

# Semiparametric Methods for Estimating the Effect of a Longitudinal Covariate and Time-Dependent Treatment on Survival Using Observational Data with Dependent Censoring

by  
Qi Gong

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Biostatistics)  
in The University of Michigan  
2012

Doctoral Committee:

Associate Professor Douglas E. Schaebel, Chair  
Professor John D. Kalbfleisch  
Professor Robert M. Merion  
Assistant Professor Min Zhang

# TABLE OF CONTENTS

<b>LIST OF FIGURES</b> . . . . .	<b>iv</b>
<b>LIST OF TABLES</b> . . . . .	<b>v</b>
<b>LIST OF APPENDICES</b> . . . . .	<b>vi</b>
<b>CHAPTER</b>	
<b>I. Introduction</b> . . . . .	<b>1</b>
<b>II. Partly Conditional Estimation of the Effect of a Time-Dependent Factor in the Presence of Dependent Censoring</b> . . . . .	<b>4</b>
2.1 Introduction . . . . .	4
2.2 Proposed Methods . . . . .	9
2.3 Asymptotic Properties . . . . .	15
2.4 Simulation . . . . .	17
2.5 Application . . . . .	20
2.6 Discussion . . . . .	27
<b>III. Estimating Average Treatment Effects on Mean Survival Time when Treatment is Time-dependent and Censoring is Dependent</b> . . . . .	<b>30</b>
3.1 Introduction . . . . .	30
3.2 Proposed Methods . . . . .	33
3.3 Asymptotic Properties . . . . .	43
3.4 Simulation . . . . .	45
3.5 Application . . . . .	48
3.6 Discussion . . . . .	53
<b>IV. Semiparametric Cumulative Contrasts of Group-Specific Mortality in the Presence of Dependent Censoring</b> . . . . .	<b>58</b>
4.1 Introduction . . . . .	58

4.2	Proposed Methods . . . . .	61
4.3	Asymptotic Properties . . . . .	65
4.4	Simulation . . . . .	66
4.5	Application . . . . .	68
4.6	Discussion . . . . .	75
<b>V. Conclusion . . . . .</b>		<b>78</b>
<b>APPENDICES . . . . .</b>		<b>81</b>
<b>BIBLIOGRAPHY . . . . .</b>		<b>138</b>

## LIST OF FIGURES

### Figure

2.1	Examples of the relationship between cross-section time and follow-up time. . . . .	10
3.1	Examples of the relationship between cross-section time and follow-up time. . . . .	38
3.2	Estimated post- and pre-transplant survival curves by MELD score at transplant. . . . .	56
4.1	Cumulative hazard ratio of high-MELD patients compared to Status 1 patients. . . . .	73
4.2	Cumulative hazard ratio for MELD>40 patients compared to Status 1 patients, with 95% confidence interval. . . . .	74
4.3	Cumulative hazard ratio for MELD 36-40 patients compared to Status 1 patients, with 95% confidence interval. . . . .	74
4.4	Cumulative hazard ratio for MELD 21-23 patients compared to Status 1 patients, with 95% confidence interval. . . . .	75
B.1	Estimated post- and pre-transplant survival curves for all MELD groups . . . . .	122

## LIST OF TABLES

### Table

2.1	Simulation results for $\hat{\beta}$ computed using Type A Weight; $n = 1000$ .	21
2.2	Simulation results for $\hat{\beta}$ computed using Type B Weight given in (2.4); $n = 1000$ . . . . .	22
2.3	Simulation results for $\hat{\beta}$ computed using Type C Weight given in (2.5); $n = 1000$ . . . . .	23
2.4	Analysis of liver wait-list mortality (using Type B Weight) . . . . .	25
2.5	Analysis of liver wait-list mortality; Comparison of results using different weight. . . . .	26
3.1	Simulation result for $\Delta = 0$ with $\{W_{ik}^B(t)\}$ . . . . .	49
3.2	Simulation result for $\Delta > 0$ with $\{W_{ik}^B(t)\}$ . . . . .	50
3.3	Estimate of $\Delta(\text{SE})$ by MELD groups. . . . .	54
3.4	Estimate of $S_{\Delta}(\text{SE})$ by MELD groups. . . . .	55
4.1	Simulation results for $\hat{\phi}_1(t)$ when $\phi_1(t)$ increases in $t$ . . . . .	69
4.2	Simulation results for $\hat{\phi}_1(t)$ when $\phi_1(t)$ decreases in $t$ . . . . .	70
4.3	Simulation results for $\hat{\phi}_1(t)$ when $\phi_1(t)$ is constant in $t$ . . . . .	71

## LIST OF APPENDICES

### Appendix

A.	Appendix A: Proof of Theorem II.1 . . . . .	82
B.	Appendix B: Proof of Theorem III.1 . . . . .	92
C.	Appendix C: Proof of Theorem VI.1 . . . . .	123

# CHAPTER I

## Introduction

This dissertation develops novel methodology for three challenging problems in the estimation of treatment and group effects in the presence of dependently censored survival data. In Chapter II, we propose semiparametric methods for estimating the effect of a time-dependent covariate on treatment-free survival. The data structure of interest consists of a longitudinal sequence of measurements and a potentially censored survival time. The factor of interest is time-dependent. Treatment-free survival is of interest and is dependently censored by the receipt of treatment. Patients may be removed from consideration for treatment, temporarily or permanently. The proposed methods involve landmark analysis and partly conditional hazard regression. Dependent censoring is overcome through a variant of Inverse Probability of Censoring Weighting (IPCW). The regression parameter of interest is marginal in the sense that time-varying covariates are taken as fixed at each landmark, with the modeled mortality hazard function then implicitly averaging across future covariate trajectories. The proposed methods circumvent the need for explicit modeling of the longitudinal covariate process.

Mean survival time can be used to quantify both the patient-specific and average effect of a time-dependent treatment. In Chapter III, we propose semiparamet-

ric methods for estimating the mean difference between treatment-free and post-treatment restricted mean lifetime. The data structure of the treatment-free period consists of a longitudinal sequence of measurements and a potentially censored survival time, with treatment-free survival dependently censored by the receipt of treatment. Landmark analysis, partly conditional hazard regression, and IPCW are used to model treatment-free survival. The post-treatment death hazard is modeled using the measurement history leading up to the treatment time. The average treatment effect is obtained by averaging over subjects who received treatment, in a manner which accounts for the independent censoring of treatment times.

In time to event data observed in medical studies, nonproportional hazards and dependent censoring are common issues when comparing group-specific mortality. The group effect on mortality may vary over time, as opposed to being constant. One remedy is adopting a parametric form to model the time-dependent pattern. However, it is generally difficult to verify the chosen parametric function. Moreover, in the settings where the proportional hazard assumption fails, investigators tend to be more interested in cumulative effects on mortality rather than the instantaneous effect. Estimators using standard approaches are generally biased in the presence of dependent censoring, which may occur when both censoring and death depend on the same time-dependent covariates. Therefore, in Chapter IV, we propose an estimator for the cumulative group effect on survival in the presence of nonproportional hazards and dependent censoring. The proposed estimator is based on the cumulative hazard function, assumed to follow a stratified Cox model. No functional form needs to be assumed for the nature of the nonproportionality.

Each of the proposed methods is shown to be consistent and asymptotically normal, with consistent covariance estimators provided. Simulation studies reveal that



the proposed estimation procedures are appropriate for practical use. We apply the proposed methods to pre- and post-transplant data on End-stage Liver Disease patients.

## CHAPTER II

# Partly Conditional Estimation of the Effect of a Time-Dependent Factor in the Presence of Dependent Censoring

### 2.1 Introduction

Longitudinal and survival data are often observed simultaneously in biomedical studies. For instance, measurements related to patient health may be collected over time as a longitudinal process during the course of a patient's disease, while time to a failure event (e.g., death) is of chief interest. Information on time-varying covariates is usually collected at multiple follow-up times through the time to the event. For example, the number of CD4-lymphocyte counts is frequently employed as a surrogate marker for HIV; the glomerular filtration rate has been utilized as an indicator of kidney failure; the Model for End-stage Liver Disease (MELD) score, (Wiesner et al., 2001) is a very sensitive indicator of liver dysfunction among End-stage Liver Disease (ESLD) patients. In particular, the liver failure setting is the motivation for the methods we propose in this report.

In the presence of longitudinal and survival data, it is often of interest to estimate the effect on survival time of a time-dependent factor hypothesized to be an important indicator for disease progression. One approach to link the time-varying

covariate with the time to event is proportional hazards model (Cox, 1972). To jointly model survival and longitudinal data, a regression model for the time dependent covariate process is usually adopted. Recent literature includes, for example, methods proposed by Tsiatis, Degruittola, and Wulfsohn (1995); Henderson, Diggle, and Dobson (1997); Xu and Zeger (2001); Song, Davidian, and Tsiatis (2002) and Taylor (2011). In joint modeling, valid inference on the time-to-event component generally requires that the longitudinal process be modeled accurately. This is difficult to accomplish in many cases. For example, in practice, one may only measure the covariate process at discrete times instead of continuously.

Moreover, each of the joint modeling approaches cited in the preceding paragraph models the death hazard as a function of the covariate at time  $t$ . Although these time-dependent approaches have a long history in survival analysis and are applicable in many situations, they may not be consistent with the investigator's objectives in many practical settings. For example, in the liver failure setting, it is often of interest to determine which of several patients awaiting liver transplantation will die soonest in the absence of a transplant; i.e., based on each patient's history up until that date, and averaging over the possible scenarios that could occur in the future without a transplant. An available donor liver is allocated once, meaning that the decision about which patient is expected to die fastest without a liver transplant would need to be based on information only up to the date the organ is allocated.

Zheng and Heagerty (2005) proposed a partly conditional model applicable to some settings like that described in the preceding paragraph. Typically in modeling survival data, the event time,  $D_i$ , is from study entry (marking the beginning of follow-up) to the occurrence of the failure event. In Zheng and Heagerty (2005), it is duration since measurement time  $S_i$ , i.e.,  $(D_i - S_i)$  that is modeled. The method of

Zheng and Heagerty (2005) is referred as “partly conditional” in the sense that the hazard function being modeled (i.e., that of  $D_i - S_i$ ) only conditions on the covariate history through time  $S_i$ , rather than the full covariate history. The time-varying covariate is “frozen” at each measurement time, as opposed to using information on  $\{t : t > S_i\}$ . There would typically be multiple event times for the same subject, each corresponding to a different measurement time. The authors approach this element of the data structure through a multivariate survival analysis framework analogous to Wei, Lin, and Weissfeld (1989). The method does not require modeling the longitudinal covariate process and there is no imposed dependence structure between different survival times from the same individual. With respect to related work, Van Houwelingen (2007) proposed a landmark model based on the partly conditional method. In this case, the time clock is not reset to zero every time a measurement is taken.

Each of the partly conditional methods described this far requires that censoring be independent of death time. However, dependent censoring frequently occurs in observational studies. A particular case is when survival in the absence of treatment (hereafter referred to as treatment-free survival) is of interest and both death and treatment assignment depend on the same time-varying covariates. If the model being fitted had conditioned on the entire history of the time-dependent covariates, then independent censoring could be assumed. However, since landmark methods freeze the time-varying measurement at the landmark time, hence only using part of the covariate history, dependent censoring can result due to the mutual correlation between future treatment assignment date, treatment-free death hazard, and the unmodeled portion of the covariate process occurring after the landmark time. Naturally, dependent censoring can result in greatly biased estimation.

As stated previously, the data which motivated our current research arise from the liver failure setting. The preferred method of treatment for liver failure is deceased-donor liver transplantation. There are thousands more patients awaiting liver transplantation than there are deceased-donor livers. As a result, patients requiring liver transplantation who are deemed medically suitable are placed on a liver transplant waiting list. Currently in the United States, the guiding principle in prioritizing patients for liver transplantation is that patients who are predicted to die the fastest without a transplant should be given the highest priority for transplantation. Patients with acute liver disease (known as Status 1 patients) are at the top of the wait list with highest priority for liver transplantation. They are followed by chronic liver failure patients, who are sequenced in decreasing order of MELD score. Note that MELD scores are updated over time, such that MELD is a time dependent process. In addition, if a patient becomes too sick, then he/she is removed from the transplant waiting list. Or, inactive status can be issued but possibly canceled afterward. During an inactive period, the patient is not eligible for transplantation and will not receive offers of deceased-donor livers. In practice, usually within the organ arrival day, doctors wish to allocate the donor liver to the patient who will die soonest without a transplant based on information up until, not beyond that calendar date. In reality, liver allocation uses updated MELD scores, which is similar to our model freezing the covariate. Therefore, although it is clear that liver transplantation censors pre-transplant death, such censoring amounts to dependent censoring in the context of a partly conditional model using landmark methods.

Note that, due to dependent censoring and the potential for subjects to become ineligible for inclusion in the sample, existing methods described previously (including existing partly conditional models and landmark methods) cannot be applied

directly to liver wait list data. Because the donor organs arrive in calendar time, it is more natural to choose cross sections based on calendar date as opposed to follow-up time.

In this report, we propose landmark methods featuring a partly conditional model to estimate the effect of a time-dependent covariate, in the presence of dependent censoring. Inverse Probability of Censoring Weighting (IPCW; Robins and Rotnitzky, 1992; Robins and Finkelstein, 2000) is used to obtain consistent estimators in the presence of dependent censoring. To increase precision, we propose two weight stabilizers that are different than those in the existing IPCW literature. Each landmark is based on a common calendar date, not follow-up time, consistent with the motivating example. At each cross-section (landmark) date, patients who are under observation (alive, uncensored), untreated, and treatment-eligible are included in the cross-section. Survival time, with respect to a cross section, is measured from the landmark date forward; such that the time clock is essentially reset to zero at each cross section date. We assume that the baseline hazards may differ by cross-section, such that a stratified Cox model (1972) is appropriate. Instead of modeling the time-varying covariate process on continuous time, multiple cross sections are chosen over the observation period to collect measurements from numerous discrete time points. The proposed methods do not require modeling of the longitudinal process.

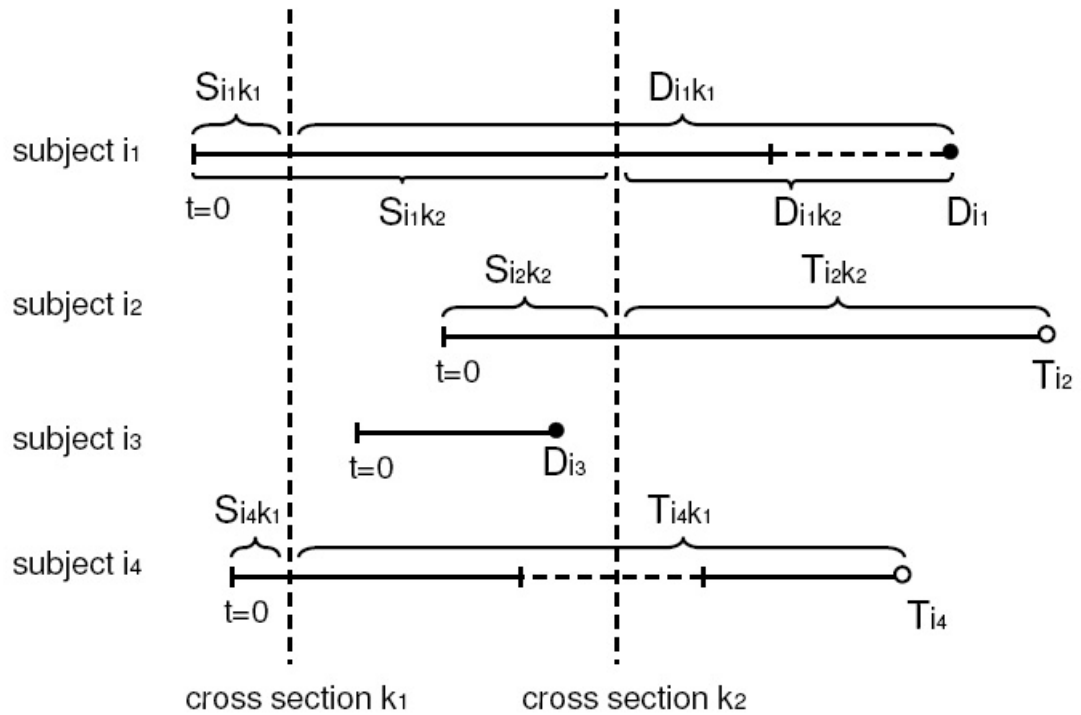
The remainder of this article is organized as follows. In Section 2.2, we formulate the previously described characteristics of the motivating data structure, then describe the proposed methods. Asymptotic properties of the proposed estimators are given in Section 2.3. A simulation study is provided in Section 2.4. Results of applying the proposed method to the afore-described liver failure data are presented in Section 2.5. In Section 2.6, we provide some concluding remarks and discussion.

Asymptotic derivations are provided in Appendix A.

## 2.2 Proposed Methods

We begin by setting up the required notation. Let  $D_i$  be the treatment-free time to failure for subject  $i$ , with  $i = 1, \dots, n$ . We assume that  $D_i$  may be censored at treatment time,  $T_i$ , or independent censoring time,  $C_i$ , and therefore we define the treatment-free observation time as  $X_i = \min(D_i, T_i, C_i)$ . We also define the associated indicators,  $\Delta_i = I(X_i = D_i)$  and  $\Delta_i^T = I(X_i = T_i)$ , where  $I(A) = 1$  when condition  $A$  is true and 0 otherwise. We define  $Y_i(t) = I(X_i \geq t)$ . We choose  $K$  cross section times  $CS_k$  with  $k = 1, \dots, K$ , where the cross-section times represent calendar dates. The  $\{CS_1, \dots, CS_K\}$  will typically be equally spaced. The number of cross-sections,  $K$ , will generally depend on the length of the study's observation period. Let  $S_{ik}$  denote the follow-up time of the  $i$ th subject on the calendar date of the  $k$ th cross section,  $CS_k$ . As we describe shortly, we will be modeling survival times from the cross-section dates, which essentially accounts to re-setting the time clock to 0 on each cross-section date. Correspondingly, since  $D_i$ ,  $T_i$  and  $C_i$  are measured in study time (i.e., time since subject  $i$  started follow-up), we then define death, treatment and censoring times (each measured as time post-cross-section) as follows:  $D_{ik} = D_i - S_{ik}$ ,  $T_{ik} = T_i - S_{ik}$  and  $C_{ik} = C_i - S_{ik}$ . Thus  $D_{ik}$ ,  $T_{ik}$  and  $C_{ik}$  are the death, treatment and censoring times, respectively, as measured from the  $k$ th cross section date. Let  $A_i(t)$  take value 1 if patient  $i$  is eligible to receive treatment as of follow-up time  $t$ , and 0 otherwise.

Figure 2.1 provides a graphical depiction of how each subject's treatment-free observation time is transformed into a set of time-since-cross-section dates. Four subjects are shown  $(i_1, i_2, i_3, i_4)$  and two cross sections,  $(k_1, k_2)$ . The four subjects



Note: Vertical dashed lines denote cross-section dates, while horizontal dashed lines denote treatment-ineligible period. Subject  $i_1$  has death times  $D_{i_1k_1}$  and  $D_{i_1k_2}$ , corresponding to cross sections  $k_1$  and  $k_2$ , respectively. Note that, even though subject  $i_1$  is not censored after becoming treatment-ineligible. Subject  $i_2$  is treated at time  $T_{i_2k_2}$ , and hence censored (perhaps dependently) at that time, with respect to cross section  $k_2$ . Subject  $i_3$  is not included in either cross section since  $i_3$  starts then finishes follow-up in between cross-sections. Subject  $i_4$  first passes cross section  $k_1$  and then becomes inactive for a while until a time after cross section  $k_2$ . Subject  $i_4$  is treatment-ineligible at cross section  $k_2$  and, therefore, is not included in this cross section. With respect to cross section  $k_1$ , the transplant time  $T_{i_4k_1}$  is not censored at the beginning of the treatment-ineligible period.

Figure 2.1: Examples of the relationship between cross-section time and follow-up time.



begin follow-up at times which are staggered in calendar time. Subject  $i_1$  has failure times  $D_{i_1k_1}$  corresponding to cross section  $k_1$  and  $D_{i_1k_2}$  with respect to cross section  $k_2$ . Note that even though subject  $i_1$  is deemed treatment-ineligible after cross section  $k_2$ , the subsequent death is not censored. Subject  $i_2$  is treated (and therefore censored) at time  $T_{i_2k_2}$  with respect to cross section  $k_2$ . Subject  $i_3$  is not included in either cross section since  $i_3$  starts and then finishes follow-up in between cross-sections. Subject  $i_4$  is included in cross section  $k_1$ , then becomes treatment-ineligible for a while until a time after cross section  $k_2$ . With respect to cross section  $k_1$ ,  $i_4$  is censored at treatment time  $T_{i_4k_1}$ , as opposed to earlier (at the beginning of the treatment-ineligible period). Since subject  $i_4$  is treatment-ineligible at cross section  $k_2$ ,  $i_4$  is not included in this cross section.

With respect to cross-section  $k$ , one observes a vector for subject  $i$ ,  $(X_{ik}, \Delta_{ik}, \Delta_{ik}^T)$ , where  $X_{ik} = \min(D_{ik}, T_{ik}, C_{ik})$ ,  $\Delta_{ik} = I(X_{ik} = D_{ik})$  and  $\Delta_{ik}^T = I(X_{ik} = T_{ik})$ . Note that, for a censored subject,  $\Delta_{ik} = \Delta_{ik}^T = 0$ . For ease of presentation, define  $A_{ik} = A_i(S_{ik})$ , an indicator for subject  $i$  being treatment-eligible at the time of the  $k$ th cross-section. We now set up a modified version of counting process notation. In particular, we let  $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)A_{ik}$ , and write  $dN_{ik}(t)$  for the increment  $N_{ik}\{(t+dt)^-\} - N_{ik}(t)$ . The at risk process is defined as  $Y_{ik}(t) = I(X_{ik} \geq t)A_{ik}$ ; i.e., in addition to subject  $i$  being alive and not treated as of time  $S_{ik}$  (i.e., that  $X_i > S_{ik}$ ), to be included in the  $k$ th cross section, it is also required that the subject is treatment-eligible at time  $S_{ik}$  (i.e., that  $A_{ik} = 1$ ). However, if  $A_{ik} = 1$ , subject  $i$  is not censored if he/she later becomes inactive at time  $t > S_{ik}$ . Thus, being treatment-eligible at time  $S_{ik}$  is a cross-section inclusion criterion, but subsequently becoming ineligible for treatment is not a censoring criterion. With respect to the treatment process, we define  $N_i^T(t) = I(X_i \leq t, \Delta_i^T = 1)$ .

Let  $Z_i(t)$  denote the covariate vector for subject  $i$  at follow-up time  $t$ ; and let  $Z_{ik} \equiv Z_i(S_{ik})$ , the covariate vector as measured at the  $k$ th cross section date,  $CS_k$ . The  $Z_{ik}$  vector would typically include baseline covariates, such as race and gender, and time-varying elements as of time  $S_{ik}$ . For notational simplicity, we write  $Z_{ik} = Z_i(S_{ik})$ , although, more generally  $Z_{ik}$  can be any function of the covariate history  $\{Z_i(s); s \in (0, S_{ik}]\}$ . The hazard function of interest can be expressed as:

$$\lambda_{ik}(t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P[t \leq D_{ik} < t + \delta | D_{ik} \geq t, Z_{ik}, A_{ik} = 1].$$

Note that the  $t$  argument pertains to time after the  $k$ th cross-section date, with the covariate “frozen” at its cross-section date value. The objective is to determine the relationship between the covariate (as known on the  $k$ th cross-section date) and future treatment-free survival time. Since the underlying goal is to determine what factors are associated with treatment urgency, only subjects who are treatment-eligible at the  $k$ th cross-section date are of interest; hence the conditioning on  $[A_{ik} = 1]$ .

Death times are modeled using stratified Cox regression,

$$(2.1) \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp\{\beta_0' Z_{ik}\},$$

where the baseline hazards are allowed to be cross-section-specific, although covariate effects are assumed to be equal across all cross-sections. We make the standard independent censoring assumption which, in the context of the observed data, is given by:

$$\lambda_{ik}(t | \tilde{Z}_i(S_{ik} + t), A_{ik} = 1, T_{ik} > t, C_{ik} > t) = \lambda_{ik}(t | \tilde{Z}_i(S_{ik} + t), A_{ik} = 1),$$

where  $\tilde{Z}_i(S_{ik} + t) = \{Z_i(s); s \in (0, S_{ik} + t)\}$ . However, a model for  $D_{ik}$  conditioning on  $Z_{ik}$  does not incorporate  $\{Z_i(r); r \in (S_{ik}, S_{ik} + t)\}$ . Further, it will generally be

the case that  $\lambda_{ik}(t) \neq \lambda_{ik}(t|Z_{ik}, A_{ik} = 1, T_{ik} > t)$  due to the correlation between  $T_{ik}$  and  $D_{ik}$  resulting from mutual dependence on  $\{Z_i(r); r > S_{ik}\}$ .

We use a variant of Inverse Probability of Censoring Weighting (IPCW) to overcome the dependent censoring of  $D_{ik}$  by  $T_{ik}$ . The model which takes treatment as the event is imposed, with hazard function

$$(2.2) \quad \lambda_i^T(t) = A_i(t)\lambda_0^T(t) \exp\{\theta'_0 Z_i(t)\},$$

where  $t$  is the time from study entry. As indicated in equation (2.2), the treatment hazard is zero at times during which the patient is treatment-ineligible. Therefore, treatment hazards among eligible patients are assumed to be proportional. The covariate in model (2.2) is written as  $Z_i(t)$  for notational convenience. More generally, the covariate could be any function of the covariate history  $\{Z_i(r); r \in (0, t)\}$ . The regression coefficient,  $\theta_0$ , is estimated by  $\hat{\theta}$ , as the root of the score function,

$$U^T(\theta) = \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(t; \theta)\} dN_i^T(t),$$

where  $\bar{Z}(t; \theta) = R_T^{(1)}(t; \theta)/R_T^{(0)}(t; \theta)$ ,  $R_T^{(p)}(t; \theta) = n^{-1} \sum_{i=1}^n A_i(t)Y_i(t)Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}$ ,  $p = 0, 1, 2$ , where, for a vector  $z$ ,  $z^{\otimes 0} = 1$ ,  $z^{\otimes 1} = z$ ,  $z^{\otimes 2} = zz'$ . The Breslow estimator of  $\Lambda_0^T(t)$  is given by  $\hat{\Lambda}_0^T(t) = n^{-1} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \hat{\theta})^{-1} dN_i^T(u)$ .

The IPCW method allows us to obtain consistent estimators by weighting each subject's experience by the inverse of (what can be thought of heuristically as) the probability of remaining untreated. In particular, the covariate effect,  $\beta_0$ , can be estimated as the root of the stratified inverse-weighted score function,

$$(2.3) \quad U(\beta, W) = \sum_{k=1}^K \sum_{i=1}^n \int_0^{\tau_k} A_{ik} \{Z_{ik} - \bar{Z}_k(t; \beta, W)\} W_{ik}^A(t) dN_{ik}(t),$$

where  $\tau_k$  satisfies  $P(X_{ik} \geq \tau_k) > 0$ , and the weight function is given by  $W_{ik}^A(t) = Y_{ik}(t) \exp\{\Lambda_i^T(S_{ik} + t) - \Lambda_i^T(S_{ik})\}$ ,  $\bar{Z}_k(t; \beta, W) = R_k^{(1)}(t; \beta, W)/R_k^{(0)}(t; \beta, W)$ , with

$R_k^{(p)}(t; \beta, W) = n^{-1} \sum_{i=1}^n A_{ik} W_{ik}^A(t) Y_{ik}(t) Z_{ik}^{\otimes p} \exp(\beta' Z_{ik})$  for  $p = 0, 1, 2$ . We refer to  $W_{ik}^A(t)$  as the Type A weight. Typically,  $\tau_k$  is set to the largest observation time from the  $k$ th cross section.

With some algebra,  $dN_{ik}(t)$  in (2.3) can be replaced by  $dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t)d\Lambda_{ik}(t)$ , such that the score function can also be written as

$$U(\beta, W) = \sum_{k=1}^K \sum_{i=1}^n \int_0^{\tau_k} A_{ik} \{Z_{ik} - \bar{Z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t).$$

A consistent estimator of  $\beta_0$  should satisfy  $E[W_{ik}^A(t)dM_{ik}(t)|Z_{ik}] = 0$ . We can write  $dM_{ik}(t) = I(T_{ik} > t)I(C_{ik} > t)dM_{ik}^*(t)$ , where  $dM_{ik}^*(t) = I(D_{ik} \geq t)[dN_{ik}^*(t) - d\Lambda_{ik}(t)]$  and  $N_{ik}^*(t) = I(D_{ik} \leq t)$ . Under the assumed model,  $E[dM_{ik}^*(t)|Z_{ik}] = 0$ , which leads to  $E[I(C_{ik} > t)dM_{ik}^*(t)|Z_{ik}] = 0$  in the case of independent censoring. In the presence of dependent censoring, without the IPCW term  $W_{ik}^A(t)$ ,  $E[dM_{ik}(t)|Z_{ik}] \neq 0$  since  $E[dM_{ik}(t)|Z_{ik}] \neq E[I(T_{ik} > t)I(C_{ik} > t)|Z_{ik}]E[dM_{ik}^*(t)|Z_{ik}]$ . However, it can be shown that

$$\begin{aligned} & E[W_{ik}^A(t)dM_{ik}(t)|Z_{ik}] \\ &= E[E[W_{ik}^A(t)dM_{ik}(t)|Z_{ik}, Z_i(S_{ik} + t)]|Z_{ik}] \\ &= E\left[\frac{E[I(T_i > S_{ik} + t)I(T_i > S_{ik})]}{\exp\{-\Lambda_i^T(S_{ik} + t) + \Lambda_i^T(S_{ik})\}} E[I(C_{ik} > t)dM_{ik}^*(t)|Z_{ik}, Z_i(S_{ik} + t)] \Big| Z_{ik}\right] \\ &= E[E[I(C_{ik} > t)dM_{ik}^*(t)|Z_{ik}, Z_i(S_{ik} + t)]|Z_{ik}] \\ &= E[I(C_{ik} > t)dM_{ik}^*(t)|Z_{ik}] = 0, \end{aligned}$$

such that the weighted score function, (2.3), has mean 0.

However, in practice, some values of  $W_{ik}^A(t)$  could be very large due to large values of  $\Lambda_i^T(t)$ . As a result, such unstable weights could lead to imprecise estimation of  $\beta_0$ . Similar to Robins and Finkelstein (2000) and Miloslavsky et al. (2004), to make the estimator more precise, a stabilizer is incorporated into the weight function. We

explore two versions of the stabilized weight. The first, which we refer to as the Type B weight, is given by

$$(2.4) \quad W_{ik}^B(t) = Y_{ik}(t) \frac{\exp\{\Lambda_i^T(S_{ik} + t)\}}{\exp\{\Lambda_i^T(S_{ik})\} \exp\{\Lambda_{ik}^T(t)\}},$$

where  $\Lambda_{ik}^T(t)$  is the cumulative hazard from a proportional hazard model which takes transplant as the event,  $\lambda_{ik}^T(t) = \lambda_{0k}^T(t) \exp\{\theta_1' Z_{ik}\}$  and  $t$  is the time from entry of study with  $\theta_1$  as covariate effect. For the Type B weight, the time intervals covered by stabilizer and Type A estimator are the same, such that  $W_{ik}^B(t)$  may serve to reduce the variability in the estimator of  $\beta_0$ . Another weight we evaluate is more intuitive and yields the Type C weight,

$$(2.5) \quad W_{ik}^C(t) = Y_{ik}(t) \exp\{\Lambda_i^T(S_{ik} + t)\},$$

which tracks the history from entry to study up to time  $t$ , without conditioning on the fact that  $T_i > S_{ik}$ . Since the stabilizing components of both  $W_{ik}^B(t)$  and  $W_{ik}^C(t)$  only based on the covariate history up to the  $k$ th cross section time,  $\{Z_{ik}\}$ , it can be shown that  $E[W_{ik}^B(t) dM_{ik}(t) | Z_{ik}] = 0$  and  $E[W_{ik}^C(t) dM_{ik}(t) | Z_{ik}] = 0$ , such that the weighted score function (2.3) still has mean 0.

## 2.3 Asymptotic Properties

We assume that the random vectors  $\{X_i, \Delta_i, \Delta_i^T, \tilde{Z}_i(X_i), \tilde{A}_i(X_i)\}$ , for  $i = 1 \dots n$ , are independent and identically distributed, with  $Z_i(t)$  bounded for  $t \in (0, \tau]$ , where  $\tau$  satisfies  $P(X_i \geq \tau) > 0$ . We summarize the asymptotic properties of the proposed methods in the following theorem. The regularity conditions are listed in Appendix A.

**Theorem II.1.** *Under certain regularity conditions,  $n^{1/2}(\hat{\beta} - \beta_0)$  converges asymptotically to a zero-mean Gaussian process with covariance function  $E[\varphi_i \varphi_i']$ , where*

$\{\varphi_1, \dots, \varphi_n\}$  are i.i.d. with mean 0 asymptotically, with

$$\begin{aligned} \varphi_i = & \Omega(\beta_0)^{-1} \left[ \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta_0, W)\} W_{ik}^A(t) dM_{ik}(t) \right. \\ & \left. + H'(t; \beta_0, W) \Omega_T(\theta_0)^{-1} U_i^T(\theta_0) + \int_0^{\tau} G(t, \tau; \beta_0) r_T^{(0)}(t; \theta_0)^{-1} dM_i^T(t) \right], \end{aligned}$$

where

$$\begin{aligned} dM_i^T(t) &= dN_i^T(t) - Y_i(t) \exp\{\theta' Z_i(t)\} \Lambda_0^T(t), \\ \bar{z}_k(t; \beta, W) &= r_k^{(1)}(t; \beta, W) / r_k^{(0)}(t; \beta, W), \\ \bar{z}(t; \theta) &= r_T^{(1)}(t; \theta) / r_T^{(0)}(t; \theta), \\ r_k^{(p)}(t; \beta, W) &= E[A_{ik} W_{ik}^A(t) Y_{ik}(t) Z_{ik}^{\otimes p} \exp(\beta' Z_{ik})], p = 0, 1, 2, \\ r_T^{(p)}(t; \theta) &= E[Y_i(t) Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}], p = 0, 1, 2, \end{aligned}$$

with  $\Omega(\beta)$ ,  $H(t; \beta, W)$ ,  $\Omega_T(\theta)$ ,  $U_i^T(\theta)$  and  $G(t_1, t_2; \beta)$  defined in Appendix A.

The covariance can be estimated consistently by  $n^{-1} \sum_{i=1}^n \widehat{\varphi}_i \widehat{\varphi}_i'$ , where  $\widehat{\varphi}_i$  is obtained by replacing all limiting values in  $\varphi_i$  by their empirical counterparts. A proof of Theorem II.1 is provided in Appendix A. The proof proceeds by demonstrating that, asymptotically,  $n^{1/2}(\widehat{\beta} - \beta_0) = n^{-1/2} \sum_{i=1}^n \varphi_i + o_p(1)$  through a sequence of Taylor series expansions.

The proof is given for the Type A weight,  $W_{ik}^A(t) = Y_{ik}(t) \exp\{\Lambda_i^T(S_{ik} + t) - \Lambda_i^T(S_{ik})\}$ . In practice, a stabilized version would usually be preferred. As implied by Theorem II.1, the computation of the variance is quite complicated, and is more complicated with the Type B weight. Such considerations motivate a computationally simpler form for the variance estimator. That is, to take the weight  $W_{ik}^A(t)$ , or  $W_{ik}^B(t)$  or  $W_{ik}^C(t)$ , as the case may be as fixed. Then, the variance estimator simplifies to  $n^{-1} \sum_{i=1}^n \widehat{\varphi}_i^* \widehat{\varphi}_i^{*'}$ , where

$$(2.6) \quad \widehat{\varphi}_i^* = \widehat{\Omega}(\widehat{\beta})^{-1} \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{Z}_k(t; \widehat{\beta}, \widehat{W})\} \widehat{W}_{ik}(t) d\widehat{M}_{ik}(t).$$

This simplified variance estimator can be calculated using Cox regression software that allows weights and a robust variance estimator; e.g., `proc phreg` in SAS, `coxph` in R.

## 2.4 Simulation

We modify the algorithm developed by Zheng and Heagerty (2005) to generate data which follow a partly conditional proportional hazards model. We first generate a binary treatment group indicator,  $Z_{ia}$ , taking values 0 and 1 with probability 0.5. A longitudinal marker  $Z_i(S_{ik})$  measured at a common set of cross section dates  $CS_1, CS_2, \dots, CS_K$  is constructed. To generate data  $[D_i, Z_{ia}, Z_{ib}]$  where  $Z_{ib} = \text{vec}\{Z_i(S_{ik})\}$ , we first create  $Z_{ib0} = b_i + \sum_{k=1}^K \log(V_{ik})/\gamma_2$ , where  $b_i \sim N(\mu, \sigma^2)$  and  $V_{ik} \sim P(\rho)$ , independent positive stable random variables with index  $\rho$  (Samoridnitsky and Taqqu, 1994). A pre-treatment death time  $D_i$ , is then generated with hazard  $\lambda_i(t) = V_{i0}^{1/\rho} \lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_{ib0}\}$ , where  $V_{i0} \sim P(\rho)$  and is independent of  $V_{ik}$ , with  $\Lambda_0(t) = (t/a)^{1/\rho^2}$  and  $a$  is a constant. Let  $Z_i(S_{ik}) = Z_{ib0} - \log(V_{ik})/\gamma_2$ . Then the death hazard can be written as

$$(2.7) \quad \lambda_i(t) = V_{i0}^{1/\rho} \lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik}) + \log(V_{ik})\}.$$

Subject  $i$  enters the study on calendar date  $L_i$ , where  $L_i$  is a *Uniform*(0,  $b$ ) variate. Treatment time,  $T_i$ , is generated from the proportional hazards model,

$$(2.8) \quad \lambda_i^T(t) = \lambda_0^T(t) \exp\{\theta_{01} Z_{ia} + \theta_{02} I(t > R_i)\},$$

where  $\lambda_0^T(t) = d_3$ ,  $\theta_0 = (\theta_{01}, \theta_{02})$ , and  $R_i$  is time of treatment-ineligibility which is generated with hazard  $\lambda_i^R(t) = \lambda_0^R(t) \exp\{d_1 V_{i0}\}$ , with  $\lambda_0^R(t) = 1/d_2$ . Thus,  $R_i$  and  $D_i$  are positively correlated, which is a reflection of the data which motivated the

proposed methods. Note that treatment time and pre-treatment death time,  $T_i$ , and  $D_i$ , are dependent since both depend on time of treatment ineligibility  $R_i$ .

To see that the prescribed set-up yields proportional hazards, integrating both sides of model (2.7), gives

$$\Lambda_i(t) = V_{i0}^{1/\rho} \Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{ik},$$

such that the pre-treatment survival function is given by

$$\exp[-\Lambda_i(t)] = \exp[-\Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{ik} V_{i0}^{1/\rho}].$$

Transforming the time scale to reflect time since cross section, define  $t_k = t - S_{ik}$ .

Then, take the expectation with respect to  $V_{ik}$  first and using the properties of the positive stable distribution, we have

$$\exp[-\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}, V_{i0})] = \exp[-(\Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{i0}^{1/\rho})^\rho / \cos(\pi\rho/2)].$$

Then, taking the expectation with respect to  $V_{i0}$ , we have

$$\exp[-\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik})] = \exp[-(\Lambda_0(t)^\rho \exp\{\rho^2 \gamma_1 Z_{ia} + \rho^2 \gamma_2 Z_i(S_{ik})\} / \cos(\pi\rho/2)^{(\rho+1)})],$$

which implies the following equation after taking logarithm and negative of both sides

$$\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \Lambda_0(t)^\rho \exp\{\rho^2 \gamma_1 Z_{ia} + \rho^2 \gamma_2 Z_i(S_{ik})\} / \cos(\pi\rho/2)^{(\rho+1)}.$$

Differentiating with respect to  $t_k$ ,

$$\lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \frac{\lambda_0(t_k + S_{ik}) \rho^2 \{\Lambda_0(t_k + S_{ik})\}^{(\rho^2-1)}}{\cos(\pi\rho/2)^{(\rho+1)}} \exp\{\rho^2 \gamma_1 Z_{ia} + \rho^2 \gamma_2 Z_i(S_{ik})\}.$$

Using this construction, the hazard for  $D_{ik} = D_i - S_{ik}$  will generally depend on  $S_{ik}$  and therefore stratified models similar to those considered by (Wei, Lin, and



Weissfeld, 1989) would be appropriate. With  $\Lambda_0(t) = (t/a)^{1/\rho^2}$ ,  $\lambda_0(t_k + S_{ik})\rho^2\{\Lambda_0(t_k + S_{ik})\}^{(\rho^2-1)} = 1/a$ , we obtain

$$\lambda_i(t_k|Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \exp\{\rho^2\gamma_1 Z_{ia} + \rho^2\gamma_2 Z_i(S_{ik})\}/[a \cos(\pi\rho/2)^{(\rho+1)}].$$

If we define  $\lambda_{ik}(t; S_{ik}) = \lambda_i(t_k|Z_{ia}, Z_i(S_{ik}), D_i > S_{ik})$ ,  $\lambda_{0k}(t) = [a \cos(\pi\rho/2)^{(\rho+1)}]^{-1}$  and  $\beta_0 = (\beta_{01}, \beta_{02}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ , then the proportional hazard model on treatment-free survival is achieved,

$$(2.9) \quad \lambda_{ik}(t; S_{ik}) = \lambda_{0k}(t) \exp\{\beta_{01}Z_{ia} + \beta_{02}Z_i(S_{ik})\}.$$

After generating the data, we only include for analysis those  $Z_i(S_{ik})$  with  $L_i < S_{ik} < \min(X_i, R_i)$ . Data pertaining to survival time since cross section  $\{X_{ik}, \Delta_{ik}, Z_{ia}, Z_i(S_{ik})\}$  is used to fit model (2.9), with time to transplant data  $\{X_i, \Delta_i^T, Z_{ia}, Z_i(t)\}$  used to fit model (2.8).

We evaluate samples with  $n = 1000$  subjects and obtain 10%, 20%, and 40% censoring by varying  $a$  from  $10^4$  to  $4 \times 10^7$ . The value of  $d_2$  varies from 300 to 3000, resulting in ineligibility rates from 10% to 30%. There are  $K = 10$  cross section dates. We set  $b = 500$ ,  $[\theta_{01}, \theta_{02}] = [-1, -1]$ ,  $\mu = 18$ ,  $\sigma = 1$ ,  $[\gamma_1, \gamma_2] = [-1, -0.5]$ ,  $[-0.5, -0.25]$ ,  $[0, 0]$ ,  $d_1 = d_3 = 0.001$ , with  $CS_k = 100 \times k$ . For all our simulated situations, 1000 Monte Carlo data sets are used. We present results using  $\rho = 0.8$ , thus  $[\beta_{01}, \beta_{02}] = [-0.64, -0.32]$  when  $[\gamma_1, \gamma_2] = [-1, -0.5]$ . With number of cross sections set to  $K = 10$ , the average number of cross sections per subject is 0.7 to 2.4, depending on the censoring level. We apply the simplified variance estimate which treats the estimated weights as fixed; i.e., as given in (2.6).

Table 2.1 presents simulation results based on Type A weight, while Tables 2.2 and 2.3 present results for Type B and Type C, respectively. Estimates of  $\beta_0$  appear to be consistent based on all weights. The variance of the Type B estimator is smaller

than that of Type A, which is likely the result of the added stabilizer. Coverage probabilities using the proposed (simplified) variance estimator are close to the nominal 95% level, with those of the Type B estimator being slightly higher than those of Type A. It appears that the variance of the Type C estimator is greater than that of Type A. This could result from the noise added by incorporating history before the cross section time in  $W_{ik}^C(t)$ . The added piece covers a different time interval and thus does not actually function as a stabilizer. Coverage probabilities for the Type C estimator are slightly lower than those of Type A.

## 2.5 Application

We applied the proposed methods in order to compare pre-transplant mortality between acute and chronic End-Stage Liver Disease (ESLD). The study population included patients initially wait listed for deceased-donor liver transplantation between March 1, 2002 and December 31, 2009 in United States. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR), a national, population based organ transplant registry. Only patients age  $\geq 18$  at listing and not previously transplanted (i.e., not repeat transplant candidates) were included in the study population. Cross-section dates were chosen every 7 days from 03/01/2002 to 12/31/2009, such that there were  $K = 409$  cross sections in all. At any given cross section date, any subject who was still on the wait-list (not inactive and not removed) was included in the cross section since, in practice, patients who got removed or were made inactive were no longer eligible to receive offers for deceased-donor livers. Given the objectives of our analysis, it is appropriate to compare only patients who, in a given cross-section date, are in fact eligible to receive a liver transplant. However, after being included into a given cross-section, such patients are not censored if they are

Table 2.1: Simulation results for  $\hat{\beta}$  computed using Type A Weight;  $n = 1000$ .

Censored	$\hat{\beta}_{01}$				$\hat{\beta}_{02}$				
	$\beta_{01}^*$	Bias	ESE	CP	$\beta_{02}^*$	Bias	ESE	CP	
10%	-0.64	-0.009	0.132	0.125	0.93	-0.32	-0.001	0.020	0.95
20%		-0.004	0.143	0.128	0.92		-0.002	0.020	0.94
40%		-0.008	0.145	0.129	0.93		0.001	0.018	0.94
10%	-0.32	0.002	0.146	0.132	0.93	-0.16	0.001	0.013	0.93
20%		-0.004	0.140	0.129	0.92		-0.001	0.010	0.94
40%		-0.001	0.144	0.130	0.92		-0.001	0.010	0.94
10%	0	0.001	0.140	0.130	0.93	$-10^{-4}\ddagger$	-0.003	0.048	0.94
20%		-0.002	0.136	0.127	0.94		-0.001	0.042	0.95
40%		-0.007	0.143	0.128	0.93		-0.002	0.038	0.94

\* $\beta_0 = (\beta_{01}, \beta_{02}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ , where  $\rho = 0.8$ .†The Bias, ESE and ASE shown in this block are in  $10^{-4}$  scale.

Table 2.2: Simulation results for  $\hat{\beta}$  computed using Type B Weight given in (2.4);  $n = 1000$ .

	$\hat{\beta}_{01}$				$\hat{\beta}_{02}$						
	Censored	$\beta_{01}^*$	Bias	ESE	ASE	CP	$\beta_{02}^*$	Bias	ESE	ASE	CP
10%	-0.64	0.004	0.130	0.121	0.121	0.94	-0.32	-0.001	0.021	0.020	0.94
20%		-0.008	0.121	0.116	0.116	0.94		-0.001	0.018	0.018	0.95
40%		-0.008	0.113	0.112	0.112	0.94		-0.001	0.016	0.016	0.95
10%	-0.32	-0.005	0.136	0.130	0.130	0.94	-0.16	-0.001	0.012	0.011	0.94
20%		-0.005	0.122	0.117	0.117	0.94		-0.001	0.010	0.009	0.94
40%		0.003	0.112	0.109	0.109	0.93		-0.001	0.008	0.008	0.94
10%	0	0.005	0.135	0.126	0.126	0.93	$-10^{-4}\ddagger$	-0.001	0.044	0.043	0.94
20%		-0.005	0.123	0.118	0.118	0.94		-0.002	0.041	0.038	0.94
40%		0.004	0.109	0.109	0.109	0.95		-0.002	0.031	0.031	0.95

\* $\beta_0 = (\beta_{01}, \beta_{02}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ , where  $\rho = 0.8$ .†The Bias, ESE and ASE shown in this block are in  $10^{-4}$  scale.

Table 2.3: Simulation results for  $\hat{\beta}$  computed using Type C Weight given in (2.5);  $n = 1000$ .

Censored	$\hat{\beta}_{01}$				$\hat{\beta}_{02}$					
	$\beta_{01}^*$	Bias	ESE	CP	$\beta_{02}^*$	Bias	ESE	CP		
10%	-0.64	0.002	0.142	0.129	0.93	-0.32	-0.002	0.022	0.021	0.94
20%		-0.005	0.144	0.136	0.93		-0.001	0.022	0.020	0.92
40%		-0.010	0.159	0.136	0.90		-0.002	0.020	0.019	0.94
10%	-0.32	-0.003	0.146	0.135	0.94	-0.16	-0.001	0.012	0.012	0.94
20%		-0.012	0.154	0.138	0.91		-0.001	0.012	0.011	0.92
40%		-0.003	0.157	0.138	0.91		-0.001	0.011	0.010	0.94
10%	0	-0.008	0.153	0.136	0.92	$-10^{-4\dagger}$	-0.002	0.047	0.046	0.94
20%		-0.001	0.146	0.134	0.92		-0.002	0.045	0.043	0.94
40%		0.007	0.155	0.137	0.92		-0.002	0.042	0.040	0.93

\* $\beta_0 = (\beta_{01}, \beta_{02}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ , where  $\rho = 0.8$ .†The Bias, ESE and ASE shown in this block are in  $10^{-4}$  scale.

subsequently deactivated or removed from the wait-list. Deactivation and removal (and the associated death that may follow) are potential consequences of not receiving a liver transplant. For the death model, the failure time was defined from date of cross section to the date of death, transplant or censoring whichever occurred first.

In order to construct the IPCW weight,  $\Lambda_i^T(t)$  was estimated based on a time-dependent Cox model in which transplant was the event. For the time-to-transplant model, time  $t$  starts from the beginning of the follow-up (the date of wait listing), as opposed to cross section time. The model was stratified, such that

$$\lambda_{ir}^T(t) = A_i(t)\lambda_{0r}^T(t) \exp\{\theta'_0 Z_i(t)\},$$

where the subscript  $r = 1, \dots, 11$  stands for region. The presence of the indicator,  $A_i(t)$ , reflects the fact that a patient's time while inactive or removed does not contribute to the estimation of  $\theta_0$  or  $\Lambda_{0r}^T(t)$ . The patient level covariate,  $Z_i(t)$ , included MELD score, Status 1, albumin, age, gender, race, diagnosis of Hepatitis C, body mass index, diabetes, hospitalization, blood type, dialysis within prior week, encephalopathy, ascites and serum creatinine. We evaluated several different versions of weight, including  $W_{ikr}(t) = Y_{ikr}(t)$  (unweighted),  $W_{ikr}^A(t) = Y_{ikr}(t) \exp\{\Lambda_{ikr}^T(t + S_{ik}) - \Lambda_{ikr}^T(S_{ik})\}$ ,  $W_{ikr}^B(t)$  and  $W_{ikr}^C(t)$ . Even for the stabilized weight, some very large values occurred. Since we found that 99% of weights were less than 10, weights were then capped at 10.

The model of primary interest, the pre-transplant death model, was also stratified

$$\lambda_{ikr}(t) = \lambda_{0kr}(t) \exp\{\beta'_0 Z_{ik}\},$$

where the subscript  $r = 1, \dots, 11$  stands for region and  $k = 1, \dots, 409$  stands for cross section. The subject level covariates at cross section  $k$ ,  $Z_{ik}$ , included MELD score (21-23, 24-26, 27-29, 30-32, 33-35, 36-40), Status 1 (as the reference, to which all MELD

Table 2.4: Analysis of liver wait-list mortality (using Type B Weight)

Group	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$e^{\hat{\beta}}$	p-value
Status 1	0	.	1	.
MELD 21-23	0.05	0.267	1.05	0.87
MELD 24-26	0.18	0.272	1.20	0.50
MELD 27-29	0.52	0.276	1.68	0.06
MELD 30-32	0.25	0.334	1.29	0.45
MELD 33-35	0.96	0.301	2.62	0.001
MELD 36-40	0.95	0.306	2.58	0.002

categories are compared), albumin, age, gender, race, diagnosis, body mass index, diabetes, hospitalization status at listing and previous malignancy. Also included in  $Z_{ik}$  were average change in MELD score (pertaining to the time interval between the date of listing and cross-section  $k$  date, and estimated using ordinary least squares) and average change in albumin (estimated analogously). Other elements included the percentage of time spent in inactive status, and percent of time receiving dialysis. Since 99% of MELD and albumin slope values before cross sections fell in the  $[-1,1]$  interval, the slopes were bounded by -1 and 1.

We focused on comparing each MELD category  $> 20$  with Status 1 (reference). Our final sample consisted of  $n = 23,657$  patients.

Results based on the Type B weight are listed in Table 2.4. MELD group 33-35 has a partly conditional pre-transplant death hazard of 2.62 ( $p = 0.001$ ) times that of Status 1; the corresponding hazard ratio for MELD 36-40 is 2.58 ( $p = 0.002$ ).

Both unweighted and weighted results are listed in Table 2.5. After weighting the

Table 2.5: Analysis of liver wait-list mortality; Comparison of results using different weight.

Group	Unweighted		Type C Weight		Type A Weight		Type B Weight	
	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$
Status 1	0	.	0	.	0	.	0	.
MELD 21-23	-0.81	0.210	0.01	0.271	0.07	0.270	0.05	0.267
MELD 24-26	-0.75	0.215	-0.002	0.286	0.11	0.281	0.18	0.272
MELD 27-29	-0.29	0.220	0.31	0.287	0.42	0.283	0.52	0.276
MELD 30-32	-0.32	0.256	0.10	0.348	0.11	0.339	0.25	0.334
MELD 33-35	0.26	0.246	0.91	0.345	0.92	0.321	0.96	0.301
MELD 36-40	0.33	0.272	0.79	0.335	0.73	0.324	0.95	0.306

model, the parameter estimates of MELD group became larger, in each case. Similar to the findings from simulation studies, the standard errors in Table 2.5 were the lowest for the Type B weight, while those for Type C were the largest.

It is interesting that MELD (i.e., chronic liver disease) patients had higher mortality than Status 1 (i.e., acute liver disease) patients, both from a clinical and from a public health perspective. Supplementary analysis revealed that if only based on cross section status, acute patients died very fast in the early stage, i.e., the Status 1 Kaplan-Meier curves dropped more quickly than MELD groups. However, Status 1 survival leveled off eventually, while the survival curves for the MELD groups kept dropping.



## 2.6 Discussion

In this chapter, we propose semiparametric methods to estimate the effect of a time dependent covariate in the presence of a possibly censored survival time. Treatment-free survival is of our interest, and pre-treatment death may be dependently censored by receipt of treatment. Subjects can also experience permanent or temporary periods of ineligibility for receiving treatment. Landmark analysis with a partly conditional hazard model are combined in the proposed methods. Multiple cross-section dates are chosen, with subjects included in the sample corresponding to a cross-section if they are alive uncensored, treatment-free, and treatment-eligible. Survival time is measured from a given cross-section date forward, rather than from the start of follow-up. Covariates values are “frozen” at cross-section dates and not updated afterward, which is the reason why the method is called partly conditional. However, fixing the covariate values on cross-section dates results in dependent censoring, since death and treatment times are independent only if conditioning on the whole time-dependent covariate history. A modified version of Inverse Probability of Censoring produces consistent estimators in the presence of dependent censoring. Stabilized versions of the weights are used to estimate the covariate effect more precisely. The same subject could have multiple survival times with respect to different cross-section dates, and no specific covariance structure is imposed for such times. A stratified IPCW weighted proportional hazards model is applied with each cross-section date data serving as a stratum. The proposed methods circumvent modeling of the longitudinal covariate process explicitly. The model is in part marginal, since fixing time-varying covariate values at cross-section dates could be seen as averaging mortality hazard function implicitly across future covariate trajectories. The pro-

posed estimators are demonstrated to be consistent and asymptotically normal, with consistent covariance estimators provided. Through simulation studies, the proposed estimator is revealed to be appropriate for practical use. We applied the proposed methods to wait-list mortality among end-stage liver disease patients.

Zheng and Heagerty (2005) proposed partly conditional Cox regression methods. Van Houwelingen (2007) proposed landmark models based on partly conditional methods. In Zheng and Heagerty (2005), the time clock is re-set at covariate measurement times, unlike our methods, wherein the clock is re-set at cross-section dates. Neither the Zheng and Heagerty (2005) or Van Houwelingen (2007) methods deal with dependent censoring or accommodate treatment ineligibility.

Comparisons of pre-transplant death rates between Status 1 and MELD patients have rarely been conducted previously; largely because the assumption that Status 1 patients have the highest death rate is widely accepted by the liver transplant community. However, in a recent study using a traditional time-dependent model, death rates of high-MELD patients were shown to be comparable to those of Status 1. Specifically, (Sharma et al., 2012) compared the mortality of Status 1 and high-MELD patients using a time-dependent model to ESLD patients and a 14-day follow-up period. The most important result of their study was that  $\text{MELD} \geq 40$  patients have higher pre-transplant mortality than Status 1 patients. Results of our proposed methods show that  $\text{MELD} \geq 33$  is associated with significantly higher pre-transplant mortality than Status 1. Unlike traditional time-dependent models, the proposed method compares the mortality from a given cross-section date forward; inspired by the real-world situation in which an organ arrives (on a given calendar date) and one wishes to identify which patient would die next in the absence of liver transplantation, using only information up until that particular calendar date.

The methods in this paper are of easy implementation and computation. Multiple landmarks are chosen to make use of the data more fully. Too few cross sections will lead imprecision, while too many cross-sections may lead to computation difficulties. Sampling cross section dates or sampling subjects within each cross section may be one solution. Another problem is non-proportional hazards, in which case the ratio of cumulative hazards or the difference in restricted mean life time are possible measures by which to summarize the cumulative effect. Extensions of the proposed methods to estimate such quantities would be quite useful.

## CHAPTER III

# Estimating Average Treatment Effects on Mean Survival Time when Treatment is Time-dependent and Censoring is Dependent

### 3.1 Introduction

In clinical settings, subjects often begin follow-up untreated, with some going to receive the treatment at some time in the future. In various settings, patients may experience treatment ineligibility, temporarily or permanently. In the setting of our interest, receiving treatment is time-dependent, not randomized and depends on time-varying covariates for which serial data are measured. As a result, joint data on the longitudinal covariate and possibly censored survival time are observed simultaneously. The effect of the time-dependent treatment is of interest and could be quantified through mean survival time, which is easy to interpret and needs no assumption on the shape of survival curves, such as proportionality in the context of (Cox, 1972) model. Due of finite duration of study and unclosed survival curves, it is more practical to use restricted mean life time, which is the area under the survival curve from 0 up to a specific time  $L$ , as opposed to infinity.

When approached through a joint modeling approach, valid inference on the time-to-event component generally requires the longitudinal process be accurately

modeled, which is often very difficult to achieve. Zheng and Heagerty (2005) proposed a partly conditional model, in which the duration since measurement time  $S_i$ ,  $(D_i - S_i)$  is modeled, rather than the event time  $D_i$  from study entry. The time-varying covariate value is then “frozen” at time  $S_i$ . The resulting model is referred to as “partly conditional” since the hazard function only conditions on the covariate history on measurement up to time  $S_i$ , as opposed to the full information which includes covariate values for  $\{t : t > S_i\}$ . Therefore, rather than modeling the time-varying covariate explicitly to predict future covariate values, the partly conditional model implicitly averages over the possible paths. However, adjusting for only part of the covariate history causes treatment initiation to be dependent censoring for treatment-free death. Consistent estimation for parameters pertaining to the treatment-free death hazard can be obtained through Inverse Probability of Censoring Weighting (IPCW), proposed originally by Robins and Rotnitzky (1992); see also Robins and Finkelstein (2000). The IPCW method aims to essentially create pseudo-data free of dependent censoring through weighting. Of our chief interest, the average effect of treatment on the treated, is a weighted average of subject-specific treatment effect estimates, with each observed treatment weighted by inverse of the probability of remaining uncensored (Horvitz and Thompson, 1952).

The motivation of our research is the liver transplant setting. The number of available deceased-donor livers is always less than the number of patients awaiting liver transplantation. As a result, patients who are medically suitable are placed on a deceased-donor liver transplant waiting list. Patients typically begin follow-up on the wait list (untreated), such that transplantation is a time-dependent treatment. At the top of the list are acute liver disease (Status 1) patients, followed by chronic liver failure patients sequenced in decreasing order of Model for End-Stage Liver Disease

(MELD) score, a very strong predictor of pre-transplant mortality. Transplantation results in dependent censoring to pre-transplant death since MELD scores predict both wait list mortality and transplant rates. Note that patients may be removed from the wait list, or made inactive and, in such cases, are per permanently or temporarily ineligible to receive a transplant.

Various methods in the causal inference literature have estimated the average causal effect through the hazard ratio based on a marginal structural model (e.g., Hernan, Brumback, and Robins, 2000, 2001; Robins, Hernan, and Brumback, 2000). Other authors estimated treatment effect via restricted mean survival time (Chen and Tsiatis, 2001; Schaubel and Wei, 2011; Zhang and Schaubel, 2010). But very few methods have contrasted pre- and post-treatment mean survival time when treatment is time-dependent. Due to the complexity caused by the time-dependent treatment, the potential for patients to become ineligible for treatment consideration and dependent censoring, the afore-mentioned methods cannot be applied to our motivating data directly. Schaubel et al. (2009) proposed methods for estimating pre-treatment restricted mean lifetime. Of chief interest was the application to national liver failure data. Specifically, a patient-specific score was developed to replace the MELD score as the basis for ranking wait-listed patients. The focus of this work was on the fitted values. Asymptotic properties were not developed, nor were simulations conducted. Moreover, methods for estimating the average treatment effect were not proposed.

In this article, we propose methods for estimating the average effect on the treated of a time-dependent treatment by taking an appropriately weighted average of estimated subject-specific differences between restricted mean survival time, post- versus pre-treatment. In particular, for each observed treatment time, we project out a predicted post-treatment survival curve based on parameter estimates from post-

treatment survival model, which conditions on measurement history leading up to the treatment time. We also project a treatment-free survival curve, using fitted values from the cross-section-based partly conditional treatment-free survival model, which is analogous to the methods from Chapter II. Then, the areas under the survival curves up to a pre-specified truncation time (e.g., 5 years) represent the restricted mean lifetime estimates. The difference between post- and pre-treatment restricted mean lifetimes represent the subject level treatment effect estimate, and is computed for each observed treatment. At last, we then weight each observed treatment by the probability of remaining uncensored. This average is taken using normalized weights by IPCW (based on a model for independent censoring), such that the weighted time-until-transplant distribution represents that which would have been observed in the absence of independent censoring.

The remainder of this article is organized as follows. In Section 3.2, we set up the notation and describe the proposed methods. Asymptotic properties of the proposed estimators are listed in Section 3.3. A simulation study is provided in Section 3.4. Results of applying proposed methods to data obtained from the Scientific Registry of Transplant Recipients (SRTR) will be given in Section 3.5. Discussion is provided in Section 3.6. Details regarding the asymptotic derivations are given in Appendix B.

## 3.2 Proposed Methods

First, we define the necessary notation. Let  $D_i$  be the death time for subject  $i$ , with  $i = 1, \dots, n$ . We set  $T_i$  to be the treatment time for subject  $i$ . It is possible that  $D_i$  and/or  $T_i$  may not be observed due to independent censoring at time  $C_i$ . Observation time is then given by  $X_i = D_i \wedge C_i$ , where  $a \wedge b = \min(a, b)$ . Subjects

begin follow-up untreated, with some going to receive treatment unless they die first.

We define the observed data counting processes for death, censoring, and treatment as  $N_i(t) = I(D_i \leq t, D_i < C_i)$ ,  $N_i^C(t) = I(C_i \leq t, C_i < D_i)$ , and  $N_i^T(t) = I(T_i \leq t, T_i < D_i \wedge C_i)$  respectively. The covariate vector  $Z_i(t)$  contains both time-independent and time-dependent factors and perhaps some functions of  $t$  such as  $(t, t^2, t^3)$ . The covariate history on  $(0, t]$  is defined as  $\tilde{Z}_i(t) = \{Z_i(u); u \in (0, t]\}$ . The indicator for treatment eligibility is defined as  $A_i(t)$  which equals 1 when the  $i$ th subject qualifies for the treatment as of follow-up time  $t$  and 0 otherwise. Treatment eligibility history is represented by  $\tilde{A}_i(t) = \{A_i(u); u \in (0, t]\}$ . Let  $D_i^0$  be the death time for subject if treatment were unavailable. Consider a subject treated at time  $T_i$ , and define the death time in the presence of treatment as  $D_i^1$ . For a treated subject, the death time is necessarily post-treatment and the post-treatment survival time (since treatment) is given by  $D_i^1 - T_i$ . It is useful to consider when the treated subject would have died had, in fact, treatment not been received, and we denote this time by  $D_i^0 - T_i$ .

Suppose that treatment is assigned in calendar time. On any given day, we can consider the benefit of assigning treatment to specific patients. In particular, it would be useful to predict the patient's prognosis survival from that day forward, under two cases: (0) the subject does not receive treatment ever (1) the subject receives treatment starting that day. Then the difference between two survival predictions can be used to quantify the benefit of treatment.

Examining this setting more formally, suppose that patient  $i$  is being considered for treatment on a given calendar date. On that calendar date, the subject has previous follow-up time of  $S_i$  time units. As stated previously, only subjects who are alive and treatment eligible are going to receive the treatment. Suppose that subject



$i$  has covariate  $Z_i(S_i)$  and is eligible; i.e.,  $\{A_i(S_i) = 1, D_i > S_i, Z_i(S_i)\}$ . Of interest are both the restricted mean (residual) lifetime if subject  $i$  is treated at time  $S_i$ ,

$$\mu_{i1}(S_i) = E[\{D_i^1 - S_i\} \wedge L | D_i > S_i, A_i(S_i) = 1, Z_i(S_i), T_i = S_i],$$

versus the corresponding mean residual lifetime if the subject remained untreated,

$$\mu_{i0}(S_i) = E[\{D_i^0 - S_i\} \wedge L | D_i > S_i, A_i(S_i) = 1, Z_i(S_i)].$$

The treatment effect is then defined as,

$$\Delta_i(S_i) = \mu_{i1}(S_i) - \mu_{i0}(S_i).$$

Note that  $L$  is some specific length of time chosen in light of the available follow-up time. The arguments of  $\mu_{ij}(\cdot)$ , with  $j = 0, 1$ , and  $\Delta_i(\cdot)$  reflect that a specific follow-up time,  $S_i$ , and covariate vector,  $Z_i(S_i)$ , are being considered. Thus subject specific treatment effects,  $\Delta_i(S_i)$ , is predicted via  $\mu_{i1}(S_i)$  and  $\mu_{i0}(S_i)$ .

The overall treatment effect could be measured by averaging over the subject level estimate,  $\Delta = E[\Delta_i(S_i)]$ , where the expectation is with respect to  $\{T, Z(T)\}$ , the joint distribution of treatment time and covariate vector (at the time of treatment) among treated patients.

Instead of estimating the pertinent means directly, the following hazard functions are modeled,

$$\lambda_{i1}(t; S_i) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P\{t \leq D_i - S_i < t + \delta | D_i - S_i \geq t, A_i(S_i) = 1, Z_i(S_i), dN_i^T(S_i) = 1\},$$

$$\lambda_{i0}(t; S_i) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P\{t \leq D_i - S_i < t + \delta | D_i - S_i \geq t, A_i(S_i) = 1, Z_i(S_i), N_i^T(S_i + t) = 0\},$$

with  $\mu_{ij}(S_i) = \int_0^L \exp\{-\int_0^u \lambda_{ij}(t; S_i) dt\} du$ . Note that the conditioning specifically on  $A_i(S_i)$  and  $Z_i(S_i)$  is for ease of presentation. More generally, one could condition

on  $\tilde{A}_i(S_i)$  and  $\tilde{Z}_i(S_i)$ , the treatment-eligibility and covariate histories, respectively, if the application at hand required.

Note that the just-defined hazard functions are assumed to be independent of the actual receipt of treatment; i.e.,  $\lambda_{ij}(t; S_i | \{N_i^T(u); u \geq S_i\}) = \lambda_{ij}(t; S_i)$ . This no-unmeasured-confounders assumption is required in order to consistently estimate the quantities of interest via observed data. The observed-data version of the hazard function is given by  $\lambda_i(t; S_i) = \lambda_{i1}(t; S_i) dN_i^T(S_i) + \lambda_{i0}(t; S_i) \{1 - N_i^T(S_i + t)\}$ . We assume that  $C_i$  is independent censoring in the sense that  $\lambda_i(t; S_i) = \lambda_i(t; S_i | C_i > S_i + t)$  and that  $T_i$  is also independently censored by  $C_i$ . With these assumptions,  $\lambda_{i1}(t; S_i) = \lambda_{i1}(t; S_i | dN_i^T(S_i) = 1)$ , which implies that one can validly model the post-treatment hazards, using only the observed treatments.

We now turn our attention to the model for  $\lambda_{i0}(t; S_i)$ . Since the hazard function explicitly conditions on  $Z_i(S_i)$ , and since  $\{Z_i(u); u > S_i\}$  may affect both the treatment-free death hazard and the treatment hazard at time  $\{u > S_i\}$ , treatment-free death is dependently censored by  $T_i$ . We overcome this issue using IPCW, which requires a model for

$$\lambda_i^T(t; Z_i(t)) = \lim_{\delta \downarrow 0} \delta^{-1} P\{t \leq T_i < t + \delta | D_i \wedge T_i \geq t, A_i(t), Z_i(t)\}.$$

Note that the conditioning is on  $Z_i(t)$  and  $A_i(t)$  instead of  $Z_i(S_i)$  and  $A_i(S_i)$ . The treatment time hazard is assumed to follow a time-dependent model, with time axis  $t$  (follow-up time) as opposed to time since cross section  $S_i$ . We also assume that  $\lambda_i^T(t) = \lambda_i^T(t; Z_i(t), D_i)$ , which allows estimation of  $\lambda_{i0}(t; S_i)$  based on pre-transplant data in the presence dependent censoring at  $T_i$  through IPCW.

We assume the following proportional hazards models for pre- and post-treatment

respectively,

$$\lambda_{ij}(t; S_i) = \lambda_{0j}(t) \exp\{\beta'_j Z_i(S_i)\}, j = 0, 1,$$

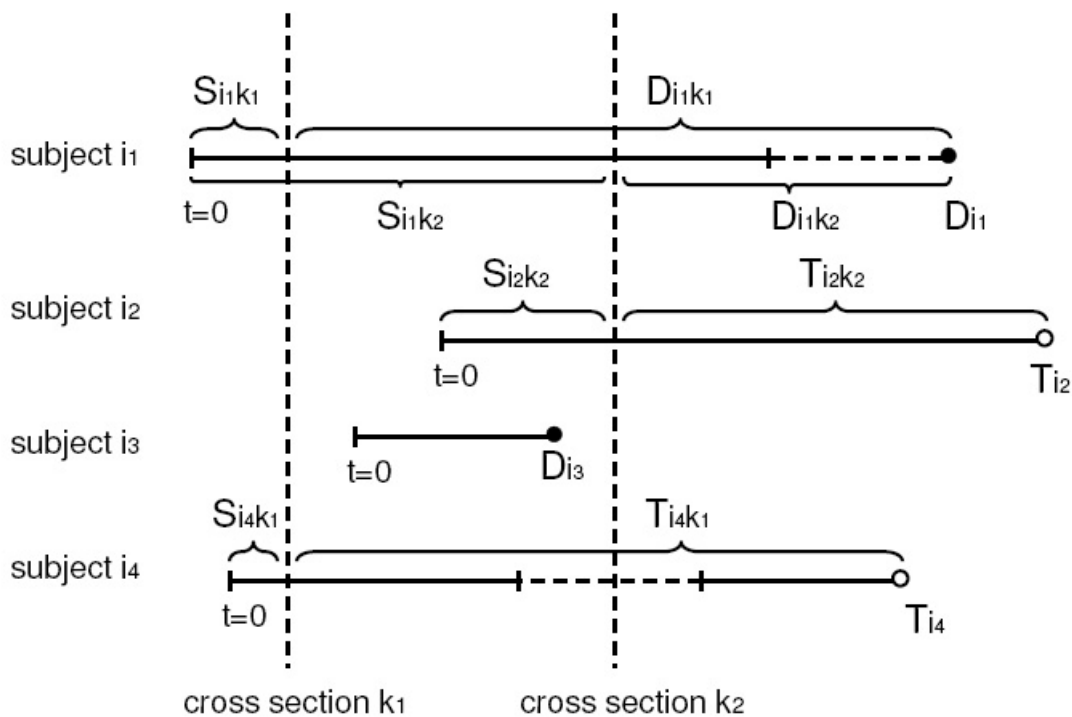
where  $\lambda_{0j}(t)$ , are unspecified baseline hazards,  $\beta_j$  are unknown regression coefficients; and  $j = 0$  is with respect to pre-transplant and  $j = 1$  refers to post-treatment.

We now describe how to fit the pre-treatment death model,

$$\lambda_{i0}(t; S_i) = \lambda_{00}(t) \exp\{\beta'_0 Z_i(S_i)\}.$$

We choose  $K$  calendar date based cross-section times  $CS_k$  with  $k = 1, \dots, K$ . The  $\{CS_1, \dots, CS_K\}$  can be equally spaced; but, this is not required. Suppose  $S_{ik}$  is the follow-up time for the  $i$ th patient at the  $k$ th cross-section date,  $CS_k$ . Since we will model the survival time since cross-section, the time clock is essentially reset to 0 at each cross section date. Subject  $i$  is included in the cross-section sample at  $CS_k$  if alive, untreated, uncensored and treatment-eligible; i.e., if  $T_i \wedge C_i \wedge D_i > S_{ik}$  and  $A_i(S_{ik}) = 1$ . Since survival time from cross-section is modeled, define  $D_{ik} = D_i - S_{ik}$ ,  $T_{ik} = T_i - S_{ik}$  and  $C_{ik} = C_i - S_{ik}$  as the death, transplant and censoring time respectively corresponding to the  $i$ th patient and measured from the  $k$ th cross section date. A modified counting process is also defined, where  $N_{i0k}(t) = N_i(S_{ik} + t)I(T_i > S_{ik} + t)$  with increment  $dN_{i0k}(t)$  as  $N_{i0k}\{(t + dt)^-\} - N_{i0k}(t)$ . The pertinent at-risk process is given as  $Y_{i0k}(t) = Y_i(S_{ik} + t)I(T_i > S_{ik} + t)$ . To simplify the notation, we define  $A_{ik} = A_i(S_{ik})$ ,  $Z_{i0k} = Z_i(S_{ik})$  and  $Z_{i0k}(t) = Z_i(S_{ik} + t)$ .

Figure 3.1 provides an illustration of how the treatment-free observation time is transformed into time-since-cross-section times. Four subjects ( $i_1, i_2, i_3, i_4$ ) and two cross sections ( $k_1, k_2$ ) are shown. The four subjects begin follow-up at different calendar time. For subject  $i_1$ , failure times  $D_{i_1 k_1}$  corresponds to cross section  $k_1$  and  $D_{i_1 k_2}$  refers to cross section  $k_2$ . Note subject  $i_1$  is not censored at the treatment-ineligible



Note: Vertical dashed lines denote cross-section dates, while horizontal dashed lines denote treatment-ineligible period. Subject  $i_1$  has death times  $D_{i_1k_1}$  and  $D_{i_1k_2}$ , corresponding to cross sections  $k_1$  and  $k_2$ , respectively. Note that, even though subject  $i_1$  is not censored after becoming treatment-ineligible. Subject  $i_2$  is treated at time  $T_{i_2k_2}$ , and hence censored (perhaps dependently) at that time, with respect to cross section  $k_2$ . Subject  $i_3$  is not included in either cross section since  $i_3$  starts then finishes follow-up in between cross-sections. Subject  $i_4$  first passes cross section  $k_1$  and then becomes inactive for a while until a time after cross section  $k_2$ . Subject  $i_4$  is treatment-ineligible at cross section  $k_2$  and, therefore, is not included in this cross section. With respect to cross section  $k_1$ , the transplant time  $T_{i_4k_1}$  is not censored at the beginning of the treatment-ineligible period.

Figure 3.1: Examples of the relationship between cross-section time and follow-up time.

time after cross section  $k_2$ . Subject  $i_2$  is treated, and dependently censored at time  $T_{i_2k_2}$  since cross section  $k_2$ . Subject  $i_3$  is not included in either cross section due to late start and early finish of the follow-up within between two cross-sections. Subject  $i_4$  is included in cross section  $k_1$ , but becomes and remains treatment-ineligible until a time after cross section  $k_2$ . With respect to cross section  $k_1$ ,  $i_4$  is censored at treatment time  $T_{i_4k_1}$ , as opposed to being censored earlier at the beginning of the treatment-ineligible period. Since subject  $i_4$  is treatment-ineligible at cross section  $k_2$ ,  $i_4$  is not included in this cross section.

We will estimate  $\beta_0$  through a stratified model,

$$\lambda_{i0k}(t; S_i) = \lambda_{00k}(t) \exp\{\beta'_0 Z_{i0k}\},$$

where  $\lambda_{00k}(t)$  are the cross-section-specific baseline hazards. The (overall) baseline hazard function of interest will be estimated by appropriately averaging the cross section specific  $\lambda_{00k}(t)$  estimators, as will be described later.

As discussed previously, we require a model for the treatment hazard,

$$\lambda_i^T(t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P\{t \leq T_i < t + \delta | T_i \geq t, A_i(t), Z_i(t)\},$$

in order to carry out IPCW for the pre-treatment experience. The following proportional hazards model is assumed,

$$\lambda_i^T(t) = A_i(t) \lambda_0^T(t) \exp\{\theta'_0 Z_i(t)\},$$

where  $t$  is the time from entry to the study. Define the treatment counting process  $N_i^T(t) = I(T_i \leq t \wedge X_i)$  and its corresponding increment  $dN_i^T(t) = N_i^T\{(t + dt)^-\} - N_i^T(t)$ . The regression coefficient,  $\theta_0$ , is estimated by  $\hat{\theta}$ , the root of the score function,

$$U^T(\theta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t; \theta)\} dN_i^T(t),$$

where  $\bar{Z}(t; \theta) = R_T^{(1)}(t; \theta)/R_T^{(0)}(t; \theta)$ ,  $R_T^{(p)}(t; \theta) = n^{-1} \sum_{i=1}^n Y_i(t) Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}$ ,  $p = 0, 1, 2$ , where, for a vector  $z$ ,  $z^{\otimes 0} = 1$ ,  $z^{\otimes 1} = z$ ,  $z^{\otimes 2} = zz'$ . The Breslow estimator of  $\Lambda_0^T(t)$  is given by  $\hat{\Lambda}_0^T(t) = n^{-1} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \hat{\theta})^{-1} dN_i^T(u)$ .

After estimating the treatment hazards, we come back to the death hazard models. An estimator for  $\beta_0$ , denoted by  $\hat{\beta}_0$ , is obtained through solving the IPCW score function,

$$U_0(\beta) = \sum_{i=1}^n \sum_{k=1}^K \int_0^{\tau_{0k}} A_{ik} \{Z_{i0k} - \bar{Z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dN_{i0k}(t),$$

with  $\bar{Z}_{0k}(t; \beta_0) = R_{0k}^{(1)}(t; \beta_0)/R_{0k}^{(0)}(t; \beta_0)$  and  $R_{0k}^{(d)}(t; \beta_0) = n^{-1} \sum_{i=1}^n A_{ik} W_{ik}^A(t) Z_{i0k}^{\otimes d} \exp\{\beta_0' Z_{i0k}\}$  with  $d = 0, 1, 2$  and where  $\tau_{0k}$  satisfies  $P\{Y_{i0k}(\tau_{0k}) = 1\} > 0$ , and can in practice be set to the largest observation time start from the  $k$ th cross section. As defined in Chapter II, an appropriate IPCW weight is given by  $W_{ik}^A(t) = Y_{i0k}(t) \exp\{\Lambda_i^T(S_{ik} + t) - \Lambda_i^T(S_{ik})\}$ , with  $\Lambda_i^T(t; \theta_0) = \int_0^t A_i(u) \lambda_0^T(u; \theta_0) \exp\{\theta_0' Z_i(u)\} du$ ,

However, in practice, some values of  $W_{ik}^A(t)$  could be very large and result in increased variance of  $\hat{\beta}_0$ . In order to improve the precision of  $\hat{\beta}_0$ , a stabilized weight is suggested,

$$W_{ik}^B(t) = Y_{i0k}(t) \frac{\exp\{\Lambda_i^T(S_{ik} + t) - \Lambda_i^T(S_{ik})\}}{\exp\{\Lambda_{ik}^T(t)\}},$$

analogous to procedures suggested by Robins and Finkelstein (2000) and Miloslavsky et al. (2004), where  $\Lambda_{ik}^T(t)$  is the cumulative hazard from a proportional hazards model which takes transplant as the event and adjusts only for covariates up to the cross-section time,  $\lambda_{ik}^T(t) = \lambda_{0k}^T(t) \exp\{\theta_1' Z_{i0k}\}$  and  $t$  is the time from entry to study with  $\theta_1$  as covariate effect.

With stratum-specific baseline hazards estimated, a weighted version of the Breslow-Aalen estimator pooled across strata could be obtained,

$$\hat{\Lambda}_{00}(t; \hat{\beta}_0) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^t R_0^{(0)}(u; \hat{\beta}_0)^{-1} A_{ik} W_{ik}^A(u) dN_{i0k}(u)$$

for  $t \in (0, L]$ , where  $R_0^{(0)}(u; \beta_0) = \sum_{k=1}^K R_{0k}^{(0)}(u; \beta_0)$ .

After estimating all the necessary parameters, for previous follow-up time  $S_i$ , we have  $\Lambda_{i0}(t; S_i) = \Lambda_{00}(t) \exp\{\beta'_0 Z_i(S_i)\}$  and  $\mu_{i0}(S_i) = \int_0^L \exp\{-\Lambda_{i0}(t; S_i)\} dt$ .

Compared to treatment-free survival, the estimation of the post-treatment parameters is more conventional. The following model is assumed,

$$\lambda_{i1}(t; T_i) = \lambda_{01}(t) \exp\{\beta' Z_i(T_i)\},$$

and fitted using observed post-transplant data. Note  $t$  here is from treatment initiation forward. Correspondingly, we define the counting process  $N_{i1}(t) = N_i(T_i + t)$  with increment  $dN_{i1}(t) = N_{i1}\{(t + dt)^-\} - N_{i1}(t)$  and at risk indicator  $Y_{i1}(t) = Y_i(T_i + t)$ . Define  $Z_{i1} = Z_i(T_i)$ .  $N_{i1}(t)$  and  $Y_{i1}(t)$  are defined to be 0 if patient  $i$  never receives treatments; i.e.,  $N_i^T(X_i) = 0$ . Since  $T_i$  is only subject to independent censoring, standard partial likelihood method can be applied to estimate  $\beta_1$ , that is, by solving  $U_1(\beta) = 0$ , where

$$U_1(\beta) = \sum_{i=1}^n \int_0^{\tau_1} \{Z_{i1} - \bar{Z}_1(t; \beta)\} dN_{i1}(t),$$

and  $\tau_1$  satisfies  $P\{Y_{i1}(\tau_1) = 1 > 0\}$ ,  $\bar{Z}_1(t; \beta_1) = R_1^{(1)}(t; \beta_1)/R_1^{(0)}(t; \beta_1)$  and  $R_1^{(p)}(t; \beta_1) = n^{-1} \sum_{i=1}^n Y_{i1}(t) Z_{i1}^{\otimes p} \exp\{\beta'_1 Z_{i1}\}$ , for  $p = 0, 1, 2$ . The Breslow estimator has the form  $\hat{\Lambda}_{01}(t; \beta_1) = n^{-1} \sum_{i=1}^n \int_0^t R_1^{(0)}(u; \beta_1)^{-1} dN_{i1}(u)$ . We then have  $\Lambda_{i1}(t; T_i) = \Lambda_{01}(t) \exp\{\beta'_1 Z_{i1}\}$  and  $\mu_{i1}(T_i) = \int_0^L \exp\{-\Lambda_{i1}(t; T_i)\} dt$ .

As a result, the subject specific treatment effect can be expressed as

$$\Delta_i(S_i) = \mu_{i1}(S_i) - \mu_{i0}(S_i),$$

which can be estimated by replacing all parameters with their corresponding afore-listed estimators. The subject specific difference in survival probability at time  $t$  is defined as

$$S_{\Delta i}(t; S_i) = S_{i1}(t; S_i) - S_{i0}(t; S_i).$$

Our main objective is to evaluate the average treatment effect under the current treatment assignment patterns. As implied previously, patients are selected for treatment non-randomly. We wish to estimate the average difference in restricted mean survival time, given the true (i.e., uncensored) distribution of  $\{T_i, Z_i(T_i)\}$ . This involves taking a weighted average of  $\Delta_i(T_i)$  estimators for subjects who received treatment  $\{T_i, Z_i(T_i) : T_i < X_i\}$ . A simple unweighted average will make the average treatment effect a function of censoring distribution, which is not desirable. Similar to a Horvitz and Thompson (1952) estimator, we weight each observed transplant by the inverse of the probability of remaining uncensored. The weighted data represent what would have been observed in the absence of independent censoring by  $C_i$ . The average treatment effect is defined as  $\Delta = E[\Delta(T)]$  where the expectation sign is with respect to the joint distribution of  $\{T_i, Z_i(T_i), T_i < \tau, T_i < D_i^0\}$ , with corresponding estimator given by

$$\widehat{\Delta} = \frac{\sum_{i=1}^n \int_0^\tau \widehat{\Delta}_i(t) \widehat{G}_i(t)^{-1} dN_i^T(t)}{\sum_{i=1}^n \int_0^\tau \widehat{G}_i(t)^{-1} dN_i^T(t)},$$

where  $G_i(t) = P\{C_i > t | Z_i(0)\}$ . We estimate  $G_i(t)$  via a standard Cox model,

$$\lambda_i^C(t) = \lambda_0^C(t) \exp\{\alpha_0' Z_i(0)\},$$

fitted to  $\{X_i, I(C_i < D_i), Z_i(0)\}$ , where  $t$  represents follow-up time. The quantity  $\Delta$  represents the average number of time units (e.g., years) saved by treatment (out of the next  $L$  years).

Analogously, the average difference in survival probability at time  $t$  is defined as  $S_\Delta(t) = E[S_\Delta(t; T)]$ , with corresponding estimator given by

$$\widehat{S}_\Delta(t) = \frac{\sum_{i=1}^n \int_0^\tau \widehat{S}_{\Delta_i}(t; u) \widehat{G}_i(u)^{-1} dN_i^T(u)}{\sum_{i=1}^n \int_0^\tau \widehat{G}_i(u)^{-1} dN_i^T(u)}.$$



### 3.3 Asymptotic Properties

We assume that the random vectors  $\{X_i, N_i(\cdot), N_i^T(\cdot), \tilde{Z}_i(X_i)\}$ , for  $i = 1 \dots n$ , are independent and identically distributed, with  $Z_i(t)$  bounded for  $t \in (0, \tau]$ , where  $\tau$  satisfies  $P(X_i \geq \tau) > 0$ . We summarize the asymptotic properties of the proposed methods in Theorem III.1 with all regularity conditions listed in Appendix B.

**Theorem III.1.** *Under certain regularity conditions,  $n^{1/2}[\hat{S}_\Delta(t) - S_\Delta(t)]$  and  $n^{1/2}(\hat{\Delta} - \Delta)$  converge asymptotically to a zero-mean Gaussian processes with covariance functions  $E[\xi_j(t)^2]$  and  $E[\eta_j^2]$ , respectively, where  $\{\xi_1(t), \dots, \xi_n(t)\}$  and  $\{\eta_1, \dots, \eta_m\}$  are i.i.d. with mean 0 asymptotically, and*

$$\begin{aligned}\xi_j(t) &= V(\tau)^{-1} \left\{ V_{1j}(t) + V_{2j}(t) + \int_0^\tau \{S_{\Delta j}(t; u) - S_\Delta(t)\} G_j(u)^{-1} dN_j^T(u) \right\}, \\ \eta_j &= \int_0^L \xi_j(t) dt,\end{aligned}$$

where

$$\begin{aligned}V(\tau) &= E \left[ \int_0^\tau G_i(u)^{-1} dN_i^T(u) \right], \\ V_{1j}(t) &= E \left[ \int_0^\tau \varphi_{ij}^S(t) G_i(u)^{-1} dN_i^T(u) \right], \\ V_{2j}(t) &= E \left[ \int_0^\tau S_{\Delta i}(t; u) \varphi_{ij}^C(u) dN_i^T(u) \right], \\ G_i(t) &= \exp\{-\Lambda_i^C(t)\}, \\ \varphi_{ij}^C(t) &= G_i(t) \{D_i^{C'}(t) \Omega_C(\alpha_0)^{-1} U_j^C(\alpha_0) + J_{ij}^C(t)\}, \\ \varphi_{ij}^S(t) &= S_{i0}(t) \{ \Lambda_{i0}(t) Z_i(S_i)' \Omega_0^{-1}(\beta_0) U_{j0}(\beta_0) - e^{\beta_0' Z_i(t)} \Phi_{j0}(t) \}, \\ &\quad - S_{i1}(t) \{ \Lambda_{i1}(t) Z_{i1}' \Omega_1^{-1}(\beta_1) U_{j1}(\beta_1) - e^{\beta_1' Z_{i1}(t)} \Phi_{j1}(t) \},\end{aligned}$$

with  $\eta_j$ ,  $\Omega_1(\beta_1)$ ,  $\Omega_0(\beta_0)$ ,  $U_{j1}(\beta_1)$ ,  $U_{j0}(\beta_0)$ ,  $\mu_{i1}(t)$ ,  $\mu_{i0}(t)$ ,  $d\Phi_{j1}(t)$ ,  $d\Phi_{j0}(t)$ ,  $D_i^C(t)$ ,  $\Omega_C(\alpha_0)$ ,  $U_j^C(\alpha_0)$  and  $J_{ij}^C(t)$  defined in Appendix B.

The covariance estimator of  $\widehat{\Delta}$  is given by  $n^{-1} \sum_{i=1}^n \widehat{\eta}_i^2$ , that of  $\widehat{S}_\Delta(t)$  is given by  $n^{-1} \sum_{i=1}^n \widehat{\xi}_i(t)^2$ , where  $\widehat{\eta}_i$  and  $\widehat{\xi}_i(t)$  are estimated by changing all limiting values in  $\varphi_i$  by their empirical counterparts. A proof of Theorem III.1 is given in Appendix B. The main idea of the proof is demonstrating that, asymptotically,  $n^{1/2}(\widehat{\Delta} - \Delta) = n^{-1/2} \sum_{i=1}^n \eta_i + o_p(1)$  and  $n^{1/2}\{\widehat{S}_\Delta(t) - S_\Delta(t)\} = n^{-1/2} \sum_{i=1}^n \xi_i(t) + o_p(1)$  through a sequence of Taylor series expansions.

The proof is provided for the weight,  $W_{ik}^A(t) = Y_{i0k}(t) \exp\{\Lambda_i^T(S_{ik} + t) - \Lambda_i^T(S_{ik})\}$ . In practice, a stabilized weight is preferred to improve precision. As shown by Theorem III.1, the computation of the variance is already very complicated, and is more complicated when a stabilizer is incorporated. Such concerns motivate a computationally simpler form for the variance estimator. That is, to take the weight  $W_{ik}^A(t)$ , or  $W_{ik}^B(t)$  and  $G_i(t)^{-1}$ , as the case may be as fixed. Then, the variance estimator of  $S_\Delta(t)$  and  $\Delta$  could simplify to  $n^{-1} \sum_{i=1}^n \widehat{\xi}_i^*(t)^2$  and  $n^{-1} \sum_{i=1}^n \widehat{\eta}_i^{*2}$ , where

$$(3.1) \quad \begin{aligned} \widehat{\xi}_j^*(t) &= \widehat{V}(\tau)^{-1} \sum_{i=1}^n \int_0^\tau \widehat{\varphi}_{ij}^{S^*}(t) \widehat{G}_i(u)^{-1} dN_i^T(u) \\ &\quad + \widehat{V}(\tau)^{-1} \int_0^\tau \{\widehat{S}_{\Delta j}(t; u) - \widehat{S}_\Delta(t)\} \widehat{G}_j(u)^{-1} dN_j^T(t), \end{aligned}$$

$$(3.2) \quad \begin{aligned} \widehat{\eta}_j^* &= \int_0^L \widehat{\xi}_j^*(t) dt \\ \widehat{V}(\tau) &= n^{-1} \sum_{i=1}^n \int_0^\tau \widehat{G}_i(t)^{-1} dN_i^T(t), \end{aligned}$$

where

$$\begin{aligned} \varphi_{ij}^{S^*}(t) &= S_{i0}(t) \{\Lambda_{i0}(t) Z_i(S_i)' \Omega_0^{-1}(\beta_0) U_{j0}(\beta_0) - e^{\beta_0' Z_i(t)} \Phi_{j0}^*(t)\} \\ &\quad - S_{i1}(t) \{\Lambda_{i1}(t) Z_{i1}' \Omega_1^{-1}(\beta_1) U_{j1}(\beta_1) - e^{\beta_1' Z_{i1}(t)} \Phi_{j1}^*(t)\} \\ \Phi_{i0}^*(t) &= - \int_0^t \overline{z}'_0(u; \beta, W) d\Lambda_{00}(u) \Omega_0^{-1}(\beta_0) U_{i0}(\beta_0) + \sum_{k=1}^K \int_0^t \frac{A_{ik} W_{ik}^A(u) dM_{i0k}(u)}{r_0^{(0)}(u; \beta_0, W)}, \\ \Phi_{i1}^*(t) &= - \int_0^t \overline{z}'_1(u; \beta_1) d\Lambda_{01}(u) \Omega_1^{-1}(\beta_1) U_{i1}(\beta_1) + \int_0^t \frac{\delta_i^T dM_{i1}(u)}{r_1^{(0)}(u; \beta_1)}, \end{aligned}$$

$$dM_{i0k}(t) = dN_{i0k}(t) - Y_{i0k}(t)d\Lambda_{i0}(t),$$

$$dM_{i1}(t) = dN_{i1}(t) - Y_{i1}(t)d\Lambda_{i1}(t),$$

where  $\bar{z}_0(t; \beta, W)$ ,  $r_0^{(0)}(t; \beta_0, W)$ ,  $\bar{z}_1(t; \beta_1)$  and  $r_1^{(0)}(t; \beta_1)$  are the infinity values of  $\bar{Z}_0(t; \beta, W)$ ,  $R_0^{(0)}(t; \beta_0, W)$ ,  $\bar{Z}_1(t; \beta_1)$  and  $R_1^{(0)}(t; \beta_1)$ , respectively.

### 3.4 Simulation

For treatment-free survival, we modify the settings in Zheng and Heagerty (2005) to generate data following the assumed partly conditional model. First, we generate a single binary treatment group indicator  $Z_{ia}$ , with value 0 and 1 with probability 0.5. A single longitudinal covariate  $Z_i(S_{ik})$  measured at a common set of cross section dates  $CS_1, CS_2, \dots, CS_K$  is then created. To generate data  $\{D_i, Z_{ia}, Z_{ib}\}$  where  $Z_{ib} = \text{vec}\{Z_i(S_{ik})\}$ , we first let  $Z_{ib0} = b_i + \sum_{k=1}^K \log(V_{ik})/\gamma_2$ , where  $b_i \sim N(\mu, \sigma^2)$  and  $V_{ik} \sim P(\rho)$ , independent positive stable random variables with index  $\rho$  (Samoridnitsky and Taqqu, 1994). A pre-treatment death time  $D_i$ , is then generated with a hazard  $\lambda_i(t) = V_{i0}^{1/\rho} \lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_{ib0}\}$ , where  $V_{i0} \sim P(\rho)$  and is independent of  $V_{ik}$ , with  $\Lambda_0(t) = (t/a)^{1/\rho^2}$  and  $a$  is a constant. Let  $Z_i(S_{ik}) = Z_{ib0} - \log(V_{ik})/\gamma_2$ . Then, the death hazard can be written as

$$(3.3) \quad \lambda_i(t) = V_{i0}^{1/\rho} \lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik}) + \log(V_{ik})\}.$$

The  $i$ th subject enters the study on calendar date,  $L_i$ , which is generated from a  $Uniform(0, b)$  distribution. Treatment time,  $T_i$ , is generated from the proportional hazards model,

$$(3.4) \quad \lambda_i^T(t) = \lambda_0^T(t) \exp\{\theta_{01} Z_{ia} + \theta_{02} I(t > R_i)\},$$

where  $\lambda_0^T(t) = d_3$  and  $\theta_0 = (\theta_{01}, \theta_{02})$  and the time of treatment-ineligibility,  $R_i$ , is generated with hazard  $\lambda_i^R(t) = \lambda_0^R(t) \exp\{d_1 V_{i0}\}$ , where  $\lambda_0^R(t) = d_2$ . Thus,  $R_i$  and

$D_i$  are positively correlated, which is consistent with the data which motivated the proposed methods. Administrative censoring time,  $C_i$ , is generated from hazard

$$(3.5) \quad \lambda_i^C(t) = \lambda_0^C(t) \exp\{\alpha_0 Z_{ia}\},$$

where  $\lambda_0^C(t) = d_4$ . Note that treatment time and pre-treatment death time,  $T_i$ , and  $D_i$  are dependent since both depend on treatment-ineligibility time  $R_i$ . However, independent censoring time  $C_i$  is independent with  $D_i$  conditional on  $Z_{ia}$ .

To see how the preceding set-ups generate proportional hazards in the partly conditional setting, integrate both sides of the expression (3.3),

$$\Lambda_i(t) = V_{i0}^{1/\rho} \Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{ik},$$

such that the pre-treatment survival function is given by

$$\exp[-\Lambda_i(t)] = \exp[-\Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{ik} V_{i0}^{1/\rho}].$$

Now, transforming the time scale to reflect time since cross section, define  $t_k = t - S_{ik}$ .

Then, take the expectation with respect to  $V_{ik}$  first and from the property of positive stable distribution, we have

$$\exp[-\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}, V_{i0})] = \exp[-(\Lambda_0(t) \exp\{\gamma_1 Z_{ia} + \gamma_2 Z_i(S_{ik})\} V_{i0}^{1/\rho})^\rho / \cos(\pi\rho/2)].$$

Then, taking the expectation with respect to  $V_{i0}$ , we have

$$\exp[-\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik})] = \exp[-(\Lambda_0(t)^\rho \exp\{\rho^2 \gamma_1 Z_{ia} + \rho^2 \gamma_2 Z_i(S_{ik})\} / \cos(\pi\rho/2))^{(\rho+1)}],$$

which implies the following equation after taking the logarithm and negating both sides,

$$\Lambda_i(t_k | Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \Lambda_0(t)^\rho \exp\{\rho^2 \gamma_1 Z_{ia} + \rho^2 \gamma_2 Z_i(S_{ik})\} / \cos(\pi\rho/2)^{(\rho+1)}.$$

Transforming the time scale, we then obtain,

$$\lambda_i(t_k|Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \frac{\lambda_0(t_k + S_{ik})\rho^2\{\Lambda_0(t_k + S_{ik})\}^{(\rho^2-1)}}{\cos(\pi\rho/2)^{(\rho+1)}} \exp\{\rho^2\gamma_1 Z_{ia} + \rho^2\gamma_2 Z_i(S_{ik})\}.$$

Using this construction, the hazard for  $D_{ik} = D_i - S_{ik}$  will generally depend on  $S_{ik}$  and, therefore, stratified models similar to those considered by Wei, Lin, and Weissfeld (1989) would be appropriate. With  $\Lambda_0(t) = (t/a)^{1/\rho^2}$  and  $\lambda_0(t_k + S_{ik})\rho^2\{\Lambda_0(t_k + S_{ik})\}^{(\rho^2-1)} = 1/a$ , it follows that

$$\lambda_i(t_k|Z_{ia}, Z_i(S_{ik}), D_i > S_{ik}) = \exp\{\rho^2\gamma_1 Z_{ia} + \rho^2\gamma_2 Z_i(S_{ik})\}/[a \cos(\pi\rho/2)^{(\rho+1)}].$$

If we define  $\lambda_{i0k}(t; S_{ik}) = \lambda_i(t_k|Z_{ia}, Z_i(S_{ik}), D_i > S_{ik})$ ,  $\lambda_{00k}(t) = [a \cos(\pi\rho/2)^{(\rho+1)}]^{-1}$  and  $\beta_0 = (\beta_{01}, \beta_{02}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ , then the proportional hazards model for pre-treatment death time is given by

$$(3.6) \quad \lambda_{i0k}(t; S_{ik}) = \lambda_{00k}(t) \exp\{\beta_{01} Z_{ia} + \beta_{02} Z_i(S_{ik})\}.$$

We only include for analysis those  $Z_i(S_{ik})$  with  $L_i < S_{ik} < \min(X_i, R_i)$ .

For patients who received treatment prior to dying ( $D_i > T_i$ ), a post-treatment death time  $D_i^1 - T_i$ , is then generated via hazard

$$(3.7) \quad \lambda_{i1}(t; T_i) = \lambda_{01}(t) \exp\{\beta_{11} Z_{ia} + \beta_{12} Z_i(T_i)\},$$

where  $t$  represents time from treatment date  $T_i$  forward and  $\beta_1 = (\beta_{11}, \beta_{12}) = (\rho^2\gamma_1, \rho^2\gamma_2)$ . We set  $\lambda_{01}(t) = a_1$ .

After generating the data, the pre-treatment death data since cross section,  $\{X_{ik}, N_{ik}(\cdot), Z_{ia}, Z_i(S_{ik})\}$  are used to fit model (3.6); treatment data  $\{X_i, N_i^T(X_i), Z_{ia}, Z_i(t)\}$  are used to fit model (3.4); censoring data  $\{X_i, N_i^C(\cdot), Z_{ia}, Z_i(0)\}$  are used to fit model (3.5); post-treatment death data  $\{X_i - T_i, Z_{ia}, Z_i(T_i); N_i^T(X_i) = 1\}$  are used in the fitting of (3.7).

We evaluate samples with  $n = 500$  subjects and obtain censoring and transplant rates as approximately [10%, 10%], [15%, 15%] and [20%, 20%] with  $a = 2000, 5000, 7000$  respectively. There are  $K = 10$  cross section dates. We set  $b = 500$ ,  $(\theta_{01}, \theta_{02}) = (-1, -1)$ ,  $\mu = 18$ ,  $\sigma = 1$ ,  $(\gamma_1, \gamma_2) = (-1, -0.5)$  and  $d_1 = d_2 = d_3 = d_4 = 0.001$ . We set  $CS_k = 100 \times k$ . For all our simulated data configurations, 1000 Monte Carlo data sets are used. We present results using  $\rho = 0.8$ , thus  $(\beta_{01}, \beta_{02}) = (\beta_{11}, \beta_{12}) = (-0.64, -0.32)$ . With the number of cross sections set to  $K = 10$ , the average number of records per subject is 0.7 to 2.4, depending on the censoring level. Since we consider a 3-year restricted mean survival time and try to simulate time data in “day” scale, we set  $L = 365 \times 3 = 1095$ . Two scenarios are given, positive effect of treatment on mean survival ( $\Delta > 0$ ) with  $a_1 = 0.5 \times 10^{-4}$ , and no treatment effect ( $\Delta = 0$ ) with  $a_1 = [a \cos(\pi\rho/2)^{(\rho+1)}]^{-1}$ .

We applied the simplified variance estimator in which  $\{\widehat{W}_{ik}^A(t)\}$ ,  $\{\widehat{W}_{ik}^B(t)\}$  and  $\{\widehat{G}_i(t)^{-1}\}$  are all treated as fixed; i.e., in (3.2) and (3.1). Table 3.1 presents the simulation results for the case with  $\Delta = 0$ , while Table 3.2 provides results for  $\Delta > 0$  cases. The quantity  $\Delta$  equals the difference of 3-year restricted mean survival time post- versus pre-treatment. The proposed estimators appear to be approximately unbiased and coverage probabilities of  $\Delta$  are close to the nominal 95% level.

### 3.5 Application

We applied the proposed methods to estimate the average effect of liver transplantation among the transplanted, by Model for End-stage Liver Disease (MELD) score. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR), a national population based organ transplant registry. The study population consists of patients initially wait listed for deceased-donor liver transplantation

Table 3.1: Simulation result for  $\Delta = 0$  with  $\{W_{ik}^B(t)\}$ .

$n=500$		$\Delta = 0$				
[Censor, Treat]	True Value	Term(yr)	Bias	ESE	ASE	CP
[10%, 10%]	0	$\Delta$	0.040	0.204	0.190	0.92
	0	$S_{\Delta}(1)$	0.012	0.089	0.082	0.92
	0	$S_{\Delta}(2)$	0.016	0.092	0.085	0.93
	0	$S_{\Delta}(3)$	0.022	0.094	0.082	0.91
[15%, 15%]	0	$\Delta$	0.022	0.164	0.154	0.93
	0	$S_{\Delta}(1)$	0.007	0.065	0.061	0.93
	0	$S_{\Delta}(2)$	0.010	0.077	0.072	0.93
	0	$S_{\Delta}(3)$	0.010	0.083	0.077	0.91
[20%, 20%]	0	$\Delta$	0.009	0.144	0.141	0.94
	0	$S_{\Delta}(1)$	0.001	0.056	0.054	0.93
	0	$S_{\Delta}(2)$	0.004	0.067	0.066	0.94
	0	$S_{\Delta}(3)$	0.005	0.074	0.073	0.94

Table 3.2: Simulation result for  $\Delta > 0$  with  $\{W_{ik}^B(t)\}$ .

$n=500$		$\Delta > 0$				
[Censor,Treat]	True Value	Term(yr)	Bias	ESE	ASE	CP
[10%, 10%]	0.871	$\Delta$	0.030	0.204	0.190	0.92
	0.294	$S_{\Delta}(1)$	0.009	0.088	0.074	0.92
	0.346	$S_{\Delta}(2)$	0.009	0.100	0.088	0.92
	0.350	$S_{\Delta}(3)$	0.008	0.110	0.097	0.92
[15%, 15%]	0.614	$\Delta$	0.017	0.150	0.145	0.94
	0.193	$S_{\Delta}(1)$	0.006	0.054	0.052	0.94
	0.253	$S_{\Delta}(2)$	0.008	0.070	0.068	0.94
	0.278	$S_{\Delta}(3)$	0.005	0.082	0.077	0.92
[20%, 20%]	0.427	$\Delta$	0.020	0.135	0.133	0.94
	0.133	$S_{\Delta}(1)$	0.006	0.048	0.048	0.94
	0.177	$S_{\Delta}(2)$	0.009	0.064	0.062	0.93
	0.197	$S_{\Delta}(3)$	0.006	0.077	0.072	0.93



between March 1, 2002 and December 31, 2009 in the United States. Only adult patients (age  $\geq 18$  at listing) not previously transplanted (i.e., not repeat transplant candidates) were included in the study cohort. Patients who were Status 1 at  $t = 0$  were excluded. Cross-section dates were chosen every 7 days, 30 days or 90 days from 03/01/2002 to 12/31/2009, which lead to  $K = 409, 96, 32$  cross sections respectively. Three different cross-section frequencies dates were chosen in order to study the behavior of the proposed methods.

In order to construct the IPCW weight,  $\Lambda_i^T(t)$  was estimated through a time-dependent Cox model in which transplant was the event. For the time-to-transplant model, time  $t$  starts from the beginning of the follow-up (the date of wait listing), as opposed to cross section date. The model was stratified, such that

$$\lambda_{ir}^T(t) = A_i(t)\lambda_{0r}^T(t) \exp\{\theta'_0 Z_i(t)\},$$

where the subscript  $r = 1, \dots, 11$  stands for United Network for Organ Sharing (UNOS) region. The presence of the indicator,  $A_i(t)$ , reflects the fact that a patient's time while inactive or removed does not contribute to the estimation of  $\theta_0$  or  $\Lambda_{0r}^T(t)$ . The covariate,  $Z_i(t)$ , included MELD score, Status 1, albumin, age, gender, race, diagnosis of Hepatitis C, body mass index, diabetes, hospitalization, blood type, dialysis within prior week, encephalopathy, ascites and serum creatinine.

To obtain the weight  $G_i(t)^{-1}$ ,  $\Lambda_i^C(t)$  was estimated through a baseline Cox model in which administrative censoring was the event. As in the time-to-transplant model, time  $t$  starts from the beginning of the follow-up. Covariates evaluated at baseline,  $Z_i(0)$ , included MELD score, albumin, age, gender, race, diagnosis, body mass index, diabetes, hospitalization status at listing and previous malignancy.

For the weight, some very large values occurred. Since we found that 99% of

$W_{ikr}^B(t)$  and  $G_i(t)^{-1}$  were less than 10, both weights were then capped at 10. The time-to-transplant and time-to-censoring models are regressed on the entire study cohort (i.e., one single model). Separate pre-transplant death models were fitted for each of the following MELD score categories: 6-8, 9-11, 12-14, 15-17, 18-19, 20-22, 23-25, 26-29, 30-35 and 36-40. One single post-transplant model was fitted to all MELD score categories. Then the proposed death models were applied to the chosen cohort. The pre-transplant data were divided into subgroups by MELD due to computational burden.

The pre-transplant death model, was also stratified

$$\lambda_{i0kr}(t) = \lambda_{00kr}(t) \exp\{\beta'_0 Z_{ik}\},$$

where the subscript  $r = 1, \dots, 11$  again stands for UNOS region and  $k = 1, \dots, K$  stands for cross section. The subject level covariates at cross section  $k$ ,  $Z_{ik}$ , included MELD score, albumin, age, gender, race, diagnosis, body mass index, diabetes, hospitalization status at listing, previous dialysis and malignancy. Also included in  $Z_{ik}$  were time on wait-list, average change in MELD score (pertaining to the time interval between the date of listing and cross-section  $k$  date, and estimated using ordinary least squares) and average change in albumin (estimated analogously). Other elements included the percentage of time spent in inactive status, and percent of time receiving dialysis. Since 99% of MELD and albumin slope values before cross sections fell in the  $[-1,1]$  interval, the slopes were bounded by -1 and 1.

The post-transplant death model is expressed as

$$\lambda_{i1}(t) = \lambda_{01}(t) \exp\{\beta'_1 Z_i(T_i)\}.$$

The subject level covariates at transplant time  $T_i$ ,  $Z_i(T_i)$ , included MELD score, albumin, age, gender, race, diagnosis, body mass index, diabetes, hospitalization

status at listing, previous dialysis and malignancy, donor risk index (DRI; Feng et al., 2006) and time on wait-list.

The sample in pre-transplant period consisted of  $n = 66,884$  patients, of which 36,260 were observed to receive a deceased-donor liver transplant. Estimates of the difference in restricted mean survival time,  $\Delta$ , based on the death model using the Type B weight are listed in Table 3.3. Estimates of the difference in survival probability,  $S_{\Delta}(t)$  are shown in Table 3.4 for  $t = 1, 3, 5$  years. For the MELD 30-40 subgroup, weekly cross section dates were chosen. For MELD 18-29 cross sections were drawn monthly. For MELD 6-17, cross sections were drawn every 3 months. Note that, we also tried weekly cross section dates for MELD 6-29 patients, which yielded almost identical results. There was an obvious monotone pattern in that, as MELD and restricted time  $L$  went up, estimates of differences in both restricted mean lifetime and survival probability increased.

Figure 3.2 shows the estimated average pre- and post-transplant survival curves for MELD group 6-8, 15-17, 20-22 and 36-40. For MELD 36-40 group, the wait list survival curve drops dramatically during the first couple of months, then steadily declines thereafter. As MELD score decreases, the difference between the two survival curves generally diminishes and the dropping slope of the wait list curve becomes closer to a constant; a phenomenon is quite obvious in the plot of the MELD 6-8 group.

### 3.6 Discussion

In this chapter, we proposed pre- and post-treatment models for estimating the average effect of a time-dependent treatment via differences in survival probability and restricted mean lifetime. A proportional hazards model which uses covariate

Table 3.3: Estimate of  $\Delta(\text{SE})$  by MELD groups.

MELD	1 Year	3 Years	5 Years
36-40	0.47 (0.04)	1.43 (0.18)	2.38 (0.34)
30-35	0.30 (0.04)	0.88 (0.34)	1.40 (0.75)
26-29	0.20 (0.03)	0.63 (0.11)	0.99 (0.20)
23-25	0.12 (0.01)	0.57 (0.07)	1.07 (0.14)
20-22	0.10 (0.02)	0.62 (0.07)	1.23 (0.14)
18-19	0.07 (0.01)	0.51 (0.04)	1.06 (0.08)
15-17	0.03 (0.01)	0.40 (0.05)	1.00 (0.10)
12-14	-0.01 (0.01)	0.18 (0.04)	0.59 (0.08)
9-11	-0.03 (0.01)	0.03 (0.03)	0.29 (0.07)
6-8	-0.04 (0.01)	-0.04 (0.03)	0.11 (0.07)

Table 3.4: Estimate of  $S_{\Delta}$ (SE) by MELD groups.

MELD	1 Year	3 Years	5 Years
36-40	0.48 (0.04)	0.48 (0.06)	0.45 (0.06)
30-35	0.33 (0.04)	0.27 (0.10)	0.25 (0.12)
26-29	0.25 (0.04)	0.19 (0.04)	0.16 (0.05)
23-25	0.19 (0.02)	0.23 (0.02)	0.26 (0.04)
20-22	0.19 (0.02)	0.29 (0.03)	0.30 (0.03)
18-19	0.15 (0.01)	0.26 (0.01)	0.27 (0.02)
15-17	0.09 (0.01)	0.26 (0.02)	0.32 (0.02)
12-14	0.02 (0.01)	0.16 (0.02)	0.23 (0.02)
9-11	-0.02 (0.01)	0.09 (0.01)	0.17 (0.01)
6-8	-0.03 (0.01)	0.03 (0.01)	0.11 (0.02)

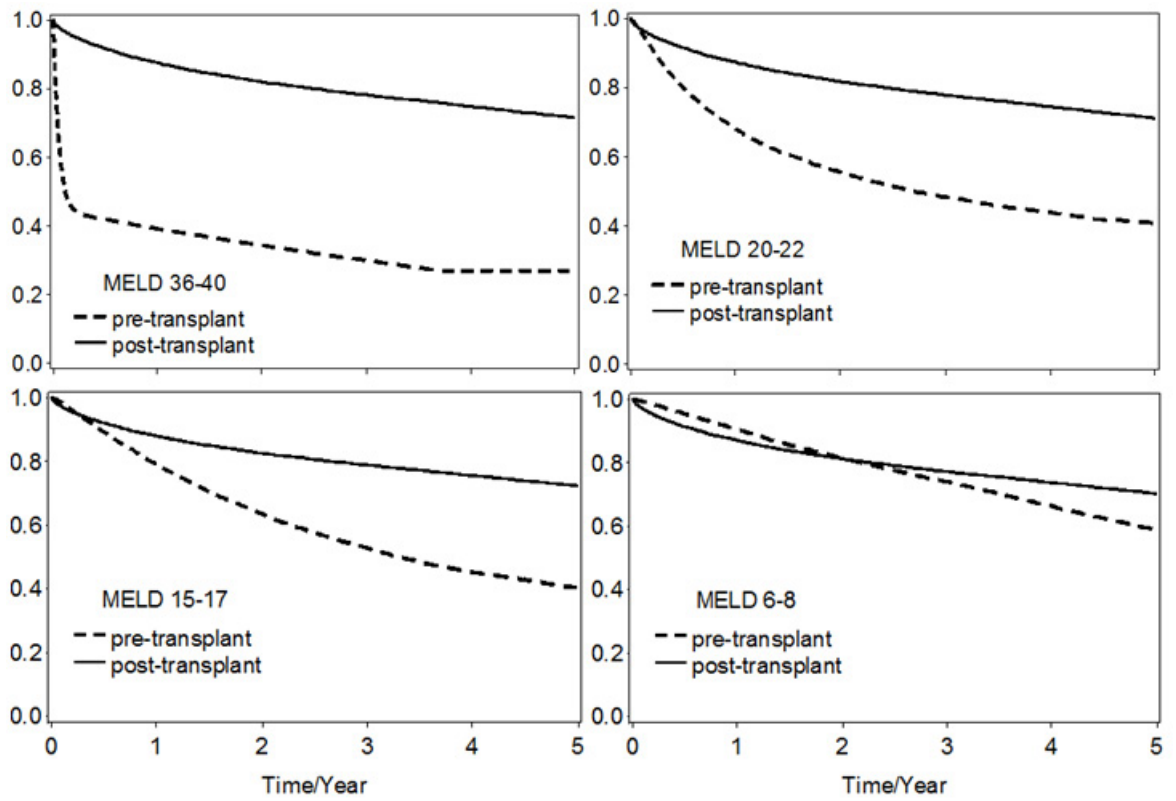


Figure 3.2: Estimated post- and pre-transplant survival curves by MELD score at transplant.

history up to the treatment time is assumed for the post-treatment death hazard. A cross-section-based partly conditional survival model which conditions on covariate history up to and time since the cross-section times is used for the treatment-free death hazard. This cross-section-based model to allows the estimation of treatment-free survival for treated patients. Thus the comparison in survival probability or restricted mean lifetime with versus without treatment at the same time point, the treatment time, is made possible. In the partly conditional survival model, values of covariates are fixed at cross-section time and multiple cross-sections are drawn with data of each serves as a stratum. Freezing the covariate value could be seem as a marginal way of estimation, which averages over all the possible future covariate paths. However, since only part of the covariate history is conditioned on, receiving treatment results in the dependent censoring of death. A revised version of IPCW is applied in aiming to estimate the effect consistently in the presence of dependent censoring. With pre- and post-treatment survival curves estimated, treatment effect could be obtained at the subject level based on each patient's covariate values at treatment time. To construct the average treatment benefit, another IPCW weighted average of subject level estimates is used to weight out the effect of administrative censoring.

## CHAPTER IV

# Semiparametric Cumulative Contrasts of Group-Specific Mortality in the Presence of Dependent Censoring

### 4.1 Introduction

In medical and clinical studies, nonproportional hazards and dependent censoring are two common issues. For example, in comparing mortality across different groups of patients, discrepancies in mortality by group may not be constant over follow-up time. In the context of the Cox (1972) model, nonproportionality could be handled by assuming that the group-specific hazard ratio effect changes with a specific pattern that can be represented parametrically. For example, the effect could be piecewise constant, or follow a continuous function. However, usually it is generally quite challenging to confirm the correctness of the selected parametric form. Moreover, when the effect is time-varying, investigators tend to be more interested in the cumulative effect; i.e. if and when the treatment-specific survival (or, equivalently the cumulative hazard) functions cross.

Another frequently arising issue is dependent censoring, which can occur when both the death and censoring process depend on overlapping time-varying covariates not captured in the death hazard model. Effect estimators that rely on censoring



and death being independent are generally no longer consistent in the presence of dependent censoring. If adjustment is made for the covariates that vary over time, the effect estimates then represent differences between subjects with equal covariate paths over time, which is usually of secondary interest at most.

Wei and Schaibel (2008) proposed estimating the cumulative treatment effect using a stratified Cox model with treatment groups as strata. The methods require no assumptions on the pattern of the group effect on mortality over time, although proportionality is assumed to hold within-group for the adjustment covariates. The ratio of stratum-specific cumulative baseline hazards is used to estimate the cumulative treatment effect over time.

The method of Wei and Schaibel (2008) assumes that censoring is conditionally independent of death given treatment and the adjustment covariates. This assumption frequently fails in practice, particularly for observational studies. Robins and Rotnitzky (1992), Robins and Finkelstein (2000) proposed Inverse Probability of Censoring Weighting (IPCW) to obtain consistent estimators in the presence of dependent censoring.

The motivation for the proposed methods arises from wait-list mortality comparisons among End-Stage Liver Disease (ESLD) patients. The preferred treatment for ESLD is deceased-donor liver transplantation. However, the number of available organs is always less than the number of patients on the liver transplant waiting list. According to the current liver allocation guidelines, the sickest patients are prioritized to the top of the waiting list, where Status-1 patients are in front of the queue, followed by chronic patients in decreasing order of Model for End-Stage Liver Disease (MELD) score. Note that MELD scores are updated over time and, thus, MELD is a time-varying covariate. MELD scores are highly predictive of both wait list mortality

and liver transplant rates; the latter being due to the liver allocation policy which has been in place since 2002. Therefore, liver transplantation represents dependent censoring for wait-list death unless adjustment is made for the time-dependent predictors, such as MELD score. However, adjusting for time-dependent covariates will bias the group effect estimate towards the null. As a result, the receipt of a liver transplant results in the dependent censoring of wait list death time, in the presence of adjustment for baseline (time 0) covariates only. Preliminary analysis has revealed that the Kaplan-Meier curves for high-MELD patients later cross that for Status-1 patients, which implies that the among-group proportionality assumption fails. Another issue is that due to health conditions, patients can be permanently removed from or temporarily inactive on the wait-list, during which time intervals they are not eligible to receive an organ transplant.

When proportionality fails, some authors estimated the causal inference, or average hazard ratio, using marginal structural models (e.g., Hernan, Brumback, and Robins, 2000, 2001; Robins, Hernan, and Brumback, 2000). Various authors estimated restricted mean survival time (Chen and Tsiatis, 2001; Zhang and Schaubel, 2010). However, contrasts between restricted mean lifetime do not describe how the effect on survival changes over time. Wei and Schaubel (2008) proposed a ratio of the cumulative hazards to quantify the group effect without considering dependent censoring. Schaubel and Wei (2011) later proposed the ratio of cumulative hazard over time and used IPCW to overcome the dependent censoring. However, Inverse Probability of Treatment Weighting (IPTW) was used instead of covariate adjustment in the death model, which estimates a quantity analogous to the average causal effect, which may not be of chief interest in certain applications. In summary, due to interest in the shape of the time-varying effect, the presence of dependent cen-

soring and the potential for subjects to become treatment ineligible, none of the afore-mentioned methods can be applied to our ESLD motivating example directly.

In this chapter, we propose the ratio of IPCW-weighted baseline cumulative hazards as an estimator of the group effect on mortality over time. Essentially, the proposed method extends that of Wei and Schaebel (2008) to the dependent censoring setting. Since only baseline, opposed to time-dependent covariates, are adjusted for, dependent censoring occurs and is overcome by IPCW, with a stabilized version of weight applied to increase precision. To construct the IPCW weight, a time-dependent Cox model with treatment as the event is adopted. The time-to-treatment model is conditional on being treatment-eligible, such that treatment-ineligible time intervals do not contribute to the treatment hazards; reflective of the fact that organs are not allocated to removed or inactive patients in the motivating example.

The remainder of this chapter is organized as follows. In Section 4.2, we provide the notation and describe the proposed estimator. How consistent estimation is obtained in presence of dependent censoring through IPCW method is also explained. Asymptotic properties of the proposed estimator are given in Section 4.3. Simulation studies are provided in Section 4.4 to evaluate the performance of the proposed methods in finite samples. The proposed methods are applied to liver wait list mortality using data from a national registry in Section 4.5. In Section 4.6, we provide some concluding remarks and discussion. Asymptotic derivations are given in Appendix C.

## 4.2 Proposed Methods

First, we set up the necessary notation. Let  $D_i$  be the time to failure for subject  $i$ , with  $i = 1, \dots, n$ . We assume that  $D_i$  may be right censored at treatment time,  $T_i$ , or

independent censoring time,  $C_i$ , and therefore we only observe  $X_i = \min(D_i, T_i, C_i)$  and associated indicators  $\Delta_i = I(X_i = D_i)$  and  $\Delta_i^T = I(X_i = T_i)$ , where  $I(A) = 1$  if event  $A$  is true, and 0 otherwise. Let  $J$  be the number of groups with index  $j = 0, 1, 2, \dots, J - 1$ . The first group ( $j = 0$ ) serves as the reference category, to which all other groups are compared. Define the death counting process,  $N_i(t) = I_i(X_i \leq t, \Delta_i = 1)$  with associated increment  $dN_i(t) = N_i\{(t + dt)^-\} - N_i(t)$  and at-risk process,  $Y_i(t) = I(X_i \geq t)$ . Let  $G_i$  be the group variable for  $i$ th subject and define  $G_{ij} = I(G_i = j)$ . Similarly, we have  $Y_{ij}(t) = Y_i(t)G_{ij}$  and  $dN_{ij}(t) = dN_i(t)G_{ij}$  as the at risk and counting process for  $i$ th subject and  $j$ th group. Note that group is defined at  $t = 0$ . With respect to the treatment process, define  $N_i^T(t) = I_i(X_i \leq t, \Delta_i^T = 1)$  with increment  $dN_i^T(t) = N_i^T\{(t + dt)^-\} - N_i^T(t)$ . Let  $A_i(t)$  be 1 if patient  $i$  is eligible to receive treatment as of time  $t$ , and 0 otherwise, and set  $\tilde{A}_i(t) = \{A_i(s); s \in [0, t)\}$ . Note that, since treatment-initiation never occurs when a patient is treatment-ineligible,  $dN_i^T(t) = A_i(t)dN_i^T(t)$ . Thus  $n$  independent vectors are observed from subjects,  $\{X_i, \Delta_i, \Delta_i^T, G_i, \tilde{Z}_i(X_i), \tilde{A}_i(X_i)\}$ , where  $Z_i(t)$  is the time-varying covariate and  $\tilde{Z}_i(t) = \{Z_i(s); s \in [0, t)\}$  is the covariate history on  $[0, t)$ . Set  $Z_{i0} \equiv Z_i(0)$  as the baseline covariates.

A variant of Inverse Probability of Censoring Weighting (IPCW) is used to overcome the dependent censoring of  $D_i$  via  $T_i$ . First, we assume that the treatment time  $T_i$  follows the Cox model,

$$\lambda_i^T(t) = A_i(t)\lambda_0^T(t) \exp\{\theta_0' Z_i(t)\},$$

where the inclusion of  $A_i(t)$  reflects the fact that the treatment hazard is zero at times during which the subject is treatment-ineligible. The IPCW weight is given by  $W_i(t, \theta) = Y_i(t) \exp\{\Lambda_i^T(t, \theta)\}$ , where  $\Lambda_i^T(t) = \int_0^t A_i(s)\lambda_i^T(s)ds$ . The Breslow-type

estimator is then given by  $\widehat{\Lambda}_0^T(t, \theta) = n^{-1} \sum_{i=1}^n \int_0^t dN_i^T(s) / R_T^{(0)}(s, \theta)$  with  $\widehat{\Lambda}_i^T(t, \theta) = \int_0^t A_i(s) \exp\{\theta' Z_i(s)\} d\widehat{\Lambda}_0^T(s, \theta)$ , where  $R_T^{(d)}(t, \theta) = n^{-1} \sum_{i=1}^n A_i(t) Y_i(t) Z_i(t)^{\otimes d} \exp\{\theta' Z_i(t)\}$ ,  $d = 0, 1, 2$ .  $z^{\otimes 0} = 1$ ,  $z^{\otimes 1} = z$  and  $z^{\otimes 2} = zz'$ , if  $z$  is a vector. The coefficient  $\theta_0$  is estimated via partial likelihood (Cox, 1975), by  $\widehat{\theta}$ , as the root of the score function,

$$U^T(\theta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t; \theta)\} dN_i^T(t),$$

where  $\overline{Z}(t; \theta) = R_T^{(1)}(t; \theta) / R_T^{(0)}(t; \theta)$ , and the constant  $\tau$  satisfies  $P(X_i \geq \tau) > 0$ .

We assume that death time  $D_i$  follows a stratified proportional hazards model, with hazard function

$$\lambda_{ij}(t) \equiv \lambda(t | G_i = j, Z_{i0}) = \lambda_{0j}(t) \exp\{\beta_0' Z_{i0}\},$$

where  $Z_{i0} = Z_i(0)$  and  $\lambda_{0j}(t)$  is the unspecified baseline hazard function for group  $j$ . Proportional hazards are not assumed across groups, but are assumed for the adjustment covariate,  $Z_{i0}$ .

The regression parameter for the adjustment covariate,  $\beta_0$ , can be estimated as the root of the stratified inverse-weighted score function,

$$(4.1) \quad U(\beta, W) = \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \overline{Z}_j(t; \beta, W)\} W_i(t) dN_{ij}(t),$$

where  $\overline{Z}_j(t; \beta, W) = R_j^{(1)}(t; \beta, W) / R_j^{(0)}(t; \beta, W)$ ,  $R_j^{(d)}(t; \beta, W) = n^{-1} \sum_{i=1}^n W_i(t) Y_{ij}(t) Z_{i0}^{\otimes d} \exp(\beta' Z_{i0})$  for  $d = 0, 1, 2$ . With some algebra,  $dN_{ij}(t)$  in (4.1) can be replaced by  $dM_{ij}(t) = dN_{ij}(t) - Y_{ij}(t) d\Lambda_{ij}(t)$ , such that the score function can be written as

$$U(\beta, W) = \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \overline{Z}_j(t; \beta, W)\} W_i(t) dM_{ij}(t).$$

A consistent estimator of  $\beta_0$  should satisfy  $E[W_i(t) dM_{ij}(t) | Z_{i0}] = 0$ . We can write  $dM_{ij}(t) = I(T_i > t) I(C_i > t) dM_{ij}^*(t)$ , where  $dM_{ij}^*(t) = I(D_i \geq t) [dN_{ij}^*(t) - d\Lambda_{ij}(t)]$  and  $N_{ij}^*(t) = I(D_i \leq t) G_{ij}$ . Under the assumed model,  $E[dM_{ij}^*(t) | Z_{i0}, G_i] = 0$ ,

which leads to  $E[I(T_i > t)dM_{ij}^*(t)|Z_{i0}, G_i] = 0$  in the case of independent censoring. In the presence of dependent censoring, without the IPCW term  $W_i(t)$ ,  $E[dM_{ij}(t)|Z_{i0}, G_i] \neq 0$  since  $E[dM_{ij}(t)|Z_{i0}, G_i] \neq E[I(T_i > t)I(C_i > t)|Z_{i0}, G_i]E[dM_{ij}^*(t)|Z_{i0}, G_i]$ . However, it can be shown that  $E[W_i(t)dM_{ij}(t)|Z_{i0}, Z_i(t), G_i] = 0$  and, after iterating the expectation, we have  $E[W_i(t)dM_{ij}(t)|Z_{i0}, G_i] = 0$ , such that the weighted score function, (4.1), is unbiased. The weighted Breslow estimator is then computed as

$$\widehat{\Lambda}_{0j}(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t \frac{W_i(s)dN_{ij}(s)}{R_j^{(0)}(t; \beta, W)},$$

and is a consistent estimator of  $\Lambda_{0j}(t)$  in the presence of dependent censoring.

However, in practice, some values of  $W_i(t)$  can be very large due to  $\widehat{\Lambda}_i^T(t)$  being unduly large, which can result in  $\widehat{\beta}$  and  $\widehat{\Lambda}_{0j}(t)$  being quite imprecise. In order to improve the precision of the estimate, similar to Robins and Finkelstein (2000) and Miloslavsky et al. (2004), a stabilized weight is define as

$$W_i^S(t) = Y_i(t) \frac{\exp\{\Lambda_i^T(t, \theta)\}}{\exp\{\Lambda_i^T(t|Z_{i0}, G_i)\}},$$

where  $\Lambda_i^T(t|Z_{i0}, G_i)$  is the cumulative hazard from a treatment hazards model which accounts for baseline covariate values only,  $\lambda_i^T(t|Z_{i0}, G_i) = \lambda_0^T(t|Z_{i0}, G_i) \exp\{\theta_1' Z_{i0}\}$  with  $\theta_1$  the effect coefficient. We refer to  $W_i(t)$  as “unstabilized” hereafter.

To compare each group to the reference group, the following measure is proposed,

$$\phi_j(t) = \frac{\Lambda_{ij}(t|Z_{i0} = z)}{\Lambda_{i0}(t|Z_{i0} = z)} = \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)},$$

for  $j = 0, 1, \dots, J - 1$ . The measure  $\phi_j(t)$  contrasts patients with same covariate pattern but in different groups. Note that if proportionality holds for groups, then  $\phi_j(t)$  reduces to a time invariant constant, which is commonly referred to as the hazard ratio. After obtaining the cumulative baseline hazard estimators and substituting

them into  $\phi_j(t)$ , we obtain the estimator

$$\widehat{\phi}_j(t) = \frac{\widehat{\Lambda}_{0j}(t, \widehat{\beta})}{\widehat{\Lambda}_{00}(t, \widehat{\beta})}, t \in [t_L, t_U],$$

where  $t_L$  is chosen to be large enough to avoid  $\widehat{\Lambda}_{00}(t, \widehat{\beta}) = 0$  and  $t_U$  is chosen to avoid instability estimate at the tail of the observation time distribution.

### 4.3 Asymptotic Properties

Regularity conditions are listed explicitly in Appendix C. In particular, we assume that the random vectors  $\{X_i, \Delta_i, \Delta_i^T, \widetilde{Z}_i(X_i), \widetilde{A}_i(X_i)\}$ , for  $i = 1 \dots n$ , are independent and identically distributed, with  $Z_i(t)$  bounded for  $t \in (0, \tau]$ , where  $\tau$  satisfies  $P(X_i \geq \tau) > 0$ . The asymptotic properties of the proposed methods are summarized in the following theorem.

**Theorem IV.1.** *Under certain regularity conditions,  $n^{1/2}\{\widehat{\phi}_j(t) - \phi_j(t)\}$  converges asymptotically to a zero-mean Gaussian process with covariance function  $\sigma_j(s, t) = E[\xi_{ij}(s)\xi_{ij}(t)]$ , where  $\{\xi_{1j}(t), \dots, \xi_{nj}(t)\}$  are i.i.d. with 0 mean asymptotically, with*

$$\xi_{ij}(t) = \frac{1}{\Lambda_{00}(t)}\Phi_{ij}(t) - \frac{\Lambda_{0j}(t)}{\Lambda_{00}^2(t)}\Phi_{i0}(t),$$

where

$$\begin{aligned} \Phi_{ij}(t) &= h_j'(t; \beta_0, W)\Omega^{-1}(\beta_0)U_i(\beta_0) + \int_0^t r_j^{(0)}(u; \beta_0, W)^{-1}W_i(u)dM_{ij}(u) \\ &\quad + [B_j'(t) + E_j'(t)]\Omega_T(\theta_0)^{-1}U_i^T(\theta_0) + \int_0^t [K_j(u, t) + P_j(u, t)]r_T^{(0)}(u; \theta_0)^{-1}dM_i^T(u), \\ h_j(t; \beta, W) &= - \int_0^t \bar{z}_j'(u; \beta, W)d\Lambda_{0j}(u), \\ r_j^{(0)}(t; \beta, W) &= E_i[W_i(t)Y_{ij}(t) \exp\{\beta'Z_{i0}\}], \\ r_T^{(0)}(t; \theta) &= E_i[Y_i(t) \exp\{\theta'Z_i(t)\}], \\ dM_i^T(t) &= dN_i^T(t) - Y_i(t)d\Lambda_i^T(t), \end{aligned}$$

with  $\Omega(\beta)$ ,  $U_i(\beta)$ ,  $B_j(t)$ ,  $E_j(t)$ ,  $\Omega_T(\theta)$ ,  $U_i^T(\theta)$ ,  $K_j(t_1, t_2)$  and  $P_j(t_1, t_2)$  defined in Appendix C.

The covariance can be estimated consistently by  $\hat{\sigma}_j(s, t) = n^{-1} \sum_{i=1}^n \hat{\xi}_{ij}(s) \hat{\xi}_{ij}(t)$ , where  $\hat{\xi}_i(t)$  is obtained by replacing all limiting values in  $\varphi_i$  by their empirical counterparts. A proof of Theorem III.1 is given in Appendix C. The proof basically shows that, asymptotically,  $n^{1/2}\{\hat{\phi}_j(t) - \phi_j(t)\} = n^{-1/2} \sum_{i=1}^n \hat{\xi}_{ij}(t) + o_p(1)$  through a sequence of Taylor series expansions.

The proof is for the unstabilized weight,  $W_i(t)$  while a stabilized version  $W_i^S(t)$  is preferred. As implied by Theorem III.1, the computation of the variance is quite complicated, and is obviously more complicated with the stabilizer. A computationally simpler variance treats the weight  $W_{ik}(t)$ , or  $W_{ik}^S(t)$ , as fixed in the asymptotics derivation. Then, the covariance estimator could simplify to  $\hat{\sigma}_j^*(s, t) = n^{-1} \sum_{i=1}^n \hat{\xi}_{ij}^*(s) \hat{\xi}_{ij}^*(t)$ , where

$$(4.2) \quad \xi_{ij}^*(t) = \frac{1}{\Lambda_{00}(t)} \Phi_{ij}^*(t) - \frac{\Lambda_{0j}(t)}{\Lambda_{00}^2(t)} \Phi_{i0}^*(t),$$

$$\Phi_{ij}^*(t) = h'_j(t; \beta_0, W) \Omega^{-1}(\beta_0) U_i(\beta_0) + \int_0^t r_j^{(0)}(u; \beta_0, W)^{-1} W_i(u) dM_{ij}(u).$$

This is the variance estimator that we suggest and evaluate through simulations described in the next section.

## 4.4 Simulation

We modify the setups of Zhang and Schaubel (2010) to generate data for the simulation study. First, a single baseline covariate  $Z_i$  is created as a truncated standard normal variable, truncated at -4 and 4. The group indicator  $G_i$  is then generated from a Bernoulli distribution with parameter  $\exp(-0.6Z_i)/\{1 + \exp(-0.6Z_i)\}$ . Death time  $D_i$  is generated by transforming  $\epsilon_{1i} \sim Uniform(0, 1)$  using the inverse of the



cumulative distribution function of a Weibull distribution with hazard function

$$(4.3) \quad \lambda_{ij}(t) = \alpha_j \gamma_j t^{\gamma_j - 1} \exp\{\beta'_0 Z_i\},$$

for subject  $i = 1, \dots, n$  and group  $j = 0, 1$ . Unless  $\gamma_0 = \gamma_1$ , proportionality does not hold across groups. Different values of  $\gamma_j$ 's are chosen to lead to a constant, increasing or decreasing true cumulative hazard ratio function  $\phi_j(t)$ . After that, we generate a time-dependent covariate which correlates with both death time  $D_i$  and treatment time  $T_i$  conditional on  $(G_i, Z_i)$ . First let  $V_{ti} = -2 \log\{G_i \epsilon_{1i} + (1 - G_i)(1 - \epsilon_{1i})\} + \epsilon_{2i}$ , where  $\epsilon_{2i} \sim \text{Uniform}(0, 1)$ . Define  $V_i(t) = I(V_{ti} \geq t)$ . Thus  $V_i(t)$  is correlated with event time  $D_i$  when conditional on  $(G_i, Z_i)$ . Then,  $T_i$  is generated to follow the proportional hazards model,

$$(4.4) \quad \lambda_i^T(t) = \lambda_0^T(t) \exp\{\theta_0 + \theta_G G_i + \theta_Z Z_i + \theta_1 V_i(t)\},$$

such that  $V_i(t)$  correlates with  $T_i$ . Consequently  $D_i$  is dependently censored by  $T_i$ , even conditional on  $(G_i, Z_i)$ . To simplify the setup, administrative censoring time  $C_i$  is not involved in the simulation study. Data pertaining to death time  $\{X_i, \Delta_i, Z_i\}$  is used to fit model (4.3), with time to transplant data  $\{X_i, \Delta_i^T, Z_i, V_i(X_i)\}$  used to fit model (4.4).

We evaluate samples of size  $n = 500, 250, 100$  and treatment percentages as 30% and 10%. The parameters  $[\beta_0, \alpha_0, \alpha_1, \theta_G, \theta_Z, \theta_1]$  are set to  $[0.3, 0.2, 0.4, 0.2, 0.2, 0.3]$ . Different values of  $\theta_0$  are used to vary the censoring percentage. The vector  $[\gamma_0, \gamma_1]$  was set to each of  $[1.5, 1.5]$ ,  $[1.5, 1.25]$  and  $[1, 1.5]$ , to make  $\phi_1(t) = \Lambda_{01}(t)/\Lambda_{00}(t)$  a constant, an increasing function and a decreasing function respectively. The properties of  $\phi_1(t)$  are studied at time  $t = 1$ ,  $t = 2$  and  $t = 3$ . Both the unstabilized  $W_i(t)$  and stabilized  $W_i^S(t)$  are applied to the simulated data. We use the simplified variance estimator which treats the IPCW weights as fixed; i.e., as given in (3.3).

Tables 4.1, 4.2 and 4.3, show the results when the true ratio is increasing, decreasing and constant, respectively. All estimates are approximately unbiased. The magnitude of the finite-sample bias generally increases as the sample size decreases. There is no obvious association between the degree of bias and treatment percentage. On one hand, one might expect that higher treatment percentages should increase bias since the dependent censoring aspect figures move prominently in the estimating function. However, on the other hand, greater number of observed treatments should generally result in modeling  $\Lambda_i^T(t)$  more precisely and hence to diminish the bias. Generally, the average asymptotic standard error (ASE) is close to empirical standard error (ESE). The coverage probabilities (CP) with  $W_i^S(t)$  are generally closer to 95% than those based on  $W_i(t)$ . The CP tended to decrease as the sample size decreased.

## 4.5 Application

We applied the proposed methods to ESLD patients initially placed on the wait list for diseased donor liver transplantation between March 1, 2002 and December 31, 2009 in United States. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). Only patients age  $\geq 18$  at listing were included in our study sample, and we excluded retransplant candidates.

In order to construct the IPCW weight, a time-dependent stratified Cox model which takes transplant as the event was adopted, such that

$$\lambda_{ir}^T(t) = A_i(t)\lambda_{0r}^T(t) \exp\{\theta'_0 Z_i(t)\},$$

where the subscript  $r = 1, \dots, 11$  represents United Network for Organ Sharing (UNOS) Region. The indicator  $A_i(t)$  reflects that patients do not contribute to the estimation of  $\theta_0$  or  $\Lambda_{0r}^T(t)$ , while they are removed from or inactive on the waiting

Table 4.1: Simulation results for  $\widehat{\phi}_1(t)$  when  $\phi_1(t)$  increases in  $t$ .

$n$	$E(\Delta_i^T)$	$t$	$\widehat{W}_i(t)$						$W_i^S(t)$					
			BIAS	ASE	ESE	CP	BIAS	ASE	ESE	CP	BIAS	ASE	ESE	CP
500	10%	1	0.008	0.095	0.099	0.94	0.008	0.094	0.097	0.94	0.008	0.094	0.097	0.94
		2	0.003	0.080	0.078	0.95	0.003	0.078	0.078	0.95	0.003	0.078	0.078	0.95
		3	0.001	0.081	0.081	0.95	0.001	0.075	0.081	0.93	0.001	0.075	0.081	0.93
	30%	1	0.007	0.098	0.103	0.95	0.007	0.096	0.103	0.93	0.007	0.096	0.103	0.93
		2	0.004	0.088	0.091	0.94	0.004	0.082	0.091	0.92	0.004	0.082	0.091	0.92
		3	0.004	0.093	0.095	0.94	0.004	0.081	0.095	0.90	0.004	0.081	0.095	0.90
250	10%	1	0.012	0.135	0.138	0.94	0.012	0.134	0.138	0.94	0.012	0.134	0.138	0.94
		2	0.010	0.115	0.116	0.95	0.010	0.111	0.116	0.94	0.010	0.111	0.116	0.94
		3	0.009	0.115	0.122	0.93	0.009	0.108	0.122	0.91	0.009	0.108	0.122	0.91
	30%	1	0.018	0.142	0.146	0.95	0.018	0.140	0.146	0.94	0.018	0.140	0.146	0.94
		2	0.016	0.126	0.128	0.94	0.016	0.118	0.127	0.93	0.016	0.118	0.127	0.93
		3	0.011	0.132	0.136	0.94	0.011	0.115	0.135	0.91	0.011	0.115	0.135	0.91
100	10%	1	0.043	0.227	0.256	0.91	0.044	0.225	0.256	0.91	0.044	0.225	0.256	0.91
		2	0.029	0.186	0.204	0.91	0.030	0.180	0.204	0.91	0.030	0.180	0.204	0.91
		3	0.021	0.183	0.200	0.92	0.021	0.171	0.201	0.90	0.021	0.171	0.201	0.90
	30%	1	0.039	0.237	0.258	0.90	0.039	0.232	0.258	0.90	0.039	0.232	0.258	0.90
		2	0.023	0.201	0.214	0.92	0.024	0.189	0.213	0.92	0.024	0.189	0.213	0.92
		3	0.027	0.209	0.222	0.92	0.027	0.186	0.221	0.90	0.027	0.186	0.221	0.90

Table 4.2: Simulation results for  $\widehat{\phi}_1(t)$  when  $\phi_1(t)$  decreases in  $t$ .

$n$	$E(\Delta_i^T)$	$t$	$W_i(t)$			$W_i^S(t)$			CP	CP
			BIAS	ASE	ESE	BIAS	ASE	ESE		
500	10%	1	0.009	0.093	0.095	0.009	0.093	0.095	0.95	0.95
		2	0.004	0.049	0.051	0.004	0.048	0.051	0.94	0.94
		3	0.003	0.037	0.039	0.003	0.036	0.039	0.93	0.93
	30%	1	0.005	0.096	0.096	0.005	0.095	0.096	0.95	0.95
		2	0.003	0.053	0.054	0.003	0.050	0.054	0.93	0.93
		3	0.002	0.041	0.042	0.002	0.037	0.041	0.91	0.91
250	10%	1	0.013	0.133	0.138	0.014	0.133	0.139	0.94	0.94
		2	0.005	0.070	0.072	0.005	0.069	0.072	0.94	0.94
		3	0.003	0.052	0.053	0.003	0.050	0.053	0.93	0.93
	30%	1	0.008	0.137	0.136	0.009	0.135	0.137	0.94	0.94
		2	0.005	0.075	0.078	0.005	0.071	0.077	0.93	0.93
		3	0.001	0.057	0.059	0.001	0.052	0.058	0.91	0.91
100	10%	1	0.039	0.223	0.234	0.039	0.222	0.234	0.92	0.92
		2	0.008	0.111	0.118	0.008	0.109	0.118	0.91	0.91
		3	0.007	0.082	0.089	0.007	0.080	0.089	0.92	0.92
	30%	1	0.048	0.235	0.272	0.048	0.232	0.270	0.90	0.90
		2	0.017	0.121	0.129	0.017	0.115	0.129	0.92	0.92
		3	0.014	0.093	0.102	0.014	0.084	0.100	0.90	0.90

Table 4.3: Simulation results for  $\widehat{\phi}_1(t)$  when  $\phi_1(t)$  is constant in  $t$ .

$n$	$E(\Delta_i^T)$	$t$	$\widehat{W}_i(t)$						$\widehat{W}_i^S(t)$					
			BIAS	ASE	ESE	CP	BIAS	ASE	ESE	CP	BIAS	ASE	ESE	CP
500	10%	1	0.007	0.094	0.098	0.94	0.007	0.093	0.098	0.94	0.007	0.093	0.098	0.94
		2	0.002	0.066	0.066	0.95	0.002	0.064	0.066	0.93	0.002	0.064	0.066	0.93
		3	0.007	0.062	0.065	0.93	0.006	0.058	0.065	0.92	0.006	0.058	0.065	0.92
	30%	1	0.011	0.099	0.100	0.94	0.012	0.097	0.100	0.94	0.012	0.097	0.100	0.94
		2	0.006	0.073	0.076	0.94	0.007	0.068	0.076	0.93	0.007	0.068	0.076	0.93
		3	0.007	0.071	0.076	0.93	0.007	0.062	0.075	0.90	0.007	0.062	0.075	0.90
250	10%	1	0.020	0.137	0.146	0.94	0.020	0.136	0.146	0.94	0.020	0.136	0.146	0.94
		2	0.009	0.094	0.099	0.93	0.009	0.091	0.099	0.92	0.009	0.091	0.099	0.92
		3	0.005	0.084	0.091	0.93	0.005	0.081	0.091	0.92	0.005	0.081	0.091	0.92
	30%	1	0.014	0.142	0.149	0.93	0.014	0.139	0.149	0.93	0.014	0.139	0.149	0.93
		2	0.010	0.104	0.106	0.95	0.010	0.096	0.105	0.94	0.010	0.096	0.105	0.94
		3	0.012	0.101	0.107	0.94	0.012	0.088	0.105	0.90	0.012	0.088	0.105	0.90
100	10%	1	0.027	0.219	0.228	0.92	0.027	0.218	0.228	0.92	0.027	0.218	0.228	0.92
		2	0.022	0.151	0.155	0.93	0.022	0.147	0.155	0.93	0.022	0.147	0.155	0.93
		3	0.013	0.136	0.149	0.91	0.013	0.128	0.149	0.90	0.013	0.128	0.149	0.90
	30%	1	0.049	0.239	0.265	0.91	0.050	0.235	0.265	0.91	0.050	0.235	0.265	0.91
		2	0.020	0.165	0.174	0.90	0.021	0.154	0.174	0.90	0.021	0.154	0.174	0.90
		3	0.022	0.157	0.181	0.91	0.022	0.138	0.181	0.87	0.022	0.138	0.181	0.87

list. The covariate,  $Z_i(t)$ , included MELD score, Status 1, albumin, age, gender, race, diagnosis of Hepatitis C, body mass index, diabetes, hospitalization, blood type, dialysis within prior week, encephalopathy, ascites and serum creatinine. We used two types of weights  $W_i(t)$  and  $W_i^S(t)$ . Even for the stabilized weight, some extreme values existed. Since 99% of the estimated weights were less than 10, the weights were then capped at 10.

The model of main interest (that for the wait list death hazard) is given by

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\beta'_0 Z_{i0}\},$$

where the subscript  $j = 0, \dots, 7$  corresponding to Status-1 ( $j = 0$ ; reference) and MELD score (21-23, 24-26, 27-29, 30-32, 33-35, 36-40, >40). The baseline covariate vector,  $Z_{i0}$ , included terms representing albumin, age, gender, race, diagnosis, body mass index, diabetes, hospitalization status, if dialysis within prior week, and previous malignancy. Although the MELD score is capped at 40 for allocation purposes, in our analysis, calculated MELD scores over 40 were collected into a > 40 group.

A total of 66,884 subjects were eligible for our study. We focused on comparing the MELD > 20 to Status 1 patients. The cumulative hazard ratio of high-MELD patient to Status-1 patient are estimated up to 5 years. Restricting attention to Status 1 and MELD > 20, the sample size was  $n = 16,684$ .

Since the results with  $W_i(t)$  are similar to those with  $W_i^S(t)$ , we only show the stabilized version of hazard ratios. The estimated curves of hazard ratio, comparing high-MELD to Status 1, are shown in Figure 4.1. Figures 4.2, 4.3 and 4.4 provide the estimate together with the point-wise confidence interval for MELD > 40, MELD 36-40 and MELD 21-23 versus Status-1 patients respectively. The curve of MELD > 40 patients is below the reference line of Status 1 in the short run; e.g., within 14

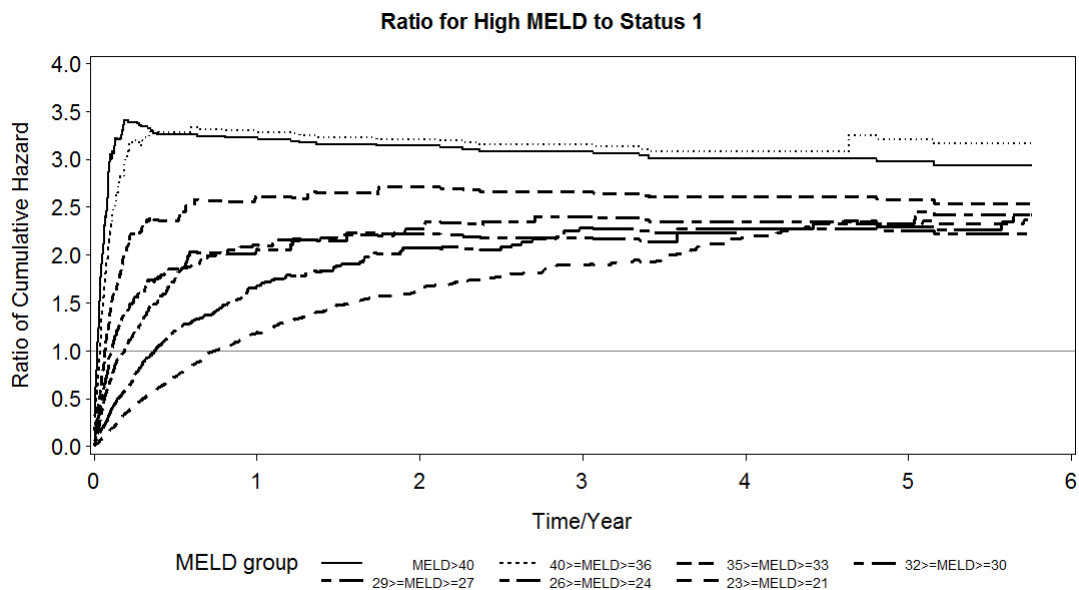


Figure 4.1: Cumulative hazard ratio of high-MELD patients compared to Status 1 patients.

days. But afterward, the curve crosses the reference line and climbs to  $\hat{\phi}_i(t) \approx 3$  and remains flat for the remainder study follow-up. Other MELD categories have similar curve patterns over time, but with different crossing times, varying from 14 to 250 days.

It is interesting that Status 1 patients have a lower mortality compared to high-MELD patients. In practice, acute patients automatically lose their Status 1 designation after 14 days, since they are very likely to die within 2 weeks. In fact, it is patients must apply to have their Status 1 classification re-certified after 7 days. It is assumed that if they were ill enough the warrant Status 1 designation, they would die within one week in the absence of a liver transplant. However, it appears that a large portion of patients who were Status 1 at baseline live for years after wait listing.

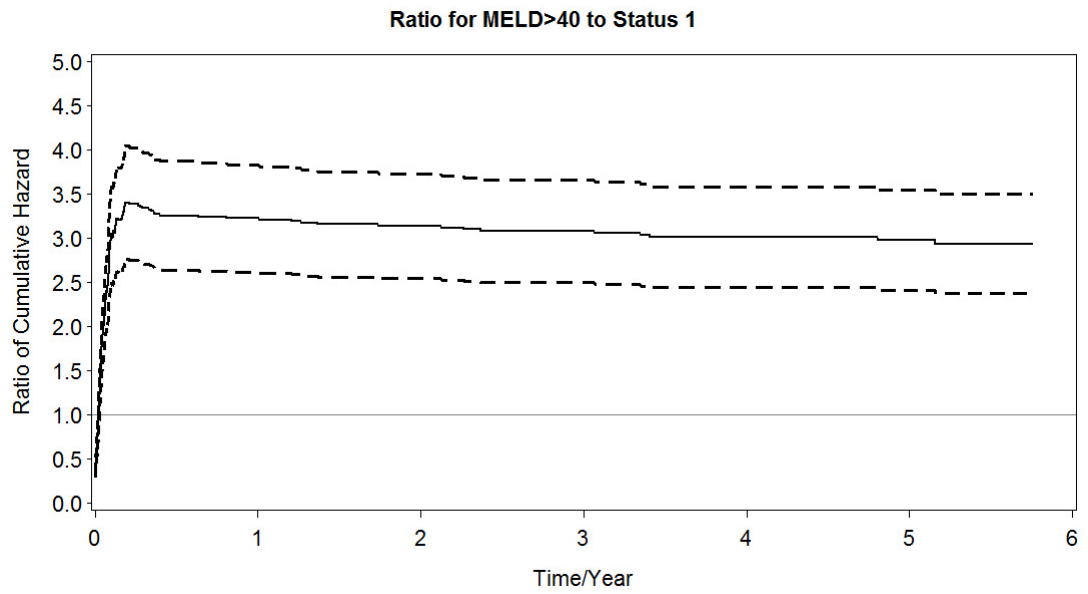


Figure 4.2: Cumulative hazard ratio for MELD > 40 patients compared to Status 1 patients, with 95% confidence interval.

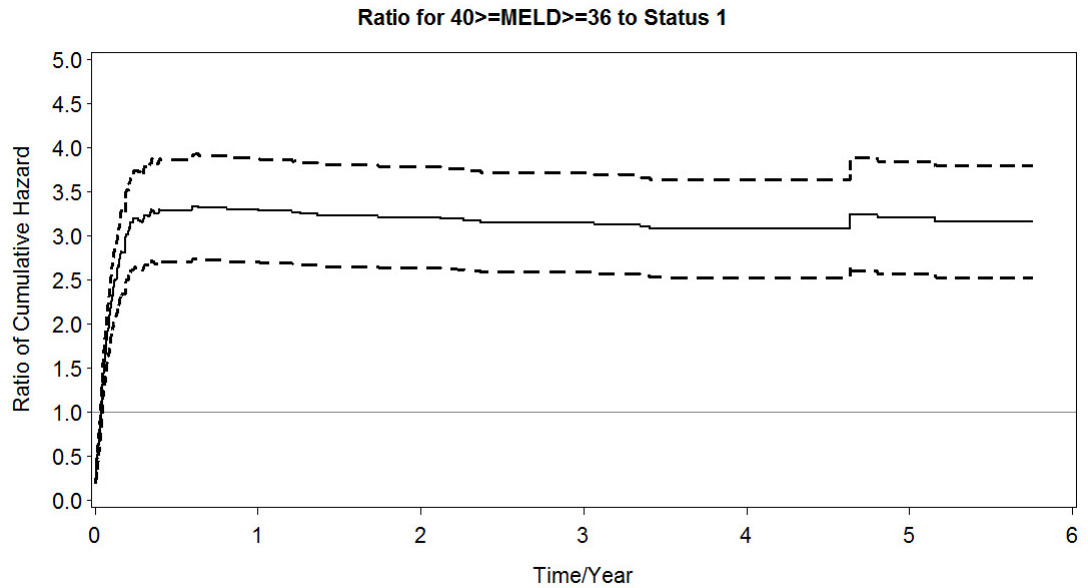


Figure 4.3: Cumulative hazard ratio for MELD 36-40 patients compared to Status 1 patients, with 95% confidence interval.



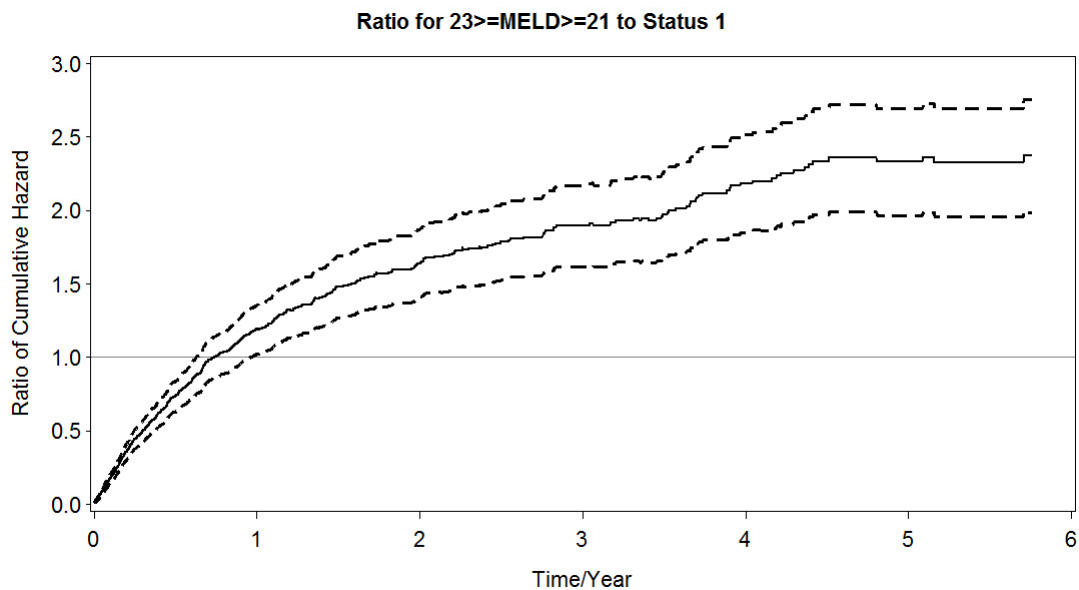


Figure 4.4: Cumulative hazard ratio for MELD 21-23 patients compared to Status 1 patients, with 95% confidence interval.

## 4.6 Discussion

In this article, we use the ratio of group-specific cumulative baseline hazard functions as a time-dependent measure of the covariate-adjusted effect of group on mortality. Under the target data structure, covariates are time-varying. The death hazard model of interest adjusts for only time 0 covariates, such that death may be dependently censored due to mutual correlation between the values of the time-varying covariates (after time 0) with death and censoring times. IPCW is used to obtain unbiased estimators, with a stabilized weight suggested. The fact that subjects may experience time intervals of treatment-ineligibility is handled by fitting a treatment hazard model that is conditional on being treatment-eligible, with a subject's ineligible periods then zeroed out in the probability-of-remaining-untreated calculation. Simulations show that the proposed methods work well in reasonable sized samples. Finally, the proposed methods are applied in order to contrast liver

wait list mortality between acute and chronic liver failure patients.

Very few comparisons between the mortality of Status 1 and high-MELD patients have been carried out, primarily since it is generally assumed by the liver transplant community that mortality is much higher for acute relative to chronic liver failure patients. However, based on a conventional time-dependent model and capping follow-up at 14 days, Sharma et al. (2012) reported that wait list mortality for high-MELD patients was very comparable with that of Status 1 patients. Moreover, Sharma et al (2012) found that  $\text{MELD} \geq 40$  was associated with significantly higher two-week mortality than Status 1. The analysis in this chapter provides further evidence that the prioritization for deceased-donor liver transplantation of Status 1 over  $\text{MELD} > 40$  patients needs to be reconsidered.

Various methods related to those proposed in this chapter can be found in the existing literature, although none estimate the quantity of interest under the assumed data structure. In particular, Hernan, Brumback, and Robins (2000, 2001) developed methods for estimating the average causal effect in terms of the hazard ratio. Although the methods allow for treatment to be time-varying, proportional hazards are assumed under the marginal structure model employed. Chen and Tsiatis (2001); Schaubel and Wei (2011); Zhang and Schaubel (2010) used restricted mean survival time as a measure; such that the cumulative effect is captured. Each of these methods estimates an average causal effect, which would be of interest in many but not all settings. In many applications (such as our motivating example), a conditional group effect is of greater interest. The estimator proposed in this Chapter is analogous to that of Wei and Schaubel (2008), but allows for a richer covariate structure and more general censoring patterns.

The motivating example considers treatment-free survival. It would be of interest

to extend the proposed methods to estimate treatment effects.

## CHAPTER V

### Conclusion

In this dissertation, three novel semiparametric methods were developed for estimating the effect of either a longitudinal covariate or a time-dependent treatment on survival. Each of the methods was motivated by issues in the analysis of liver transplant data. The methods are intended for the analysis of observational data and can accommodate dependent censoring.

In Chapter II, we developed semiparametric methods for estimating the effect of a time-dependent covariate on treatment-free survival. The data structure of interest consists of a longitudinal sequence of measurements and a potentially censored survival time. The factor of chief interest is time-dependent. Treatment-free survival is of interest and is dependently censored by the receipt of treatment. A further complication is that patients may be removed from consideration for treatment, either temporarily or permanently. The proposed methods involve landmark analysis and partly conditional hazard regression. Dependent censoring was overcome by a variant of Inverse Probability of Censoring Weighting (IPCW). Application of the proposed methods to national organ transplant registry data revealed that the sickest chronic end-stage liver disease patients have significantly greater pre-transplant mortality than acute liver failure patients. The method may be challenging to apply

to large data sets with long follow-up, due to computational intensity.

In Chapter III, we developed semiparametric methods for estimating the average difference between treatment-free and post-treatment restricted mean lifetime. The underlying data structure was the same as in Chapter II, except that post-treatment mortality was also considered. Methods from Chapter II were used to model treatment-free survival. The post-treatment death hazard was modeled using the time-dependent covariate history leading up to the treatment time. The average treatment effect was obtained by averaging over treated subjects, in a manner which accounts for the competing risks structure and the independent censoring of time to treatment. Through the proposed methods, the mean survival benefit of deceased-donor liver transplantation, based on 5-year restricted mean lifetime, was estimated by Model for End-stage Liver Disease (MELD) score. The methods entail inverse weights for each of the treatment and independent censoring time; with consistency requiring the correctness of both such models.

In time to event data observed in medical studies, nonproportional hazards and dependent censoring are common issues when comparing group-specific mortality. The group effect on mortality might vary over time, as opposed to being constant. In settings where the proportional hazard assumption fails, investigators tend to be more interested in cumulative (as opposed to instantaneous) effects on mortality. Therefore, in Chapter IV, we developed an estimator for the cumulative group effect on survival in the presence of nonproportional hazards and dependent censoring. The proposed estimator is based on the cumulative hazard function, assumed to follow a stratified Cox model. No functional form need be assumed for the nonproportionality. Through the proposed methods, it was revealed that acute liver failure patients have lower pre-transplant survival than high-MELD patients, but only in the short-term.

The use of stratification for the proposed methods provides much flexibility, but may lead to instability for smaller data sets.

Each of the proposed methods was shown to be consistent and asymptotically normal, with consistent covariance estimators provided. Simulation studies revealed that the proposed estimation procedures were appropriate for practical use.

## APPENDICES

## APPENDIX A

### Appendix A: Proof of Theorem II.1

#### Appendix A: Proof of Theorem II.1

##### A.I. Notation

We begin by reviewing the essential notation:

$i$ : subject ( $i = 1, \dots, n$ )

$n$ : number of subjects

$k$ : cross section ( $k = 1, \dots, K$ )

$D_i$ : death time for the  $i$ th subject

$C_i$ : censoring time for the  $i$ th subject

$T_i$ : treatment time for the  $i$ th subject

$X_i = \min(D_i, C_i, T_i)$ , observed time for the  $i$ th subject

$\Delta_i = I(X_i = D_i)$

$\Delta_i^T = I(X_i = T_i)$

$CS_k$ :  $k$ th cross-section

$S_{ik}$ : follow-up time for  $i$ th subject at calendar date of the  $k$ th cross-section date

$A_i(t)$ : treatment eligibility indicator of  $i$ th subject at time  $t$ ,  $A_{ik} = A_i(S_{ik})$



$\tilde{A}_i(t) = \{A_i(s); s \in [0, t]\}$ : treatment eligibility history up to time  $t$

$Z_i(t)$ : time dependent covariate of the  $i$ th subject

$$Z_{ik} = Z_i(S_{ik})$$

$\tilde{Z}_i(t) = \{Z_i(s); s \in [0, t]\}$ : covariate history up to time  $t$

$D_{ik} = D_i - S_{ik}$ , death time for the  $i$ th subject at date of the  $k$ th cross section

$T_{ik} = T_i - S_{ik}$ , treatment time for the  $i$ th subject at date of the  $k$ th cross section

$C_{ik} = C_i - S_{ik}$ , independent censoring time for the  $i$ th subject at date of the  $k$ th cross section

$\beta_0$ : parameter coefficient of the death model

$\theta_0$ : parameter coefficient of the treatment model

Death hazard;  $\lambda_{ik}(t) = \lambda_{0k}(t) \exp\{\beta'_0 Z_{ik}\}$

Treatment hazard;  $\lambda_i^T(t) = A_i(t) \lambda_0^T(t) \exp\{\theta'_0 Z_i(t)\}$

## A.II. Regularity Conditions

In deriving the asymptotic properties of the proposed estimators the following conditions are assumed for  $i = 1, \dots, n$  and  $k = 1, \dots, K$

(a)  $\{X_i, \Delta_i, \Delta_i^T, \tilde{Z}_i(X_i), \tilde{A}_i(X_i \wedge T_i)\}$  are independent and identically distributed random vectors.

(b)  $|Z_{il}(t)| < \kappa_l$ , for  $t \in [0, \tau]$  and  $Z_{il}(t)$  is the  $l$ th element of  $Z_i(t)$ .

(c)  $\int_0^\tau \lambda_{0k}(t) dt < \infty$  and  $\int_0^\tau \lambda_0^T(t) dt < \infty$  where  $\tau$  is the maximum follow-up time.

(d) Continuity of the following functions:

$$\begin{aligned} r_k^{(1)}(t; \beta, W) &= \frac{\partial}{\partial \beta} r_k^{(0)}(t; \beta, W), \\ r_k^{(2)}(t; \beta, W) &= \frac{\partial^2}{\partial \beta \partial \beta'} r_k^{(0)}(t; \beta, W), \end{aligned}$$

and  $r_k^{(0)}(t; \beta, W)$ , where

$$r_k^{(p)}(t; \beta, W) = E[W_{ik}^A(t) Y_{ik}(t) Z_{ik}^{\otimes p} \exp(\beta' Z_{ik})],$$

is the limiting value of

$$R_k^{(p)}(t; \beta, W) = n^{-1} \sum_{i=1}^n W_{ik}^A(t) Y_{ik}(t) Z_{ik}^{\otimes p} \exp(\beta' Z_{ik}),$$

for  $p = 0, 1, 2$ , with  $r_k^{(1)}(t; \beta, W)$  and  $r_k^{(2)}(t; \beta, W)$  bounded and  $r_k^{(0)}(t; \beta, W)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\beta$  in an open set.

(e) Continuity of the following functions:

$$\begin{aligned} r_T^{(1)}(t; \theta) &= \frac{\partial}{\partial \theta} r_T^{(0)}(t; \theta), \\ r_T^{(2)}(t; \theta) &= \frac{\partial^2}{\partial \theta \partial \theta'} r_T^{(0)}(t; \theta), \end{aligned}$$

and  $r_T^{(0)}(t; \theta)$ , where

$$r_T^{(p)}(t; \theta) = E[A_i(t) Y_i(t) Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}],$$

is the limiting value of

$$R_T^{(p)}(t; \theta) = n^{-1} \sum_{i=1}^n A_i(t) Y_i(t) Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\},$$

for  $p = 0, 1, 2$ , with  $r_T^{(1)}(t; \theta)$  and  $r_T^{(2)}(t; \theta)$  bounded and  $r_T^{(0)}(t; \theta)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\theta$  in an open set.

(f) Positive-definiteness of the matrices  $\Omega_T(\theta)$  and  $\Omega(\beta)$ , where

$$\begin{aligned} \Omega_T(\theta) &= E \left[ \int_0^\tau \left\{ \frac{r_T^{(2)}(t; \theta)}{r_T^{(0)}(t; \theta)} - \bar{z}(t; \theta)^{\otimes 2} \right\} dN_i^T(t) \right], \\ \bar{z}(t; \theta) &= r_T^{(1)}(t; \theta) / r_T^{(0)}(t; \theta), \\ \Omega(\beta) &= E \left[ \sum_{k=1}^K \int_0^{\tau_k} \left\{ \frac{r_k^{(2)}(t; \beta, W)}{r_k^{(0)}(t; \beta, W)} - \bar{z}_k(t; \beta, W)^{\otimes 2} \right\} dN_{ik}(t) \right], \\ \bar{z}_k(t; \beta, W) &= r_k^{(1)}(t; \beta, W) / r_k^{(0)}(t; \beta, W), \end{aligned}$$

(g)  $P(Y_{ik}(t) = 1) > 0$  for  $t \in [0, \tau_k]$ .

### A.III. Outline of Asymptotic Derivation

We derive the influence functions of terms of interest as summations of independent and identical distributed (i.i.d.) terms plus a term which converges to zero in probability. Inverse weighting is involved in the below derivation and the proof focuses on the Type C weight. The derivation consists of several parts in which the quantities are approximated by a summation of i.i.d. terms.

1.  $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$
2.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\}$
3.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_i^T(t) - \Lambda_i^T(t)\}$
4.  $n^{\frac{1}{2}}\{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\}$
5.  $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0)$

### A.IV. Derivation of Asymptotic Properties

Several parts of the proof regarding the proportional hazards model are well-established results. Therefore, they are simply listed without proof. For details, please refer to Anderson and Gill (1982), Fleming and Harrington (1991) and Andersen et al. (1993).

#### A.IV.1 $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$

As  $n \rightarrow \infty$ , we have

$$n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) = \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) + o_p(1),$$

where

$$U_i^T(\theta) = \int_0^\tau \{Z_i(t) - \bar{z}(t; \theta)\} dM_i^T(t; \theta),$$

$$dM_i^T(t) = dN_i^T(t) - Y_i(t) d\Lambda_i^T(t),$$

$$\mathbf{A.IV.2} \quad n^{\frac{1}{2}} \{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\}$$

We induce the following decomposition:

$$(A.1) \quad \begin{aligned} & n^{\frac{1}{2}} \{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\} \\ &= n^{\frac{1}{2}} \{\widehat{\Lambda}_0^T(t; \widehat{\theta}) - \Lambda_0^T(t; \theta_0)\} \end{aligned}$$

$$(A.2) \quad + n^{\frac{1}{2}} \{\widehat{\Lambda}_0^T(t; \theta_0) - \Lambda_0^T(t)\}.$$

We can express the first term as

$$(A.1) \quad \begin{aligned} &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_T^{(0)}(u; \widehat{\theta})^{-1} - R_T^{(0)}(u; \theta_0)^{-1}\} dN_i^T(u) \\ &= \widehat{h}'_T(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) \\ &= h'_T(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) + o_p(1). \end{aligned}$$

where the third line follows from the convergence in probability of

$$\widehat{h}'_T(t; \theta) = -\frac{1}{n} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \theta)^{-1} \bar{Z}(u; \theta) dN_i^T(u) = -\int_0^t \bar{Z}(u; \theta) d\widehat{\Lambda}_0^T(u; \theta),$$

$$\widehat{\Omega}_T(\theta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{R_T^{(2)}(t; \theta)}{R_T^{(0)}(t; \theta)} - \bar{Z}(t, \theta)^{\otimes 2} \right\} dN_i^T(t),$$

where  $\bar{Z}(t; \theta) = R_T^{(1)}(t; \theta) / R_T^{(0)}(t; \theta)$ , to the quantities

$$h'_T(t; \theta) = -\int_0^t \bar{z}(u; \theta) d\Lambda_0^T(u),$$

and  $\Omega_T(\theta)$  respectively, with  $\Omega_T(\theta)$  defined in Regularity Condition (f).

With respect to the second term in the decomposition, we have,

$$\begin{aligned}
(A.2) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) + o_p(1),
\end{aligned}$$

where the second line follows from continuity and Condition (f). Combining results, for the decomposition, we have

$$n^{\frac{1}{2}} \{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_i^T(t; \theta_0) + o_p(1),$$

where

$$\Phi_i^T(t; \theta) = h'_T(t; \theta) \Omega_T(\theta)^{-1} U_i^T(\theta) + \int_0^t r_T^{(0)}(u; \theta)^{-1} dM_i^T(u) = \int_0^t d\Phi_i^T(u; \theta),$$

and

$$d\Phi_i^T(u; \theta) = -\mathcal{Z}'(u; \theta) d\Lambda_0^T(u) \Omega_T(\theta)^{-1} U_i^T(\theta) + r_T^{(0)}(u; \theta)^{-1} dM_i^T(u).$$

$$\mathbf{A.IV.3} \quad n^{\frac{1}{2}} \{\widehat{\Lambda}_i^T(t) - \Lambda_i^T(t)\}$$

We begin with another decomposition,

$$\begin{aligned}
&n^{\frac{1}{2}} \{\widehat{\Lambda}_i^T(t) - \Lambda_i^T(t)\} \\
(A.3) \quad &= n^{\frac{1}{2}} \left\{ \int_0^t e^{\widehat{\theta}' Z_i(u)} d\widehat{\Lambda}_0^T(u) - \int_0^t e^{\theta'_0 Z_i(u)} d\widehat{\Lambda}_0^T(u) \right\} \\
(A.4) \quad &+ n^{\frac{1}{2}} \left\{ \int_0^t e^{\theta'_0 Z_i(u)} d\widehat{\Lambda}_0^T(u) - \int_0^t e^{\theta'_0 Z_i(u)} d\Lambda_0^T(u) \right\}.
\end{aligned}$$

Considering the first term,

$$(A.3) = n^{\frac{1}{2}} \int_0^t \{e^{\widehat{\theta}' Z_i(u)} - e^{\theta'_0 Z_i(u)}\} d\widehat{\Lambda}_0^T(u).$$

By a Taylor series expansion,

$$\begin{aligned}
n^{\frac{1}{2}} \{e^{\widehat{\theta}' Z_i(u)} - e^{\theta'_0 Z_i(u)}\} &= Z'_i(u) e^{\theta'_0 Z_i(u)} n^{\frac{1}{2}} (\widehat{\theta} - \theta) + o_p(1) \\
&= Z'_i(u) e^{\theta'_0 Z_i(u)} \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) + o_p(1).
\end{aligned}$$

Since  $\widehat{\Lambda}_0^T(t) \xrightarrow{p} \Lambda_0^T(t)$  for  $t \in [0, \tau]$ , we obtain

$$(A.3) = \int_0^t Z_i'(u) d\Lambda_i^T(u) \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) + o_p(1).$$

By using Result 4.2, the second term can be written as

$$(A.4) = n^{\frac{1}{2}} \int_0^t e^{\theta_0' Z_i(u)} d\{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\} \\ = \int_0^t e^{\theta_0' Z_i(u)} n^{-\frac{1}{2}} \sum_{l=1}^n d\Phi_l^T(u; \theta_0) + o_p(1).$$

Combining results from the decomposition leads to

$$n^{\frac{1}{2}} \{\widehat{\Lambda}_i^T(t) - \Lambda_i^T(t)\} = \int_0^t \{Z_i(u) - \bar{z}(u; \theta_0)\}' d\Lambda_i^T(u) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\ + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t e^{\theta_0' Z_i(u)} r_{Tl}^{(0)}(u; \theta_0)^{-1} dM_l^T(u) + o_p(1) \\ = D_i'(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{il}^T(t; \theta_0) + o_p(1),$$

where we define

$$D_i(t; \theta) = \int_0^t \{Z_i(u) - \bar{z}(u; \theta)\}' d\Lambda_i^T(u) = \int_0^t dD_i(u; \theta), \\ J_{il}^T(t; \theta) = \int_0^t e^{\theta' Z_i(u)} r_{Tl}^{(0)}(u; \theta)^{-1} dM_l^T(u).$$

$$\mathbf{A.IV.4} \quad n^{\frac{1}{2}} \{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\}$$

When the subscript of quantities doesn't involve the cross section notation  $k$ ,  $t$  refers the time from study entry. If  $k$  is present in the subscript, then  $t$  denotes the time from cross section date.

Since  $W_{ik}^A(t) = e^{\Lambda_i^T(t+S_{ik}) - \Lambda_i^T(S_{ik})}$  and  $\widehat{W}_{ik}^A(t) = e^{\widehat{\Lambda}_i^T(t+S_{ik}) - \widehat{\Lambda}_i^T(S_{ik})}$ , we then have

$$n^{\frac{1}{2}} \{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\} \\ = n^{\frac{1}{2}} \{e^{\widehat{\Lambda}_i^T(t+S_{ik}) - \widehat{\Lambda}_i^T(S_{ik})} - e^{\Lambda_i^T(t+S_{ik}) - \Lambda_i^T(S_{ik})}\} \\ = W_{ik}^A(t) n^{\frac{1}{2}} [\{\widehat{\Lambda}_i^T(t+S_{ik}) - \Lambda_i^T(t+S_{ik})\} - \{\widehat{\Lambda}_i^T(S_{ik}) - \Lambda_i^T(S_{ik})\}] + o_p(1) \\ = W_{ik}^A(t) n^{-\frac{1}{2}} \sum_{l=1}^n \{D_{ik}'(t; \theta_0) \Omega_T(\theta_0)^{-1} U_l^T(\theta_0) + J_{ikl}^T(t; \theta_0)\} + o_p(1),$$

where we define

$$D_{ik}(t; \theta) = \int_{S_{ik}}^{S_{ik}+t} \{Z_i(u) - \bar{z}(u; \theta)\}' d\Lambda_i^T(u) = \int_0^t dD_i(u; \theta),$$

$$J_{ikl}^T(t; \theta) = \int_{S_{ik}}^{S_{ik}+t} e^{\theta' Z_i(u)} r_T^{(0)}(u; \theta)^{-1} dM_l^T(u).$$

**A.IV.5**  $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0)$

It is straightforward to show that

$$n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) = \Omega^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{ik}(\beta_0, \widehat{W}) + o_p(1),$$

where we define

$$U_{ik}(\beta, W) = \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t),$$

$$dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t) d\Lambda_{ik}(t).$$

The term  $n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{ik}(\beta, \widehat{W})$  can be decomposed as follows,

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{ik}(\beta, \widehat{W}) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, \widehat{W})\} \widehat{W}_{ik}^A(t) dM_{ik}(t) \\ \text{(A.5)} \quad &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t) \\ \text{(A.6)} \quad &- n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{\bar{z}_k(t; \beta, \widehat{W}) - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t) \\ \text{(A.7)} \quad &+ n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, \widehat{W})\} \{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\} dM_{ik}(t) \\ &+ o_p(1). \end{aligned}$$

Now, through the Functional Delta Method, combined with a lot of tedious algebra,

(A.6) converges in probability to 0.

$$(A.7) = n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) n^{-1}$$

$$(A.8) \quad \times \sum_{l=1}^n D'_{ik}(t; \theta) \Omega_T(\theta)^{-1} U_l^T(\theta) dM_{ik}(t)$$

$$(A.9) \quad + n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) n^{-1} \sum_{l=1}^n J_{ikl}^T(t; \theta) dM_{ik}(t).$$

Switching the order of summation, we have

$$\begin{aligned} (A.8) &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta) dM_{ik}(t) \\ &\quad \times \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) \\ &= \widehat{H}'(t; \beta, W) \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) \\ &= H'(t; \beta, W) \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) + o_p(1), \end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{H}'(t; \beta, W) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{Z}_k(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta) dM_{ik}(t),$$

to the quantity

$$H'(t; \beta, W) = E \left[ \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta) dM_{ik}(t) \right].$$

Switching the order of summation and integration

$$\begin{aligned} (A.9) &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_{S_{ik}}^{\tau} \left[ n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\theta' Z_i(u)} \int_{u-S_{ik}}^{\tau-S_{ik}} \{Z_{ik} - \bar{z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t) \right] \\ &\quad \times r_T^{(0)}(u; \theta)^{-1} dM_l^T(u) \\ &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_{S_{ik}}^{\tau} \widehat{G}(u, \tau; \beta) R_T^{(0)}(u; \theta)^{-1} dM_l^T(u) \\ &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_{S_{ik}}^{\tau} G(u, \tau; \beta) r_T^{(0)}(u; \theta)^{-1} dM_l^T(u) + o_p(1), \end{aligned}$$



where the last equality follows from the convergence in probability of

$$\widehat{G}(t_1, t_2; \beta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\theta' Z_i(t_1)} \int_{t_1 - S_{ik}}^{t_2 - S_{ik}} \{Z_{ik} - \overline{Z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t),$$

to the quantity

$$G(t_1, t_2; \beta) = E \left[ \sum_{k=1}^K A_{ik} e^{\theta' Z_i(t_1)} \int_{t_1 - S_{ik}}^{t_2 - S_{ik}} \{Z_{ik} - \overline{z}_k(t; \beta, W)\} W_{ik}^A(t) dM_{ik}(t) \right].$$

Combining equations (A.5) (A.8) and (A.9), we obtain

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) &= \Omega(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \overline{z}_k(t; \beta_0, W)\} W_{ik}^A(t) dM_{ik}(t) \\ &\quad + \Omega(\beta_0)^{-1} H'(t; \beta_0, W) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) \\ &\quad + \Omega(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \int_{S_{ik}}^{\tau} G(t, \tau; \beta_0) r_T^{(0)}(t; \theta_0)^{-1} dM_i^T(t) + o_p(1). \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \varphi_i + o_p(1). \end{aligned}$$

where

$$\begin{aligned} \varphi_i &= \Omega(\beta_0)^{-1} \left[ \sum_{k=1}^K A_{ik} \int_0^{\tau_k} \{Z_{ik} - \overline{z}_k(t; \beta_0, W)\} W_{ik}^A(t) dM_{ik}(t) \right. \\ &\quad \left. + H'(t; \beta_0, W) \Omega_T(\theta_0)^{-1} U_i^T(\theta_0) + \int_{S_{ik}}^{\tau} G(t, \tau; \beta_0) r_T^{(0)}(t; \theta_0)^{-1} dM_i^T(t) \right]. \end{aligned}$$

## APPENDIX B

### Appendix B: Proof of Theorem III.1

#### Appendix B: Proof of Theorem III.1

##### B.I. Notation

We begin by reviewing the essential notation:

$i$ : subject ( $i = 1, \dots, n$ )

$n$ : number of subjects

$k$ : cross section ( $k = 1, \dots, K$ )

$D_i$ : death time for the  $i$ th subject

$C_i$ : independent censoring time for the  $i$ th subject

$T_i$ : treatment time for the  $i$ th subject

$X_i$ :  $\min\{D_i, C_i\}$ : observation time for the  $i$ th subject

$Z_i(t)$ : covariate for  $i$ th subject at follow-up time  $t$

$A_i(t)$ : treatment eligibility indicator of  $i$ th subject at time  $t$

$N_i^T(t) = I(T_i \leq t, T_i < X_i)$ ; note that  $dN_i^T(t) = A_i(t)dN_i^T(t)$

$\tilde{Z}_i(t) = \{Z_i(s); s \in [0, t]\}$ : covariate history up to time  $t$

$\tilde{A}_i(t) = \{A_i(s); s \in [0, t]\}$ : treatment eligibility history up to time  $t$

$\tau$ : pre-specified constant satisfying  $P(X_i \geq \tau) > 0$  for all  $i$ .

$S_{ik}$ : follow-up time at calendar date of the  $k$ th cross section

$D_{ik} = D_i - S_{ik}$ , death time measured from date of  $k$ th cross section

$T_{ik} = T_i - S_{ik}$ , treatment time measured from date of  $k$ th cross section

$C_{ik} = C_i - S_{ik}$ , independent censoring time measured from date of  $k$ th cross section

$Z_{i0k} = Z_i(S_{ik})$ : covariate for subject  $i$  at date of  $k$ th cross-section

$N_{i0k}(t) = N_i(S_{ik} + t)I(T_i > S_{ik} + t)$

$A_{ik} = A_i(S_{ik})$

$\tau_{0k}$ : pre-specified constant satisfying  $P(D_{ik} \wedge T_{ik} \wedge C_{ik} \geq \tau_{0k}) > 0$

$N_{i1}(t) = I(T_i < X_i)N_i(T_i + t)$

$\tau_1$ : pre-specified constant satisfying  $P(D_i - T_i \geq \tau_1 | T_i, T_i < D_i) > 0$

$N_i^C(t) = I(C_i \leq t, C_i < D_i)$

Treatment-free death hazard:  $\lambda_{i0}(t; S_i) = \lambda_{00}(t) \exp\{\beta'_0 Z_{i0k}\}$

Post-treatment death hazard:  $\lambda_{i1}(t; T_i) = \lambda_{01}(t) \exp\{\beta'_1 Z_i(T_i)\}$

Treatment initiation hazard:  $\lambda_i^T(t) = A_i(t)\lambda_0^T(t) \exp\{\theta'_0 Z_i(t)\}$

Independent censoring hazard:  $\lambda_i^C(t) = \lambda_0^C(t) \exp\{\alpha'_0 Z_i(0)\}$

## B.II. Regularity Conditions

In deriving the asymptotic properties of the proposed estimators the following conditions are assumed for  $i = 1, \dots, n$  and  $k = 1, \dots, K$

(a)  $\{X_i, \Delta_i, \Delta_i^T, \tilde{Z}_i(X_i), \tilde{A}_i(X_i \wedge T_i)\}$  are independent and identically distributed random vectors.

(b)  $|Z_{il}(t)| < \kappa_l$ , for  $t \in [0, \tau]$  and  $Z_{il}(t)$  is the  $l$ th element of  $Z_i(t)$ .

(c)  $\int_0^{\tau_{0k}} \lambda_{0k}(t)dt < \infty$ ,  $\int_0^{\tau_1} \lambda_{01}(t)dt < \infty$ ,  $\int_0^{\tau} \lambda_0^T(t)dt < \infty$  and  $\int_0^{\tau} \lambda_0^C(t)dt < \infty$ .

(d) Continuity of the following functions:

$$\begin{aligned} r_T^{(1)}(t; \theta) &= \frac{\partial}{\partial \theta} r_T^{(0)}(t; \theta), \\ r_T^{(2)}(t; \theta) &= \frac{\partial^2}{\partial \theta \partial \theta'} r_T^{(0)}(t; \theta), \end{aligned}$$

and  $r_T^{(0)}(t; \theta)$ , where

$$r_T^{(p)}(t; \theta) = E[A_i(t)Y_i(t)Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}],$$

is the limiting value of

$$R_T^{(p)}(t; \theta) = n^{-1} \sum_{i=1}^n A_i(t)Y_i(t)Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\},$$

for  $p = 0, 1, 2$ , with  $r_T^{(1)}(t; \theta)$  and  $r_T^{(2)}(t; \theta)$  bounded and  $r_T^{(0)}(t; \theta)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\theta$  in an open set, with  $z^{\otimes 0} = 1$ ,  $z^{\otimes 1} = z$  and  $z^{\otimes 2} = zz'$  for a vector  $z$ .

Continuity of the following functions:

$$\begin{aligned} r_{0k}^{(1)}(t; \beta, W) &= \frac{\partial}{\partial \beta} r_{0k}^{(0)}(t; \beta, W), \\ r_{0k}^{(2)}(t; \beta, W) &= \frac{\partial^2}{\partial \beta \partial \beta'} r_{0k}^{(0)}(t; \beta, W), \end{aligned}$$

and  $r_{0k}^{(0)}(t; \beta, W)$ , where

$$r_{0k}^{(p)}(t; \beta, W) = E[A_{ik}W_{ik}^A(t)Z_{i0k}^{\otimes p} \exp(\beta' Z_{i0k})],$$

is the limiting value of

$$R_{0k}^{(p)}(t; \beta, W) = n^{-1} \sum_{i=1}^n A_{ik}W_{ik}^A(t)Z_{i0k}^{\otimes p} \exp(\beta' Z_{i0k}),$$

for  $p = 0, 1, 2$ , with  $r_{0k}^{(1)}(t; \beta, W)$  and  $r_{0k}^{(2)}(t; \beta, W)$  bounded and  $r_{0k}^{(0)}(t; \beta, W)$  bounded away from 0 for  $t \in [0, \tau_{0k}]$  and  $\beta$  in an open set.

Continuity of the following functions:

$$\begin{aligned} r_1^{(1)}(t; \beta) &= \frac{\partial}{\partial \beta} r_1^{(0)}(t; \beta), \\ r_1^{(2)}(t; \beta) &= \frac{\partial^2}{\partial \beta \partial \beta'} r_1^{(0)}(t; \beta), \end{aligned}$$

and  $r_1^{(0)}(t; \beta)$  where

$$r_1^{(p)}(t; \beta_1) = E[Y_{i1}(t) Z_{i1}^{\otimes p} \exp(\beta_1' Z_{i1})],$$

is the limiting value of

$$R_1^{(p)}(t; \beta_1) = n^{-1} \sum_{i=1}^n Y_{i1}(t) Z_{i1}^{\otimes p} \exp(\beta_1' Z_{i1}),$$

for  $p = 0, 1, 2$ , with  $r_1^{(1)}(t; \beta_1)$  and  $r_1^{(2)}(t; \beta_1)$  bounded and  $r_1^{(0)}(t; \beta_1)$  bounded away from 0 for  $t \in [0, \tau_1]$  and  $\beta_1$  in an open set.

Continuity of the following functions:

$$\begin{aligned} r_C^{(1)}(t; \alpha) &= \frac{\partial}{\partial \alpha} r_C^{(0)}(t; \alpha), \\ r_C^{(2)}(t; \alpha) &= \frac{\partial^2}{\partial \alpha \partial \alpha'} r_C^{(0)}(t; \alpha), \end{aligned}$$

and  $r_C^{(0)}(t; \alpha)$  where

$$r_C^{(p)}(t; \alpha) = E[Y_i(t) Z_i(0)^{\otimes p} \exp\{\alpha' Z_i(0)\}],$$

is the limiting value of

$$R_C^{(p)}(t; \alpha) = n^{-1} \sum_{i=1}^n Y_i(t) Z_i(0)^{\otimes p} \exp\{\alpha' Z_i(0)\},$$

for  $p = 0, 1, 2$ , with  $r_C^{(1)}(t; \alpha)$  and  $r_C^{(2)}(t; \alpha)$  bounded and  $r_C^{(0)}(t; \alpha)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\alpha$  in an open set.

(e) Positive-definiteness of the matrices  $\Omega_T(\theta_0)$ ,  $\Omega_0(\beta_0)$ ,  $\Omega_1(\beta_1)$  and  $\Omega_C(\alpha_0)$ , where

$$\begin{aligned}\Omega_T(\theta) &= E \left[ \int_0^\tau \left\{ \frac{r_T^{(2)}(t; \theta)}{r_T^{(0)}(t; \theta)} - \bar{z}(t; \theta)^{\otimes 2} \right\} dN_i^T(t) \right], \\ \bar{z}(t; \theta) &= r_T^{(1)}(t; \theta) / r_T^{(0)}(t; \theta), \\ \Omega_0(\beta) &= E \left[ \sum_{k=1}^K \int_0^{\tau_{0k}} \left\{ \frac{r_{0k}^{(2)}(t; \beta, W)}{r_{0k}^{(0)}(t; \beta, W)} - \bar{z}_{0k}(t; \beta, W)^{\otimes 2} \right\} dN_{i0k}(t) \right], \\ \bar{z}_{0k}(t; \beta, W) &= r_{0k}^{(1)}(t; \beta, W) / r_{0k}^{(0)}(t; \beta, W), \\ \Omega_1(\beta) &= E \left[ \int_0^{\tau_1} \left\{ \frac{r_1^{(2)}(t; \beta)}{r_1^{(0)}(t; \beta)} - \bar{z}_1(t; \beta)^{\otimes 2} \right\} dN_{i1}(t) \right], \\ \bar{z}_1(t; \beta) &= r_1^{(1)}(t; \beta) / r_1^{(0)}(t; \beta), \\ \Omega_C(\alpha) &= E \left[ \int_0^\tau \left\{ \frac{r_C^{(2)}(t; \alpha)}{r_C^{(0)}(t; \alpha)} - \bar{z}_C(t; \alpha)^{\otimes 2} \right\} dN_i^C(t) \right], \\ \bar{z}_C(t; \alpha) &= r_C^{(1)}(t; \alpha) / r_C^{(0)}(t; \alpha),\end{aligned}$$

(f)  $P\{Y_i(t) = 1\} > 0$  for  $t \in (0, \tau]$

### B.III. Outline of Asymptotic Derivation

We derive the influence functions of terms of interest as summations of independent and identical distributed (i.i.d.) terms plus a term which converges to zero in probability. The terms are as follows:

1.  $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$
2.  $n^{\frac{1}{2}}\{\hat{\Lambda}_0^T(t) - \Lambda_0^T(t)\}$
3.  $n^{\frac{1}{2}}\{\hat{\Lambda}_i^T(t) - \Lambda_i^T(t)\}$
4.  $n^{\frac{1}{2}}\{\hat{W}_{ik}^A(t) - W_{ik}^A(t)\}$
5.  $n^{\frac{1}{2}}(\hat{\beta}_0 - \beta_0)$

6.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{00}(t) - \Lambda_{00}(t)\}$

7.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i)\}$

8.  $n^{\frac{1}{2}}\{\widehat{S}_{i0}(t; S_i) - S_{i0}(t; S_i)\}$

9.  $n^{\frac{1}{2}}\{\widehat{\mu}_{i0}(S_i) - \mu_{i0}(S_i)\}$

10.  $n^{\frac{1}{2}}(\widehat{\beta}_1 - \beta_1)$

11.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{01}(t) - \Lambda_{01}(t)\}$

12.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\}$

13.  $n^{\frac{1}{2}}\{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\}$

14.  $n^{\frac{1}{2}}\{\widehat{\mu}_{i1}(T_i) - \mu_{i1}(T_i)\}$

15.  $n^{\frac{1}{2}}(\widehat{\alpha} - \alpha_0)$

16.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$

17.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\}$

18.  $n^{\frac{1}{2}}\{\widehat{G}_i(t)^{-1} - G_i(t)^{-1}\}$

19.  $n^{\frac{1}{2}}\{\widehat{S}_{\Delta_i}(t) - S_{\Delta_i}(t)\}$

20.  $n^{\frac{1}{2}}\{\widehat{\Delta}_i(t) - \Delta_i(t)\}$

21.  $n^{\frac{1}{2}}\{\widehat{S}_{\Delta}(t) - S_{\Delta}(t)\}$

22.  $n^{\frac{1}{2}}(\widehat{\Delta} - \Delta)$

### B.IV. Derivation of Asymptotic Properties

Several parts of the proof regarding the proportional hazards model are well-established results. Therefore, they are simply listed without proof. For details, please refer to Anderson and Gill (1982), Fleming and Harrington (1991) and Andersen et al. (1993).

#### B.IV.1 $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$

As  $n \rightarrow \infty$ , we have

$$n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) = \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) + o_p(1),$$

where

$$U_i^T(\theta) = \int_0^\tau \{Z_i(t) - \bar{z}(t; \theta)\} dM_i^T(t; \theta),$$

$$dM_i^T(t) = dN_i^T(t) - Y_i(t) d\Lambda_i^T(t),$$

This is now a well-established Cox model result, derived through Martingale theory.

#### B.IV.2 $n^{\frac{1}{2}}\{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\}$

We induce the following decomposition:

$$(B.1) \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_0^T(t) - \Lambda_0^T(t)\} = n^{\frac{1}{2}}\{\widehat{\Lambda}_0^T(t; \widehat{\theta}) - \widehat{\Lambda}_0^T(t; \theta_0)\}$$

$$(B.2) \quad + n^{\frac{1}{2}}\{\widehat{\Lambda}_0^T(t; \theta_0) - \Lambda_0^T(t)\}.$$



We can express the first term as

$$\begin{aligned}
(B.1) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_T^{(0)}(u; \hat{\theta})^{-1} - R_T^{(0)}(u; \theta_0)^{-1}\} dN_i^T(u) \\
&= \hat{h}'_T(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) \\
&= h'_T(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) + o_p(1).
\end{aligned}$$

where the third line follows from the convergence in probability of

$$\begin{aligned}
\hat{h}'_T(t; \theta) &= -n^{-1} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \theta)^{-1} \bar{Z}(u; \theta) dN_i^T(u) = - \int_0^t \bar{Z}(u; \theta) d\hat{\Lambda}_0^T(u; \theta), \\
\hat{\Omega}_T(\theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{R_T^{(2)}(t; \theta)}{R_T^{(0)}(t; \theta)} - \bar{Z}(t; \theta)^{\otimes 2} \right\} dN_i^T(t),
\end{aligned}$$

where  $\bar{Z}(t; \theta) = R_T^{(1)}(t; \theta) / R_T^{(0)}(t; \theta)$ , to the quantities

$$h'_T(t; \theta) = - \int_0^t \bar{z}(u; \theta) d\Lambda_0^T(u),$$

and  $\Omega_T(\theta)$  respectively, with  $\Omega_T(\theta)$  defined in Regularity Condition (e).

With respect to the second term in the decomposition, we have,

$$\begin{aligned}
(B.2) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t R_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) + o_p(1),
\end{aligned}$$

where the second line follows from continuity and Condition (d). Combining results,

for the decomposition, we have

$$n^{\frac{1}{2}} \{\hat{\Lambda}_0^T(t) - \Lambda_0^T(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_i^T(t; \theta_0) + o_p(1),$$

where

$$\Phi_i^T(t; \theta) = h'_T(t; \theta) \Omega_T(\theta)^{-1} U_i^T(\theta) + \int_0^t r_T^{(0)}(u; \theta)^{-1} dM_i^T(u) = \int_0^t d\Phi_i^T(u; \theta),$$

and

$$d\Phi_i^T(u; \theta) = -\bar{z}'(u; \theta) d\Lambda_0^T(u) \Omega_T(\theta)^{-1} U_i^T(\theta) + r_T^{(0)}(u; \theta)^{-1} dM_i^T(u).$$

$$\mathbf{B.IV.3} \quad n^{\frac{1}{2}} \{ \widehat{\Lambda}_i^T(t) - \Lambda_i^T(t) \}$$

We begin with another decomposition,

$$\begin{aligned} & n^{\frac{1}{2}} \{ \widehat{\Lambda}_i^T(t) - \Lambda_i^T(t) \} \\ (B.3) \quad & = n^{\frac{1}{2}} \left\{ \int_0^t e^{\widehat{\theta}' Z_i(u)} d\widehat{\Lambda}_0^T(u) - \int_0^t e^{\theta'_0 Z_i(u)} d\widehat{\Lambda}_0^T(u) \right\} \\ (B.4) \quad & + n^{\frac{1}{2}} \left\{ \int_0^t e^{\theta'_0 Z_i(u)} d\widehat{\Lambda}_0^T(u) - \int_0^t e^{\theta'_0 Z_i(u)} d\Lambda_0^T(u) \right\}. \end{aligned}$$

Considering the first term,

$$(B.3) = n^{\frac{1}{2}} \int_0^t \{ e^{\widehat{\theta}' Z_i(u)} - e^{\theta'_0 Z_i(u)} \} d\widehat{\Lambda}_0^T(u).$$

By a Taylor series expansion,

$$\begin{aligned} n^{\frac{1}{2}} \{ e^{\widehat{\theta}' Z_i(u)} - e^{\theta'_0 Z_i(u)} \} & = Z'_i(u) e^{\theta'_0 Z_i(u)} n^{\frac{1}{2}} (\widehat{\theta} - \theta) + o_p(1) \\ & = Z'_i(u) e^{\theta'_0 Z_i(u)} \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) + o_p(1). \end{aligned}$$

Since  $\widehat{\Lambda}_0^T(t) \xrightarrow{p} \Lambda_0^T(t)$  for  $t \in [0, \tau]$ , we obtain

$$(B.3) = \int_0^t Z'_i(u) d\Lambda_i^T(u) \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta) + o_p(1).$$

By using Result B.IV.2, the second term can be written as

$$\begin{aligned} (B.4) & = n^{\frac{1}{2}} \int_0^t e^{\theta'_0 Z_i(u)} d\{ \widehat{\Lambda}_0^T(t) - \Lambda_0^T(t) \} \\ & = \int_0^t e^{\theta'_0 Z_i(u)} n^{-\frac{1}{2}} \sum_{l=1}^n d\Phi_l^T(u; \theta_0) + o_p(1). \end{aligned}$$

Combining results from the decomposition leads to

$$\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\Lambda}_i^T(t) - \Lambda_i^T(t)\} &= \int_0^t \{Z_i(u) - \bar{z}(u; \theta_0)\}' d\Lambda_i^T(u) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t e^{\theta_0' Z_i(u)} r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) + o_p(1) \\
&= D_i'(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{il}^T(t; \theta_0) + o_p(1),
\end{aligned}$$

where we define

$$\begin{aligned}
D_i(t; \theta) &= \int_0^t \{Z_i(u) - \bar{z}(u; \theta)\}' d\Lambda_i^T(u) = \int_0^t dD_i(u; \theta), \\
J_{il}^T(t; \theta) &= \int_0^t e^{\theta' Z_i(u)} r_T^{(0)}(u; \theta)^{-1} dM_l^T(u).
\end{aligned}$$

**B.IV.4**  $n^{\frac{1}{2}}\{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\}$

Consistent with the notation set in Section B.I, when the subscript of quantities does not involve the cross section notation  $k$ ,  $t$  refers the time from study entry. If  $k$  is present in the subscript, then  $t$  denotes the time from the  $k$ th cross section date.

Since  $W_{ik}^A(t) = e^{\Lambda_i^T(t+S_{ik}) - \Lambda_i^T(S_{ik})}$  and  $\widehat{W}_{ik}^A(t) = e^{\widehat{\Lambda}_i^T(t+S_{ik}) - \widehat{\Lambda}_i^T(S_{ik})}$ , we then have

$$\begin{aligned}
&n^{\frac{1}{2}}\{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\} \\
&= n^{\frac{1}{2}}\{e^{\widehat{\Lambda}_i^T(t+S_{ik}) - \widehat{\Lambda}_i^T(S_{ik})} - e^{\Lambda_i^T(t+S_{ik}) - \Lambda_i^T(S_{ik})}\} \\
&= W_{ik}^A(t) n^{\frac{1}{2}}[\{\widehat{\Lambda}_i^T(t+S_{ik}) - \Lambda_i^T(t+S_{ik})\} - \{\widehat{\Lambda}_i^T(S_{ik}) - \Lambda_i^T(S_{ik})\}] + o_p(1) \\
&= W_{ik}^A(t) n^{-\frac{1}{2}} \sum_{l=1}^n \{D_{ikl}'(t; \theta_0) \Omega_T(\theta_0)^{-1} U_l^T(\theta_0) + J_{ikl}^T(t; \theta_0)\} + o_p(1),
\end{aligned}$$

where we define

$$\begin{aligned}
D_{ik}(t; \theta) &= \int_{S_{ik}}^{S_{ik}+t} \{Z_i(u) - \bar{z}(u; \theta)\}' d\Lambda_i^T(u) = \int_0^t dD_i(u; \theta), \\
J_{ikl}^T(t; \theta) &= \int_{S_{ik}}^{S_{ik}+t} e^{\theta' Z_i(u)} r_T^{(0)}(u; \theta)^{-1} dM_l^T(u).
\end{aligned}$$

**B.IV.5**  $n^{\frac{1}{2}}(\widehat{\beta}_0 - \beta_0)$ 

It is straightforward to show that

$$n^{\frac{1}{2}}(\widehat{\beta}_0 - \beta_0) = \Omega_0^{-1}(\beta_0)n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{i0k}(\beta_0, \widehat{W}) + o_p(1),$$

where we define

$$U_{i0k}(\beta, W) = \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t),$$

$$dM_{i0k}(t) = dN_{i0k}(t) - Y_{i0k}(t) d\Lambda_{i0k}(t).$$

The term  $n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{i0k}(\beta, \widehat{W})$  can be decomposed as follows,

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} U_{i0k}(\beta, \widehat{W}) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, \widehat{W})\} \widehat{W}_{ik}^A(t) dM_{i0k}(t) \\ \text{(B.5)} &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t) \\ \text{(B.6)} & - n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{\bar{z}_{0k}(t; \beta, \widehat{W}) - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t) \\ \text{(B.7)} & + n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, \widehat{W})\} \{\widehat{W}_{ik}^A(t) - W_{ik}^A(t)\} dM_{i0k}(t) \\ & + o_p(1). \end{aligned}$$

Now, through the Functional Delta Method, combined with a lot of tedious algebra,

(B.6) converges in probability to 0.

Using result B.IV.4

$$\begin{aligned} \text{(B.7)} &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) \\ \text{(B.8)} & \quad \times n^{-1} \sum_{l=1}^n D'_{ik}(t; \theta_0) \Omega_T(\theta_0)^{-1} U_l^T(\theta_0) dM_{i0k}(t) \\ & \quad + n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) \end{aligned}$$

$$(B.9) \quad \times n^{-1} \sum_{l=1}^n J_{ikl}^T(t; \theta_0) dM_{i0k}(t).$$

Switching the order of summation, we have

$$\begin{aligned} (B.8) &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta_0) dM_{i0k}(t) \\ &\quad \times \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\ &= \widehat{H}'_0(t; \beta, W) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\ &= H'_0(t; \beta, W) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) + o_p(1), \end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{H}'_0(t; \beta, W) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta_0) dM_{i0k}(t),$$

to the quantity

$$H'_0(t; \beta, W) = E \left[ \sum_{k=1}^K A_{ik} \int_0^{\tau_{0k}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) D'_{ik}(t; \theta_0) dM_{i0k}(t) \right].$$

Switching the order of summation and integration

$$\begin{aligned} (B.9) &= n^{-\frac{1}{2}} \sum_{j=1}^n \int_{S_{ik}}^{\tau} \left[ n^{-1} \sum_{l=1}^n \sum_{k=1}^K A_{ik} e^{\theta'_0 Z_i(u)} \int_{u-S_{ik}}^{\tau-S_{ik}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t) \right] \\ &\quad \times r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) \\ &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_{S_{ik}}^{\tau} \widehat{G}_0(u, \tau; \beta) R_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) \\ &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_{S_{ik}}^{\tau} G_0(u, \tau; \beta) r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) + o_p(1), \end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{G}_0(t_1, t_2; \beta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\theta'_0 Z_i(t_1)} \int_{t_1-S_{ik}}^{t_2-S_{ik}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t),$$

to the quantity

$$G_0(t_1, t_2; \beta) = E \left[ \sum_{k=1}^K A_{ik} e^{\theta'_0 Z_i(t_1)} \int_{t_1-S_{ik}}^{t_2-S_{ik}} \{Z_{i0k} - \bar{z}_{0k}(t; \beta, W)\} W_{ik}^A(t) dM_{i0k}(t) \right].$$

Combining the equations (B.5) (B.8) and (B.9), we obtain

$$n^{\frac{1}{2}}(\widehat{\beta}_0 - \beta_0) = \Omega_0(\beta_0)^{-1}n^{-\frac{1}{2}} \sum_{i=1}^n U_{i0}(\beta_0) + o_p(1),$$

where

$$\begin{aligned} U_{i0}(\beta_0) &= \sum_{k=1}^K \int_0^{\tau_{0k}} A_{ik} \{Z_{i0k} - \bar{z}_{0k}(t; \beta_0, W)\} W_{ik}^A(t) dM_{i0k}(t) \\ &\quad + H'_0(t; \beta_0, W) \Omega_T(\theta_0)^{-1} U_i^T(\theta_0) \\ &\quad + \int_{S_{ik}}^{\tau} G_0(t, \tau; \beta_0) r_T^{(0)}(t; \theta_0)^{-1} dM_i^T(t). \end{aligned}$$

$$\mathbf{B.IV.6} \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_{00}(t) - \Lambda_{00}(t)\}$$

We define

$$\widehat{\Lambda}_{00}(t; \beta_0) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^t R_0^{(0)}(u; \beta_0)^{-1} A_{ik} W_{ik}^A(u) dN_{i0k}(u)$$

for  $t \in (0, L]$ , where  $R_0^{(0)}(u; \beta_0) = \sum_{k=1}^K R_{0k}^{(0)}(u; \beta_0)$ .

We begin another decomposition,

$$\begin{aligned} &n^{\frac{1}{2}}\{\widehat{\Lambda}_{00}(t) - \Lambda_{00}(t)\} \\ (B.10) \quad &= n^{\frac{1}{2}}[\widehat{\Lambda}_{00}\{t; \widehat{W}, R_0(\widehat{\beta}_0, \widehat{W})\} - \widehat{\Lambda}_{00}\{t; \widehat{W}, R_0(\beta_0, \widehat{W})\}] \end{aligned}$$

$$(B.11) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{00}\{t; \widehat{W}, R_0(\beta_0, \widehat{W})\} - \widehat{\Lambda}_{00}\{t; W, R_0(\beta_0, \widehat{W})\}]$$

$$(B.12) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{00}\{t; W, R_0(\beta_0, \widehat{W})\} - \widehat{\Lambda}_{00}\{t; W, R_0(\beta_0, W)\}]$$

$$(B.13) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{00}\{t; W, R_0(\beta_0, W)\} - \Lambda_{00}(t)]$$

By using Result B.IV.5, we can express the first term as

$$\begin{aligned} (B.10) &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K \int_0^t A_{ik} \{R_0^{(0)}(u; \widehat{\beta}_0, \widehat{W})^{-1} - R_0^{(0)}(u; \beta_0, \widehat{W})^{-1}\} \widehat{W}_{ik}^A(u) dN_{i0k}(u) \\ &= - \int_0^t \overline{Z}'_0(u; \beta_0, W) d\Lambda_{00}(u) \Omega_0(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{i0}(\beta_0) + o_p(1) \\ &= h'_0(t; \beta_0, W) \Omega_0(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{i0}(\beta_0) + o_p(1), \end{aligned}$$

where we define

$$\begin{aligned}\bar{Z}_0(t; \beta, W) &= R_0^{(1)}(t; \beta, W)/R_0^{(0)}(t; \beta, W). \\ R_0^{(p)}(t; \beta, W) &= \sum_{k=1}^K R_{0k}^{(p)}(t; \beta, W), \\ \bar{z}_0(t; \beta, W) &= r_0^{(1)}(t; \beta, W)/r_0^{(0)}(t; \beta, W). \\ r_0^{(p)}(t; \beta, W) &= \sum_{k=1}^K r_{0k}^{(p)}(t; \beta, W), \\ h_0(t; \beta, W) &= - \int_0^t \bar{z}'_0(u; \beta, W) d\Lambda_{00}(u).\end{aligned}$$

By using Result B.IV.4, we have

$$\begin{aligned}(B.11) \quad &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t \{\widehat{W}_{ik}^A(u) - W_{ik}^A(u)\} R_0^{(0)}(u; \beta_0, \widehat{W})^{-1} dN_{i0k}(u) \\ &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t R_0^{(0)}(u; \beta_0, W)^{-1} W_{ik}^A(u) D'_{ik}(u; \theta_0) \Omega_T(\theta_0)^{-1}\end{aligned}$$

$$(B.14) \quad \times n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) dN_{i0k}(u)$$

$$+ n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t R_0^{(0)}(u; \beta_0, W)^{-1} W_{ik}^A(u)$$

$$(B.15) \quad \times n^{-\frac{1}{2}} \sum_{l=1}^n J_{ikl}^T(u; \theta_0) dN_{i0k}(u) + o_p(1).$$

Switching the order of summation, we have

$$\begin{aligned}(B.14) \quad &= \widehat{B}'_0(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\ &= B'_0(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) + o_p(1),\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{B}_0(t; \beta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t R_0^{(0)}(u; \beta, W)^{-1} W_{ik}^A(u) D'_{ik}(u; \theta) dN_{i0k}(u)$$

to the quantity

$$B_0(t; \beta) = E \left[ A_{ik} \int_0^t r_0^{(0)}(u; \beta, W)^{-1} W_{ik}^A(u) D'_{ik}(u; \theta) dN_{i0k}(u) \right].$$

Switching the order of summation and integration

$$\begin{aligned}
(B.15) &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t \widehat{K}_0(u, t; \beta_0) r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t K_0(u, t; \beta_0) r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) + o_p(1),
\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{K}_0(t_1, t_2; \beta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K e^{\theta' Z_i(t_1)} \int_{t_1 - S_{ik}}^{t_2 - S_{ik}} A_{ik} W_{ik}^A(u) R_0^{(0)}(u; \beta, W)^{-1} dN_{i0k}(u),$$

to the quantity

$$K_0(t_1, t_2; \beta) = E \left[ e^{\theta_0' Z_i(t_1)} A_{ik} \int_{t_1 - S_{ik}}^{t_2 - S_{ik}} W_{ik}^A(u) r_0^{(0)}(u; \beta, W)^{-1} dN_{i0k}(u) \right].$$

Combining equations (B.14) and (B.15), we obtain

$$\begin{aligned}
(B.11) &= B_0'(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t K_0(u, t; \beta_0) r_T^{(0)}(u; \theta_0)^{-1} dM_l^T(u) + o_p(1).
\end{aligned}$$

We can have

$$(B.12) = n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t W_{ik}^A(u) \{R_0^{(0)}(u; \beta_0, \widehat{W})^{-1} - R_0^{(0)}(u; \beta_0, W)^{-1}\} dN_{i0k}(u).$$

Now, through the Function Delta Method,

$$\begin{aligned}
&n^{\frac{1}{2}} \{R_0^{(0)}(u; \beta, \widehat{W})^{-1} - R_0^{(0)}(u; \beta, W)^{-1}\} \\
&= -R_0^{(0)}(u; \beta, W)^{-2} n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\beta' Z_{i0k}} n^{\frac{1}{2}} \{\widehat{W}_{ik}^A(u) - W_{ik}^A(u)\} \\
&= -R_0^{(0)}(u; \beta, W)^{-2} n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\beta' Z_{i0k}} W_{ik}^A(u) n^{-\frac{1}{2}} \\
&\quad \times \sum_{l=1}^n \{D'_{ik}(u) \Omega_T(\theta_0)^{-1} U_l^T(\theta_0) + J_{ikl}^T(u)\} \\
&= R_0^{(0)}(u; \beta, W)^{-2} \widehat{F}'_0(u; \beta) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0)
\end{aligned}$$



$$\begin{aligned}
& +R_0^{(0)}(u; \beta, W)^{-2}n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^{u+S_{ik}} \widehat{Q}'_0(s, u; \theta_0)r_T^{(0)}(s, \theta_0)^{-1}dM_l^T(s) \\
= & R_0^{(0)}(u; \beta, W)^{-2}F'_0(u; \beta)\Omega_T(\theta_0)^{-1}n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\
& +R_0^{(0)}(u; \beta, W)^{-2}n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^{u+S_{ik}} Q'_0(s, u; \theta_0)r_T^{(0)}(s, \theta_0)^{-1}dM_l^T(s) + o_p(1),
\end{aligned}$$

where the last line follows from the convergence in probability of

$$\begin{aligned}
\widehat{F}_0(u; \beta) &= -n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\beta' Z_{i0k}} W_{ik}^A(u) D'_{ik}(u; \theta), \\
\widehat{Q}_0(t_1, t_2; \theta) &= -n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} e^{\theta' Z_i(t_1)} e^{\beta' Z_{i0k}} W_{ik}^A(t_2),
\end{aligned}$$

to the quantities

$$\begin{aligned}
F_0(u; \beta) &= -E \left[ A_{ik} e^{\beta' Z_{i0k}} W_{ik}^A(u) D'_{ik}(u; \theta) \right], \\
Q_0(t_1, t_2; \theta) &= -E \left[ A_{ik} e^{\theta' Z_i(t_1)} e^{\beta' Z_{i0k}} W_{ik}^A(t_2) \right].
\end{aligned}$$

Substituting this result into the expansion of (B.12), we obtain

$$\begin{aligned}
(B.12) &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t W_{ik}^A(u) R_0^{(0)}(u; \beta_0, W)^{-2} F'_0(u; \beta_0) \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \\
&\quad \times \sum_{l=1}^n U_l^T(\theta_0) dN_{i0k}(u) \\
&\quad + n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t W_{ik}^A(u) R_0^{(0)}(u; \beta_0, W)^{-2} n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^{u+S_{ik}} Q'_0(s, u; \theta_0) \\
&\quad \times r_T^{(0)}(s, \theta_0)^{-1} dM_l^T(s) dN_{i0k}(u).
\end{aligned}$$

Switching the order of summation for the first term, and the order of summation and integration in the second term, we have

$$\begin{aligned}
(B.12) &= \widehat{E}_0(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t \widehat{P}_0(u, t; \beta_0) r_T^{(0)}(u, \theta_0)^{-1} dM_l^T(u)
\end{aligned}$$

$$\begin{aligned}
&= E_0(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t P_0(u, t; \beta_0) r_T^{(0)}(u, \theta_0)^{-1} dM_l^T(u) + o_p(1),
\end{aligned}$$

where the last line follows from the convergence in probability of

$$\begin{aligned}
\widehat{E}_0(t; \beta) &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t \frac{W_{ik}^A(u) F_0(u; \beta)}{R_0^{(0)}(u; \beta, W)^2} dN_{i0k}(u), \\
\widehat{P}_0(t_1, t_2; \beta) &= n^{-1} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_{t_1-S_{ik}}^{t_2-S_{ik}} \frac{W_{ik}^A(u) Q_0(t_1, u; \theta)}{R_0^{(0)}(u; \beta, W)^2} dN_{i0k}(u),
\end{aligned}$$

to the quantities

$$\begin{aligned}
E_0(t; \beta) &= E \left[ A_{ik} \int_0^t \frac{W_{ik}^A(u) F_0(u; \beta)}{r_0^{(0)}(u; \beta, W)^2} dN_{i0k}(u) \right], \\
P_0(t_1, t_2; \beta) &= E \left[ A_{ik} \int_{t_1-S_{ik}}^{t_2-S_{ik}} \frac{W_{ik}^A(u) Q_0(t_1, u; \theta)}{r_0^{(0)}(u; \beta, W)^2} dN_{i0k}(u) \right].
\end{aligned}$$

We can also express

$$\begin{aligned}
(B.13) &= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t \frac{W_{ik}^A(u)}{R_0^{(0)}(u; \beta_0, W)} dM_{i0k}(u), \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t \frac{W_{ik}^A(u)}{r_0^{(0)}(u; \beta_0, W)} dM_{i0k}(u) + o_p(1),
\end{aligned}$$

Combining the results of equations (B.10) (B.11) (B.12) and (B.13), we obtain

$$\begin{aligned}
&n^{\frac{1}{2}} \{ \widehat{\Lambda}_{00}(t) - \Lambda_{00}(t) \} \\
&= h_0'(t; \beta_0, W) \Omega_0^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{i=1}^n U_{i0}(\beta_0) \\
&\quad + [B_0'(t; \beta_0) + E_0'(t; \beta_0)] \Omega_T(\theta)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t [K_0(u, t; \beta_0) + P_0(u, t; \beta_0)] r_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) \\
&\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K A_{ik} \int_0^t r_0^{(0)}(u; \beta_0, W)^{-1} W_{ik}^A(u) dM_{i0k}(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t d\Phi_{i0}(u),
\end{aligned}$$

where

$$\begin{aligned}
\Phi_{i0}(t) &= h'_0(t; \beta_0, W) \Omega_0^{-1}(\beta_0) U_{i0}(\beta_0) \\
&\quad + [B'_0(t; \beta_0) + E'_0(t; \beta_0)] \Omega_T(\theta_0)^{-1} U_i^T(\theta_0) \\
&\quad + \int_0^t [K_0(u, t; \beta_0) + P_0(u, t; \beta_0)] r_T^{(0)}(u; \theta_0)^{-1} dM_i^T(u) \\
&\quad + \sum_{k=1}^K A_{ik} \int_0^t r_0^{(0)}(u; \beta_0, W)^{-1} W_{ik}^A(u) dM_{i0k}(u) \\
&= \int_0^t d\Phi_{i0}(u).
\end{aligned}$$

$$\mathbf{B.IV.7} \quad n^{\frac{1}{2}} \{ \widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i) \}$$

We begin with another decomposition

$$\begin{aligned}
&n^{\frac{1}{2}} \{ \widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i) \} \\
\text{(B.16)} \quad &= n^{\frac{1}{2}} \{ \widehat{\Lambda}_{i0}(t, \widehat{\beta}; S_i) - \widehat{\Lambda}_{i0}(t, \beta_0; S_i) \}
\end{aligned}$$

$$\text{(B.17)} \quad + n^{\frac{1}{2}} \{ \widehat{\Lambda}_{i0}(t, \beta_0; S_i) - \Lambda_{i0}(t) \}.$$

Consider the first term and using Result B.IV.5

$$\begin{aligned}
\text{(B.16)} &= \widehat{\Lambda}_{00}(t) n^{\frac{1}{2}} \{ e^{\widehat{\beta}'_0 Z_i(S_i)} - e^{\beta_0' Z_i(S_i)} \} \\
&= \Lambda_{00}(t) e^{\beta_0' Z_i(S_i)} Z_i(S_i)' n^{\frac{1}{2}} (\widehat{\beta}_0 - \beta_0) + o_p(1) \\
&= \Lambda_{00}(t) e^{\beta_0' Z_i(S_i)} Z_i(S_i)' \Omega_0^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{j=1}^n U_{j0}(\beta_0) + o_p(1).
\end{aligned}$$

By using Result B.IV.6, the second term can be written as

$$\begin{aligned}
\text{(B.17)} &= e^{\beta_0' Z_i(S_i)} n^{\frac{1}{2}} \{ \widehat{\Lambda}_{00}(t) - \Lambda_{00}(t) \} \\
&= e^{\beta_0' Z_i(S_i)} n^{-\frac{1}{2}} \sum_{j=1}^n \Phi_{j0}(t) + o_p(1).
\end{aligned}$$

Combining equations (B.16) and (B.17), we obtain

$$\begin{aligned}
&n^{\frac{1}{2}} \{ \widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i) \} \\
&= \Lambda_{i0}(t; S_i) Z_i(S_i)' \Omega_0^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{j=1}^n U_{j0}(\beta_0) + e^{\beta_0' Z_i(S_i)} n^{-\frac{1}{2}} \sum_{j=1}^n \Phi_{j0}(t) + o_p(1).
\end{aligned}$$

$$\mathbf{B.IV.8} \quad n^{\frac{1}{2}}\{\widehat{S}_{i0}(t; S_i) - S_{i0}(t; S_i)\}$$

Using the Functional Delta Method and Result B.IV.7, we have

$$n^{\frac{1}{2}}\{\widehat{S}_{i0}(t; S_i) - S_{i0}(t; S_i)\} = -S_{i0}(t; S_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i)\} + o_p(1).$$

$$\mathbf{B.IV.9} \quad n^{\frac{1}{2}}\{\widehat{\mu}_{i0}(S_i) - \mu_{i0}(S_i)\}$$

Define  $\mu_{i0}(S_i) = \int_0^L S_{i0}(u; S_i)du$ , where  $L$  is restricted time point. By continuity and Result B.IV.7 and B.IV.8, we have

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\mu}_{i0}(S_i) - \mu_{i0}(S_i)\} \\ &= n^{\frac{1}{2}} \int_0^L \{\widehat{S}_{i0}(t; S_i) - S_{i0}(t; S_i)\} dt \\ &= - \int_0^L S_{i0}(t; S_i) n^{\frac{1}{2}} \{\widehat{\Lambda}_{i0}(t; S_i) - \Lambda_{i0}(t; S_i)\} dt + o_p(1) \\ \text{(B.18)} \quad &= - \int_0^L S_{i0}(t; S_i) \Lambda_{i0}(t; S_i) Z_i(S_i)' \Omega_0^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{j=1}^n U_{j0}(\beta_0) dt \end{aligned}$$

$$\text{(B.19)} \quad - \int_0^L S_{i0}(t; S_i) e^{\beta_0' Z_i(S_i)} n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^t d\Phi_{j0}(u) dt + o_p(1).$$

For the second term, switching the order of integration and summation

$$\begin{aligned} \text{(B.19)} \quad &= -n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L \int_u^L S_{i0}(t; S_i) e^{\beta_0' Z_i(S_i)} dt d\Phi_{j0}(u) \\ &= -n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L e^{\beta_0' Z_i(S_i)} \{\mu_{i0}(S_i) - \int_0^t S_{i0}(u; S_i) du\} d\Phi_{j0}(t). \end{aligned}$$

Combining equations (B.18) and (B.19), we have

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{\mu}_{i0}(S_i) - \mu_{i0}(S_i)\} &= \int_0^L S_{i0}(t; S_i) \Lambda_{i0}(t; S_i) Z_i(S_i)' \Omega_0^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{j=1}^n U_{j0}(\beta_0) dt \\ &\quad - n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L e^{\beta_0' Z_i(S_i)} \{\mu_{i0}(S_i) - \int_0^t S_{i0}(u; S_i) du\} d\Phi_{j0}(t) \\ &= n^{-\frac{1}{2}} \sum_{j=1}^n \varphi_{ij0}(S_i). \end{aligned}$$

where

$$\begin{aligned}\varphi_{ij_0}(S_i) &= Z_i(S_i)' \Omega_0^{-1}(\beta_0) U_{j_0}(\beta_0) \int_0^L S_{i0}(t; S_i) \Lambda_{i0}(t; S_i) dt \\ &\quad - e^{\beta_0' Z_i(S_i)} \int_0^L \left\{ \mu_{i0}(S_i) - \int_0^t S_{i0}(u; S_i) du \right\} d\Phi_{j_0}(t).\end{aligned}$$

$$\mathbf{B.IV.10} \quad n^{\frac{1}{2}}(\widehat{\beta}_1 - \beta_1)$$

It is straight forward to show that

$$n^{\frac{1}{2}}(\widehat{\beta}_1 - \beta_1) = \Omega_1^{-1}(\beta_1) n^{-\frac{1}{2}} \sum_{i=1}^n U_{i1}(\beta_1) + o_p(1),$$

where

$$\begin{aligned}U_{i1}(\beta_1) &= \int_0^{\tau_1} \{Z_{i1} - \bar{z}_1(t; \beta_1)\} dM_{i1}(t), \\ dM_{i1}(t) &= dN_{i1}(t) - Y_{i1}(t) d\Lambda_{i1}(t).\end{aligned}$$

This is now a well-established Cox model result, derived through Martingale theory.

$$\mathbf{B.IV.11} \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_{01}(t) - \Lambda_{01}(t)\}$$

We begin with another decomposition,

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_{01}(t) - \Lambda_{01}(t)\} \\ \text{(B.20)} \quad &= n^{\frac{1}{2}}[\widehat{\Lambda}_{01}(t; \widehat{\beta}_1) - \widehat{\Lambda}_{01}(t; \beta_1)]\end{aligned}$$

$$\text{(B.21)} \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{01}(t; \beta_1) - \Lambda_{01}(t)]$$

Consider the first term,

$$\begin{aligned}\text{(B.20)} &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_1^{(0)}(u; \widehat{\beta}_1)^{-1} - R_1^{(0)}(u; \beta_1)^{-1}\} dN_{i1}(u) \\ &= \widehat{h}'_1(t; \beta_1) \Omega_1(\beta_1)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{i1}(\beta_1) \\ &= h'_1(t; \beta_1) \Omega_1(\beta_1)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{i1}(\beta_1) + o_p(1),\end{aligned}$$

where the third line follows from the convergence in probability of

$$\begin{aligned}\widehat{h}'_1(t; \beta_1) &= - \int_0^t \widehat{z}'_1(u; \beta_1) d\widehat{\Lambda}_{01}(u), \\ \widehat{\Omega}_1(\beta_1) &= n^{-1} \sum_{i=1}^n \int_0^{\tau_1} \left\{ \frac{R_1^{(2)}(t; \beta_1)}{R_1^{(0)}(t; \beta_1)} - \widehat{z}_1(t; \beta_1)^{\otimes 2} \right\} dN_{i1}(t),\end{aligned}$$

to the quantities

$$h'_1(t; \beta_1) = - \int_0^t z'_1(u; \beta_1) d\Lambda_{01}(u),$$

and  $\Omega_1(\beta_1)$  respectively, with  $\Omega_1(\beta_1)$  defined in Regularity Condition (e).

With respect to the second term in the decomposition, we have,

$$\begin{aligned}(B.21) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{i1}(u)}{R_1^{(0)}(u; \beta_1)}, \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{i1}(u)}{r_1^{(0)}(u; \beta_1)} + o_p(1),\end{aligned}$$

where the second line follows from continuity and Condition (d). Combining equations (B.20) and (B.21), for the decomposition, we have

$$\begin{aligned}& n^{\frac{1}{2}} \{ \widehat{\Lambda}_{01}(t) - \Lambda_{01}(t) \} \\ &= h'_1(t; \beta_1) \Omega_1(\beta_1)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{i1}(\beta_1) \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_1^{(0)}(u; \beta_1)^{-1} dM_{i1}(u) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t d\Phi_{i1}(u),\end{aligned}$$

where

$$\begin{aligned}\Phi_{i1}(t) &= h'_1(t; \beta_1) \Omega_1(\beta_1)^{-1} U_{i1}(\beta_1) \\ &\quad + \int_0^t r_1^{(0)}(u; \beta_1)^{-1} dM_{i1}(u) \\ &= \int_0^t d\Phi_{i1}(u).\end{aligned}$$

$$\mathbf{B.IV.12} \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\}$$

We begin with another decomposition

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} \\ (B.22) \quad &= n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t, \widehat{\beta}_1; T_i) - \widehat{\Lambda}_{i1}(t, \beta_1; T_i)\} \end{aligned}$$

$$(B.23) \quad + n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t, \beta_1; T_i) - \Lambda_{i1}(t)\}.$$

Considering the first term, by a Taylor series expansion and Result B.IV.10, and  $\widehat{\Lambda}_{01}(t) \xrightarrow{p} \Lambda_{01}(t)$  for  $t \in [0, \tau]$ , we obtain

$$\begin{aligned} (B.22) \quad &= \widehat{\Lambda}_{01}(t)n^{\frac{1}{2}}\{e^{\widehat{\beta}'_1 Z_{i1}} - e^{\beta'_1 Z_{i1}}\} \\ &= \Lambda_{01}(t)e^{\beta'_1 Z_{i1}}Z'_{i1}n^{\frac{1}{2}}(\widehat{\beta}_1 - \beta_1) + o_p(1) \\ &= \Lambda_{01}(t)e^{\beta'_1 Z_{i1}}Z'_{i1}\Omega_1(\beta_1)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n U_{j1}(\beta_1) + o_p(1). \end{aligned}$$

By using Result B.IV.11, the second term can be written as

$$\begin{aligned} (B.23) \quad &= e^{\beta'_1 Z_{i1}}n^{\frac{1}{2}}\{\widehat{\Lambda}_{01}(t) - \Lambda_{01}(t)\} \\ &= e^{\beta'_1 Z_{i1}}n^{-\frac{1}{2}}\sum_{j=1}^n \Phi_{j1}(t) + o_p(1). \end{aligned}$$

Combining results from the decomposition leads to,

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} \\ &= \Lambda_{i1}(t; T_i)Z'_{i1}\Omega_1(\beta_1)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n U_{j1}(\beta_1) + e^{\beta'_1 Z_{i1}}n^{-\frac{1}{2}}\sum_{j=1}^n \Phi_{j1}(t) + o_p(1). \end{aligned}$$

$$\mathbf{B.IV.13} \quad n^{\frac{1}{2}}\{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\}$$

Using the Functional Delta Method

$$n^{\frac{1}{2}}\{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\} = -S_{i1}(t; T_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} + o_p(1).$$

$$\mathbf{B.IV.14} \quad n^{\frac{1}{2}}\{\widehat{\mu}_{i1}(T_i) - \mu_{i1}(T_i)\}$$

Define  $\mu_{i1}(T_i) = \int_0^L S_{i1}(u; T_i) du$  and where  $L$  is the restriction time point. By continuity and Result B.IV.12 and B.IV.13

$$\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_{i1}(T_i) - \mu_{i1}(T_i)\} &= n^{\frac{1}{2}} \int_0^L \{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\} dt \\
&= - \int_0^L S_{i1}(t; T_i) n^{\frac{1}{2}} \{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} dt + o_p(1) \\
\text{(B.24)} \quad &= - \int_0^L S_{i1}(t; T_i) \Lambda_{i1}(t; T_i) Z'_{i1} \Omega_1(\beta_1)^{-1} n^{-\frac{1}{2}} \sum_{j=1}^n U_{j1}(\beta_1) dt
\end{aligned}$$

$$\text{(B.25)} \quad - \int_0^L S_{i1}(t; T_i) e^{\beta'_1 Z_{i1}} n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^t d\Phi_{j1}(u) dt.$$

For the second term, switching the order of integration and summation

$$\begin{aligned}
\text{(B.25)} &= -n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L \int_u^L S_{i1}(t; T_i) e^{\beta'_1 Z_{i1}} dt d\Phi_{j1}(u) \\
&= -n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L e^{\beta'_1 Z_{i1}} \left\{ \mu_{i1}(T_i) - \int_0^t S_{i1}(u; T_i) du \right\} d\Phi_{j1}(t).
\end{aligned}$$

Combining equations (B.24) and (B.25), we obtain

$$\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_{i1}(T_i) - \mu_{i1}(T_i)\} &= \int_0^L S_{i1}(t; T_i) \Lambda_{i1}(t; T_i) Z'_{i1} \Omega_1(\beta_1)^{-1} n^{-\frac{1}{2}} \sum_{j=1}^n U_{j1}(\beta_1) dt \\
&\quad - n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^L e^{\beta'_1 Z_{i1}} \left\{ \mu_{i1}(T_i) - \int_0^t S_{i1}(u; T_i) du \right\} d\Phi_{j1}(t) \\
&= n^{-\frac{1}{2}} \sum_{j=1}^n \varphi_{ij1}(T_i),
\end{aligned}$$

where

$$\begin{aligned}
\varphi_{ij1}(T_i) &= Z'_{i1} \Omega_1(\beta_1)^{-1} U_{j1}(\beta_1) \int_0^L S_{i1}(t; T_i) \Lambda_{i1}(t; T_i) dt \\
&\quad - e^{\beta'_1 Z_{i1}} \int_0^L \left\{ \mu_{i1}(T_i) - \int_0^t S_{i1}(u; T_i) du \right\} d\Phi_{j1}(t).
\end{aligned}$$

$$\mathbf{B.IV.15} \quad n^{\frac{1}{2}}(\widehat{\alpha} - \alpha_0)$$

It is straightforward to show that

$$n^{\frac{1}{2}}(\widehat{\alpha} - \alpha_0) = \Omega_C(\alpha_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\alpha_0) + o_p(1),$$



where we define

$$U_i^C(\alpha) = \int_0^\tau \{Z_i(0) - \bar{z}_C(t; \alpha)\} dM_i^C(t; \alpha),$$

$$dM_{i1}(t) = dN_{i1}(t) - Y_{i1}(t)d\Lambda_{ik}(t).$$

This is now a well-established Cox model result, derived through Martingale theory.

$$\mathbf{B.IV.16} \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$$

We start the following decomposition

$$(B.26) \quad \begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\} \\ &= n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t; \widehat{\alpha}) - \widehat{\Lambda}_0^C(t; \alpha_0)\} \end{aligned}$$

$$(B.27) \quad +n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t; \alpha_0) - \Lambda_0^C(t)\}.$$

We can express the first term as

$$(B.26) \quad \begin{aligned} &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_C^{(0)}(u; \widehat{\alpha})^{-1} - R_C^{(0)}(u; \alpha_0)^{-1}\} dN_i^C(u) \\ &= \widehat{h}'_C(t; \alpha_0) \Omega_C(\alpha_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\alpha_0) \\ &= h'_C(t; \alpha_0) \Omega_C(\alpha_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\alpha_0) + o_p(1), \end{aligned}$$

where the third line follows from the convergence in probability of

$$\widehat{h}'_C(t; \alpha) = -n^{-1} \sum_{i=1}^n \int_0^t R_C^{(0)}(u; \alpha)^{-1} \bar{z}_C(u; \alpha) dN_i^C(u) = - \int_0^t \bar{z}_C(u; \alpha) d\widehat{\Lambda}_0^C(u; \alpha),$$

$$\widehat{\Omega}_C(\alpha) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{R_C^{(2)}(t; \alpha)}{R_C^{(0)}(t; \alpha)} - \bar{z}_C(t; \beta_1)^{\otimes 2} \right\} dN_i^C(t),$$

to the quantity

$$h'_C(t; \alpha) = - \int_0^t \bar{z}_C(u; \alpha) d\Lambda_0^C(u),$$

and  $\Omega_C(\alpha)$  respectively, with  $\Omega_C(\alpha)$  defined in Regularity Condition (e).

With respect to the second term in the decomposition, we have

$$\begin{aligned}
(B.27) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t R_C^{(0)}(u; \alpha_0)^{-1} dM_i^C(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_C^{(0)}(u; \alpha_0)^{-1} dM_i^C(u) + o_p(1),
\end{aligned}$$

where the second line follows from continuity and Condition (d). Combining equations (B.26) and (B.27), for the decomposition, we have

$$n^{\frac{1}{2}} \{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_i^C(t; \alpha_0) + o_p(1),$$

where

$$\Phi_i^C(t; \alpha) = h'_C(t; \alpha) \Omega_C(\alpha_0)^{-1} U_i^C(\alpha_0) + \int_0^t r_C^{(0)}(u; \alpha_0)^{-1} dM_i^C(u) = \int_0^t d\Phi_i^C(u; \alpha),$$

and

$$d\Phi_i^C(u; \alpha) = -\bar{z}_C(u; \alpha) d\Lambda_0^C(u) \Omega_C(\alpha)^{-1} U_i^C(\alpha) + r_C^{(0)}(u; \alpha)^{-1} dM_i^C(u).$$

$$\mathbf{B.IV.17} \quad n^{\frac{1}{2}} \{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\}$$

We start with another decomposition,

$$\begin{aligned}
(B.28) \quad & n^{\frac{1}{2}} \{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} \\
&= n^{\frac{1}{2}} \left\{ \widehat{\Lambda}_i^C(t; \widehat{\alpha}) - \widehat{\Lambda}_i^C(t; \alpha) \right\}
\end{aligned}$$

$$(B.29) \quad + n^{\frac{1}{2}} \left\{ \widehat{\Lambda}_i^C(t; \alpha) - \Lambda_i^C(t) \right\}.$$

Considering the first term,

$$(B.28) = \widehat{\Lambda}_0^C(t) n^{\frac{1}{2}} \{e^{\widehat{\alpha}' Z_i(0)} - e^{\alpha_0' Z_i(0)}\}.$$

By a Taylor series expansion,

$$\begin{aligned}
n^{\frac{1}{2}} \{e^{\widehat{\alpha}' Z_i(0)} - e^{\alpha_0' Z_i(0)}\} &= Z_i'(0) e^{\alpha_0' Z_i(0)} n^{\frac{1}{2}} (\widehat{\alpha} - \alpha_0) + o_p(1) \\
&= Z_i'(0) e^{\alpha_0' Z_i(0)} \Omega_C(\alpha_0)^{-1} n^{-\frac{1}{2}} \sum_{j=1}^n U_j^C(\alpha_0) + o_p(1).
\end{aligned}$$

As  $\widehat{\Lambda}_0^C(t) \xrightarrow{p} \Lambda_0^C(t)$  for  $t \in [0, \tau]$ , we obtain

$$(B.28) = Z'_i(0)\Lambda_i^C(t)e^{\alpha'_0 Z_i(0)}\Omega_C(\alpha_0)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n U_j^C(\alpha_0) + o_p(1).$$

By using result B.IV.16, the second term can be written as

$$(B.29) = e^{\alpha'_0 Z_i(0)}n^{\frac{1}{2}}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\} \\ = e^{\alpha'_0 Z_i(0)}n^{-\frac{1}{2}}\sum_{j=1}^n d\Phi_j^C(u; \alpha_0) + o_p(1).$$

Combining result leads to

$$n^{\frac{1}{2}}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} = \int_0^t \{Z_i(0) - \bar{z}_C(u; \alpha_0)\}' d\Lambda_i^C(u)\Omega_C(\alpha_0)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n U_j^C(\alpha_0) \\ + n^{-\frac{1}{2}}\sum_{j=1}^n \int_0^t e^{\alpha'_0 Z_i(0)}r_C^{(0)}(u; \alpha_0)^{-1}dM_j^C(u) + o_p(1) \\ = D_i^{C'}(t)\Omega_C(\alpha_0)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n U_j^C(\alpha_0) + n^{-\frac{1}{2}}\sum_{j=1}^n J_{ij}^C(t) + o_p(1),$$

where we define

$$D_i^{C'}(t) = \int_0^t \{Z_i(0) - \bar{z}_C(u; \alpha_0)\}' d\Lambda_i^C(u) = \int_0^t dD_i^{C'}(t), \\ J_{ij}^C(t) = \int_0^t e^{\alpha'_0 Z_i(0)}r_C^{(0)}(u; \alpha_0)^{-1}dM_j^C(u).$$

$$\mathbf{B.IV.18} \quad n^{\frac{1}{2}}\{\widehat{G}_i(t)^{-1} - G_i(t)^{-1}\}$$

Since  $G_i(t)^{-1} = e^{\Lambda_i^C(t)}$  and  $\widehat{G}_i(t)^{-1} = e^{\widehat{\Lambda}_i^C(t)}$ , we then have

$$n^{\frac{1}{2}}\{\widehat{G}_i(t)^{-1} - G_i(t)^{-1}\} = n^{\frac{1}{2}}\{e^{\widehat{\Lambda}_i^C(t)} - e^{\Lambda_i^C(t)}\} \\ = G_i(t)^{-1}n^{\frac{1}{2}}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} + o_p(1) \\ = G_i(t)^{-1}n^{-\frac{1}{2}}\sum_{j=1}^n \{D_i^{C'}(t)\Omega_C(\alpha_0)^{-1}U_j^C(\alpha) + J_{ij}^C(t)\} + o_p(1) \\ = n^{-\frac{1}{2}}\sum_{j=1}^n \varphi_{ij}^C(t) + o_p(1),$$

where

$$\varphi_{ij}^C(t) = G_i(t)^{-1}\{D_i^{C'}(t)\Omega_C(\alpha_0)^{-1}U_j^C(\alpha_0) + J_{ij}^C(t)\}.$$

$$\mathbf{B.IV.19} \quad n^{\frac{1}{2}}\{\widehat{S}_{\Delta_i}(t) - S_{\Delta_i}(t)\}$$

From Result B.IV.8 and B.IV.13, we have

$$n^{\frac{1}{2}}\{\widehat{S}_{i0}(t; T_i) - S_{i0}(t; T_i)\} = -S_{i0}(t; T_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i0}(t; T_i) - \Lambda_{i0}(t; T_i)\} + o_p(1),$$

$$n^{\frac{1}{2}}\{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\} = -S_{i1}(t; T_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} + o_p(1),$$

then using Result B.IV.7 and B.IV.12, we obtain

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{S}_{\Delta_i}(t) - S_{\Delta_i}(t)\} \\ &= n^{\frac{1}{2}}\{\widehat{S}_{i1}(t; T_i) - S_{i1}(t; T_i)\} - n^{\frac{1}{2}}\{\widehat{S}_{i0}(t; T_i) - S_{i0}(t; T_i)\} + o_p(1) \\ &= S_{i0}(t; T_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i0}(t; T_i) - \Lambda_{i0}(t; T_i)\} - S_{i1}(t; T_i)n^{\frac{1}{2}}\{\widehat{\Lambda}_{i1}(t; T_i) - \Lambda_{i1}(t; T_i)\} + o_p(1) \\ &= n^{-\frac{1}{2}} \sum_{j=1}^n \varphi_{ij}^S(t) + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \varphi_{ij}^S(t) &= S_{i0}(t)\{\Lambda_{i0}(t)Z_i(S_i)'\Omega_0^{-1}(\beta_0)U_{j0}(\beta_0) - e^{\beta_0'Z_i(t)}\Phi_{j0}(t)\} \\ &\quad - S_{i1}(t)\{\Lambda_{i1}(t)Z_{i1}'\Omega_1^{-1}(\beta_1)U_{j1}(\beta_1) - e^{\beta_1'Z_{i1}}\Phi_{j1}(t)\}. \end{aligned}$$

$$\mathbf{B.IV.20} \quad n^{\frac{1}{2}}\{\widehat{\Delta}_i(t) - \Delta_i(t)\}$$

Since  $\widehat{\Delta}_i(t) = \int_0^t \widehat{S}_{\Delta_i}(u)du$  and  $\Delta_i(t) = \int_0^t S_{\Delta_i}(u)du$ , and Result B.IV.19, we have

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{\Delta}_i(t) - \Delta_i(t)\} &= n^{\frac{1}{2}} \int_0^t \{\widehat{S}_{\Delta_i}(u) - S_{\Delta_i}(u)\}du, \\ &= n^{-\frac{1}{2}} \int_0^t \sum_{j=1}^n \varphi_{ij}^S(u)du + o_p(1), \end{aligned}$$

where switch the integration and summation sign, we obtain

$$n^{\frac{1}{2}}\{\widehat{\Delta}_i(t) - \Delta_i(t)\} = n^{-\frac{1}{2}} \sum_{j=1}^n \varphi_{ij}^D(t) + o_p(1),$$

where

$$\varphi_{ij}^D(t) = \int_0^t \varphi_{ij}^S(u)du.$$

$$\mathbf{B.IV.21} \quad n^{\frac{1}{2}}\{\widehat{S}_\Delta(t) - S_\Delta(t)\}$$

First define

$$\begin{aligned}\widehat{V}(\tau) &= n^{-1} \sum_{i=1}^n \int_0^\tau \widehat{G}_i(u)^{-1} dN_i^T(u), \\ V(\tau) &= P(T_i \leq t, T_i < D_i),\end{aligned}$$

and

$$\widehat{V}(t) \xrightarrow{p} \int_0^\tau E\left(\frac{dN_i^T(t)}{G_i(t)}\right) = P(T_i \leq t, T_i < D_i) = V(\tau).$$

Then by Slutsky's Theorem, we can write

$$\widehat{S}_\Delta(t) = V(\tau)^{-1} n^{-1} \sum_{i=1}^n \int_0^\tau \widehat{S}_{\Delta i}(t; u) \widehat{G}_i(u)^{-1} dN_i^T(u) + o_p(1).$$

Since we define  $S_\Delta(t) = E[S_\Delta(t; T, Z(T))]$ ,  $\widehat{S}_\Delta(t) - S_\Delta(t)$  can then be decomposed as follows:

$$\begin{aligned}& n^{\frac{1}{2}}\{\widehat{S}_\Delta(t) - S_\Delta(t)\} \\ = & n^{\frac{1}{2}} \left[ \frac{\sum_{i=1}^n \int_0^\tau \widehat{S}_{\Delta i}(t; u) \widehat{G}_i(u)^{-1} dN_i^T(u)}{nV(\tau)} - \frac{\sum_{i=1}^n \int_0^\tau S_{\Delta i}(t; u) \widehat{G}_i(u)^{-1} dN_i^T(u)}{nV(\tau)} \right] \\ & + n^{\frac{1}{2}} \left[ \frac{\sum_{i=1}^n \int_0^\tau S_{\Delta i}(t; u) \widehat{G}_i(u)^{-1} dN_i^T(u)}{nV(\tau)} - \frac{\sum_{i=1}^n \int_0^\tau S_{\Delta i}(t; u) G_i(u)^{-1} dN_i^T(u)}{nV(\tau)} \right] \\ & + n^{\frac{1}{2}} \left[ \frac{\sum_{i=1}^n \int_0^\tau S_{\Delta i}(t; u) G_i(u)^{-1} dN_i^T(u)}{nV(\tau)} - S_\Delta(t) \right] + o_p(1).\end{aligned}$$

Then we can write

$$(B.30) \quad n^{\frac{1}{2}}\{\widehat{S}_\Delta(t) - S_\Delta(t)\} = V(\tau)^{-1} n^{-1} \sum_{i=1}^n \int_0^\tau n^{\frac{1}{2}}\{\widehat{S}_{\Delta i}(t; u) - S_{\Delta i}(t; u)\} \widehat{G}_i(u)^{-1} dN_i^T(u)$$

$$(B.31) \quad + V(\tau)^{-1} n^{-1} \sum_{i=1}^n \int_0^\tau S_{\Delta i}(t; u) n^{\frac{1}{2}}\{\widehat{G}_i(u)^{-1} - G_i(u)^{-1}\} dN_i^T(u)$$

$$(B.32) \quad + n^{-\frac{1}{2}} V(\tau)^{-1} \sum_{i=1}^n \int_0^\tau \{S_{\Delta i}(t; u) - S_\Delta(t)\} G_i(u)^{-1} dN_i^T(u) + o_p(1).$$

By Results B.IV.19 and Slutsky Theorem, we have the following decomposition

$$\begin{aligned}
(B.30) &= n^{-\frac{1}{2}}V(\tau)^{-1}n^{-1}\sum_{i=1}^n\int_0^\tau\sum_{j=1}^n\varphi_{ij}^S(t)G_i(u)^{-1}dN_i^T(u)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^nn^{-1}\sum_{i=1}^n\int_0^\tau\varphi_{ij}^S(t)G_i(u)^{-1}dN_i^T(u)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^n\widehat{V}_{1j}(t)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^nV_{1j}(t)+o_p(1)
\end{aligned}$$

where

$$\begin{aligned}
\widehat{V}_{1j}(t) &= n^{-1}\sum_{i=1}^n\int_0^\tau\varphi_{ij}^S(t)G_i(u)^{-1}dN_i^T(u), \\
V_{1j}(t) &= E\left[\int_0^\tau\varphi_{ij}^S(t)G_i(u)^{-1}dN_i^T(u)\right].
\end{aligned}$$

By Results B.IV.18 we have

$$\begin{aligned}
(B.31) &= n^{-\frac{1}{2}}V(\tau)^{-1}n^{-1}\sum_{i=1}^n\int_0^\tau S_{\Delta_i}(t;u)\sum_{j=1}^n\varphi_{ij}^C(u)dN_i^T(u)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^nn^{-1}\sum_{i=1}^n\int_0^\tau S_{\Delta_i}(t;u)\varphi_{ij}^C(u)dN_i^T(u)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^n\widehat{V}_{2j}(t)+o_p(1) \\
&= n^{-\frac{1}{2}}V(\tau)^{-1}\sum_{j=1}^nV_{2j}(t)+o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{V}_{2j}(t) &= n^{-1}\sum_{i=1}^n\int_0^\tau S_{\Delta_i}(t;u)\varphi_{ij}^C(u)dN_i^T(u), \\
V_{2j}(t) &= E\left[\int_0^\tau S_{\Delta_i}(t;u)\varphi_{ij}^C(u)dN_i^T(u)\right].
\end{aligned}$$

Combining all the results above, we can have

$$\begin{aligned}
& n^{\frac{1}{2}}\{\widehat{S}_\Delta(t) - S_\Delta(t)\} \\
= & n^{-\frac{1}{2}} \sum_{j=1}^n V(\tau)^{-1} \left\{ V_{1j}(t) + V_{2j}(t) + \int_0^\tau \{S_{\Delta_j}(t; u) - S_\Delta(t)\} G_j(u)^{-1} dN_j^T(u) \right\} + o_p(1) \\
= & n^{-\frac{1}{2}} \sum_{j=1}^n \xi_j(t) + o_p(1),
\end{aligned}$$

where

$$\xi_j(t) = V(\tau)^{-1} \left\{ V_{1j}(t) + V_{2j}(t) + \int_0^\tau \{S_{\Delta_j}(t; u) - S_\Delta(t)\} G_j(u)^{-1} dN_j^T(u) \right\}.$$

**B.IV.22**  $n^{\frac{1}{2}}\{\widehat{\Delta} - \Delta\}$

Since  $\widehat{\Delta} = \int_0^L \widehat{S}_\Delta(t) dt$  and  $\Delta = \int_0^L S_\Delta(t) dt$ , we can have

$$n^{\frac{1}{2}}\{\widehat{\Delta} - \Delta\} = n^{-\frac{1}{2}} \sum_{j=1}^n \eta_j + o_p(1),$$

where

$$\eta_j = \int_0^L \xi_j(t) dt.$$

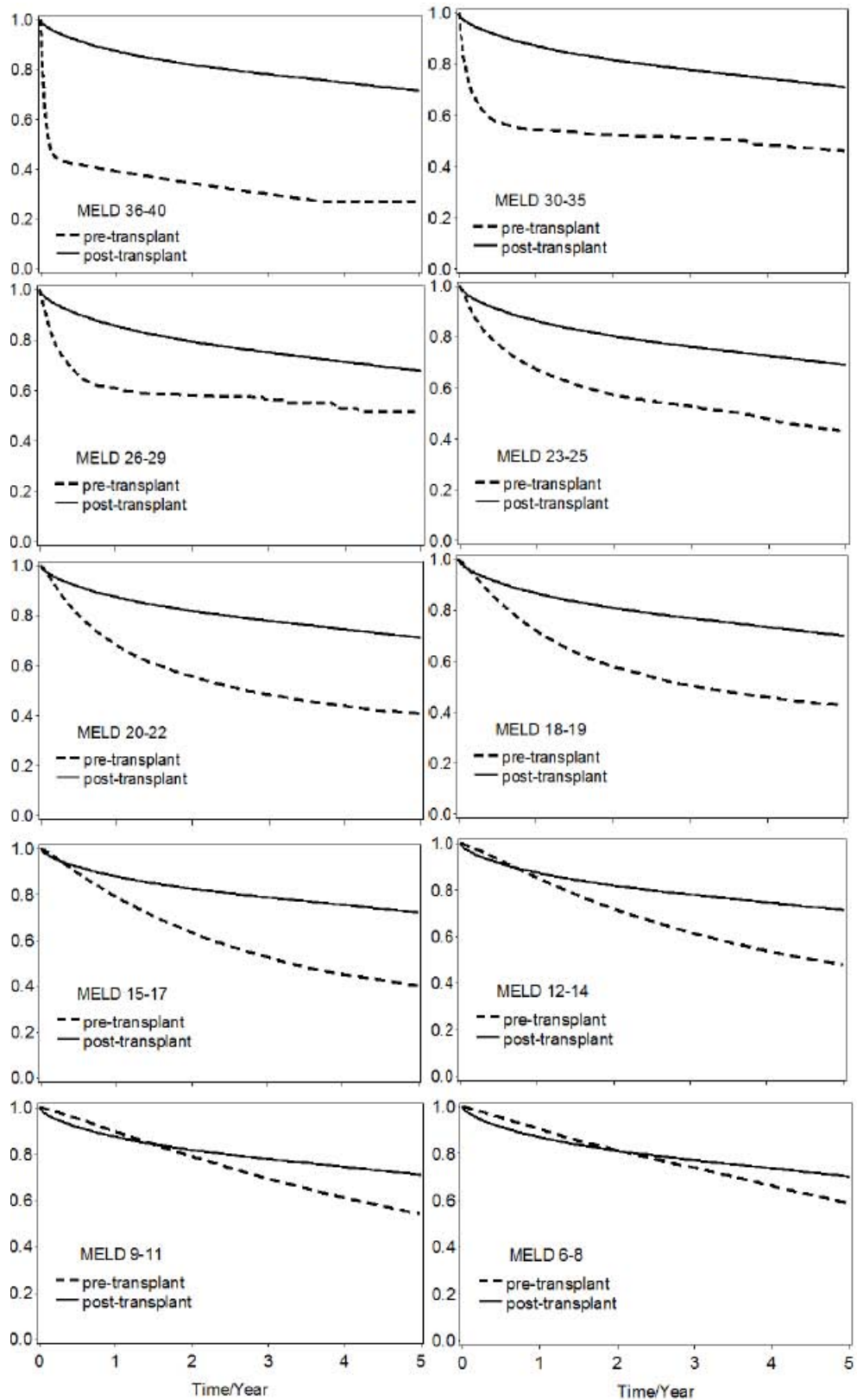


Figure B.1: Estimated post- and pre-transplant survival curves for all MELD groups



## APPENDIX C

### Appendix C: Proof of Theorem VI.1

#### Appendix C: Proof of Theorem IV.1

##### C.I. Notation

We begin by reviewing the essential notation:

$i$ : subject ( $i = 1, \dots, n$ )

$j$ : strata ( $j = 0, 1, \dots, J - 1$ )

$D_i$ : death time for  $i$ th subject

$C_i$ : censoring time for  $i$ th subject

$T_i$ : treatment time for  $i$ th subject

$X_i = \min\{D_i, C_i, T_i\}$ : observation time for  $i$ th subject

$\Delta_i = I(X_i = D_i)$

$\Delta_i^T = I(X_i = T_i)$

$A_i(t)$ : treatment eligibility indicator of  $i$ th subject at time  $t$

$Z_i(t)$ : covariate for  $i$ th subject at follow-up time  $t$

$Z_{i0} = Z_i(0)$ : covariate for  $i$ th subject at follow-up time  $t = 0$

$\tilde{Z}_i(t) = \{Z_i(s); s \in [0, t]\}$ : covariate history up to time  $t$

$\tilde{A}_i(t) = \{A_i(s); s \in [0, t]\}$ : treatment eligibility history up to time  $t$

$G_i$ : group variable for  $i$ th subject

$$Y_i(t) = I(X_i \geq t)$$

$$Y_{ij}(t) = Y_i(t)I(G_i = j)$$

$\beta_0$ : regression parameter, death model

$\theta_0$ : regression parameter, treatment model

Death baseline hazard:  $\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\beta'_0 Z_{i0}\}$

Treatment baseline hazard:  $\lambda_{ij}^T(t) = \lambda_{0j}^T(t) \exp\{\theta'_0 Z_i(t)\}$

## C.II. Regularity Conditions

In deriving the asymptotic properties of the proposed estimators the following conditions are assumed for  $i = 1, \dots, n$  and  $j = 0, \dots, J - 1$

(a)  $\{X_i, \Delta_i, \Delta_i^T, \tilde{Z}_i(X_i), \tilde{A}_i(X_i)\}$  are independent and identically distributed random vectors.

(b)  $Z_i(t)$  has bounded variation, i.e.,  $|Z_i(t)| < \kappa$ , where  $\kappa$  is a constant for  $t \in [0, \tau]$ .

(c)  $\int_0^\tau \lambda_{0j}(t)dt < \infty$  and  $\int_{0j}^\tau \lambda_{0j}^T(t)dt < \infty$  where  $\tau$  is the maximum follow-up time.

(d) Continuity of the following functions:

$$\begin{aligned} r_j^{(1)}(t; \beta, W) &= \frac{\partial}{\partial \beta} r_j^{(0)}(t; \beta, W), \\ r_j^{(2)}(t; \beta, W) &= \frac{\partial^2}{\partial \beta \partial \beta'} r_j^{(0)}(t; \beta, W), \end{aligned}$$

and  $r_j^{(0)}(t; \beta, W)$ , where

$$r_j^{(p)}(t; \beta, W) = E[W_i(t)Y_{ij}(t)Z_{i0}^{\otimes p} \exp(\beta' Z_{i0})],$$

is the limiting value of

$$R_j^{(p)}(t; \beta, W) = n^{-1} \sum_{i=1}^n W_i(t) Y_{ij}(t) Z_{i0}^{\otimes p} \exp(\beta' Z_{i0}),$$

for  $p = 0, 1, 2$ , with  $r_j^{(1)}(t; \beta, W)$  and  $r_j^{(2)}(t; \beta, W)$  bounded and  $r_j^{(0)}(t; \beta, W)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\beta$  in an open set.

(e) Continuity of the following functions:

$$\begin{aligned} r_{T_j}^{(1)}(t; \theta) &= \frac{\partial}{\partial \theta} r_{T_j}^{(0)}(t; \theta), \\ r_{T_j}^{(2)}(t; \theta) &= \frac{\partial^2}{\partial \theta \partial \theta'} r_{T_j}^{(0)}(t; \theta), \end{aligned}$$

and  $r_{T_j}^{(0)}(t; \theta)$ , where

$$r_{T_j}^{(p)}(t; \theta) = E[Y_{ij}(t) Z_i(t)^{\otimes p} \exp\{\theta' Z_i(t)\}],$$

is the limiting value of

$$R_{T_j}^{(d)}(t; \theta) = n^{-1} \sum_{i=1}^n Y_{ij}(t) Z_i(t)^{\otimes d} \exp\{\theta' Z_i(t)\},$$

for  $p = 0, 1, 2$ ; with  $r_{T_j}^{(1)}(t; \theta)$  and  $r_{T_j}^{(2)}(t; \theta)$  bounded and  $r_{T_j}^{(0)}(t; \theta)$  bounded away from 0 for  $t \in [0, \tau]$  and  $\theta$  in an open set.

(f) Positive-definiteness of the matrices  $\Omega_T(\theta)$  and  $\Omega(\beta)$ , where

$$\begin{aligned} \Omega_T(\theta) &= E \left[ \sum_{j=0}^{J-1} \int_0^\tau \left\{ \frac{r_{T_j}^{(2)}(t; \theta)}{r_{T_j}^{(0)}(t; \theta)} - \bar{x}_j(t; \theta)^{\otimes 2} \right\} dN_{ij}^T(t) \right], \\ \bar{x}_j(t; \theta) &= r_{T_j}^{(1)}(t; \theta) / r_{T_j}^{(0)}(t; \theta), \\ \Omega(\beta) &= E \left[ \sum_{j=0}^{J-1} \int_0^\tau \left\{ \frac{r_j^{(2)}(t; \beta, W)}{r_j^{(0)}(t; \beta, W)} - \bar{z}_j(t; \beta, W)^{\otimes 2} \right\} dN_{ij}(t) \right], \\ \bar{z}_j(t; \beta, W) &= r_j^{(1)}(t; \beta, W) / r_j^{(0)}(t; \beta, W). \end{aligned}$$

(g)  $P(Y_i(t) = 1) > 0$  for  $t \in [0, \tau]$ .

### C.III. Outline of Asymptotic Derivation

The proof below focuses on the unstabilized weight. We derive the influence functions of various terms of interest as summations of independent and identically distributed (i.i.d.) terms plus a term which converges to zero in probability. The terms are as follows:

1.  $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$
2.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}^T(t) - \Lambda_{0j}^T(t)\}$
3.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}^T(t) - \Lambda_{ij}^T(t)\}$
4.  $n^{\frac{1}{2}}\{\widehat{W}_i(t) - W_i(t)\}$
5.  $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0)$
6.  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\}$
7.  $n^{\frac{1}{2}}\{\widehat{\phi}_j(t) - \phi_j(t)\}$

### C.IV. Derivation of Asymptotic Properties

#### Asymptotic Normality:

Some parts of the proof regarding the proportional hazards model are results well-established. Therefore, they are simply listed without proof. For details, please refer to Anderson and Gill (1982), Fleming and Harrington (1991) and Andersen et al. (1993).

**C.IV.1**  $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$ 

As  $n \rightarrow \infty$ , we have

$$n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) = \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{j=0}^{J-1} U_{ij}^T(\theta_0) + o_p(1),$$

where

$$\begin{aligned} U_{ij}^T(\theta) &= \int_0^\tau \{Z_i(t) - \bar{x}_j(t; \theta)\} dM_{ij}^T(t; \theta), \\ dM_{ij}^T(t) &= dN_{ij}^T(t) - Y_{ij}(t) d\Lambda_i^T(t), \end{aligned}$$

This is now a well-established Cox model result, derived through Martingale theory.

**C.IV.2**  $n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}^T(t) - \Lambda_{0j}^T(t)\}$ 

We induce the following decomposition:

$$(C.1) \quad \begin{aligned} &n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}^T(t) - \Lambda_{0j}^T(t)\} \\ &= n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}^T(t; \widehat{\theta}) - \Lambda_{0j}^T(t; \theta_0)\} \end{aligned}$$

$$(C.2) \quad + n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}^T(t; \theta_0) - \Lambda_{0j}^T(t)\}.$$

We can express the first term as

$$\begin{aligned} (C.1) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_{Tj}^{(0)}(u; \widehat{\theta})^{-1} - R_{Tj}^{(0)}(u; \theta_0)^{-1}\} dN_{ij}^T(u) \\ &= \widehat{h}'_{Tj}(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{j=0}^{J-1} U_{ij}^T(\theta_0) \\ &= h'_{Tj}(t; \theta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{j=0}^{J-1} U_{ij}^T(\theta_0) + o_p(1). \end{aligned}$$

where where the third line follows from the convergence in probability of

$$\begin{aligned} \widehat{h}'_{Tj}(t; \theta) &= -\frac{1}{n} \sum_{i=1}^n \int_0^t R_{Tj}^{(0)}(u; \theta)^{-1} \bar{X}_j(u; \theta) dN_{ij}^T(u) = -\int_0^t \bar{X}_j(u; \theta) d\widehat{\Lambda}_{0j}^T(u; \theta), \\ \widehat{\Omega}_T(\theta) &= n^{-1} \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^\tau \left\{ \frac{R_{Tj}^{(2)}(t; \theta)}{R_{Tj}^{(0)}(t; \theta)} - \bar{X}_j(t, \theta)^{\otimes 2} \right\} dN_{ij}^T(t), \end{aligned}$$

where  $\bar{X}_j(t; \theta) = R_{T_j}^{(1)}(t; \theta)/R_{T_j}^{(0)}(t; \theta)$ , to the quantities

$$h'_{T_j}(t; \theta) = - \int_0^t \bar{x}_j(u; \theta) d\Lambda_{0j}^T(u),$$

and  $\Omega_T(\theta)$  respectively, with  $\Omega_T(\theta)$  defined in Regularity Condition (f).

With respect to the second term in the decomposition, we have,

$$\begin{aligned} (C.2) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t R_{T_j}^{(0)}(u; \theta_0)^{-1} dM_{ij}^T(u) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_{T_j}^{(0)}(u; \theta_0)^{-1} dM_{ij}^T(u) + o_p(1), \end{aligned}$$

where the second line follows from continuity and Condition (f). Combining results, for the decomposition, we have

$$n^{\frac{1}{2}} \{\widehat{\Lambda}_{0j}^T(t) - \Lambda_{0j}^T(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_{ij}^T(t; \theta_0) + o_p(1),$$

where

$$\Phi_{ij}^T(t; \theta) = h'_{T_j}(t; \theta) \Omega_T(\theta)^{-1} U_{ij}^T(\theta) + \int_0^t r_{T_j}^{(0)}(u; \theta)^{-1} dM_{ij}^T(u) = \int_0^t d\Phi_{ij}^T(u; \theta),$$

and

$$d\Phi_{ij}^T(u; \theta) = -\bar{x}'_j(u; \theta) d\Lambda_{0j}^T(u) \Omega_T(\theta)^{-1} U_{ij}^T(\theta) + r_{T_j}^{(0)}(u; \theta)^{-1} dM_{ij}^T(u).$$

### C.IV.3 $n^{\frac{1}{2}} \{\widehat{\Lambda}_{ij}^T(t) - \Lambda_{ij}^T(t)\}$

We begin with another decomposition,

$$\begin{aligned} (C.3) \quad & n^{\frac{1}{2}} \{\widehat{\Lambda}_{ij}^T(t) - \Lambda_{ij}^T(t)\} \\ &= n^{\frac{1}{2}} \left\{ \int_0^t e^{\widehat{\theta}' Z_i(u)} d\widehat{\Lambda}_{0j}^T(u) - \int_0^t e^{\theta_0' Z_i(u)} d\widehat{\Lambda}_{0j}^T(u) \right\} \\ (C.4) \quad & + n^{\frac{1}{2}} \left\{ \int_0^t e^{\theta_0' Z_i(u)} d\widehat{\Lambda}_{0j}^T(u) - \int_0^t e^{\theta_0' Z_i(u)} d\Lambda_{0j}^T(u) \right\}. \end{aligned}$$

Considering the first term,

$$(C.3) = n^{\frac{1}{2}} \int_0^t \{e^{\widehat{\theta}' Z_i(u)} - e^{\theta_0' Z_i(u)}\} d\widehat{\Lambda}_{0j}^T(u).$$

By a Taylor series expansion,

$$\begin{aligned} n^{\frac{1}{2}}\{e^{\widehat{\theta}'Z_i(u)} - e^{\theta_0'Z_i(u)}\} &= Z_i(u)e^{\theta_0'Z_i(u)}n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) + o_p(1) \\ &= Z_i(u)e^{\theta_0'Z_i(u)}\Omega_T(\theta_0)^{-1}n^{-\frac{1}{2}}\sum_{l=1}^n\sum_{k=0}^{J-1}U_{lk}^T(\theta_0) + o_p(1). \end{aligned}$$

As  $\widehat{\Lambda}_{0j}^T(t) \xrightarrow{p} \Lambda_{0j}^T(t)$  for  $t \in [0, \tau]$ , we get

$$(C.3) = \int_0^t Z_i(u)d\Lambda_{ij}^T(u)\Omega_T(\theta_0)^{-1}n^{-\frac{1}{2}}\sum_{l=1}^n\sum_{k=0}^{J-1}U_{lk}^T(\theta_0) + o_p(1).$$

By using result C.IV.2, the second term can be written as

$$\begin{aligned} (C.4) &= n^{\frac{1}{2}}\int_0^t e^{\theta_0'Z_i(u)}d\{\widehat{\Lambda}_{0j}^T(t) - \Lambda_{0j}^T(t)\} \\ &= \int_0^t e^{\theta_0'Z_i(u)}n^{-\frac{1}{2}}\sum_{l=1}^n d\Phi_{lj}^T(u; \theta_0) + o_p(1). \end{aligned}$$

Combining result from the decomposition, then we have

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}^T(t) - \Lambda_{ij}^T(t)\} &= \int_0^t \{Z_i(u) - \bar{x}_j(u; \theta_0)\}'d\Lambda_{ij}^T(u)\Omega_T(\theta_0)^{-1}n^{-\frac{1}{2}}\sum_{l=1}^n\sum_{k=0}^{J-1}U_{lk}^T(\theta_0) \\ &\quad + n^{-\frac{1}{2}}\sum_{l=1}^n\int_0^t e^{\theta_0'Z_i(u)}r_{Tj}^{(0)}(u; \theta_0)^{-1}dM_{lj}^T(u) + o_p(1) \\ &= D'_{ij}(t; \theta_0)\Omega_T(\theta_0)^{-1}n^{-\frac{1}{2}}\sum_{l=1}^n\sum_{k=0}^{J-1}U_{lk}^T(\theta_0) + n^{-\frac{1}{2}}\sum_{l=1}^n J_{ilj}^T(t; \theta_0) + o_p(1), \end{aligned}$$

where we define

$$\begin{aligned} D_{ij}(t; \theta) &= \int_0^t \{Z_i(u) - \bar{x}_j(u; \theta)\}'d\Lambda_{ij}^T(u) = \int_0^t dD_{ij}(t; \theta), \\ J_{ilj}^T(t; \theta) &= \int_0^t e^{\theta'Z_i(u)}r_{Tj}^{(0)}(u; \theta)^{-1}dM_{lj}^T(u). \end{aligned}$$

#### C.IV.4 $n^{\frac{1}{2}}\{\widehat{W}_i(t) - W_i(t)\}$

As  $W_i(t|G_i) = \exp\{\Lambda_{iG_i}^T(t)\}$  and  $\widehat{W}_i(t|G_i) = \exp\{\widehat{\Lambda}_{iG_i}^T(t)\}$ , we have

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{W}_i(t) - W_i(t)\} &= n^{\frac{1}{2}}\{\exp\{\widehat{\Lambda}_{iG_i}^T(t)\} - \exp\{\Lambda_{iG_i}^T(t)\}\} \\ &= W_i(t)n^{\frac{1}{2}}\{\widehat{\Lambda}_{iG_i}^T(t) - \Lambda_{iG_i}^T(t)\} + o_p(1) \\ &= W_i(t)n^{-\frac{1}{2}}\sum_{l=1}^n\left\{D'_{iG_i}(t; \theta_0)\Omega_T(\theta_0)^{-1}\sum_{k=0}^{J-1}U_{lk}^T(\theta_0) + J_{ilG_i}^T(t; \theta_0)\right\} + o_p(1). \end{aligned}$$

**C.IV.5**  $n^{\frac{1}{2}}\{\widehat{\beta} - \beta_0\}$

It is straight forward to show that

$$n^{\frac{1}{2}}\{\widehat{\beta} - \beta_0\} = \Omega^{-1}(\beta_0)n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}U_{ij}(\beta_0, \widehat{W}) + o_p(1),$$

where

$$U_{ij}(\beta, W) = \int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\}W_i(t)dM_{ij}(t),$$

$$dM_{ij}(t) = dN_{ij}(t) - Y_{ij}(t)d\Lambda_{ij}(t).$$

The term  $n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}U_{ij}(\beta, \widehat{W})$  can be decomposed as follows

$$n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}U_{ij}(\beta, \widehat{W}) = n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, \widehat{W})\}\widehat{W}_i(t)dM_{ij}(t)$$

(C.5)  $= n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\}W_i(t)dM_{ij}(t)$

(C.6)  $-n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{\bar{Z}_j(t; \beta, \widehat{W}) - \bar{z}_j(t; \beta, W)\}W_i(t)dM_{ij}(t)$

(C.7)  $+n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, \widehat{W})\}\{\widehat{W}_i(t) - W_i(t)\}dM_{ij}(t) + o_p(1).$

Now, through the Functional Delta Method, combined with a lot of tedious algebra,

(C.6) converges in probability to 0.

Using result C.IV.4

(C.7)  $= n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\}W_i(t)$

(C.8)  $\times n^{-1}\sum_{l=1}^n D'_{iG_i}(t; \theta_0)\Omega_T(\theta_0)^{-1}\sum_{k=0}^{J-1}U_{lk}^T(\theta_0)dM_{ij}(t)$

$+n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{j=0}^{J-1}\int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\}W_i(t)$

(C.9)  $\times n^{-1}\sum_{l=1}^n J_{ilG_i}^T(t; \theta_0)dM_{ij}(t).$



Switching the order of summation, we have

$$\begin{aligned}
(C.8) &= n^{-1} \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) D'_{iG_i}(t; \theta_0) dM_{ij}(t) \Omega_T(\theta_0)^{-1} \\
&\quad \times n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) \\
&= \widehat{H}'(t; \beta, W) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) \\
&= H'(t; \beta, W) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0),
\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{H}'(t; \beta, W) = n^{-1} \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) D'_{iG_i}(t; \theta_0) dM_{ij}(t),$$

to the quantity

$$H'(t; \beta, W) = E \left[ \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) D'_{iG_i}(t; \theta_0) dM_{ij}(t) \right].$$

Switching the order of summation and integration

$$\begin{aligned}
(C.9) &= n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^\tau \left[ n^{-1} \sum_{i=1}^n \sum_{j=0}^{J-1} e^{\theta'_0 Z_i(u)} \int_u^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) dM_{ij}(t) \right] \\
&\quad \times r_{TG_i}^{(0)}(u; \theta_0)^{-1} dM_{lG_i}^T(u) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \widehat{V}_l(\beta) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n V_l(\beta) + o_p(1),
\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\begin{aligned}
\widehat{V}_l(\beta) &= \int_0^\tau \left[ n^{-1} \sum_{i=1}^n \sum_{j=0}^{J-1} e^{\theta'_0 Z_i(u)} \int_u^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) dM_{ij}(t) \right] \\
&\quad \times r_{TG_i}^{(0)}(u; \theta_0)^{-1} dM_{lG_i}^T(u),
\end{aligned}$$

to the quantity

$$V_l(\beta) = E \left[ \int_0^\tau \sum_{j=0}^{J-1} e^{\theta'_0 Z_i(t_1)} \int_u^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) dM_{ij}(t) r_{TG_i}^{(0)}(u; \theta_0)^{-1} dM_{lG_i}^T(u) \right].$$

Combining equations (C.5) (C.8) and (C.9), we obtain

$$n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) = \Omega(\beta_0)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_i(\beta_0) + o_p(1),$$

where we set

$$\begin{aligned} U_i(\beta) &= \sum_{j=0}^{J-1} \int_0^\tau \{Z_{i0} - \bar{z}_j(t; \beta, W)\} W_i(t) dM_{ij}(t) \\ &\quad + H'(t; \beta, W) \Omega_T(\theta)^{-1} \sum_{j=0}^{J-1} U_{ij}^T(\theta) + V_i(\beta). \end{aligned}$$

$$\mathbf{C.IV.6} \quad n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\}$$

We begin another decomposition,

$$\begin{aligned} &n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\} \\ (C.10) \quad &= n^{\frac{1}{2}}[\widehat{\Lambda}_{0j}\{t; \widehat{W}, R_j(\widehat{\beta}, \widehat{W})\} - \widehat{\Lambda}_{0j}\{t; \widehat{W}, R_j(\beta_0, \widehat{W})\}] \end{aligned}$$

$$(C.11) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{0j}\{t; \widehat{W}, R_j(\beta_0, \widehat{W})\} - \widehat{\Lambda}_{0j}\{t; W, R_j(\beta_0, \widehat{W})\}]$$

$$(C.12) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{0j}\{t; W, R_j(\beta_0, \widehat{W})\} - \widehat{\Lambda}_{0j}\{t; W, R_j(\beta_0, W)\}]$$

$$(C.13) \quad + n^{\frac{1}{2}}[\widehat{\Lambda}_{0j}\{t; W, R_j(\beta_0, W)\} - \Lambda_{0j}(t)]$$

By using Result C.IV.5, we can express the first term as

$$\begin{aligned} (C.10) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{R_j^{(0)}(u; \widehat{\beta}, \widehat{W})^{-1} - R_j^{(0)}(u; \beta_0, \widehat{W})^{-1}\} \widehat{W}_i(u) dN_{ij}(u) \\ &= - \int_0^t \bar{z}'_j(u; \beta_0, W) d\Lambda_{0j}(u) \Omega(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i(\beta_0) + o_p(1) \\ &= h'_j(t; \beta_0, W) \Omega(\beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i(\beta_0) + o_p(1), \end{aligned}$$

where we define

$$h_j(t; \beta, W) = - \int_0^t \bar{z}'_j(u; \beta, W) d\Lambda_{0j}(u).$$

By using Result C.IV.4, we have

$$\begin{aligned}
(C.11) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{\widehat{W}_i(u) - W_i(u)\} R_j^{(0)}(u; \beta_0, \widehat{W})^{-1} dN_{ij}(u) \\
&= n^{-1} \sum_{i=1}^n \int_0^t R_j^{(0)}(u; \beta_0, W)^{-1} W_i(u) D'_{iG_i}(u) \Omega_T(\theta_0)^{-1} \\
(C.14) &\quad \times n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) dN_{ij}(u) \\
(C.15) &\quad + n^{-1} \sum_{i=1}^n \int_0^t R_j^{(0)}(u; \beta_0, W)^{-1} W_i(u) n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilG_i}^T(u) dN_{ij}(u) + o_p(1).
\end{aligned}$$

Switching the order of summation, we obtain

$$\begin{aligned}
(C.14) &= \widehat{B}'_j(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) \\
&= B'_j(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) + o_p(1),
\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{B}_j(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t R_j^{(0)}(u; \beta, W)^{-1} W_i(u) D'_{iG_i}(u) dN_{ij}(u)$$

to the quantity

$$B_j(t; \beta) = E \left[ \int_0^t r_j^{(0)}(u; \beta, W)^{-1} W_i(u) D'_{iG_i}(u) dN_{ij}(u) \right].$$

Switching the order of summation and integration

$$\begin{aligned}
(C.15) &= n^{-\frac{1}{2}} \sum_{l=1}^n \widehat{K}_{lj}(\beta_0) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n K_{lj}(\beta_0) + o_p(1),
\end{aligned}$$

where the last equality follows from the convergence in probability of

$$\widehat{K}_{lj}(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t e^{\theta'_0 Z_i(s)} \int_s^t W_i(u) R_j^{(0)}(u; \beta, W)^{-1} dN_{ij}(u) r_{TG_i}^{(0)}(s; \theta_0)^{-1} dM_{lG_i}^T(s)$$

to the quantity

$$K_{lj}(t; \beta) = E \left[ \int_0^t e^{\theta'_0 Z_i(t_1)} \int_{t_1}^{t_2} W_i(u) r_j^{(0)}(u; \beta, W)^{-1} dN_{ij}(u) r_{TG_i}^{(0)}(u; \theta_0)^{-1} dM_{lG_i}^T(u) \right].$$

Combining equations (C.14) and (C.15), we have

$$(C.11) = \widehat{B}'_j(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) + n^{-\frac{1}{2}} \sum_{l=1}^n K_{lj}(t; \beta_0) + o_p(1).$$

We can have

$$(C.12) = n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{j=0}^{J-1} \int_0^t W_i(u) \{R_j^{(0)}(u; \beta_0, \widehat{W})^{-1} - R_j^{(0)}(u; \beta_0, W)^{-1}\} dN_{ij}(u).$$

Now through the Function Delta Method,

$$\begin{aligned} & n^{\frac{1}{2}} \{R_j^{(0)}(u; \beta, \widehat{W})^{-1} - R_j^{(0)}(u; \beta, W)^{-1}\} \\ = & -R_j^{(0)}(u; \beta, W)^{-2} n^{-1} \sum_{i=1}^n e^{\beta' Z_{i0}} n^{\frac{1}{2}} \{\widehat{W}_i(u) - W_i(u)\} \\ = & -R_j^{(0)}(u; \beta, W)^{-2} n^{-1} \sum_{i=1}^n e^{\beta' Z_{i0}} W_i(u) n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ D'_{iG_i}(u) \Omega_T(\theta_0)^{-1} \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) + J'_{iG_i}(u) \right\} \\ = & R_j^{(0)}(u; \beta, W)^{-2} \widehat{F}'(u; \beta) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) \\ & + R_j^{(0)}(u; \beta, W)^{-2} n^{-\frac{1}{2}} \sum_{l=1}^n \widehat{Q}'_l(s, u; \theta_0) + o_p(1) \\ = & R_j^{(0)}(u; \beta, W)^{-2} F'(u; \beta) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) \\ & + R_j^{(0)}(u; \beta, W)^{-2} n^{-\frac{1}{2}} \sum_{i=1}^n Q'_l(s, u; \theta_0) + o_p(1), \end{aligned}$$

where the last line follows from the convergence in probability of

$$\begin{aligned} \widehat{F}(u; \beta) &= -n^{-1} \sum_{i=1}^n e^{\beta' Z_{i0}} W_i(u) D'_{iG_i}(u; \theta), \\ \widehat{Q}'_l(u; \theta) &= -n^{-1} \sum_{i=1}^n \int_0^u e^{\theta' Z_i(s)} e^{\beta' Z_{i0}} W_i(u) r_{TG_i}^{(0)}(s, \theta_0)^{-1} dM_{lG_i}^T(s). \end{aligned}$$

to the quantities

$$F(u; \beta) = -E \left[ e^{\beta' Z_{i0}} W_i(u) D'_{iG_i}(u; \theta) \right],$$

$$Q'_l(u; \theta) = -E \left[ \int_0^u e^{\theta' Z_i(s)} e^{\beta' Z_{i0}} W_i(u) r_{TG_i}^{(0)}(s, \theta_0)^{-1} dM_{iG_i}^T(s) \right],$$

Substituting this result into the expansion of (C.12), we obtain

$$(C.12) = n^{-1} \sum_{i=1}^n \int_0^t W_i(u) R_j^{(0)}(u; \beta_0, W)^{-2} F'(u; \beta_0) \Omega_T(\theta_0)^{-1}$$

$$\times n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) dN_{ij}(u)$$

$$+ n^{-1} \sum_{i=1}^n \int_0^t W_i(u) R_j^{(0)}(u; \beta_0, W)^{-2} n^{-\frac{1}{2}} \sum_{l=1}^n Q'_l(u; \theta) dN_{ij}(u).$$

Switching the order of summation for the first term, and the order of summation and integration in the second term, we have

$$(C.12) = \widehat{E}_j(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) + n^{-\frac{1}{2}} \sum_{l=1}^n \widehat{P}_{lj}(t; \beta_0)$$

$$= E_j(t; \beta_0) \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{k=0}^{J-1} U_{lk}^T(\theta_0) + n^{-\frac{1}{2}} \sum_{l=1}^n P_{lj}(u; \beta_0) + o_p(1),$$

where the last equality follows the convergence in probability of

$$\widehat{E}_j(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t \frac{W_i(u) F(u; \beta)}{R_j^{(0)}(u; \beta, W)^2} dN_{ij}(u),$$

$$\widehat{P}_{lj}(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t \frac{W_i(u) Q_l(u; \theta_0)}{R_j^{(0)}(u; \beta, W)^2} dN_{ij}(u),$$

to the quantities

$$E_j(t; \beta) = E \left[ \int_0^t \frac{W_i(u) F(u; \beta)}{r_j^{(0)}(u; \beta, W)^2} dN_{ij}(u) \right],$$

$$P_{lj}(t; \beta) = E \left[ \int_0^t \frac{W_i(u) Q_l(u; \theta_0)}{r_j^{(0)}(u; \beta, W)^2} dN_{ij}(u) \right].$$

We can also express

$$\begin{aligned}
(C.13) &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{W_i(u)}{R_j^{(0)}(u; \beta_0, W)} dM_{ij}(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{W_i(u)}{r_j^{(0)}(u; \beta_0, W)} dM_{ij}(u) + o_p(1).
\end{aligned}$$

Combining the results of equations (C.10) (C.11) (C.12) and (C.13), we obtain

$$\begin{aligned}
&n^{\frac{1}{2}} \{ \widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t) \} \\
&= h'_j(t; \beta_0, W) \Omega^{-1}(\beta_0) n^{-\frac{1}{2}} \sum_{i=1}^n U_i(\beta_0) \\
&\quad + [B'_j(t; \beta_0) + E'_j(t; \beta_0)] \Omega_T(\theta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=0}^{J-1} U_{ik}^T(\theta_0) \\
&\quad + n^{-\frac{1}{2}} \sum_{i=1}^n [K_{ij}(t; \beta_0) + P_{ij}(t; \beta_0)] \\
&\quad + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_j^{(0)}(u; \beta_0, W)^{-1} W_i(u) dM_{ij}(u) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t d\Phi_{ij}(u),
\end{aligned}$$

where

$$\begin{aligned}
\Phi_{ij}(t) &= h'_j(t; \beta_0, W) \Omega^{-1}(\beta_0) U_i(\beta_0) \\
&\quad + [B'_j(t; \beta_0) + E'_j(t; \beta_0)] \Omega_T(\theta_0)^{-1} \sum_{k=0}^{J-1} U_{ik}^T(\theta_0) \\
&\quad + K_{ij}(t; \beta_0) + P_{ij}(t; \beta_0) \\
&\quad + \int_0^t r_j^{(0)}(u; \beta_0, W)^{-1} W_i(u) dM_{ij}(u) \\
&= \int_0^t d\Phi_{ij}(u).
\end{aligned}$$

$$\mathbf{C.IV.7} \quad n^{\frac{1}{2}} \{ \widehat{\phi}_j(t) - \phi_j(t) \}$$

By a Taylor series expansion and Result C.IV.6, we have

$$\begin{aligned}
& n^{\frac{1}{2}}\{\widehat{\phi}_j(t) - \phi_j(t)\} \\
&= n^{\frac{1}{2}}\left[\frac{1}{\Lambda_{00}(t)}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\} + \widehat{\Lambda}_{0j}(t)\left\{\frac{1}{\widehat{\Lambda}_{00}(t)} - \frac{1}{\Lambda_{00}(t)}\right\}\right] \\
&= n^{\frac{1}{2}}\left[\frac{1}{\Lambda_{00}(t)}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\} - \frac{\widehat{\Lambda}_{0j}(t)}{\Lambda_{00}^2(t)}\{\widehat{\Lambda}_{00}(t) - \Lambda_{00}(t)\} + o(n^{-1/2})\right] \\
&= n^{-\frac{1}{2}}\left[\frac{1}{\Lambda_{00}(t)}\sum_{i=1}^n\Phi_{ij}(t) - \frac{\widehat{\Lambda}_{0j}(t)}{\Lambda_{00}^2(t)}\sum_{i=1}^n\Phi_{i0}(t)\right] + o_p(1) \\
&= n^{-\frac{1}{2}}\sum_{i=1}^n\xi_{ij}(t) + o_p(1),
\end{aligned}$$

where

$$\xi_{ij}(t) = \frac{1}{\Lambda_{00}(t)}\Phi_{ij}(t) - \frac{\Lambda_{0j}(t)}{\Lambda_{00}^2(t)}\Phi_{i0}(t).$$

## **BIBLIOGRAPHY**



## BIBLIOGRAPHY

- Chen, P. and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57**, 1030–1038.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B* **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.
- Feng, S., Goodrich, N. P., Bragg-Gresham, J. L., Dykstra, D. M., Punch, J. D., DeRoy, M. A., Greenstein, S. M., and Merion, R. M. (2006). Characteristics associated with liver graft failure: The concept of a donor risk index. *American Journal of Transplant* **6**, 783–790.
- Henderson, R., Diggle, P., and Dobson, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.
- Hernan, M. A., Brumback, B., and Robins, J. M. (2000). Marginal structural models to estimate the causal effect on the survival of hiv-positive men. *Epidemiology* **11**, 561–570.
- Hernan, M. A., Brumback, B., and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association Applications and Case Studies* **96**, 440–448.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.
- 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.
- Robins, J. M. and Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics* **56**, 779–788.
- Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.

- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N., Dietz, K., and Farewell, B., editors, *AIDS Epidemiology - Methodological Issues*, pages 297–331. Boston: Birkhäuser.
- Samoridnitsky, G. and Taqqu, M. S. (1994). *Stable non-gaussian random processes, stochastic models with infinite variance*. New York: Chapman and Hall.
- Schaubel, D. E., Guidinger, M. K., Biggins, S. W., Kalbfleisch, J. D., Pomfret, E. A., Sharma, P., and Merion, R. M. (2009). Survival benefit-based deceased-donor liver allocation. *American Journal of Transplantation* **9**, 970–981.
- Schaubel, D. E. and Wei, G. (2011). Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics* **67**, 29–38.
- Sharma, P., Schaubel, D. E., Gong, Q., Guidinger, M. K., and Merion, R. M. (2012). End-stage liver disease candidates at the highest meld scores have higher waitlist mortality than status 1a candidates. *Hepatology* **55**, 192–198.
- Song, X., Davidian, M., and Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.
- Taylor, J. M. (2011). Discussion of predictive comparison of joint longitudinal-survival modeling: A case study illustrating competing approaches, by hanson, branscum and johnson. *Lifetime Data Analysis* **17**, 29–32.
- Tsiatis, A. A., Degruittola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. application to survival and cd4 counts in patients with aids. *Journal of American Statistical Association* **90**, 27–37.
- Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34**, 70–85.
- Wei, G. and Schaubel, D. E. (2008). Estimating cumulative treatment effects in the presence of nonproportional hazards. *Biometrics* **64**, 724–732.
- Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.
- Wiesner, R. H., McDiarmid, S. V., Kamath, P. S., Edwards, E. B., Malinchoc, M., Kremers, W. K., Krom, R. A. F., and Kim, W. R. (2001). Meld and peld: Application of survival models to liver allocation. *Liver Transplantation* **7**, 567–580.

- Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.
- Zhang, M. and Schaubel, D. E. (2010). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics* **67**, 740–749.
- Zheng, Y. Y. and Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics* **61**, 379–391.