of interest, $\boldsymbol{\alpha}_\star$ and $\boldsymbol{\beta}_\star$. This assumption can be explored using interactions with time of treatment, s, as well as time since treatment, t as in Chapters II and III.

CHAPTER V

Conclusions and Future Work

In the above dissertation we develop methods for estimating time-dependent treatment effects on the recurrent event rate using an extension of the method of sequential stratification developed by Schaubel et al (2009). While randomized controlled trials are the gold-standard for determining treatment effects, in many settings they are neither feasible nor ethical. In these cases observational studies, which often include treatment assignment after the beginning of follow-up and treatment by indication, are necessary for estimating treatment effects. In this setting it is important to balance conditional and marginal approaches in order to obtain unbiased estimates with appropriate interpretations.

We first presented methodology for estimating rare treatment effects on the recurrent event rate. We introduced a two-stage modeling technique in order to estimate these effects. In the first stage we fit a prognostic model on the pre-treatment experience for all subjects. Treated subjects were then matched to as-yet untreated subjects based on prognostic scores. The final model was fitted using the method of sequential stratification yielding a treatment effect estimate which can be interpreted as the effect of treatment on the recurrent event rate for a subject treated at time s compared to the rate that would have occurred after s had the subject not been

treated at that time. The use of the prognostic score for matching ensures that patients are on the same recurrent event trajectory prior to the time of treatment, but allows these trajectories to diverge post-treatment in order to allow for a potentially differing effect of treatment on the recurrent event rate.

In the next chapter we extended the method proposed in Chapter II in two ways. First we allowed for multiple treatment states by introducing the concept of conventional therapy, in which patients begin follow-up untreated and subsequently receive a standard treatment. These patients are contrasted with patients that receive an experimental treatment at time s. Second, we incorporated Inverse Probability of Censoring Weighting in order to correct informative censoring at the time of treatment. As demonstrated in Chapter II, when experimental treatment exceeds 20%, censoring of control subjects at subsequent experimental treatment initiation induces bias when treatment is associated with the recurrent event process.

The third method proposed addressed the case when the recurrent event process is correlated with a terminal event. In order to account for this correlation we use a frailty prognostic model in the first stage of modeling. Similar to the previous two chapters the prognostic models are fit on the pre-treatment experience, however, we propose the use of a random frailty to account for the correlation between the recurrent and terminal events. We assume in this setting that the terminal event is also of interest and model it in addition to the recurrent events, therefore we complete the matching based on prognostic scores twice. The method allows for matching to depend on prognostic scores based on the pre-treatment recurrent event process, the pre-treatment terminal event hazard, or some combination of the two depending on what is most appropriate for the data at hand. In addition estimated frailty values can be incorporated into the matching. The final models give treatment

effect estimates for both the recurrent event rate and terminal event hazard with similar interpretation as described above, and weighting is incorporated for common treatments.

There are several areas in which the methods proposed above could be extended. First, in the setting of common treatments we proposed the use of the bootstrap as well as a variation of a method proposed by Lin et al (2000). In this setting the question of convergence of $\overline{\boldsymbol{Z}}_i^{\star}$ remains unresolved, and deserves further investigation. One potential way to address this issue would be to extend the method in order to allow for multiple treated subjects per stratum. In the methods proposed the continuous time axis and matching scheme necessitates a single treated subject per stratum, however, grouping similar treated subjects could assure convergence of $\overline{\boldsymbol{Z}}_i^{\star}$ as well as allow for more efficient estimators of the treatment effect. Finally, another area of interest would be to extend the current methods to allow for increments in the recurrent event counter greater than 1, for potential application to cost data.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Abadie A., Imbens G.W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica* **76,** 1537-1557.

[2] Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* **10,** 129–143.

[3] Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics* **25,** 662–682.

[4] Chen B. E. and Cook R. J. (2004). Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics* **5,** 339–354.

[5] Chen, Q., Zeng, D., Ibrahim, J.G., Akacha, M. and Schmidli, H. (2013). Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrika* **100,** 339–354.

[6] Cook, R.J. and Lawless, J.F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16,** 911–924.

[7] Cook, R.J., Lawless, J.F., Lakhal-Chaieba, L. and Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association* **104,** 60–75.

[8] Cook, R.J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Event Data*, New York: Wiley.

[9] Cox, D.R. (1975) Partial likelihood. *Biometrika* **62,** 269–276.

[10] Davidson, R. and MacKinnon J.G. (2000). Bootstrap tests: how many bootstraps? *Econometric Reviews* **19,** 55–68.

[11] Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7,** 1-26.

[12] Ghosh, D. and Lin. D.Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56,** 554–562.

[13] Ghosh, D. and Lin. D.Y. (2002). Marginal regression models for recurrent and terminal event. *Statistica Sinica* **12,** 663–688.

[14] Gong, Q. and Schaubel, D.E. (2013). Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics*, **69,** 338–347.

[15] Hansen, B.B. (2008). The prognostic analogue of the propensity score. *Biometrika* **95,** 481–488.

[16] Hansen, B.B. (2008). The essential role of balance tests in propensity-matched observational studies: comments on 'A critical appraisal of propensity-score matching in the medical literatire between 1996 and 2003' by Peter Austin, Statistics in Medicine. *Statistics in Medicine* **27,** 2037–2049.

[17] Huang C.-Y., Qin, J. and Wang, M.-C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66,** 39–49.

[18] Huang, C.-Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99,** 1153–1165.

[19] , Huang X. and Liu, L.(2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics* **63,** 389–397.

[20] Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data: 2nd Edition*, New York: Wiley.

[21] Kalbfleisch, J.D., Schaubel, D.E., Ye, Y. and Gong, Q. (2013). An estimating function approach to the analysis of recurrent and terminal events. *Biometrics*, 69, 366-374.

[22] Karatzas, I. and Shreve, S.E. (1988). *Brownian Motion and Stochastic Calculus*. New York: Springer.

[23] Kim, S., Zeng, D., Chambless, L. and Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in the Biosciences* **4,** 262–281.

[24] Lancaster T. and Intrator O. (1998). Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93,** 46–53.

[25] Lawless, J. F., Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

[26] Leacy, F.P. and Stuart, E.A. (2014). On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine* **33,** 3488–3508.

[27] Li, Y., Schaubel, D. E., He, K. (2014). Matching methods for obtaining survival functions to estimate the effect of a time-dependent treatment. *Statistics in Biosciences* **6,** 105–126.

[28] Lin, D. Y., Wei, L. J., Yang, I., Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62,** 711–730.

[29] Liu, L., Wolfe, R. A. and Huang X. (2004). Shared frailty model for recurrent events and a terminal event. *Biometrics* **60,** 747–756.

[30] Lu, B. (2005). Propensity score matching with time dependent covariates. *Biometrics* **61,** 721–728.

[31] Merion, R. M., Shearon, T. H., Berg, C. L., Everhart, J. E., et al. (2010). Hospitalization rates before and after adult-to-adult living donor or deceased donor liver transplantation. *Annals of Surgery* **251,** 542–549.

[32] Miloslavsky, M., Keles, S., Van Der Laan, M. J. and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society, Series B*, **66,** 239–257.

[33] Pan Q. and Schaubel D. E. (2009). Flexible estimation of differences in treatment-specifi recurrent event means in the presence of a terminating event. *Biometrics* **65,** 753–761.

[34] Pena, E.A., Strawderman, R.L. and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association* **96,** 1299–1315.

[35] Pepe, M. S., Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

[36] Pepe, M.S. and Couper, D. (1997). Modeling partly conditional means with longitudinal data. *Journal of the American Statistical Association* **92,** 991–998.

[37] Pollard, D. (1990). *Empirical Processes: Theory and Application.* HaywardL Institute of Mathematical Statistics.

[38] Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology - Methodological Issues.* Birkhäuser, Boston, 297–331.

[39] Rondeau V., Mathoulin-Pelissier S., Jacqmin-Gadda H., Brouste V., and Soubeyran P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* **8,** 708–721.

[40] Rosenbaum P. and Rubin D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41-55.

[41] Rubin, D.B. and Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* **95,** 573–585.

[42] Schaubel, D. E., Wolfe, R. A., Sima, C. S., Merion, R. M. (2009). Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate. *Journal of the American Statistical Association* **104,** 49–59.

[43] Schaubel, D.E. and Zhang, M. (2010). Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. *Lifetime Data Analysis* **16,** 451–477.

[44] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

[45] Strawderman R. L. (2000). Estimating the mean of an increasing stochastic process at a censored stopping time. *Journal of the American Statistical Association* **95,** 1192–1208.

[46] Stuart, E.A., Lee, B.K. and Leacy, F.P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology* **66,** S84–S90.

[47] van Houwelingen H.C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34,** 70-85.

[48] van Houwelingen H.C. and Putter H. (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis* **14,** 447-463.

[49] van Houwelingen H.C. and Putter H. (2011). *Dynamic prediction in clinical survival analysis.* Boca Raton, FL: CRC Press, Inc.

[50] Wang M.-C., Qin J. and Chiang C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96,** 1057–1065.

[51] Yu Z., Liu L., Bravata D.M. and Williams L. S. (2014). Joint model of recurrent events and a terminal event with time-varying coefficients. *Biometrical Journal* **56,** 183–197.

[52] Ye, Y., Kalbfleisch, J.D. and Schaubel, D.E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63,** 78–87.

[53] Zeng D. and Lin D.Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65,** 746–752.

[54] Zhao, X., Liu, L., Liu, Y. and Xu. W. (2012). Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal* **54,** 585–599.

[55] Zheng, Y. Y. and Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics*, **61,** 379–391.