Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Xin Lu	Date

Estimation of Potential Outcomes when Treatment Assignment and Discontinuation Compete in Observational Data

By

Xin Lu

Doctor of Philosophy

Biostatistics

Brent A. Johnson, Ph.D.
Advisor

Qi Long, Ph.D.
Committee Member

Robert H Lyles, Ph.D.
Committee Member

Rob Stephenson, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Estimation of Potential Outcomes when Treatment Assignment and Discontinuation Compete in Observational Data

By

Xin Lu
M.S., University of Tennessee at Knoxville, 2010
M.S., University of Tennessee at Knoxville, 2008
B.S., Jilin University, 2006

Advisor : Brent A. Johnson, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2015

Abstract

Estimation of Potential Outcomes when Treatment Assignment and Discontinuation Compete in Observational Data

By Xin Lu

In clinical studies, randomization of treatment lengths may not be feasible in practice, resulting in the confounding of treatment effects. Moreover, treatment decisions may be missing due to treatment-terminating events. Therefore, to estimate the mean outcome across treatment lengths while accounting for the above obstacles, we propose several new estimators using causal inference theory and methods for different treatment assignment settings.

In the first project, we propose a new direct estimator for the mean outcome of a target treatment length policy using outcome regression. The estimator works well in both discrete and continuous time. We exemplify the direct estimator through small sample numerical studies and the analysis of two real data sets and show the direct estimator is more precise.

In many dynamic regimes, patients' treatment plan may vary with changes in their clinical characteristics that measured at routine clinic visits, which may also be confounded with patients' outcomes. To taking into account of the time-varying effects, in the second project, we implemented the G-computational algorithm in outcome regression with two approaches to estimate the mean potential outcome on treatment length policies. In simulation studies, our approaches are more efficient compared to an existing inverse probability weighting estimator. It could also approximate the distribution as well as the mean of the potential outcomes.

To maintain the consistency of our estimators proposed in the previous two projects, the outcome regression models must be correctly specified, which may not be always met. To achieve a consistent estimation under moderate miss-specification, under the same dynamic regime setting as project 2, we propose a doubly-robust estimator and an improved doubly-robust estimators for estimating the mean potential outcomes while adjusting for time-varying effects. They demonstrate desirable properties for small samples in simulation studies and the improved doubly robust estimator achieves minimum variance even when the outcome regression model may be miss-specified.

Estimation of Potential Outcomes when Treatment Assignment and Discontinuation Compete in Observational Data

By

Xin Lu

M.S., University of Tennessee at Knoxville, 2010

M.S., University of Tennessee at Knoxville, 2008

B.S., Jilin University, 2006

Advisor : Brent A. Johnson, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2015

Acknowledgment

I would like to gratefully acknowledge all of the support and encouragement I have received from my family, friends, and classmates. My dissertation adviser Dr. Brent Johnson provided invaluable guidance during this process and has also served as an excellent mentor. I would also like to thank my committee members Dr. Qi Long, Dr. Robert Lyles, and Dr. Rob Stephenson for their thoughtful comments that led to improvements in my research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

In clinical studies, patients takes treatment over time according to plans pre-determined by their physicians largely based on the patients demographic and clinical characteristics. However, treatment-terminating events such as adverse events may stop the treatment earlier than scheduled. Because the presence or absence as well as the timing of such treatment-terminating events are expected to be related to the clinical outcome and also influenced by patient-specific factors, causal inference methods will be appropriate to apply instead of approaches for controlled randomized trials. In this dissertation, we propose new estimators using casual inference theory to analyze datasets motivated by several medical studies. Below is an illustration on two motivating studies, as well as a brief introduction on causal inference methods.

1.1.1 The Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy (ESPRIT) trial

The Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy (ESPRIT) trial was conducted to compare eptifibatide therapy to placebo on the ba-

sis of the composite binary endpoint of death, myocardial infarction, or urgent target vessel revascularization within 30 days. The administration of the regimen consists of a integrilin bolus followed by integrilin infusion, the latter of which was expected to last 18-24 hours. If a serious complication occurred during the infusion, the infusion process was discontinued immediately so the patient can receive appropriate medical attention. Although this data was taken from a randomized controlled clinical trial of integrilin versus placebo treatments, patients were not randomized to infusion length and a comparison of outcomes across different infusion lengths is subject to confounding. Hence, we adjusted for five baseline confounders in our analysis: indicators for diabetes, percutaneous transluminal coronary angioplasty, angina, heparin use, and the continuous variable weight.

This same data set of 1040 patients randomized to integrilin therapy was analyzed in Johnson and Tsiatis (2004) using inverse weighting to adjust for potential confounding and so our re-analysis here allows for a direct comparison — see $\hat{\mu}_{jn}^{(1)}$ (assuming confounding is presented through baseline only) in Table 2 of Johnson and Tsiatis (2004).

Only 10% of patients did not experience an adverse event within the follow-up period. All of them are censored during the first 19 hours. More than half (55%) are censored within 5 hours after the infusion starts. The Kaplan-Meier survival curves are shown in Figure 1.1.

1.1.2 AIDS Clinical Trials Group Study A5095

AIDS Clinical Trials Group (ACTG) Study A5095 is a multi-center clinical trial designed to compare three antiretroviral regimens in antiretroviral therapy (ART)-naive patients with a median of follow-up time of approximately three years. After a patient experiences an virologic failure in one ART arm, he/she will switch to another regimen. In practice, the time between virologic failure to switch is usually

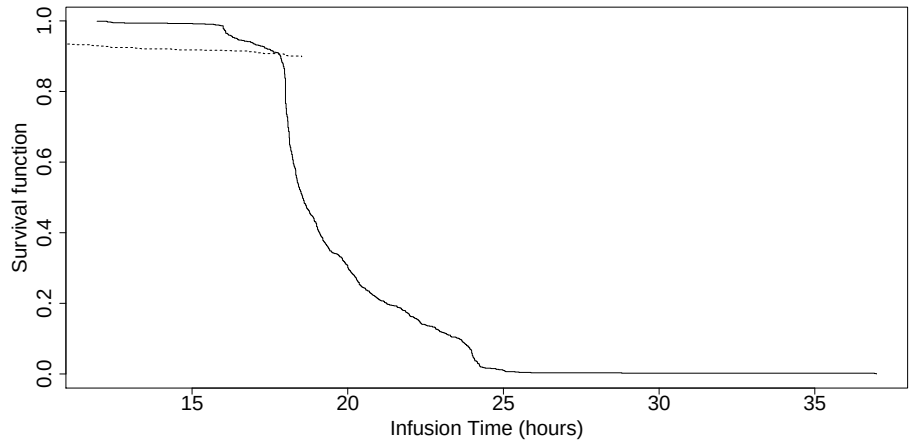

Figure 1.1: Survival curves for patients who experience adverse events in 30 days Survival probability in time regarding adverse events (solid line) and censoring (dashed line)

subject to physicans' decisions which very likely involve the patients demographic and clinical characteristics as well. However, due to the trade-off between prolonging virologic suppression and preventing clinical symptoms occurrence in patients who rebound in viral load, the best timing in regimen switch is still under investigation. Additionally, the treatment with initial regimen may stop early for some patients prior to the designed switch time because of drop out or death of the patients or the patients may not switch regimen at all during the course of the study. Thus, estimating the outcomes when switching skeme is performed exactly as designed is substantially meaningful for finding the optimal switch time point.

Our interest here is the potential outcome of viral load and cumulative CD4 cell count in the hypothetical setting that the patients who experienced virologic failure switch regimen at designed time. Specifically, we used three length-adjusted area-under-the-curve endpoints, i.e. cumulative viral load, days below the limit of detection (the proportion of time with suppressed viral load) and CD4 cell count. Only (182) patients who experienced virologic failure were considered here. Among them, 100 participants (55% of 182) did not switch within the follow-up period. Of these 100

participants with censored switching times, 42 participants were followed for at least 100 days, 27 participants followed for at least 120 days, and 11 participants followed for at least 140 days. The Kaplan-Meier survival curves are shown in Figure 1.2. The detail specifications of the endpoints and covariates are available in Li et al. (2012); Johnson et al. (2013).

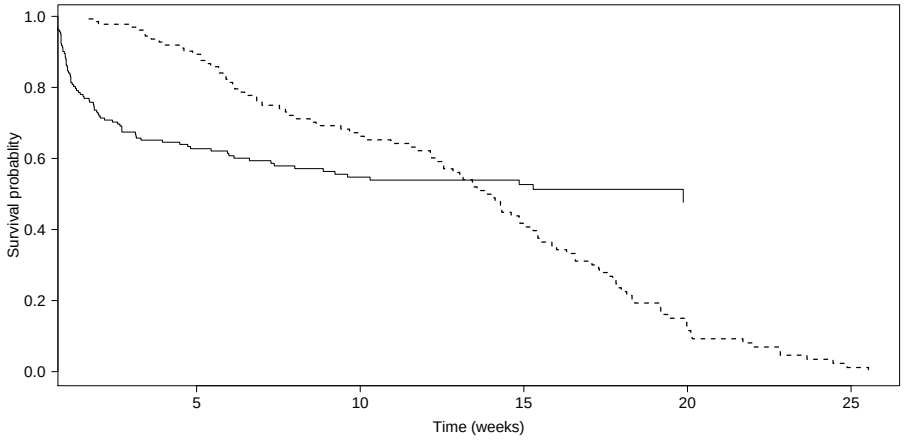

Figure 1.2: Survival curves for patients who experience virologic failure from ATCG Study A5095
Survival probability in time regarding switching (solid line) and censoring (dashed line)

1.1.3 Causal Inference

In this dissertation research, we adopt Rubin's counterfactual model (Rubin, 1974) and the ideas of potential outcomes (Neyman, 1923; Rubin, 1974). The potential outcome of a treatment plan is defined as the outcome that one will observe if one is treated according to the treatment plan.

In a simple case where we have binary treatment. Then each unit in population has observed data (Z, X, Y), and potential outcomes (Y_1^*, Y_0^*) where

- Z = indicator for treatment assignment with levels 0, 1, where 1="treated", 0="control".

- $\mathbf{X}$ = a vector of covariates
- Y = observed outcome
- Y_1^* = potential outcome that would be observed if unit assigned to $Z = 1$.
- Y_0^* = potential outcome that would be observed if unit assigned to $Z = 0$.

The causal effect for each unit is $Y_1^* - Y_0^*$. The average causal effect is defined as $T = E(Y_1^* - Y_0^*) = E(Y_1^*) - E(Y_0^*)$. That is, how much effect on the outcome in the population of interest is attributable to 1 vs. 0. In order to estimate $E(Y_1^*)$ and $E(Y_0^*)$, we need to use two fundamental assumptions in causal inference.

Under the stable unit value treatment assumption (SUTVA)(Rubin, 1978), there is no interference on response between individual units. So we have

$$Y = Y_1^* Z + Y_0^*(1 - Z).$$

Then under the sequential randomization assumption (SRA) which is also known as no unmeasured confounders assumption, choice of treatment is completely random given covariate $\mathbf{X}$.

Then we can estimate T, by

$$\begin{aligned} T &= E(Y_1^*) - E(Y_0^*) = E\{E(Y_1^* \mid \mathbf{X}) - E(Y_0^* \mid \mathbf{X})\} \\ &= E\{E(Y \mid Z = 1, \mathbf{X}) - E(Y \mid Z = 0, \mathbf{X})\} \end{aligned}$$

In a more sophisticated setting where we have binary treatment length t=t_1, t_2 and treatment terminating event C, the observed data are (U, X, Δ, Y) and potential outcomes for each individual are $(Y_{t_1}^*, Y_{t_2}^*, Y_C^*)$.

Under SUTVA,

$$Y = Y_{t \wedge C}^* = Y_{t_1}^* \mathbf{1}(U = t_1, \Delta = 1) + Y_{t_2}^* \mathbf{1}(U = t_2, \Delta = 1) + Y_C^* \mathbf{1}(U = C, \Delta = 0)$$

We then have causal estimand:

$$\begin{aligned}
E(Y^*_{t_1\wedge C}) &= E\{Y^*_{t_1}1(U=t_1,\Delta=1)+Y^*_C1(U<t_1,\Delta=0)\}\\
&= E\{E\left[Y^*_{t_1}1(U=t_1,\Delta=1)\mid\mathbf{X}\right]+E\left[Y^*_C1(U<t_1,\Delta=0)\mid\mathbf{X}\right]\}\\
&= E\{E\left[Y\mid U=t_1,\Delta=1,\mathbf{X}\right]+E\left[Y\mid U<t_1,\Delta=0,\mathbf{X}\right]\}\\
E(Y^*_{t_2\wedge C}) &= E\{Y^*_{t_2}1(U=t_2,\Delta=1)+Y^*_C1(U<t_2,\Delta=0)\}\\
&= E\{E\left[Y^*_{t_2}1(U=t_2,\Delta=1)\mid\mathbf{X}\right]+E\left[Y^*_C1(U<t_2,\Delta=0)\mid\mathbf{X}\right]\}\\
&= E\{E\left[Y\mid U=t_2,\Delta=1,\mathbf{X}\right]+E\left[Y\mid U<t_2,\Delta=0,\mathbf{X}\right]\}
\end{aligned}$$

Under the setting that the treatment length is continuous instead of categorical, we have the observed data (U, X, Δ, Y) and potential outcomes for each individual $(Y_t^*,\, Y_C^*)$

By SUTVA, $Y = Y_t^*\Delta + Y_C^*(1-\Delta)$. So we have causal estimand

$$E(Y^*_{t\wedge C}) = E\{Y_t^*\mathbf{1}(C\geq t)+Y_C^*\mathbf{1}(C<t)\},$$

which by SRA, is equivalent to

$$E\left[E\{Y\mid C\geq t,\mathbf{X}\}+E\{Y\mid C<t,\mathbf{X}\}\right].$$

Estimation approaches

To estimate causal estimand, the popular strategies include matching or stratification based on covariates or propensity score(Rosenbaum and Rubin, 1984), inverse probability weighting based on propensity score(Johnson and Tsiatis, 2004, 2005) and outcome regression modeling (Chapter 2 & 3). Here we focus on the inverse probability weighting appraoch, with propensity score modeling the probability of treatment assignment given covariates, as well as the outcome regression approach, which di-

rectly fits the outcome with covariates and treatment assignment. For the simplest case mentioned above, if using outcome regression method, one first postulate a model for the outcome regression $E\{Y \mid Z, \mathbf{X}\}$, then fit the model and average the result estimates $E(Y \mid Z = 1, \mathbf{X})$ and $E(Y \mid Z = 1, \mathbf{X})$ over all observed $\mathbf{X}$ respectively. In order to get unbiased estimates, the postulated outcome regression must be identical to the true regression model. On the other hand, if using the inverse probability weighting estimator. one first postulate and fit a logistic model for treatment selection $P(Z \mid \mathbf{X}) = \pi(\mathbf{X}, \boldsymbol{\beta}),\ j = 1, 2$. Then calculate the inverse propensity score weighted averages $\frac{1(Z=1)}{\pi(\mathbf{X},\hat{\boldsymbol{\beta}})}Y$ and $\frac{1(Z=0)}{1-\pi(\mathbf{X},\hat{\boldsymbol{\beta}})}Y$. If $\pi(\mathbf{X}, \boldsymbol{\beta}) = \pi(\mathbf{X})$ the true propensity score, then

$$\begin{aligned} E\left[\frac{ZY}{\pi(\mathbf{X})}\right] &= E\left[\frac{ZY_1^*}{\pi(\mathbf{X})}\right] \\ &= E\left[E\{\frac{ZY_1^*}{\pi(\mathbf{X})} \mid Y_1^*, \mathbf{X}\}\right] \\ &= E\left[Y_1^* \frac{P(Z \mid Y_1^*, \mathbf{X})}{\pi(\mathbf{X})}\right] \\ &= E\left[Y_1^*\right] \end{aligned}$$

Note that the first step holds because $ZY = Z(ZY_1^* + (1 - Z)Y_0^*) = Z^2Y_1^* = ZY_1^*$. Similarly,

$$\begin{aligned} E\left[\frac{(1-Z)Y}{1-\pi(\mathbf{X})}\right] &= E\left[\frac{(1-Z)Y_0^*}{1-\pi(\mathbf{X})}\right] \\ &= E\left[E\{\frac{(1-Z)Y_0^*}{\pi(\mathbf{X})} \mid Y_0^*, \mathbf{X}\}\right] \\ &= E\left[Y_0^* \frac{1 - P(Z \mid Y_0^*, \mathbf{X})}{1-\pi(\mathbf{X})}\right] \\ &= E\left[Y_0^*\right] \end{aligned}$$

In order to for the above properties to hold when using our postulated propensity score model $\pi(\mathbf{X}, \boldsymbol{\beta})$, the postulated propensity score model must be identical to the true propensity $\pi(\mathbf{X})$.

So in order for the inverse probability weighting or outcome regression estimators to work, their postulated model has be correctly specified, which may not always suffice. To improve the consistency in estimation under model miss specification, Robins et al. (1994) proposed an augmented inverse probability estimator, which is an inverse probability estimator augmented with conditional outcome regression models. This new estimator is consistent as long as either the propensity score model or the outcome regression model is correctly specified. And it is also semi-parametric efficient if both model assumptions are correct. Based on these properties, Scharfstein et al. (1999) identified this class of estimators as being doubly-robust. It is then recommended for routine use due to the merit of extra protection over model miss-specification (Bang and Robins, 2005; Li et al., 2012). In the same simple case mentioned above, the doubly robust estimators will be:

$$\hat{\mu}_{1,DR} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z_i Y_i}{\pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})} - \frac{Z_i - \pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})}{\pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})} m_1(\mathbf{X}_i, \hat{\boldsymbol{\alpha}}_1)\right]$$

$$\hat{\mu}_{0,DR} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - Z_i) Y_i}{1 - \pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})} - \frac{Z_i - \pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})}{1 - \pi(\mathbf{X}_i, \hat{\boldsymbol{\beta}})} m_0(\mathbf{X}_i, \hat{\boldsymbol{\alpha}}_0)\right].$$

Here $\pi(\mathbf{X}, \boldsymbol{\beta})$ is the postulated propensity score model; $m_0(\mathbf{X}, \boldsymbol{\alpha}_0)$ and $m_1(\mathbf{X}, \boldsymbol{\alpha}_1)$ are the postulated model for outcome regressions $E(Y \mid Z = 0, \mathbf{X})$ and $E(Y \mid Z = 1, \mathbf{X})$

respectively. Note that for the treatment group,

$$\begin{aligned}
& E\left[\frac{ZY}{\pi(\mathbf{X},\boldsymbol{\beta})} - \frac{Z-\pi(\mathbf{X},\boldsymbol{\beta})}{\pi(\mathbf{X},\boldsymbol{\beta})} m_1(\mathbf{X},\boldsymbol{\alpha}_1)\right] \\
= \; & E\left[\frac{ZY_1^*}{\pi(\mathbf{X},\boldsymbol{\beta})} - \frac{Z-\pi(\mathbf{X},\boldsymbol{\beta})}{\pi(\mathbf{X},\boldsymbol{\beta})} m_1(\mathbf{X},\boldsymbol{\alpha}_1)\right] \\
= \; & E\left[Y_1^* + \frac{Z-\pi(\mathbf{X},\boldsymbol{\beta})}{\pi(\mathbf{X},\boldsymbol{\beta})}\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\}\right] \\
= \; & E(Y_1^*) + E\left[\frac{Z-\pi(\mathbf{X},\boldsymbol{\beta})}{\pi(\mathbf{X},\boldsymbol{\beta})}\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\}\right]
\end{aligned}$$

When the propensity score is correctly specified but outcome regressions are not, we have $\pi(\mathbf{X},\boldsymbol{\beta}) = \pi(\mathbf{X})$. Then

$$\begin{aligned}
& E\left[\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\}\right] \\
= \; & E\left[E\{\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\} \mid Y_1^*, \mathbf{X}\}\right] \\
= \; & E\left[\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\} E\{\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})} \mid Y_1^*, \mathbf{X}\}\right] \\
= \; & E\left[\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\} \frac{E\{Z \mid Y_1^*, \mathbf{X}\} - \pi(\mathbf{X})}{\pi(\mathbf{X})}\right] \\
= \; & E\left[\{Y_1^* - m_1(\mathbf{X},\boldsymbol{\alpha}_1)\} \frac{\pi(\mathbf{X}) - \pi(\mathbf{X})}{\pi(\mathbf{X})}\right] \\
= \; & 0
\end{aligned}$$

Thus, $\hat{\mu}_{1,DR}$ is consistent. And similarly, so is $\hat{\mu}_{0,DR}$.

When the outcome regression is correctly specified but propensity score is not, we have $m_1(\mathbf{X},\boldsymbol{\alpha}_1) = E(Y \mid Z=1,\mathbf{X}) = E(Y_1^* \mid \mathbf{X})$ and $m_0(\mathbf{X},\boldsymbol{\alpha}_0) = E(Y \mid Z =$

$0, \mathbf{X}) = E(Y_0^* \mid \mathbf{X})$. Then

$$
\begin{aligned}
& E\left[\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{Y_1^* - E(Y_1^* \mid \mathbf{X})\}\right] \\
= \; & E\left[E\{\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{Y_1^* - E(Y_1^* \mid \mathbf{X})\} \mid Z=1, \mathbf{X}\}\right] \\
= \; & E\left[\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})} E\{\{Y_1^* - E(Y_1^* \mid \mathbf{X})\} \mid Z=1, \mathbf{X}\}\right] \\
= \; & E\left[\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{E(Y_1^* \mid Z=1, \mathbf{X}) - E(Y_1^* \mid \mathbf{X})\}\right] \\
= \; & E\left[\frac{Z-\pi(\mathbf{X})}{\pi(\mathbf{X})}\{E(Y_1^* \mid \mathbf{X}) - E(Y_1^* \mid \mathbf{X})\}\right] \\
= \; & 0
\end{aligned}
$$

Thus, $\hat{\mu}_{1,DR}$ is consistent. And similarly, so is $\hat{\mu}_{0,DR}$.

It is obvious that when both models are correct, $\hat{\mu}_{1,DR}$ and $\hat{\mu}_{0,DR}$ are consistent. And when neither model is correct, both estimators are inconsistent.

When the propensity score model is correct, the variance of the doubly robust estimator will be smaller than that of the inverse probability weighting estimator. Meanwhile, when the outcome regression model is correct, the variance of the doubly robust estimator may be larger than that of the outcome regression estimator, but will offer protection in the event that it is not.

1.2 Outline

The remainder of the dissertation proposal is organized as follows. In Chapter 2, we propose a novel outcome regression estimator for estimating continuous/discrete treatment effect when decision of treatment is confounded with both the clinical outcomes and patients demographic and clinical characteristics. Simulation studies and two real data application are provided to illustrate the performance of our method. In Chapter 3, we extended the method in Chapter 2 to estimate discrete treatment effect

allowing for time-dependent confounders, where simulation studies are provided. At last in Chapter 4, we proposed a doubly robust estimator which takes time-varying confounders into account as well as an improved version of this estimator that minimizes its variance when the outcome regression model may be miss-specified. We also examined their performance in sensitivity analysis in the presence of miss-specification in either outcome regression or propensity score model via simulation studies. Then we conclude this dissertation with a summary on the three estimators.

Chapter 2

Direct Estimation of the Mean Outcome amidst Early Treatment Stoppage

2.1 Introduction

In clinical studies where treatment is administered continuously over time, there are provisions in study protocols that dictate how a patient's treatment is to be managed in critical situations, when it is no longer feasible or ethical to continue the current treatment regimen and treatment must be terminated or discontinued prematurely. When there is no treatment-terminating event, treatment continues until such time the provider renders it appropriate to stop treatment. Thus, actual treatment is given until the provider decides to stop treatment or until a treatment-terminating event occurs, whichever comes first. Treatment for all study participants is completed in a fixed window of time and a clinical endpoint measured after that window closes. The goal is to draw inference on the mean outcome for a treatment regimen that accounts for premature treatment discontinuation.

Following Johnson and Tsiatis (2004, 2005), we define Y_t^* as the potential outcome (Rubin, 1974) if a participant is treated to the intended treatment length at time t. We define the intermediate outcome C as the time at which a randomly selected individual from our population would have a treatment-terminating event, assuming the individual has been continuously treated up to time C and Y_C^* as the potential outcome if treatment is stopped due to such treatment-terminating event. Then, the potential outcome for an individual treated to time $t \wedge C$ is $Y_{t\wedge C}^* = Y_t^* I(C \geq t) + Y_C^* I(C < t)$, and the causal estimand of interest is the population mean,

$$\mu^*(t) = E(Y_{t\wedge C}^*) = E\{Y_t^* I(C \geq t)\} + E\{Y_C^* I(C < t)\}. \tag{2.1}$$

Johnson and Tsiatis (2004) referred to the combined act of treating toward the minimum of a provider-intended treatment length and a potential treatment-terminating event as the 'treatment length policy at time t' and they showed how such policies may be viewed as instances of dynamic regimes (Robins, 1986; Murphy et al., 2001) in the presence of time-dependent confounding. In defining the policy, it is important to clarify that treatment assignment is governed by the provider's action and this action serves as the control variable of a dynamic system. Although the policy is also defined by premature discontinuation or treatment-terminating events, these events are beyond provider's control and not considered part of the treatment decision. In related work, Zhang et al. (2011) extended the above definition of treatment policy to incorporate other treatment options of which the participants may avail themselves.

The methods developed by Johnson and Tsiatis (2004, 2005) and Zhang et al. (2011) were similarly motivated by applications to infusion trials and the problem is similar to the former where treatment length is the single control variable of interest. After defining the causal estimands of interest, these authors proposed inverse probability of treatment weighted (IPW) estimators to adjust for potential confounding

in the treatment assignment mechanism. The work here differs fundamentally from that proposed by Johnson and Tsiatis (2004, 2005) or Zhang et al. (2011) in that it does not rely on a treatment model and the reasons that motivate this departure are explained below.

When treatment realizes only one of K levels, $t \in \{t_1, \ldots, t_K\}$, Johnson and Tsiatis (2004) proposed an inverse probability weighting estimator for $\mu^*(t)$ as a weighted average of observations from participants whose provider stopped treatment at time t by choice and from participants whose treatment was discontinued prematurely due to treatment-terminating events with weights equal to the inverse probability of treatment stopped by choice at time t and at some time later than the treatment-terminating event, respectively. These probabilities are synonymous with the ubiquitous propensity scores (Rosenbaum and Rubin, 1983) and modeled using a discrete-time hazards model or pooled logistic regression for failure time data. However, when treatment realizes any value in an interval, say $t \in [\tau_l, \tau_u]$, their methods do not apply because of the continuous nature of the treatment assignment mechanism, hence, the probability that a provider stops treatment at time t is nil for every $t \in [\tau_l, \tau_u]$. In this case, Johnson and Tsiatis (2005) proposed to model $\mu^*(t)$ parametrically in time through a continuous-time extension of dynamic regimes (Murphy et al., 2001). The challenge with this approach is that modeling $\mu^*(t)$ as a simple parametric model can be tricky because the effect the potential treatment-terminating event process on the estimand is complex and there are few, if any, diagnostic tools to aid in evaluating model fit on the potential outcome scale.

In the sequel, we propose and evaluate a novel estimator for the causal estimand $\mu^*(t)$ that achieves three goals. First, there is no need to model parametrically the mean potential outcome $\mu^*(t)$. Second, because there is no treatment model, the estimator works equally as well for treatment data arising in discrete or continuous time. Third, the proposed estimator relies on an entirely different set of modeling

assumptions compared to Johnson and Tsiatis (2004, 2005) yet is still regular and asymptotically linear. The procedure outlined below uses a new and intriguing combination of competing risks, survival analysis, and semiparametric inference and may be of general interest outside the particular clinical application. Theoretical details are relegated to Supplementary Material.

2.2 Methods

2.2.1 Data and likelihood

In order to describe the observed data likelihood, we describe in detail the observed data. Let A denote treatment length and suppose there are K treatment targets, $0 < t_1 < t_2 < \cdots < t_K$, at which time providers intend to stop treatment. Treatment-terminating events may occur at any time after treatment initiation but, if no terminating events have occurred prior to time t_K while the participant has been continuously treated, then they are assigned treatment t_K with probability one. In chronological order, the observed data are

$$\begin{aligned} Z \;=\; & \{X, CI(0 \le C < t_1), I(A = t_1)I(C \ge t_1), CI(t_1 \le C < t_2)I(A > t_1, C \ge t_1), \\ & \cdots, \\ & CI(t_{K-1} \le C < t_K)I(A > t_{K-1}, C \ge t_{K-1}), I(A = t_K)I(A > t_{K-1}, C \ge t_K), \\ & Y\}. \end{aligned}$$

Similar to analyses of informative dropout in longitudinal studies (cf. Scharfstein et al., 1999), we have the opportunity to collect and observe data only if a patient is still being followed. However, unlike data analyses with dropout where no data is collected after the point of dropout, the outcome Y is always observed in our case regardless of how and when treatment was discontinued. It is also worth noting that

if we regard $CI(t_{j-1} \leq C < t_j)$ and $I(A = t_j)$ as intermediate potential outcomes and treatment assignment at t_j, respectively, their competition for observation does not fit cleanly into the conventional dynamic regime setup (cf. Robins, 1986, 1997; Murphy et al., 2001). In the setup here, note that treatment assignment is only well-defined at time t_j if continuously treated up through time t_j and is missing otherwise; in Murphy et al. (2001), treatment assignment is well-defined for all $t_1, \ldots, t_K$. Hence, Johnson and Tsiatis (2005) introduced the at-risk process from survival analysis (Andersen et al., 1993; Kalbfleisch and Prentice, 2002) to provide clarity in defining the observed data over the entire study period. Along these lines, we define the observed treatment length, reason for treatment discontinuation, and at-risk indicator:

$$U = A \wedge C, \qquad \Delta = I(A \leq C), \qquad R(t) = I(U \geq t),$$

respectively, where $x \wedge y$ denotes $\min(x, y)$. Then, under the conditional independence of A and C given X and with mild abuse of notation, the observed data can be rewritten

$$\begin{aligned} Z \;=\; & \{X, UI(0 \leq U < t_1, \Delta = 0), I(U = t_1, \Delta = 1)R(t_1), \\ & UI(t_1 \leq U < t_2, \Delta = 0)R(t_1), \\ & \ldots, UI(t_{K-1} \leq U < t_K, \Delta = 0)R(t_{K-1}), I(U = t_K, \Delta = 1)R(t_K), Y\}. \end{aligned}$$

We see that the at-risk process is carried through the definitions of intermediate outcomes and treatment assignment from the beginning of the study through time t_K. In connecting the former and latter definitions of observed data, it is evident that ideas central to competing risks (Tsiatis, 1975) plays a fundamental role in our problem.

In the discrete time scenario, the likelihood for a single observation Z is $L(Z) =$

$L_1(Z)L_2^{\mathrm{D}}(Z)$, where

$$L_1(Z) = \{f_1(Y \mid U, X)G(U \mid X)\}^{\Delta}\{f_0(Y \mid U, X)g(U \mid X)\}^{1-\Delta}h(X),$$

f_1 and f_0 are the conditional densities of Y given the observed treatment length U and covariates X when treatment is stopped due to provider discretion versus terminating event, respectively, h is the marginal density of X, g and G are the conditional cause-specific density and survivor function of U given X when $\Delta = 0$, respectively, and

$$\begin{aligned} L_2^{\mathrm{D}}(Z) = & \\ & \left[\{\mathrm{pr}(U = t_1, \Delta = 1 \mid X)\}^{I(U=t_1)} \cdots \{\mathrm{pr}(U = t_K, \Delta = 1 \mid X)\}^{I(U=t_K)}\right]^{\Delta} \\ & \times \left[\{\mathrm{pr}(U \geq t_1, \Delta = 1 \mid X)\}^{I(U<t_1)} \cdots \{\mathrm{pr}(U \geq t_{K-1}, \Delta = 1 \mid X)\}^{I(t_{K-2} \leq U < t_{K-1})}\right]^{1-\Delta}. \end{aligned}$$

In the discrete time setting, $L_2^{\mathrm{D}}(Z)$ is completely identified through the probabilities, $\mathrm{pr}(U = t_j, \Delta = 1 \mid X)$, $j = 1, \ldots, K$, and is important because it implies the treatment assignment mechanism is parametric.

Now, when treatment is assigned continuously over the closed interval $[\tau_l, \tau_u]$, the likelihood is $L(Z) = L_1(Z)L_2^{\mathrm{C}}(Z)$, where $L_1(Z)$ is given above and

$$\begin{aligned} L_2^{\mathrm{C}}(Z) &= \prod_{t \in [\tau_l, \tau_u]} \left\{\mathrm{pr}(t \leq U < t + dt, \Delta = 1 \mid X)^{I(t \leq U < t+dt)}\right\}^{\Delta} \times \\ &\qquad \left\{\mathrm{pr}(U \geq t, \Delta = 1 \mid X)^{I(U \geq t)}\right\}^{1-\Delta} \\ &= \{f(U \mid X)\}^{\Delta}\{S(U \mid X)\}^{1-\Delta}, \end{aligned}$$

where $\prod$ is the product integral, $f(t \mid X)$ and $S(t \mid X)$ are the conditional cause-specific density and survivor function, respectively, of U given X when $\Delta = 1$. Unlike the discrete time setting, the treatment assignment mechanism is now infinite dimensional and modeling the mechanism without unnecessarily strong assumptions is more difficult. Naturally, the cardinality of treatment assignment affects the observed data

likelihood only through $L_2^{\mathrm{D}}(Z)$ or $L_2^{\mathrm{C}}(Z)$ while $L_1(Z)$ remains the same. The primary reason $L_1(Z)$ remains the same is because the potential treatment-terminating events are assumed to arise continuously in time in either scenario, regardless of treatment assignment.

2.2.2 The estimand

To motivate the approach below, we show how the estimand may be expressed through the conditional densities of observed data, rather than simply the expectation of a potential outcome (Johnson and Tsiatis, 2004, 2005). Let W be an intermediate random variable along the causal pathway, $X \mapsto W \mapsto Y$, and write the marginal mean of Y,

$$\int_{\mathcal{X}} \int_{\mathcal{W}} \int_{\mathcal{Y}} y \, f_Y(y \mid w, x) f_W(w \mid x) h(x) \, dy \, dw \, dx, \tag{2.2}$$

where f_Y and f_W are conditional density functions, h is the marginal density function of X as before, $\mathcal{X}$, $\mathcal{W}$, and $\mathcal{Y}$ are the domains of x, w and y, respectively. Now, set $W = t \wedge C$, where C is the event time of potential treatment-terminating events and treatment assignment at time t is fixed, $t \in [\tau_l, \tau_u]$, $\tau_l \geq 0$, $\tau_u < \infty$ chosen to satisfy suitable regularity conditions. Then, for almost every $x \in \mathcal{X}$ (Ash and Doléans-Dade, 1999), the conditional density function of W given $X = x$ is a mixture of a point mass function at t when $C \geq t$ and the conditional density function of C given $X = x$ when $C < t$, i.e. $f_W(w \mid x) = \{G(t \mid x)\}^{I(w \geq t)} \{g(w \mid x)\}^{I(w < t)}$. Importantly, although we require $G(0 \mid x) = 1$ for every x, we allow $G(\tau_l \mid x) < 1$ for any $x \in \mathcal{X}$; that is, we expect that the probability of a treatment-terminating event occurring prior to time τ_l to be non-zero for some or all covariate values in the domain. The conditional density $f_Y(y \mid w, x)$ can be written as a mixture of two conditional density functions, $f_Y(y \mid w, x) = \{f_1(y \mid t, x)\}^{I(w \geq t)} \{f_0(y \mid w, x)\}^{I(w < t)}$, with f_1 and f_0 defined earlier. This definition makes explicit the role of the terminating-event on

the observed distribution of the clinical endpoint and allows for the possibility that the terminating-event is informative. Assuming that treatment length must take positive values, setting $\mathcal{W} = [0, \infty)$, and substituting the definitions of $f_W(w \mid x)$ and $f_Y(y \mid w, x)$ into (2.2) above, we define the estimand

$$\mu(t) = \int_{\mathcal{X}} \int_{\mathcal{Y}} y\, f_1(y \mid t, x) G(t \mid x) h(x)\, dy\, dx \tag{2.3}$$

$$+ \int_{\mathcal{X}} \int_{[0,t)} \int_{\mathcal{Y}} y\, f_0(y \mid w, x) g(w \mid x) h(x)\, dy\, dw\, dx. \tag{2.4}$$

Evidently, the estimand $\mu(t)$ contains two parts: expression (2.4) pertaining to the case where the treatment-terminating event precedes time t and expression (2.3) pertaining to the case where the treatment-terminating event exceeds time t. As the probability of a treatment-terminating event preceding time t decreases to zero, $G(t \mid x)$ approaches one and $\mu(t)$ approaches the conventional definition for the population mean given treatment at time t, $\int_{\mathcal{X}} \int_{\mathcal{Y}} y f_1(y \mid t, x) h(x)\, dy\, dx$. Interestingly, the estimand $\mu(t)$ depends on the likelihood through $L_1(Z)$ only whereas the treatment assignment probabilities in $L_2^{\mathrm{D}}(Z)$ or $L_2^{\mathrm{C}}(Z)$ are nuisance parameters (Robins, 1986, 1989, 1997; Murphy et al., 2001).

To see how $\mu(t)$ relates to the causal estimand $\mu^*(t)$, we rely on two assumptions — the stable unit treatment value assumption (Rubin, 1978) and the strong ignorability assumption (Rosenbaum and Rubin, 1983) stating the conditional independence of the potential outcomes and treatment assignment given X, also known as the no unmeasured confounders assumption (Robins, 1997) — and standard arguments from causal inference. Briefly, through stable unit treatment value assumption and strong ignorability, we conclude that (2.3) is exactly equal to $E\{Y_t^* I(C \geq t)\}$. The same arguments lead us to conclude (2.4) is equal to $E\{Y_C^* I(C < t)\}$. Therefore, $\mu(t) = \mu^*(t)$ by definition. Note that $\mu^*(t)$ has a causal interpretation whereas $\mu(t)$ does not. On the other hand, $\mu(t)$ can simply be viewed as a reparameterization of the

conditional densities from the observational distribution (cf. Murphy et al., 2001). In the development below, we state all results in terms of the estimand $\mu(t)$ with the understanding that the same conclusions apply to the causal estimand $\mu^*(t)$ under two additional non-identifiable assumptions.

2.2.3 Direct Estimation

Define the conditional expectations $m_1(t,x) = \int y f_1(y \mid t,x)\, dy$, $m_0(u,x) = \int y f_0(y \mid u,x)\, dy$ in the absence and presence of treatment-terminating events, respectively. Define the cause-specific hazard function,

$$\lambda^{(0)}(u \mid X) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathrm{pr}(u \leq C < u + \epsilon \mid C \geq u, X),$$

and, by definition, $G(t \mid x) = \exp\left\{ - \int_0^t \lambda^{(0)}(u \mid x)\, du \right\}$. Throughout the manuscript, we adopt the superscript "(0)" to denote cause-specific functions where $\Delta = 0$ and other related quantities associated with the potential treatment-terminating random variable. Following the expressions in (2.3)–(2.4), if m_1, m_0, and G were known, then a direct estimator for $\mu(t)$ would be

$$\int_{\mathcal{X}} \left[m_1(t,x) G(t \mid x) - \int_{[0,t)} m_0(u,x)\, dG(u \mid x) \right] dH_n(x),$$

where H_n is the empirical distribution function of X. Because m_1, m_0, and G are not known, we estimate these functions using regression models. Therefore, a general expression for the direct estimator is

$$\widehat{\mu}_n(t) = \int_{\mathcal{X}} \left[\widehat{m}_1(t,x) \widehat{G}(t \mid x) - \int_{[0,t)} \widehat{m}_0(u,x)\, d\widehat{G}(u \mid x) \right] dH_n(x), \tag{2.5}$$

where $\widehat{m}_1$, $\widehat{m}_0$, and $\widehat{G}$ are consistent estimators for m_1, m_0, and G, respectively. If there were no treatment-terminating events such that $G(t \mid x) = 1$ for all x, x

were scalar, and we modeled $m_1(t, x)$ using nonparametric kernel methods, then (2.5) reduces to an expression analogous to Cheng's (1994) outcome regression estimator for dichotomous treatments. Thus, the general expression in (2.5) is only limited by one's ability to form predicted values from three regression modeling procedures. In many cases, such predicted values are standard output in software packages.

In the sequel, we provide details for some fundamental, ubiquitous parametric and semi-parametric models. First, we parameterize m_1 and m_0 linearly through the finite parameter vectors $\beta = (\beta_0, \beta_U, \beta_X^{\mathrm{T}})^{\mathrm{T}}$ and $\alpha = (\alpha_0, \alpha_U, \alpha_X^{\mathrm{T}})^{\mathrm{T}}$ with $m_1(U, X; \beta) = \beta_0 + \beta_U U + X^{\mathrm{T}}\beta_X$ and $m_0(U, X; \alpha) = \alpha_0 + \alpha_U U + X^{\mathrm{T}}\alpha_X$. The regression coefficients α, β are estimated via least squares,

$$\widehat{\beta}_n = \min_{\beta} \sum_{i=1}^{n} \Delta_i \left\{ Y_i - m_1(U_i, X_i; \beta) \right\}^2, \quad \widehat{\alpha}_n = \min_{\alpha} \sum_{i=1}^{n} (1 - \Delta_i) \left\{ Y_i - m_0(U_i, X_i; \alpha) \right\}^2,$$

and the predicted values computed via the linear predictor. We parameterize G through its hazard function and then we parameterize and model the cause-specific hazard function $\lambda^{(0)}(t \mid X; \gamma)$ through the finite parameter vector γ. We consider both parametric log-linear models (cf. Kalbfleisch and Prentice, 2002) and Cox's (1972) proportional hazard model for $\lambda^{(0)}(t \mid X; \gamma)$. In the parametric case, we use the extreme value model, $\lambda^{(0)}(t \mid X; \gamma) = \gamma_1 \exp(X^{\mathrm{T}}\gamma_2) t^{\gamma_1 - 1}$ and consequently $G(t \mid x; \gamma) = \exp\{-\exp(x^{\mathrm{T}}\gamma_2) t^{\gamma_1}\}$, and estimate γ by maximum likelihood,

$$\widehat{\gamma}_n = \max_{\gamma} \sum_{i=1}^{n} \left[(1 - \Delta_i) \log \left\{ \lambda^{(0)}(U_i \mid X_i; \gamma) \right\} + \log \left\{ G(U_i \mid X_i; \gamma) \right\} \right].$$

The maximum likelihood estimator for the survivor function at $X = x$ is $\widehat{G}(t \mid x) = G(t \mid x; \widehat{\gamma}_n)$. In the case of the proportional hazards model, $\lambda^{(0)}(t \mid X) = \lambda_0(t) \exp(X^{\mathrm{T}}\gamma)$, for an arbitrary baseline hazard function $\lambda_0(t)$, and we similarly

estimate the regression coefficients by maximizing the log partial likelihood,

$$\widehat{\gamma}_n = \max_{\gamma} \sum_{i=1}^{n} \left\{ (1-\Delta_i) \left[X_i^{\mathrm{T}}\gamma - \log \left\{ \sum_{j=1}^{n} \exp(X_j^{\mathrm{T}}\gamma) R_j(U_i) \right\} \right] \right\}.$$

Then, we define the predicted survivor curve at $X = x$ as $\widehat{G}(t \mid x) = G(t \mid x; \widehat{\gamma}_n) = \exp\{-\widehat{\Lambda}_0(t, \widehat{\gamma}_n) \exp(x^{\mathrm{T}}\widehat{\gamma}_n)\}$, and Breslow's estimator for the integrated baseline hazard function,

$$\widehat{\Lambda}_0(t, \widehat{\gamma}_n) = \int_0^t \frac{d\bar{N}^{(0)}(u)}{\sum_{i=1}^{n} \exp(X_i^{\mathrm{T}}\widehat{\gamma}_n) R_i(u)},$$

$N_i^{(0)}(u) = I(U_i \le u, \Delta_i = 0)$, $\bar{N}^{(0)}(u) = \sum_{i=1}^{n} N_i^{(0)}(u)$.

Finally, define the vectors $\theta = (\beta^{\mathrm{T}}, \alpha^{\mathrm{T}}, \gamma^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\theta}_n = (\widehat{\beta}_n^{\mathrm{T}}, \widehat{\alpha}_n^{\mathrm{T}}, \widehat{\gamma}_n^{\mathrm{T}})^{\mathrm{T}}$, where $\widehat{\gamma}_n$ refers to either the maximum likelihood or maximum partial likelihood estimator for γ. Then, the direct estimator is $\widehat{\mu}_n(t) = \mu(t, \widehat{\theta}_n)$,

$$\begin{aligned} \mu(t, \widehat{\theta}_n) &= \int_{\mathcal{X}} \left[m_1(t, x; \widehat{\beta}_n) G(t \mid x; \widehat{\gamma}_n) - \int_{[0,t)} m_0(u, x; \widehat{\alpha}_n)\, dG(u \mid x; \widehat{\gamma}_n) \right] dH_n(x), \\ &= \frac{1}{n} \sum_{i=1}^{n} \left[m_1(t, X_i; \widehat{\beta}_n) G(t \mid X_i; \widehat{\gamma}_n) - \int_{[0,t)} m_0(u, X_i; \widehat{\alpha}_n)\, dG(u \mid X_i; \widehat{\gamma}_n) \right]. \end{aligned} \tag{2.6}$$

Evidently, the direct estimator involves the computation of n subject-specific survivor functions $\{G(u \mid X_i; \widehat{\gamma}),\ 0 \le u \le t,\ i = 1, \ldots, n\}$ and then evaluating n integrals in the second expression of (2.6). In general, the computation can be intensive especially when considering the additional computation for statistical inference (see § 2.3). If one adopts the proportional hazards model for $\lambda^{(0)}(t \mid X; \gamma)$, then the second expression in (2.6) is a stochastic integral and can be evaluated directly. When we adopt a parametric model for $\lambda^{(0)}(t \mid X; \gamma)$, we evaluate the integrals in (2.6) using numerical integration.

2.3 Large Sample Properties

In this section, we outline the large sample properties of the proposed estimators with additional details provided in Supplementary Material. Throughout, we assume that m_1, m_0, and G are all correctly specified so that the true population mean $\mu_0(t)$ at a fixed target treatment length t is synonymous with $\mu(t, \theta_0)$.

Let $\sum_{i=1}^n \psi_\alpha(Z_i; \alpha)$ be the score function from a generalized linear regression model of the outcome on observed treatment length and covariates for those subjects whose treatment was prematurely discontinued and $\sum_{i=1}^n \psi_\beta(Z_i; \beta)$ is the analogous score function for the subjects whose treatment was not prematurely discontinued. Let $\sum_{i=1}^n \psi_\gamma(Z_i; \gamma)$ be the score function for a parametric log-linear model (cf. Kalbfleisch and Prentice, 2002, p. 69) fit to the data, $\{U_i, 1 - \Delta_i, X_i,\ i = 1, \ldots, n\}$. If we further define

$$\psi_\mu\{Z_i; \mu_0(t), \theta_0\} = m_1(t, X_i; \beta_0)G(t \mid X_i; \gamma_0) - \int_0^t m_0(u, X_i; \alpha_0)\, dG(u \mid X_i; \gamma_0) - \mu_0(t),$$

then a Z-estimator for $\vartheta = \{\mu(t), \theta^{\mathrm{T}}\}^{\mathrm{T}}$ is the solution to system of equations, $0 = \Psi_n(\vartheta)$,

$$\Psi_n(\vartheta) = \sum_{i=1}^n \psi_\vartheta(Z_i; \vartheta),\ \psi_\vartheta = (\psi_\mu, \psi_\alpha^{\mathrm{T}}, \psi_\beta^{\mathrm{T}}, \psi_\gamma^{\mathrm{T}})^{\mathrm{T}}. \tag{2.7}$$

The asymptotic properties of Z-estimators are described elsewhere (cf. van der Vaart and Wellner, 1996; Boos and Stefanski, 2013). Under conditions provided in the Supplementary Material, we conclude that $n^{1/2}(\widehat{\vartheta} - \vartheta_0)$ converges in distribution to a mean-zero normal random vector with covariance $A(\vartheta_0)^{-1}B(\vartheta_0)\{A(\vartheta_0)^{-1}\}^{\mathrm{T}}$, where $B(\vartheta_0) = E\{\psi_\vartheta(Z_1, \vartheta_0)\psi_\vartheta^{\mathrm{T}}(Z_1, \vartheta_0)\}$ and $A(\vartheta_0) = E\{-(\partial/\partial\vartheta)\psi_\vartheta(Z_1, \vartheta_0)\}$.

the asymptotic details are somewhat more interesting. In this case, one can show $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\} = n^{-1/2}\sum_{i=1}^n \varphi_\mu(Z_i; \theta_0) + o_p(1)$, where the influence function may

be expressed as the sum,

$$\varphi_\mu(Z_i;\theta) = v(Z_i,\theta) + \int_0^\infty w(u, Z_i, \theta)\, dM_i^{(0)}(u),$$

v and w are both scalar functions of the data and parameters, and $M_i^{(0)}(t) = N_i^{(0)}(t) - R_i(t)\Lambda_0(t)\exp(X_i^{\mathrm{T}}\gamma_0)$, a local martingale with respect to the filtration $\mathcal{F}_t = \sigma\{N_i(u), N_i^{(0)}(u), X_i, u \leq t,\ i = 1,\ldots,n\}$. Heuristically, $v(Z_i,\theta)$ captures the influence of $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ on $\widehat{\mu}_n(t)$ while $\int w(u, Z_i, \theta)\, dM_i^{(0)}(u)$ is a measure of the combined influence of the maximum partial likelihood estimator plus the Breslow estimator while the subject is continuously treated and then at the point of treatment discontinuation, if treatment was discontinued prematurely. By the central limit theorem and Slutzky's Theorem, $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\}$ converges in distribution to a mean-zero normal random variable with covariance $E\{\varphi_\mu(Z_1;\vartheta_0)\varphi_\mu^{\mathrm{T}}(Z_1;\vartheta_0)\}$.

Under conditions given in the Supplementary Material, the empirical sandwich matrix is a consistent estimator of the asymptotic covariance for the Z-estimator. Under the proportional hazards model, one can use martingale residuals to form an empirical estimator of the asymptotic covariance (cf. Johnson and Tsiatis, 2005). We derived an analytic expressions for the asymptotic covariance and could evaluate these expressions directly for the specific models and estimators chosen here, but we anticipate a broader interest in resampling techniques for more complicated models and estimators when the asymptotic covariance cannot be not evaluated directly. With this goal in mind, we investigated both jackknife and bootstrap resampling and found that both methods performed well under a variety of scenarios with approximately the same amount of computational burden. Note, under conditions given in the Supplementary Material, bootstrap resampling for the proposed estimators can be justified formally along the lines in Kosorok (2008, § 10.2), for example. Although we investigated the jackknife and bootstrap for all direct estimators, as well as the em-

pirical sandwich matrix for the Z-estimator, we only report in § 2.4.2 the results from jackknife and the empirical sandwich matrix for the direct estimator using counting processes and parametric survival models, respectively.

2.4 Simulation Studies

2.4.1 Treatment Assignment on a Finite Set

First, we compare the direct estimator to the estimator proposed by Johnson and Tsiatis (2004) when treatment realizes only one of K values. The simulation scenarios below are adapted from those found in Johnson and Tsiatis (2004). We begin by simulating a standard normal random variable X and subsequently generating a potential treatment terminating event C as an exponential random variable with mean, $\gamma_0 e^{\gamma_1 X}$, $\gamma_0 = 0.005,\ \gamma_1 = -2$. Next, the treatment assignment mechanism follows a discrete hazards model, where

$$\rho_j(X;\varphi) = P(U = t_j, \Delta = 1 \mid U \geq t_j, X; \varphi) = H(\varphi_{0,j} + \varphi_X X),\ j = 1, \ldots, K-1, \tag{2.8}$$

$H(t) = 1/\{1 + \exp(-t)\}$ and $\varphi = (\varphi_{0,1}, \ldots, \varphi_{0,K-1}, \varphi_X)^{\mathrm{T}}$. Specifically, we use $K = 4$, $\varphi = (-1.2, -0.75, 0, -0.5)$. Then, to simulate the observed treatment length data, we adopt a sequential algorithm starting with $j = 1$: if $C < t_j$ then $(U = C, \Delta = 0)$; otherwise, generate the intermediate Bernoulli outcomes, Q_j, with success probability $H(\varphi_{0,j} + \varphi_X X)$ and set $(U = t_j, \Delta = 1)$ if $Q_j = 1$; if $Q_j = 0$, continue to time t_{j+1}. The process is continued for all time points $t_j,\ j = 1, \ldots, K$, but if a subject does not have a treatment interruption prior to t_K, then they are stopped with probability one at time t_K. Approximately 25% of the observations are censoring using the above parameter values. Finally, the endpoint is generated according to the linear regression

model,

$$Y = \beta_0 + \beta_U \log(U) + \beta_X X + \beta_\Delta (1 - \Delta) + \epsilon, \tag{2.9}$$

where $\beta_0 = -2.5$, $\beta_U = 1$, $\beta_X = 0.5$, $\beta_\Delta = 2$, and ϵ is a standard normal random variable. Note that, in this case, $m_1(t, x) = \beta_0 + \beta_U \log(t) + \beta_X x$ and $m_0(t, x) = m_1(t, x) + \beta_\Delta$. As in Johnson and Tsiatis (2004), the true value is computed through Monte Carlo integration by fixing treatment assignment at t_j, simulating the outcome Y for a large number of realizations, and taking the sample average.

In the above simulation, all of m_1,m_0 and G are correctly specified; in addition, the treatment assignment mechanism in (2.8) is also correctly modeled in the procedure by Johnson and Tsiatis (2004). In the literature, this is commonly referred to as the scenario where both the outcome regression models (i.e. m_1, m_0, and G) and propensity score model in (2.8) are both correct. To illustrate the merits and deficiencies of the direct estimator vis-à-vis the estimator by Johnson and Tsiatis (2004), we explored settings where each estimator fails. We misspecify the propensity score model by setting $\rho_j(X; \varphi) = H(\varphi_{0,j} + \varphi_X e^X),\ j = 1, \ldots, K-1,\ \varphi_X = -1$ in (2.8), however, the estimator by Johnson and Tsiatis (2004) continues to model X on the natural scale. To miss-specify the outcome regression model, only one of m_1, m_0 or G must be modeled incorrectly. To this end, we modify m_1,m_0 by modeling Y as follows:

$$Y = \beta_0 + \beta_{U_1}(\log(U) - \lambda_U) + \beta_{U_2}(\log(U) - \lambda_U)^2 + \beta_X X^2 + \beta_\Delta (1 - \Delta) + \epsilon, \tag{2.10}$$

where $\beta_{U_1} = 2$, $\beta_{U_2} = 3$, $\lambda_U = 3$, and $\beta_X = 3$; β_0, β_Δ and ϵ remain the same as in (2.9). The direct estimator incorrectly assumes that model (2.9) is correct.

The simulation results are displayed in Table 2.1 for a sample size of $n = 100$. We use the empirical sandwich estimator to estimate the variance of the direct estimators when G is extreme value and jackknife method when we G is estimated via the Breslow

Table 2.1: Simulation results for the discrete time setting

	Direct											
	PH				Extreme Value				IPW			
	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP
	PS correct, OR correct											
t_1	0.036	0.197	0.194	0.937	0.006	0.195	0.191	0.929	-0.011	0.249	0.228	0.914
t_2	0.034	0.134	0.133	0.949	0.003	0.135	0.135	0.947	0.006	0.232	0.221	0.933
t_3	0.034	0.139	0.139	0.940	0.003	0.138	0.140	0.953	0.006	0.218	0.213	0.941
t_4	0.024	0.174	0.177	0.942	0.001	0.167	0.170	0.945	-0.005	0.209	0.205	0.945
	PS incorrect, OR correct											
t_1	0.017	0.291	0.304	0.943	-0.014	0.293	0.283	0.907	-0.099	0.412	0.289	0.757
t_2	0.026	0.178	0.183	0.945	-0.005	0.178	0.174	0.925	-0.113	0.350	0.265	0.802
t_3	0.030	0.130	0.132	0.953	<0.001	0.131	0.133	0.954	-0.095	0.265	0.231	0.874
t_4	0.037	0.138	0.145	0.952	0.005	0.141	0.146	0.955	0.014	0.156	0.161	0.950
	PS correct, OR incorrect											
t_1	-0.347	1.155	1.134	0.870	0.312	1.410	1.118	0.890	-0.042	1.534	1.437	0.883
t_2	-0.295	0.996	0.976	0.857	0.360	1.244	1.053	0.903	0.055	1.477	1.426	0.911
t_3	-0.621	0.921	0.915	0.780	0.017	1.148	1.079	0.887	0.044	1.355	1.345	0.922
t_4	-1.096	0.898	0.934	0.677	-0.504	1.098	1.152	0.834	0.018	1.220	1.266	0.928

PH, proportional hazards; IPW, inverse probability weighted estimator, Johnson and Tsiatis (2004); SD, Monte Carlo standard deviation; SEE, standard error estimate; ECP, empirical coverage probability for Wald-type 95% confidence interval; PS, propensity score model; OR, outcome regression models.

estimator and a proportional hazards model for the cause-specific hazard $\lambda^{(0)}(t \mid X)$. The sample standard deviation of the point estimates is the Monte Carlo standard deviation while the Monte Carlo average of estimated standard errors is the standard error estimate. The empirical coverage probability is derived from Wald confidence intervals using critical values at the nominal level. A total of 1000 Monte Carlo data sets were simulated for each of three scenarios.

When both the outcome regression and propensity score models are correct, the finite sample bias is small for all estimators, however, the standard errors for the direct estimators are between 20-40% smaller than those from inverse weighting. If the propensity score model in (2.8) is modeled incorrectly, then the inverse weighted estimator is shows appreciable finite sample bias; the same is true of the direct estimators when m_1, m_0 in (2.9) are modeled incorrectly. The Monte Carlo standard deviation matches well the standard error estimates for the direct estimators suggesting that the variance estimators are accurate even when the sample size is moderate and the outcome regression models incorrect. When the outcome regression models are correct, the empirical coverage probabilities of the Wald intervals for the direct estimators covered the true value at the nominal level, although we observed a modest finite sample benefit of jackknife method over the empirical sandwich estimator. To investigate the finite sample bias of the empirical sandwich variance estimator, we repeated the simulation scenarios for a larger sample size $n = 200$ and found the coverage probabilities for jackknife and the empirical sandwich estimator to match closely (results not shown).

2.4.2 Treatment Assignment in Continuous Time

Similar to the discrete time scenario, we start by simulating two independent standard normal random variables $X = (X_1, X_2)^{\mathrm{T}}$. We simulate intended treatment length A as Weibull with shape parameter 2.25 and scale parameter $\varphi_0 e^{X^{\mathrm{T}}\varphi_X}$, $\varphi_0 = 2.9$, and

potential treatment-censoring C as Weibull with shape parameter 2.25 and scale parameter $\gamma_0 e^{X^{\mathrm{T}}\gamma_X}$, $\gamma_0 = 3.4$. For both cases, the observed pair (U, Δ) are defined accordingly. The outcome Y follows the linear model,

$$Y = \beta_0 + \beta_U \log(U) + \beta_X^{\mathrm{T}} X + \beta_\Delta (1 - \Delta) + \epsilon,$$

with $\beta_0 = -2.5$, $\beta_U = 1$, $\beta_X = (1, 0.5)^{\mathrm{T}}$, $\beta_\Delta = 2$, and ϵ is a standard normal random variable. To investigate the effects of potential model misspecification of the function G, we considered a more complex model for the treatment length data. In this scenario, we simulate intended treatment length A according to the survivor distribution, $\widehat{S}_0(t)e^{X^{\mathrm{T}}\varphi_X}$, $\widehat{S}_0(t)$ is the Kaplan-Meier curve of the marginal probability that a switch to second-line regimen exceeds time t using data from § 2.5.2, $\varphi_X = (-0.5, 0.5)^{\mathrm{T}}$. The potential treatment-censoring event C is simulated according the survival distribution, $G_0(t)e^{X^{\mathrm{T}}\gamma_X}$, $G_0(t) = k_1 e^{-k_2 t} + k_3 e^{-k_4 t}$, $k_1 = 1$, $k_2 = 2$, $k_3 = 0.01$, $k_4 = 0.5$, $\gamma_X = (1, -0.3)^{\mathrm{T}}$. For each simulation scenario, we repeat the simulation process for n independent subjects, $n \in \{100, 200, 300\}$, and investigate three time points: $\{0.75, 1.5, 3\}$ under the proportional hazards model and $\{0.5, 1, 2\}$ for the extreme value distribution. In order to compare our estimator to the inverse weighted estimator by Johnson and Tsiatis (2004), we collapsed the treatment length data into intervals when $\Delta = 1$ by assigning subjects to the interval with the smallest Euclidean distance from t_j, $j = 1, 2, 3$, to the observed treatment length $\log(U)$.

In general, the bias of the direct estimator was small while that of the inverse weighted estimator was highly variable depending on the scenario and time point. In the two simulation scenarios where the finite sample bias of the inverse weighted estimator was its smallest (t_2 when $n = 100$), the direct estimator was more than twice as precise as inverse weighting when G is extreme value and nearly that level when G followed a proportional hazards model. Agreement between the Monte Carlo

Table 2.2: Simulation results from the continuous time setting

	Direct											
	PH				Extreme Value				IPW			
	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP
	PH Correct											
	n=100											
t_1	0.014	0.254	0.268	0.946	-0.006	0.253	0.240	0.926	-0.109	0.500	0.380	0.835
t_2	0.015	0.155	0.164	0.953	-0.010	0.152	0.155	0.946	0.006	0.288	0.259	0.919
t_3	0.006	0.189	0.195	0.950	-0.014	0.184	0.179	0.941	0.035	0.220	0.198	0.906
	n=200											
t_1	0.022	0.173	0.180	0.957	0.010	0.171	0.170	0.947	-0.095	0.345	0.310	0.903
t_2	0.013	0.109	0.115	0.955	-0.003	0.106	0.110	0.958	0.009	0.193	0.193	0.940
t_3	0.002	0.135	0.137	0.953	-0.007	0.128	0.127	0.952	0.041	0.151	0.143	0.924
	n=300											
t_1	0.019	0.143	0.144	0.946	0.010	0.141	0.139	0.945	-0.089	0.291	0.271	0.909
t_2	0.010	0.093	0.094	0.958	-0.004	0.091	0.090	0.954	0.011	0.163	0.161	0.950
t_3	-0.005	0.115	0.113	0.946	-0.013	0.111	0.105	0.938	0.040	0.126	0.116	0.913
	Extreme Value Correct											
	n=100											
t_1	0.002	0.357	0.378	0.947	-0.001	0.356	0.333	0.919	-0.479	0.771	0.373	0.470
t_2	0.012	0.217	0.230	0.959	0.009	0.211	0.209	0.947	-0.004	0.487	0.333	0.836
t_3	0.013	0.194	0.202	0.952	0.015	0.183	0.187	0.949	0.652	0.285	0.257	0.273
	n=200											
t_1	-0.010	0.237	0.250	0.952	-0.012	0.236	0.233	0.942	-0.332	0.763	0.456	0.547
t_2	-0.010	0.149	0.155	0.951	-0.013	0.146	0.147	0.941	0.065	0.503	0.335	0.841
t_3	-0.012	0.135	0.141	0.967	-0.012	0.130	0.133	0.958	0.605	0.220	0.197	0.199
	n=300											
t_1	-0.006	0.202	0.201	0.933	-0.008	0.201	0.192	0.919	-0.275	0.709	0.477	0.605
t_2	0.005	0.125	0.126	0.953	0.004	0.123	0.120	0.946	0.090	0.472	0.308	0.868
t_3	0.004	0.117	0.114	0.941	0.003	0.113	0.109	0.932	0.615	0.177	0.163	0.112

PH, proportional hazards; IPW, inverse probability weighted estimator, Johnson and Tsiatis (2004); SD, Monte Carlo standard deviation; SEE, standard error estimate; ECP, empirical coverage probability for Wald-type 95% confidence interval; PS, propensity score model; OR, outcome regression models.

standard deviation and standard error estimates suggest that the variance estimators appear to work well. The empirical coverage probability for the direct estimates is close to the nominal level in nearly every case. Interestingly, even when the parametric model for G is incorrect, the direct estimator with a log-linear model for G performs rather well; we do not attempt to argue that this result is generalizable, however.

In this simulation study, we observed that the inverse weighted estimator performed reasonably for t_2 and not as well for t_1 and t_3. This observation indicates some finite sample bias follows from the data preparation of simulated data for the inverse weighted estimator. We also explored alternative binning strategies when processing continuous data into discrete time intervals in §A.2. Under one binning strategy, the inverse probability weighting estimator can perform as well as our outcome regression estimators in terms of coverage probability. However, using other two strategies, the inverse probability estimators doesn't work as well and can even demonstrate quite poor performances. The dependency on an appropriate binning strategy thus limit the application of the inverse probability estimator in continuous treatment studies.

2.5 Data Applications

2.5.1 ESPRIT Infusion Trial Data

Although the intended infusion lengths was to last 18-24 hours in the absence of a terminating event, the actual range was 12-37 hours. So, Johnson and Tsiatis (2004) prepared the data by collapsing the infusion data into two-hour intervals when there were no terminating events and leaving the data as is when there was a terminating event. The first and last intervals included observations in the tails of the observed infusion duration distribution. For the direct estimators proposed here, there is no data extraneous processing required. Using the same target infusion lengths of 16,

18, 20, 22, and 24 hours as in Johnson and Tsiatis (2004), a figure of the direct estimates and corresponding 95% confidence intervals are shown in Figure 2.1. The point estimates of the direct and inverse weighted estimators are close, with the biggest difference in the last two intervals for 22 and 24 hours. However, the confidence intervals from the direct estimates are substantially smaller than inverse weighting. For infusion time less than 18 hours, the confidence intervals are approximately 20% narrower than those reported by Johnson and Tsiatis (2004) while they are 40-50% narrower when infusion is longer than 18 hours. The point estimates and standard errors from the intermediate regression models m_1, m_0, and G as well as a table of the direct estimates and their corresponding standard errors are available in Appendix A.3.

2.5.2 Switch to Second-line ART in ACTG A5095

An important, open question in therapeutic HIV studies is whether delayed regimen change after virologic failure on an initial antiretroviral therapy (ART) is beneficial. Recently, Li et al. (2012) examined the differences in clinical endpoints among patients switching early or late to second-line ART after failing an initial efavirenz-based ART in AIDS Clinical Trials Group (ACTG) study A5095, a multi-center clinical trial designed to compare three ARTs in ART-naïve study participants (Gulick et al., 2004). Li et al. (2012) showed that participants who followed a two-stage policy to switch within 8 weeks of virologic failure if they failed their initial efavirenz-based ART had lower HIV-1 RNA levels and higher CD4 T-cell counts, on average, over the course of the study. Applying methods by Johnson and Tsiatis (2005) to a subset of study participants who actually failed an initial efavirenz-based ART, Johnson et al. (2013) modeled the potential outcomes $E(Y^*_{t\wedge C})$ as a function of the intended switching time, when switching time may be right-censored by a competing event. In ACTG A5095, switching time was censored for 100 of 182 participants for the

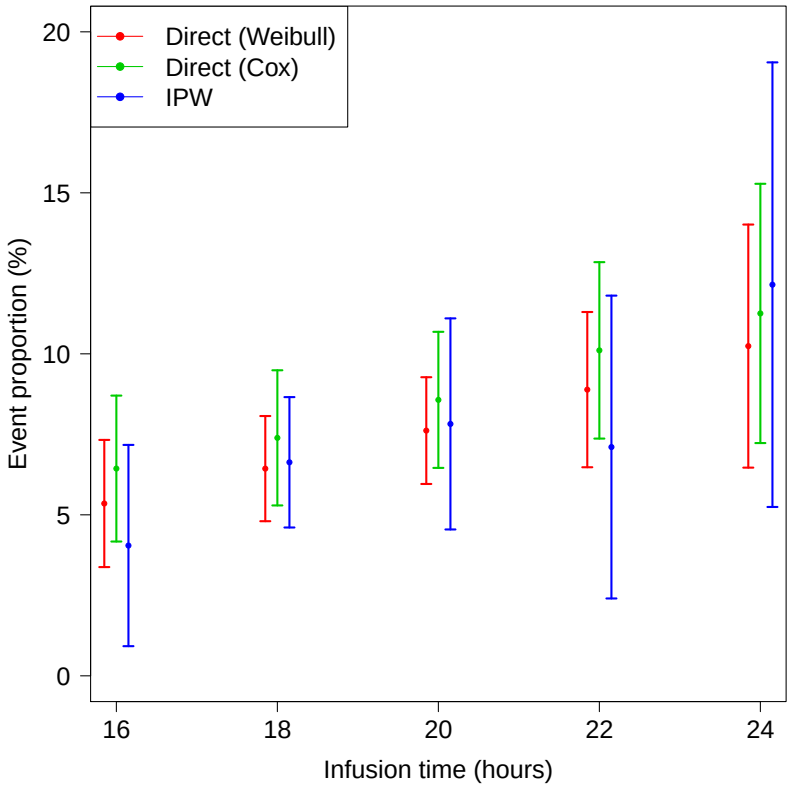

Figure 2.1: Comparisons of estimates (solid circles) and their 95% confidence intervals (line lengths) form ESPRIT trial
Our direct estimators based on Weibull distribution (red) and distribution from Cox proportional hazard model (green) are both more precise (shorter length) than the inverse probability weighting estimator (blue).

subset who failed an efavirenz-based ART. In Johnson et al. (2013), we found no statistically significant trend in switching time for the CD4 endpoint, which seemingly contradicted the result reported in Li et al. (2012). The statistical analysis below uses the same data for the same CD4 endpoint.

In the current analysis, our objective is to estimate the mean potential outcome $E(Y^*_{t\wedge C})$, where the policy is to switch to second-line regimen at time t after having already failed an initial efavirenz-based regimen, $t \in \{1, 2, 4, \ldots, 24\}$. We computed the direct estimates and their standard errors and compared them to estimates from Johnson and Tsiatis (2004). To apply the inverse weighted estimator, the data were collapsed into intervals in the same manner as in § 2.4.2 using four intervals, with

midpoints at 1, 2, 4, and 16 weeks. The results are displayed in Figure 2.2.

First, we note the obvious difference that the point estimates from the direct estimator are available for all 13 time points while those from the inverse weighted estimator are only available for four time points. This finite sample problem reflects the fact that while nearly half of all patients that failed an initial efavirenz-based ART switched to second-line ART, most of them switched within 4-6 weeks of virologic failure. As a result, there is relatively little information to model the treatment assignment mechanism beyond 8 weeks. Second, in four time points with three point estimates, the interval estimates from the direct estimator are approximately 21% smaller than interval estimates from inverse weighting. Finally, to investigate whether the new analysis had any practical impact compared to earlier analyses, we compared policies that switched at 2 and 4 weeks using a formal hypothesis test. We rejected the null hypothesis using a Wald test from the direct estimator (p-value = 0.01) but not from inverse weighting (p-value = 0.16). This finding deviates from that reported earlier in Johnson et al. (2013) and also raises new questions about the utility of the 8-week cutoff to define early vis-à-vis late regimen change.

2.6 Remark

The epidemiology and biostatistics literature is replete with examples where authors use discrete time approximations to continuous time problems. The estimator by Johnson and Tsiatis (2004) was developed for these settings or when the data are, in fact, collected at a finite number of clinic visits. In an unpublished 2011 Emory University PhD thesis, L. Li derived the doubly-robust semi-parametric efficient extension of the Johnson and Tsiatis (2004) estimator. Although the proposed direct estimator lacks the robustness of Li's estimator, an advantage is its versatility for the discrete and continuous time setup.

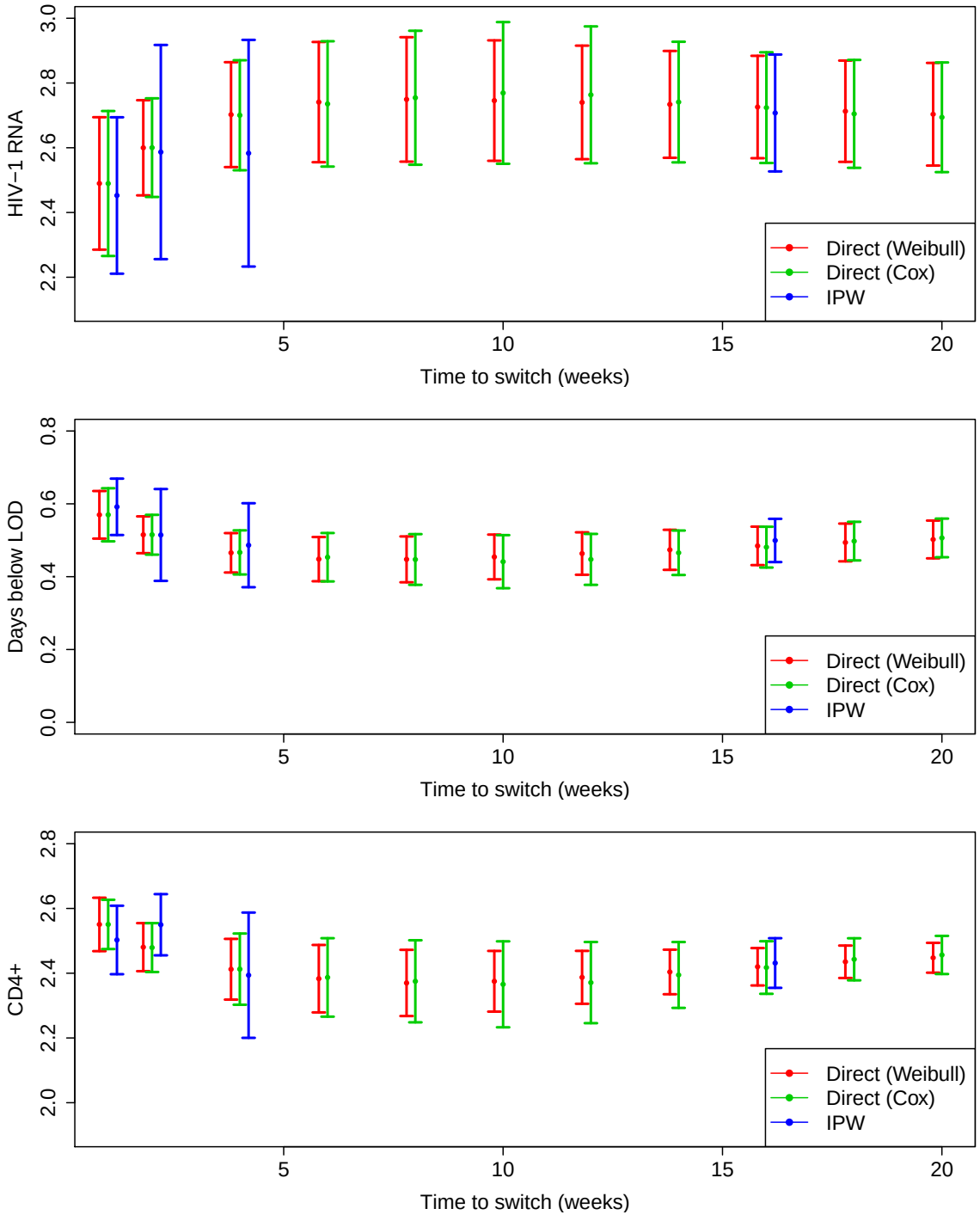

Figure 2.2: Mean outcomes across switch time estimated by direct estimator based on Weibull model (red), direct estimator based on Cox proportional hazard model (green) and inverse probability weighting estimator from Johnson and Tsiatis (2004) (blue) with their respective 95% confidence intervals.

Chapter 3

Estimation of the Distribution of Potential Outcomes amidst Early Treatment Stoppage in the Presence of Time-Varying Confounders

3.1 Introduction

In Chapter 2, we studied treatment effect with only baseline confounders. In many longitudinal studies, there often exist time-varying effects that associated with both the decision of treatment and the outcome of interest. For these senarios, one may get biased estimates if those time-dependent covariates are not accounted for appropriately. Marginal structural models (MSM), estimated by inverse probability weighting (IPW) scheme are commonly implemented to estimate treatment effects with time-varying confounders (Robins et al., 2000; Hernán et al., 2000). However, this

method can suffer from the instability and imprecision of inverse probability weighting scheme. Recently, an alternative scheme is to use G-computational algorithm, a generalized computational algorithm to approximate the distribution of potential outcomes. Several articles applied the G-computational algorithm as an alternative to IPW to estimate the parameters of marginal structural models (Robins, 1997; Snowden et al., 2011; Daniel et al., 2011) in longitudinal studies. Here in this project, we implemented the G-computational algorithm to estimate the mean potential outcome on treatment length policies, either with or without Monte Carlo integration. This application specifically models the treatment terminating event process, as what we did earlier in another project, but also accounts for time-dependent confounders.

3.2 Methods

3.2.1 Observed Data

The observed data could be illustrated in Figure 3.1 as follows:

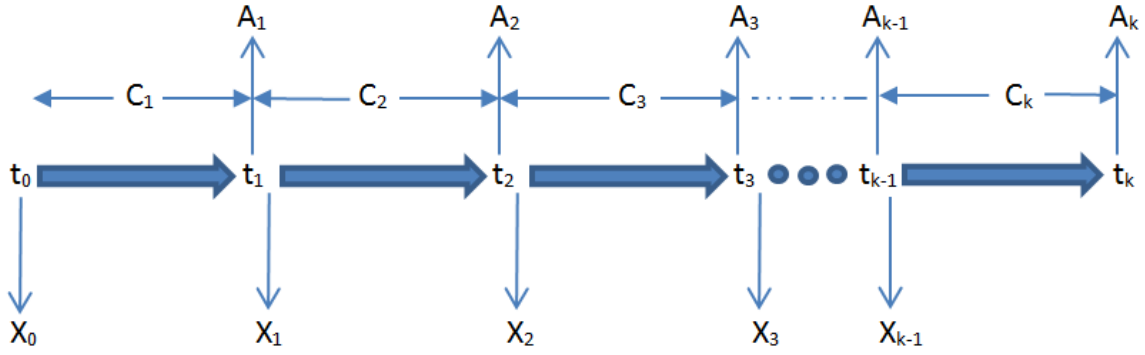

Figure 3.1: Illustration

Let $A_j = 0, 1$ denotes binary treatment decision at the jth targeted treatment length t_j ,$j = 1, \cdots, K$, $0 = t_0 < t_1 < t_2 < \cdots < t_K$; X_0 denotes all the covariate information at baseline; X_j denotes all the covariate information for the time interval $[t_j, t_{j+1})$, $j = 1, \cdots, K-1$; C_j denotes whether the potential treatment terminating

event occurs in the interval $[t_{j-1}, t_j)$, $j = 1, \cdots, K$. Patients entering the study with baseline information measured (i.e. X_0) are first treated until treatment-terminating events occur before t_1 (i.e. $C_1 = 1$), or until physicians stop the treatment at t_1 (i.e. $A_1 = 1$). In the absence of treatment-terminating events before t_1 (i.e. $C_1 = 0$) and when physicians choose to continue the treatment (i.e. $A_1 = 0$), the patients will continue to take treatment while having some of his clinical characteristics measured again (i.e. X_1). This procedure continues, if neither any treatment-terminating events occur nor physicians stop the treatment at t_j, $j = 1, \cdots, K-1$, until the end of the study t_K. Then at t_K all remaining ongoing treatment are stopped (i.e. $A_K = 1$).

Note that we always observe all the information until the first treatment length (i.e. X_0 and C_1) while we can only observed the follow-up information if the subject is still being followed. However, the outcome Y is always observed regardless of censoring status.

3.2.2 Proposed method

Define treatment decision in the presence of treatment terminating event during time interval $(t_{j-1}, t_j]$ as:

$$d_j = \begin{cases} 1, & \text{if } C_j = 1 \text{ or } A_j = 1 \\ 0, & \text{if } C_j = 0 \text{ and } A_j = 0 \end{cases}$$

and

$$\overline{d}_j = (d_1, \cdots, d_j); \quad j = 1, \cdots, K$$

Using our proposed method, the estimand could be expressed as

$$\mu_0(t_j) = E(Y^*_{\overline{d}_j}) = \int_0^\infty P(Y^*_{\overline{d}_j} \geq y) dy$$

where $P(Y^*_{\overline{d}_j} \geq y)$ could be approximated by G-computational algorithm (Robins, 1986) as follows:

Denoting the sample space (D, X). For a particular set of treatment regimes over all time points, $\overline{d}_K$, the distribution of potential outcome in the hypothetical setting that all individuals were assigned to $\overline{d}_K$ is defined as $Y^*(\overline{d}_K)$ could be approximated by

$$\begin{aligned} P(Y^*(\overline{d}_K) \leq y) &= \int_{D_1, X_0} \cdots \int_{D_K, X_{K-1}} p(Y \leq y \mid \overline{X}_{K-1}, \overline{d}_K) \\ &\quad \times \prod_{j=1}^{K} f(C_j, X_{j-1} \mid \overline{X}_{j-2}, \overline{d}_{j-1}) d\mu(C_j, X_{j-1}) \end{aligned} \tag{3.1}$$

where $\overline{X}_{j-1} = (X_0, ..., X_{j-1})$ denotes the covariate history prior to time t_j.

Additionally, the inference for time-dependent case also rely on two key assumptions like the time-fixed case: the generalization of the "stable unit treatment value assignment" and the generalization of the "sequential randomization assumption", which is a generalization of the strong ignorability assumption to time-dependent treatment assignments. The former one assumes that there is no inferences on response between patients and the value of covariates at time t_j only depends on the treatment assignment given to that individual up to time t_{j-1} and not related to any future treatment assignment. The latter assumption states that the treatment assignment at time t_k, i.e. A_k is made completely at random conditioned on the observed time-varying covariate history and time-varying treatment history up to time t_k.

Assume $X_0 \sim h(X_0, \theta_0)$, $C_1 \sim f_0(C_1 \mid X_0, \eta_0)$ and $X_j \sim h(X_j \mid \overline{X}_{j-1}, \overline{d}_j, \theta_j)$, $C_{j+1} \sim f_0(C_{j+1} \mid \overline{X}_j, \overline{d}_j, \eta_j)$, for $j = 1, \cdots, K-1$. Also assume $Y \sim f(y \mid \overline{X}_{K-1}, \overline{d}_K, \theta_y)$. When $h(.)$ follows a Markov process, $f_0(.)$ follows binomial distribution and $f(.)$ follows normal distribution, we can implement autoregressive models, discrete hazard models and linear regression model respectively to estimate $\theta'_j s$, $\eta'_j s$ and θ_y.

G computation with Monte Carlo integration

Use the model fit above, we can implement the following Monte Carlo algorithm to approximate the distribution of potential outcome (Daniel et al., 2011): For $r = 1, \cdots, M$

1. Generate $X_{0,r}$ from $h(X_0, \hat{\theta}_0)$, $C_{1,r}$ from $f_0(C_1 \mid X_{0,r}, \hat{\eta}_0)$. Let $d_{1,r} = C_{1,r}$.
2. Generate $X_{1,r}$ from $h(X_1 \mid X_{0,r}, d_{1,r}, \hat{\theta}_1)$, $C_{2,r}$ from $f_0(C_2 \mid X_{1,r}, X_{0,r}, d_{1,r}, \hat{\eta}_1)$. Let $d_{2,r} = C_{2,r}$.
3. Generate $X_{j,r}$ from $h(X_j \mid \overline{X}_{j-1,r}, \overline{d}_j, \hat{\theta}_j)$, $C_{j+1,r}$ from $f_0(C_{j+1} \mid \overline{X}_{j,r}, \overline{d}_j, \hat{\eta}_j)$ for $j = 1, \cdots, K-1$. Let $d_{j+1,r} = C_{j+1,r}$ for $j < K-1$. Let $d_{K,r} = 1$.
4. if $d_{j,r} = 1$ for any step above, then go directly to step 5.
5. Generate Y_r from $f(y \mid \overline{X}_{K-1,r}, \overline{d}_K, \hat{\theta}_y)$.
6. Go back to step 1 for $r = r + 1$.

Finally, the estimated outcome could be calculated by $\hat{\mu}_{t_j} = \hat{Y} = \frac{1}{M} \sum_{r=1}^{M} Y_r$. One can repeat the steps above to get $\hat{\mu}_{t_j}$ where $j = 1, \cdots, K$. Note that when K=0, then G-computation will be reduced to outcome regression. So it could be considered as an generalization of direct modeling that we used earlier.

Alternatively, assuming the potential outcomes follow a linear trend across time, i.e. $\mu = \delta_0 + \delta_1 t$, we can estimate this linear trend defined by δ_0 and δ_1 by combining all Y_r's for each t_j where $r = 1, \cdots, M;\ j = 1, \cdots, K$ together.

G computation with direct prediction

Instead of using the above Monte Carlo algorithm to approximate the distribution of potential outcomes, if the mean of potential outcomes is of major concern, we can also perform prediction directly based on the model fit. Similar idea is suggested before by

Snowden et al. (2011) when they were applying G computation algorithm for MSM. However, they focused on simple cases with only baseline covariates and binary treatment assignment without interruption. Our approach here utilized G computation directly and is extended to incorporate time varying effects and treatment censoring. This prediction approach avoids implementing iterations of Monte Carlo simulations and therefore saves substantial computation time, as well as accommodating sophisticated treatment mechanisms.

For example, to estimate the potential outcome when treatment is designed to stop at t_j ($j = 2, \cdots, K$) subject to treatment terminating events, the following steps can be performed:

1. Identify possible scenarios Ω which includes:

   - the treatment stopped early before t_1, i.e. $C_1 = 1$;
   - the treatment stopped at or after t_1 but before t_2, i.e. $C_1 = 0,\ C_2 = 1$;
   - $\cdots$
   - the treatment stopped at or after t_{j-1} but before t_j, i.e. $C_1 = 0,\ \cdots,\ C_{j-1} = 0,\ C_j = 1$;
   - the treatment stopped at t_j as originally designed, i.e. $C_1 = 0,\ \cdots,\ C_j = 0,\ A_j = 1$.

2. Predicted probabilities for each possible scenario for each subject: $P(\varpi \mid X_i),\ \varpi \in \Omega$.

3. Predicted outcomes if each possible scenario occurs for each subject: $E(Y \mid \varpi, X_i),\ \varpi \in \Omega$.

4. The average potential outcome could be estimated as $\frac{1}{n}\sum_{i=1}^{n}\sum_{\varpi\in\Omega} P(\varpi \mid X_i)E(Y \mid \varpi, X_i)$

Additionally, we can also assume the potential outcomes follow a linear trend across time and estimate their intercept δ_0 and slope δ_1 as mentioned in §3.2.2.

3.3 Simulation Studies

3.3.1 Data Simulation

Simulation studies were conducted to assess the performance our proposed method. We focused on with M=4 time points ($t_1,\ t_2,\ t_3,\ t_4$) and constructed a simulated dataset as follows:

1. Let $\alpha_C = (\alpha_{C,1}, \alpha_{C,2}, \alpha_{C,3}, \alpha_{C,4}) = (0.65, 1.3, 1.95, 2.6)$; $\beta_C = (-0.1, -0.1)$ and $\gamma_C = -0.9$; $\alpha = (0.5, 1, 1.5, 2)$; $\beta = (-0.2, -0.2)$ and $\gamma = -0.8$.

2. Simulate three independent normal random variables: $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$ and time-dependent effect X_t at baseline $X_0 \sim N(4,1)$.

3. Generate C_1, a potential treatment terminating event among $[0, t_1)$ as a binomial random variable with $p = \lambda_1^C = \frac{1}{1+e^{-(\alpha_{C,1}+X^T\beta_C+X_0\gamma_C)}}$. If $C_1 = 1$ then $U = t_1/2$, $\Delta = 0$. All the information afterwards except the outcome are set to missing. (i.e. $X_{t_1}, \cdots, X_{t_3}, C_2, \cdots, C_4, A_1, \cdots, A_4$)

4. If $C_1 = 0$, then generate A_1, the treatment decision at time t_1 as a binomial random variable with $p = \lambda_1 = \frac{1}{1+e^{-(\alpha_1+X^T\beta+X_0\gamma)}}$. If $A_1 = 1$, then $U = t_1$, $\Delta = 1$, and information after t_j will be set to missing (i.e. $X_{t_1}, \cdots, X_{t_3}, C_2, \cdots, C_4$, $A_2, \cdots, A_4$).

5. Otherwise, when $A_j = C_j = 0$. Generate $X_{t_1} \sim N(X_0, 1)$

6. For $j = 2, 3, 4$,

(a) Generate C_j, a potential treatment terminating event among $[t_{j-1}, t_j)$ as a binomial random variable with $p = \lambda_j^C = \frac{1}{1+e^{-(\alpha_{C,j}+X^T\beta_C+X_{t_{j-1}}\gamma_C)}}$. If $C_j = 1$ then $U = (t_{j-1} + t_j)/2$, $\Delta = 0$, and all information afterwards except the outcome are set to missing.

(b) If $C_j = 0$, generate A_j, the treatment decision at time t_j as a binomial random variable with $p = \lambda_j = \frac{1}{1+e^{-(\alpha_j+X^T\beta+X_{t_{j-1}}\gamma)}}$. If $A_j = 1$, then $U = t_j$, $\Delta = 1$, and all information afterwards except the outcome are set to missing.

(c) Otherwise, when $A_j = C_j = 0$. Generate $X_{t_j} \sim N(X_{t_{j-1}}, 1)$.

7. If $C_4 = 0$, then set $A_4 = 1$, $\Delta = 1$.

8. The continuous outcome Y is generated as

$$Y = \beta_0 + \beta_U \log(U) + \beta_{X_1} X_1 + \beta_{X_2} X_2 + \beta_\Delta \Delta + \beta_{Xt} X_t + \epsilon,$$

where $\beta_0 = 1.5$, $\beta_U = 1$, $\beta_{X_1} = 1$, $\beta_{X_2} = 0.5$, $\beta_\Delta = -2$, $\beta_{Xt} = 0.5$, X_t is the last observed time-varying covariate for each subject and ϵ is a standard normal random variable.

The population parameters of interest μ_j's, $j = 1, 2, 3, 4$, δ_0 and δ_1 were approximated by simulation. For each μ_j, the same algorithm above is followed except that the treatment decision is not simulated but forced to stop at t_j (i.e. $A_l = 0$, $l < j$; $A_j = 1$) if not already censored due to treatment terminating event. We generated the outcome Y 500,000 times and took the sample average as μ_j. And the true value of δ_0 and δ_1 were approximated by fitting a linear regression on the 500,000*4 Y's all together.

3.3.2 Monte Carlo Integration by G-computational Algorithm

Given the simulated dataset and assuming all the distribution families above are known, we estimated all the related parameters and then use these estimated parameters to conduct a set of Monte Carlo integration (M=100,000) known as the G-computational algorithm as described in §3.2.2. Finally, we got M=100,000 random draws from the potential outcome and we took their average as our estimate of the potential outcome $\hat{\mu}_j$. We fitted a linear regression on the 100,000 *4 random draws and took the estimated intercept and slope as our estimated $\hat{\delta}_0$ and $\hat{\delta}_1$. The standard errors for $\hat{\mu}_j$'s, $\hat{\delta}_0$, $\hat{\delta}_1$ were estimated by bootstrap method based on 100 bootstrap resamples. We also tested the standard errors estimated based on 500 resamples and they barely showed any differences.

3.3.3 Direct Prediction by G-computational Algorithm

After the parameters in the discrete hazard model and linear regression model are estimated from the simulation dataset, one can also predict the chance of each possible treatment schemes subject to treatment terminating events and their corresponding outcomes respectively. Sequentially, the estimate of the potential outcome $\hat{\mu}_j$, $j = 1, \cdots, K$ could be calculated as the sum of the product of each possible outcome and its occurring probability averaging over all subjects as illustrated in §3.2.2. For subjects with missing time dependent covariate information in later time, their closest covariate information in time was applied instead. We also fitted a linear regression on the n*4 estimated potential outcomes and took the estimated intercept and slope as our estimated $\hat{\delta}_0$ and $\hat{\delta}_1$. The standard errors for $\hat{\mu}_j$'s, $\hat{\delta}_0$, $\hat{\delta}_1$ were estimated by bootstrap method based on 100 bootstrap resamples.

Table 3.1: Simulation results for estimation on the potential outcome for n=300

	t	Bias	SD	SEE	ECP
Our approach	t_1	-0.003	0.156	0.154	0.947
with MC	t_2	0.001	0.111	0.109	0.937
	t_3	-0.002	0.103	0.105	0.955
	t_4	-0.001	0.112	0.122	0.965
Our approach	t_1	>-0.001	0.154	0.154	0.947
w/o MC	t_2	<0.001	0.108	0.109	0.951
	t_3	0.004	0.100	0.103	0.950
	t_4	-0.004	0.109	0.115	0.957
	t_1	0.038	0.409	0.365	0.890
IPW	t_2	0.056	0.391	0.328	0.860
	t_3	0.035	0.379	0.289	0.839
	t_4	-0.002	0.116	0.119	0.950
	t_1	0.341	0.301	0.314	0.780
$IPW_{baseline}$	t_2	0.422	0.255	0.261	0.612
	t_3	0.344	0.241	0.234	0.652
	t_4	-0.072	0.110	0.113	0.905

Truth = (3.889, 4.753, 5.377, 5.864)
SD, Monte Carlo standard deviation;
SEE, standard error estimate based on bootstrap;
ECP, empirical coverage probability for
Wald-type 95% confidence interval.
1000 Monte Carlo datasets were generated.

Table 3.2: Simulation results for estimation on the linear trend of potential outcomes across time for n=300

	δ_0				δ_1			
	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP
with MC	0.0035	0.1918	0.1892	0.948	<0.0001	0.0055	0.0056	0.949
without MC	0.0076	0.1900	0.1881	0.943	-1.00E-04	0.0055	0.0054	0.953

True value for δ_0=3.3389; true value for δ_1=0.0655
SD, Monte Carlo standard deviation; SEE, standard error estimate based on bootstrap;
ECP, empirical coverage probability for Wald-type 95% confidence interval.
1000 Monte Carlo datasets were generated.

3.3.4 Simulation Results

Using the above settings, we conducted simulation studies with our proposed methods along with the inverse probability weighting estimators for both time dependent covariates and baseline covariates as well as baseline covariates only (Johnson and Tsiatis, 2004). All approaches work well with good coverage probabilities (Table 3.1) except the IPW estimator with only baseline covariates. This IPW estimator is obviously biased most of the time due to ignoring the time varying effect. Among the other three estimators that perform well, our proposed methods demonstrate smaller variances and biases and better coverage probabilities than the inverse probability weighting estimator. When we tested these results in moderate and large sample sizes, the inverse probability weighting estimator showed slightly better coverage probabilities with increased sample sizes (results not shown), but still not as well as our approaches do at n=300.

Our two proposed methods perform equally well in estimating mean potential outcomes as well as in the linear trends (Table 3.1 & 3.2). As the computation time for our prediction approach is much less than the one using Monte Carlo integration, the direct prediction via G-computation algorithm method is more appealing. However, in situations where the distribution of potential outcomes is more of concern, (for instance, if quartiles of potential outcomes are the primary interest,) then the Monte Carlo integration by G computation algorithm method would be preferred.

3.4 Remarks

We proposed two methods in this chapter to take time-varying effects into account when estimating the mean potential outcome or its linear trend across time. Compared to the existing estimator for time-varying effects based on inverse probability weighting, our approach based on outcome regression is more efficient and could be

implemented to approximate the distribution of the potential outcome in addition to the mean of the potential outcome.

Chapter 4

Doubly Robust Estimation of Potential Outcomes amidst Early Treatment Stoppage in the Presence of Time-Varying Confounders

4.1 Introduction

Doubly-robust estimators have been recommended for routine use due to their merit of being consistent when either propensity score or the outcome regression model is correctly specified. And it is also semi-parametric efficient when both two models are correctly specified (Tan, 2006; Li et al., 2012). However, existing simulation studies (Kang and Schafer, 2007) demonstrate that this doubly robust estimator performs similarly to the inverse probability weighting estimators when OR model is miss-specified. Also, there were lack of evidence that doubly robust estimators can offer

any improvement over the outcome regression estimator by using the mildly biased propensity score model.

To advance the performance of doubly robust estimation under miss-specification, substantial work has been focused on improving the efficiency of doubly robust estimators when the outcome regression model may be biased (van der Laan and Rubin, 2006; Cao et al., 2009; Tan, 2010; Tsiatis et al., 2011; Rotnitzky et al., 2012). Under either simple or monotone coarsening settings, Cao et al. (2009); Tsiatis et al. (2011) identified versions of weighted outcome regression estimating equations that yield optimal parameters that minimize the variance of doubly robust estimators, even when the outcome regression models are miss-specified. Under missing data mechanism, Tan (2010) developed an augmented likelihood estimator of Tan (2006) by calibrating the coefficients in a linear, extended propensity score model. For cross-sectional data, Rotnitzky et al. (2012) derived improved doubly robust estimators by solving the outcome estimating equations. In spite of these developments on doubly robustness, none of these estimators were designed for dynamic regime settings and neither do they account for time-varying effects. Here in this project, we implemented the technique from Cao et al. (2009); Tsiatis et al. (2011) to derive our improved doubly robust estimators based on dynamic regime settings (Murphy et al., 2001). In addition, our proposed estimators can take into account the time-varying effects that associated with the treatment assignment selection and/or the outcome.

In this report, we first describe our study setting, a nonrandom dynamic regime scenario. Then we show how Murphy et al. (2001)'s framework would fit in our case, by examining both the specific propensity score and outcome regression models. For ease of estimation, we further studied our proposed estimator under a two-stage setting and with a simplified outcome that does not depend on treatment decision. We illustrate the derivation of our improved doubly-robust estimator for this simplified scenario and demonstrate its merit of efficiency as well as robustness under simulation

studies.

4.2 Methods

4.2.1 Observed Data

Let $A_j = 0, 1$ denotes binary treatment decision at the j_{th} targeted treatment length t_j ,$j = 1, \cdots, k$, $0 = t_0 < t_1 < t_2 < \cdots < t_k$; X_{t_0} denotes time-dependent covariate information at baseline; X_b denotes time-independent covariate information at baseline; X_{t_j} denotes the time dependent covariate information for the time interval $(t_j, t_{j+1}]$, $j = 0, \cdots, k-1$; C_j denotes whether the potential treatment terminating event occurs in the interval $[t_j, t_{j+1})$, $j = 1, \cdots, k$. Patients entering the study with baseline information measured (i.e. X_b, X_{t_0}) are first treated until treatment-terminating events occur before t_1 (i.e. $C_1 = 1$), or until physicians stop the treatment at t_1 (i.e. $A_1 = 1$). In the absence of treatment-terminating events before t_1 (i.e. $C_1 = 0$) and when physicians choose to continue the treatment (i.e. $A_1 = 0$), the patients will continue to take treatment while having his time- varying clinical characteristics measured again at t_1 (i.e. X_{t_1}). This procedure continues, if neither any treatment-terminating events occur nor physicians stop the treatment at t_j, $j = 1, \cdots, k-1$, until the end of the study t_k. Then at t_k all remaining ongoing treatment are stopped (i.e. $A_k = 1$).

Note that we always observe all the information until the first treatment length (i.e. X_b,X_{t_0}, and C_1) while we can only observed the follow-up information if the subject is still being followed. However, the outcome Y is always observed regardless of whether the treatment is stopped by physician or treatment terminating event.

We further define the time that a potential treatment-terminating event occurs as C, then the observed treatment length U and targeted treatment completion indicator Δ can be expressed as

$$\begin{array}{ll} U = t_j, & \Delta = 1, \text{ if } A_j = 1 \\ U = C, & \Delta = 0, \text{ if } C_j = 1 \end{array} \quad \text{for } j = 1, \cdots, k$$

In this case, we consider two scenarios where the exact time is observed and also the coarser view where we simply observe that a censoring event occurs in the interval $C \in [t_{j-1}, t_j)$.

Also, our findings here is based on the two essential assumptions in causal inference: the stable unit treatment value assumption (SUTVA) and the sequential randomization assumption (SRA). Define our potential outcome for an individual treated to time $t_j \wedge C$ is $Y^*_{t_j \wedge C}$, then the causal estimand of interest is the population mean,

$$\begin{aligned} \mu^*_j &= E(Y^*_{t_j \wedge C}) \\ &= E\{Y^*_{t_j} I(C \geq t_j)\} + E\{Y^*_C I(C < t_j)\} \ ^{by\ SUTVA} \\ &= E\left[E\{Y \mid U = t_j, \Delta = 1, \bar{X}_{j-1}\}\right] + E\left[E\{Y \mid U < t_j, \Delta = 0, \bar{X}_{j-1}\}\right]. \ ^{by\ SRA} \end{aligned}$$

where $\bar{X}_{j-1} = (X_b, X_{t_0}, \cdots, X_{t_{j-1}})$.

4.2.2 Doubly Robust Estimation Framework

Our new estimator is motivated by semi-parametric theory for missing data (Tsiatis, 2006). In order to show how we develop our estimator. We start with the inverse probability weighting estimator with propensity score (modeling the probability of treatment stoppage) and the outcome regression (modeling the outcome as a function of covariate information and treatment). Then we express our doubly robust estimator as an augmented inverse probability estimator, which could be view as a inverse probability weighting estimator augmented with conditional expectations of outcomes.

Propensity Score Models

Denoting the potential outcome at t_m as μ_m $(m = 1, \cdots, k)$ and the covariate information up to time t_m as $\bar{X}_{m-1} = (X_b, X_{t_0}, \cdots, X_{t_{m-1}})$ and using the setting mentioned earlier, the propensity of targeted treatment stoppage could be expressed using discrete hazards models:

$$f_m(\bar{X}_{m-1}) = \prod_{j=1}^{m-1} \left\{1 - \lambda_j(\bar{X}_{j-1})\right\} \lambda_m(\bar{X}_{m-1}),$$

with cause-specific hazard function

$$\lambda_j(\bar{X}_{j-1}) = P(U = t_j, \Delta = 1 | U \geq t_j, \bar{X}_{j-1}).$$

Then Johnson and Tsiatis (2004) inverse probability weighting estimator statistic is given as

$$\Psi(X) = \left[\frac{I(U = t_m, \Delta = 1)}{f_m(\bar{X}_{m-1})} + \frac{I(U < t_m, \Delta = 0)}{K(U, \bar{X}_{U-1})}\right](Y - \mu_m). \tag{4.1}$$

where

$$K_{\lfloor U \rfloor}(\bar{X}_{\lfloor U-1 \rfloor}) = \prod_{j=1}^{\lfloor U \rfloor} \{1 - \lambda_j(\bar{X}_{j-1})\},$$

$$\lfloor U \rfloor = \max\{j : t_j < U,\ j = 1, \ldots, k\}$$

The quantity in the braces is the inverse probability weights of the observed outcome, which are equal to one divided by the probability of treatment assignment for those subjects whose observed data is consistent with the targeted treatment length policy. Johnson and Tsiatis (2004) note that weighting scheme was equivalent to weighting subjects with unity if $(U = t_m, \Delta = 1)$ and by f_m/K_U if treatment was stopped prematurely to time t_m. In words, f_m/K_U is the probability of deciding to

j	$U \in [t_{j-1}, t_j)$	$K_{\lfloor U \rfloor}$
1	$[t_0, t_1)$	1
2	$[t_1, t_2)$	$(1-\lambda_1)$
3	$[t_2, t_3)$	$(1-\lambda_1)(1-\lambda_2)$
$\vdots$		
$m-1$	$[t_{m-2}, t_{m-1})$	$\prod_{j=1}^{m-2}(1-\lambda_j)$
m	$[t_{m-1}, t_m)$	$\prod_{j=1}^{m-1}(1-\lambda_j)$

Table 4.1: Table of possible values for $K_{\lfloor U \rfloor}$ when estimating μ_m.

stop at time t_m given the targeted treatment length was at some point later than time U.

To be clear, we can enumerate a table of possible values for k. Such a table is presented in Table 4.1. Using calculus analogous to that used in Kaplan-Meier statistics for life tables, we can show that

$$\begin{aligned} & \frac{I(U=t_m, \Delta=1)}{f_m(\bar{X}_{m-1})} + \frac{I(U<t_m, \Delta=0)}{K(U, \bar{X}_{U-1})} \\ = & \prod_{j=1}^{k} \left\{ \frac{I(t_j=t_m)}{\lambda_j(\bar{X}_{j-1})} \right\}^{I(U=t_j, \Delta=1)} \left\{ \frac{I(U \le t_m)}{1-\lambda_j(\bar{X}_{j-1})} \right\}^{I(U>t_j)} \end{aligned} \tag{4.2}$$

Outcome Regression Models

Besides the weights from propensity score models, the other essential part is the outcome regression model. Define $S_j = I(C \ge t_{j+1})$, $L_j = (S_j, X_j)$, $\bar{L}_j = (L_0, \cdots, L_j)$, $\bar{A}_{j+1} = (A_1, \cdots, A_{j+1})$, $j = 0, \cdots, k-1$. Define an indicator of whether the observed individual is following the targeted treatment plan, a nonrandom dynamic regime ($A_m = 1$, $m = 1, \cdots, k$) at time t_j as $p(a_j|S_{j-1})$, where a_j is the random variable for treatment decision at time t_j.

Note that

$$p(a_j|S_{j-1}) = \{I(t_j=t_m)\}^{I(U=t_j, \Delta=1)} \{I(U \le t_m)\}^{I(U>t_j)},$$

which constitute the numerator for $\left\{\frac{I(t_j=t_m)}{\lambda_j(\bar{X}_{j-1})}\right\}^{I(U=t_j,\Delta=1)}\left\{\frac{I(U\leq t_m)}{1-\lambda_j(\bar{X}_{j-1})}\right\}^{I(U>t_j)}$, $j=1,\cdots,k$ as in (4.2).

Then the outcome regression models can be defined as

$$g_m^{(m)}(\bar{A}_m,\bar{L}_{m-1}) = E_{obs}(Y \mid \bar{A}_m,\bar{L}_{m-1}) \quad (4.3)$$

$$g_j^{(m)}(\bar{A}_j,\bar{L}_{j-1}) = E_{obs}\left\{\sum_{a_{j+1}} p(a_{j+1}|S_j)\cdot g_{j+1}^{(m)}(a_{j+1},\bar{A}_j,\bar{L}_j) \mid \bar{A}_j,\bar{L}_{j-1}\right\} \quad (4.4)$$

Here, $g_m^{(m)}(\bar{A}_m,\bar{L}_{m-1})$ represents the expected outcome if the study ends at t_m whereas $g_j^{(m)}(\bar{A}_j,\bar{L}_{j-1})$ represents the potential outcome at t_{j+1} given the covariate and treatment decision information immediate after time t_j.

Doubly Robust Estimation with Time Varying Effects

After setting the propensity score models and outcome regression models, we construct our doubly robust estimator based on the framework from Murphy et al. (2001).

The estimating function for the doubly robust estimator could be considered as

$$\mathbb{P}_n\left[\Psi(X)-\sum_{t=1}^{k}(E\{\Psi(X)\mid\bar{A}_t,\bar{L}_{t-1}\}-E\{\Psi(X)\mid\bar{A}_{t-1},\bar{L}_{t-1}\})\right]$$

When estimating the potential outcome μ_m, our analogy to the estimating equation (5.3) from Murphy et al. (2001) could be written as

$$
\begin{aligned}
& W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) \\
& - \sum_{j=1}^{m} \{E\{W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) | \bar{A}_j, \bar{L}_{j-1}\} \qquad (4.5) \\
& - E\{W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) | \bar{A}_{j-1}, \bar{L}_{j-1}\}\} \\
= \; & W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) - \sum_{j=1}^{m} W_{\bar{p}_j}(\bar{A}_j, \bar{L}_{j-1})(g_j(\bar{A}_j, \bar{L}_{j-1}) - \mu_m) \\
& + \sum_{j=1}^{m} \sum_{a_j} \pi_j(a_j \mid \bar{A}_{j-1}, \bar{X}_{j-1}) W_{\bar{p}_j}(a_j, \bar{A}_{j-1}, \bar{L}_{j-1})(g_j(a_j, \bar{A}_{j-1}, \bar{L}_{j-1}) - \mu_m) \qquad (4.6)
\end{aligned}
$$

where

$$
\pi_j(a_j \mid \bar{A}_{j-1}, \bar{X}_{j-1}) = \begin{cases} \lambda_j(\bar{X}_{j-1}) & \text{if } a_j = 1 \\ 1 - \lambda_j(\bar{X}_{j-1}) & \text{if } a_j = 0 \end{cases}
$$

and

$$
\begin{aligned}
W_{\bar{p}_k}(\bar{A}_k, \bar{L}_{k-1}) &= (4.2), \\
W_{\bar{p}_j}(\bar{A}_j, \bar{L}_{j-1}) &= \prod_{l=1}^{j} \left\{ \frac{I(t_l = t_m)}{\lambda_l(\bar{X}_{l-1})} \right\}^{I(U=t_l, \Delta=1)} \left\{ \frac{I(U \le t_m)}{1 - \lambda_l(\bar{X}_{l-1})} \right\}^{I(U>t_l)}
\end{aligned}
$$

Note that $W_{\bar{p}_k}(\bar{A}_k, \bar{L}_{k-1})$ is the weight for all the k time points, while $W_{\bar{p}_j}(\bar{A}_j, \bar{L}_{j-1})$ is the weight for the first j time points.

Note that here we adjusted for time-varying effects when modeling the propensity score as well as the outcome regression. That is, the estimation of $\lambda_j(\bar{X}_{j-1})'s$ in $W_{\bar{p}_l}(\bar{A}_l, \bar{L}_{l-1})$ where $l = 1, \cdots, m;\ j = 1, \cdots, l$, and the estimation of $g_j(\bar{A}_j, \bar{L}_{j-1})$, $g_j(a_j, \bar{A}_{j-1}, \bar{L}_{j-1}),\ j = 1, \cdots, m$ all take time dependent $X_{t_{j-1}}$ into account.

Note that we could also write

$$
\begin{aligned}
(4.6) \quad &= W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) \\
&\quad - \sum_{j=1}^{m} \sum_{a_j} \left[I(A_j = a_j) - \pi_j(a_j | \bar{A}_{j-1}, \bar{L}_{j-1}) \right] \\
&\quad \cdot W_{\bar{p}_j}(a_j, \bar{A}_{j-1}, \bar{L}_{j-1}) \{ g_j(a_j, \bar{A}_{j-1}, \bar{L}_{j-1}) - \mu_m \} \\
&= W_{\bar{p}_m}(\bar{A}_m, \bar{L}_{m-1})(Y - \mu_m) \\
&\quad - \sum_{j=1}^{m} \sum_{a_j} \left[\frac{I(A_j = a_j) - \pi_j(a_j | \bar{A}_{j-1}, \bar{L}_{j-1}) I(U \geq t_j)}{K_j(\bar{X}_{j-1})} \right] \\
&\quad \cdot \{ g_j(a_j, \bar{A}_{j-1}, \bar{L}_{j-1}) - \mu_m \} \qquad (4.7)
\end{aligned}
$$

Which is in the format equivalent to Tsiatis et al. (2011)'s equation (2).

Due to the complex estimation procedure for $g_j(\bar{A}_j, \bar{L}_{j-1})$, we focus on the case with $k = 2$ for the present. Details of derivation of our doubly robust estimator and its connection to Murphy et al. (2001) and Tsiatis et al. (2011) is shown in Appendix B.0.1.

4.2.3 Estimation with Two-Stage Designs

In two-stage designs, the observed data in chronological order are

$$\{X_b, X_{t_0}, S_0 = I(C > t_1), A_1, X_{t_1}, S_1 = I(C > t_2), A_2, Y\},$$

or equivalently $\{L_0, A_1, L_2, A_2, Y\}$.

In this case, there are two nondynamic treatment regimes of interest: one regime that treats to $t_1 \wedge C$ and another regime that treats to $t_2 \wedge C$. Also note that because there are only two treatment decision points, A_1 and A_2, if a patient has

Δ	U	$W_{\bar{p}_1}(A_1, \bar{L}_0)$
0	$[t_0, t_1)$	$\frac{1}{1} \cdot \frac{1}{1}$=1
0	$[t_1, t_2)$	$\frac{1}{(1-\lambda_1(\bar{X}_0))} \cdot \frac{0}{1}$=0
1	t_1	$\frac{1}{\lambda_1(\bar{X}_0)} \cdot \frac{1}{1} = \frac{1}{\lambda_1(\bar{X}_0)}$
1	t_2	$\frac{0}{1-\lambda_1(\bar{X}_0)} \cdot \frac{0}{\lambda_2(\bar{X}_1)}$ =0

Table 4.2: Table of values for weights when estimating μ_1.

been continuously treated up to time t_2, then they are stopped with probability one at time t_2, i.e.$\lambda_2(\bar{X}_1) = 1$.

Regime 1. Under the first regime,

$$(4.1) = W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - \mu_1) = \left[\frac{I(U = t_1, \Delta = 1)}{\lambda_1(\bar{X}_0)} + I(U < t_1, \Delta = 0)\right](Y - \mu_1).$$

The possible values of $W_{\bar{p}_1}(A_1, \bar{L}_0)$ are illustrated in detail in Table 4.2.
By (4.3),

$$g_1^{(1)}(\bar{A}_1, \bar{L}_0) = E(Y|A_1, \bar{L}_0) = E(Y|U, \Delta, \bar{X}_0) \tag{4.8}$$

Connection of (4.8) to Chapter 1 is illustrated in the Appendix § B.0.2.

Regime 2. Under the second regime,

$$\begin{aligned}
(4.4)\ & W_{\bar{p}_2}(A_2, \bar{L}_1)(Y - \mu_2) \\
= & \left[\frac{I(U = t_2, \Delta = 1)}{\{1 - \lambda_1(\bar{X}_0)\}\lambda_2(\bar{X}_1)} + \frac{I(U \le t_2, \Delta = 0)}{K_U(\bar{X}_{U-1})}\right](Y - \mu_2) \\
= & \left[\frac{I(U = t_2, \Delta = 1)}{\{1 - \lambda_1(\bar{X}_0)\}\lambda_2(\bar{X}_1)} + I(U \le t_1, \Delta = 0) + \frac{I(t_1 < U \le t_2, \Delta = 0)}{1 - \lambda_1(\bar{X}_0)}\right](Y - \mu_2) \\
= & \left[\frac{I(U = t_2, \Delta = 1)}{1 - \lambda_1(\bar{X}_0)} + I(U \le t_1, \Delta = 0) + \frac{I(t_1 < U \le t_2, \Delta = 0)}{1 - \lambda_1(\bar{X}_0)}\right](Y - \mu_2).
\end{aligned}$$

The possible values of $W_{\bar{p}_1}(A_1, \bar{L}_0)$ and $W_{\bar{p}_2}(A_2, \bar{L}_1)$ are illustrated in detail in Table 4.3. Again, we start with definition of the conditional expectation of the outcome

Δ	U	$W_{\bar{p}_1}(A_1, \bar{L}_0)$	$W_{\bar{p}_2}(A_2, \bar{L}_1)$
0	$[t_0, t_1)$	$\frac{1}{1}$=1	$\frac{1}{1} \cdot \frac{1}{1}$=1
0	$[t_1, t_2)$	$\frac{1}{1}$=1	$\frac{1}{(1-\lambda_1(\bar{X}_0))} \cdot \frac{1}{1}$=$\frac{1}{(1-\lambda_1(\bar{X}_0))}$
1	t_1	$\frac{0}{\lambda_1(\bar{X}_0)}$= 0	$\frac{0}{\lambda_1(\bar{X}_0)} \cdot \frac{0}{1}$= 0
1	t_2	$\frac{1}{1-\lambda_1(\bar{X}_0)}$	$\frac{1}{1-\lambda_1(\bar{X}_0)} \cdot \frac{1}{\lambda_2(\bar{X}_1)}$ =$\frac{1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}$

Table 4.3: Table of values for weights when estimating μ_2.

given the entire treatment and covariate history as in (4.3),

$$g_2^{(2)}(\bar{A}_2, \bar{L}_1) = E(Y|\bar{A}_2, \bar{L}_1) = E(Y|U, \Delta, \bar{X}_U).$$

Using (4.4), we have

$$g_1^{(2)}(A_1, L_0) \quad = \quad E\left[\sum_{a_2} g_2^{(2)}(a_2, \bar{L}_1)p(a_2|S_1)\Big|A_1, L_0\right] \tag{4.9}$$

Simplification: the Endpoint Y Independent of a_t, $t = 1, \cdots, k$

For simplicity in notation, assume a first-order Markov model for all time-dependent covariates. If we imagine a model that depends only on the indicators $I(C > t_1)$, $I(C > t_2)$ time-independent covariate X_b and time-dependent covariates X_{t_0}, X_{t_1} as below:

Regime 1. Under the first regime,

$$Y = \alpha_0 + \alpha_1 I(C > t_1) + \alpha_b X_b + \alpha_t X_{t_0} + \epsilon. \tag{4.10}$$

The treatment could be either stopped by decision at t_1 or be terminated early prior to t_1. The conditional expectation of the outcome could be expressed as

- $E(Y|S_0 = 0, \bar{X}_0) = \alpha_0 + \alpha_b X_b + \alpha_t X_{t_0}$ when $U = C < t_1$

- $E(Y|S_0 = 0, \bar{X}_0) = \alpha_0 + \alpha_1 + \alpha_b X_b + \alpha_t X_{t_0}$ when $U = t_1, C > t_1$

Regime 2. Under the second regime,

$$Y = \alpha_0 + \alpha_1 I(C > t_1) + \alpha_2 I(C > t_2) + \alpha_b X_b + \alpha_t X_t + \epsilon. \tag{4.11}$$

where X_t is the most recently time-dependent covariate, i.e. either X_{t_0} or X_{t_1}.

The treatment could be stopped by decision at t_2, be terminated early prior to t_1 or be terminated early prior to t_2. The conditional expectation could be expressed as

- $E(Y|S_0 = 0, S_1 = 0, \bar{X}_0) = \alpha_0 + \alpha_b X_b + \alpha_t X_{t_0}$ when $U = C < t_1$
- $E(Y|S_0 = 1, S_1 = 0, \bar{X}_1) = \alpha_0 + \alpha_1 + \alpha_b X_b + \alpha_t X_{t_1}$ when $t_1 < U = C < t_2$
- $E(Y|S_0 = 1, S_1 = 0, \bar{X}_1) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_b X_b + \alpha_t X_{t_1}$ when $U = t_2 < C$

Assuming the (4.10) and (4.11) shares common parameters $\alpha_0, \alpha_1, \alpha_b, \alpha_t$. Then $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_b, \alpha_t)$ could be jointly modeled by

$$Y = \alpha_0 + \alpha_1 I(C > t_1) + \alpha_2 I(C > t_2) + \alpha_b X_b + \alpha_t X_t + \epsilon. \tag{4.12}$$

Using this model, we actually have four cases of conditional expectation listed below:

- $U = C < t_1$ $(S_0 = 0, S_1 = 0)$
- $U = t_1, C > t_1$ $(S_0 = 1, S_1 = 0)$
- $t_1 < U = C < t_2$ $(S_0 = 1, S_1 = 0)$
- $U = t_2 < C$ $(S_0 = 1, S_1 = 1)$

Based on (4.3) and (4.10), we have $g_1^{(1)}(S_0, \bar{X}_0) = E(Y|S_0, \bar{X}_0)$. It could be estimated from (4.12) by

$$g_1^{(1)}(\bar{L}_0; \boldsymbol{\alpha}) = \alpha_0 + \alpha_1 I(C > t_1) + \alpha_b X_b + \alpha_t X_{t_0}$$

Based on (4.3) and (4.11), we have $g_2^{(2)}(S_0, S_1, \bar{X}_1) = E(Y|S_0, S_1, \bar{X}_1)$. It could be estimated from (4.12) by

$$g_2^{(2)}(\bar{L}_1; \boldsymbol{\alpha}) = \alpha_0 + \alpha_1 I(C > t_1) + \alpha_2 I(C > t_2) + \alpha_b X_b + \alpha_t X_t$$

Based on (4.4) and (4.11), we have $g_1^{(2)}(A_1, \bar{X}_0) = E(\sum_{a_2} p_2(a_2|S_1) g_2^{(2)}(a_2, \bar{X}_1) \mid A_1, \bar{X}_0)$. Note individuals with $U = t_1$ is not following the regime 2, so $p_2(a_2|S_1) = 0$ for that case. Those individuals with observed $U = t_1$ will be treated as if they did not stop at t_1. So they could either stop at t_2 or censored between t_1 and t_2 based on

their respective covariate information. Then

$$
\begin{aligned}
g_1^{(2)}(\bar{L}_0;\boldsymbol{\alpha}) &= I(C<t_1)E(Y|S_0=S_1=0,\bar{X}_0) \\
&\quad +I(C>t_1,U\neq t_1)E\{I(t_1<C<t_2)E(Y|S_0=1,S_1=0,\bar{X}_1) \\
&\quad +I(t_2<C)E(Y|S_0=S_1=1,\bar{X}_1)\mid S_0,\bar{X}_0\} \\
&\quad +I(C>t_1,U=t_1) \\
&\quad \{P(C<t_2|C>t_1,\bar{X}_0,E(X_{t_1}|X_{t_0}))E(Y|S_0=1,S_1=0,\bar{X}_0,E(X_{t_1}|X_{t_0}) \\
&\quad +P(C>t_2|C>t_1,\bar{X}_0,E(X_{t_1}|X_{t_0}))E(Y|S_0=S_1=1,\bar{X}_0,E(X_{t_1}|X_{t_0}))\} \\
&= I(C<t_1)(\alpha_0+\alpha_b X_b+\alpha_t X_{t_0}) \\
&\quad +I(C>t_1,U\neq t_1)E\{I(t_1<C<t_2)(\alpha_0+\alpha_1+\alpha_b X_b+\alpha_t X_{t_1}) \\
&\quad +I(t_2<C)(\alpha_0+\alpha_1+\alpha_2+\alpha_b X_b+\alpha_t X_{t_1})\mid S_0,\bar{X}_0\} \\
&\quad +I(C>t_1,U=t_1) \\
&\quad \{P(C<t_2|C>t_1,\bar{X}_0,E(X_{t_1}|X_{t_0}))(\alpha_0+\alpha_1+\alpha_b X_b+\alpha_t E(X_{t_1}|X_{t_0})) \\
&\quad +P(C>t_2|C>t_1,\bar{X}_0,E(X_{t_1}|X_{t_0}))(\alpha_0+\alpha_1+\alpha_2+\alpha_b X_b \\
&\quad +\alpha_t E(X_{t_1}|X_{t_0})\} \\
&= \alpha_0+\alpha_1 I(C>t_1)+\alpha_2 I(C>t_1)P(C>t_2|C>t_1,\bar{X}_0,E(X_{t_1}|X_{t_0})) \\
&\quad +\alpha_b X_b+\alpha_t(X_{t_0}I(C<t_1)+E(X_{t_1}|X_{t_0})I(C>t_1))
\end{aligned}
$$

Note that $g_1^{(2)}$ is the projection of Y at t_2 based on the information immediately after t_1, i.e. we have

$$g_1^{(2)}(\bar{L_0}) = E\{g_2^{(2)}(\bar{L_1})|\bar{L}_0\}$$

To model the probability of being interrupted by the treatment terminating event. We use the discrete hazard model for treatment-terminating process as:

$$f_m^C(\bar{X}_{m-1}) = \prod_{j=1}^{m-1}\left\{1-\lambda_j^C(\bar{X}_{j-1})\right\}\lambda_m^C(\bar{X}_{m-1}),$$

with cause-specific hazard function

$$\lambda_j^C(\bar{X}_{j-1}) = P(t_{j-1} < C < t_j, \Delta = 0 | U \geq t_{j-1}, \bar{X}_{j-1}).$$

Using notation for discrete hazards models and the Markov assumption, we write

$$\begin{aligned} P(C > t_2 | C > t_1, \bar{L}_1) &= 1 - \lambda_2^C(\bar{X}_1) \\ P(C > t_2 | C > t_1, \bar{L}_0) &= 1 - \lambda_2^C(\bar{X}_0, E(X_{t_1} | X_{t_0})) \\ P(t_1 < C \leq t_2 | C > t_1, \bar{L}_1) &= \lambda_2^C(\bar{X}_1) \\ P(U = t_1 | C > t_1, \bar{L}_0) &= \lambda_1(\bar{X}_0) \end{aligned}$$

Thus, all the above g functions could be estimated.

After estimating all the g functions, we are able to estimate our doubly robust estimators as well as their competitors. For the two-stage design, the estimating equation (4.6) will be interpreted as
for $m = 1$:

$$W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - \mu_1) - W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(g_1^{(1)}(\bar{L}_0) - \mu_1) + (g_1^{(1)}(\bar{L}_0) - \mu_1) = 0;$$

for $m = 2$:

$$\{W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1)(Y - \mu_2) - W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(g_1^{(2)}(\bar{L}_0) - \mu_2) - W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1)(g_2^{(2)}(\bar{L}_1) - \mu_2) + (g_1^{(2)}(\bar{L}_0) - \mu_2) + W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(g_2^{(2)}(\bar{L}_1) - \mu_2)\} = 0$$

So our final *doubly robust (DR) estimators* are

$$\begin{aligned}
\mu_{DR,1} &= \mathbb{P}_n W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0)) + g_1^{(1)}(\bar{L}_0) \qquad (4.13)\\
\mu_{DR,2} &= \mathbb{P}_n\{W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1)(Y - g_2^{(2)}(\bar{L}_1)) - W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(g_1^{(2)}(\bar{L}_0) - g_2^{(2)}(\bar{L}_1)) + g_1^{(2)}(\bar{L}_0)\}
\end{aligned}$$

Note that here $W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1) = W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)$, $\mu_{DR,2}$ could be further reduced to

$$\mathbb{P}_n\{W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - g_1^{(2)}(\bar{L}_0)) + g_1^{(2)}(\bar{L}_0)\}$$

.

The *inverse probability weighting (IPW) estimators* as mentioned in (4.1) are

$$\begin{aligned}
\hat{\mu}_{IPW,1} &= \frac{\mathbb{P}_n W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0) Y}{\mathbb{P}_n W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)}\\
\hat{\mu}_{IPW,2} &= \frac{\mathbb{P}_n W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1) Y}{\mathbb{P}_n W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1)}
\end{aligned}$$

As shown in Murphy et al. (2001), the causal estimand can be written in terms of g functions as:

$$\begin{aligned}
\mu_1 &= E\left[\sum_{a_1} g_1^{(1)}(a_1, \bar{L}_0) p(a_1|S_0)\right]\\
&= E\left[g_1^{(1)}(1, S_0 = 1, \bar{X}_0) I\{S_0 = 1\} + g_1^{(1)}(., S_0 = 0, \bar{X}_0) I\{S_0 = 0\}\right],\\
&= \alpha_0 + \alpha_1 I(C > t_1) + \alpha_b X_b + \alpha_t X_{t_0}\\
\mu_2 &= E\left[\sum_{a_2} g_1^{(2)}(a_1, L_0) p(a_1|S_0)\right]\\
&= E\left[g_1^{(2)}(0, L_0) I\{S_0 = 1\} + g_1^{(2)}(., L_0) I\{S_0 = 0\}\right]\\
&= \alpha_0 + \alpha_1 I(C > t_1) + \alpha_2 I(C > t_1) P(C > t_2 | C > t_1, \bar{X}_0) + \alpha_b X_b\\
&\quad + \alpha_t (X_{t_0} I(C < t_1) + E(X_{t_1}|X_{t_0}) I(C > t_1))
\end{aligned}$$

Therefore, we also have *outcome regression(OR) estimators*:

$$\begin{aligned}\hat{\mu}_{OR,1} &= \mathbb{P}_n g_1^{(1)}(\bar{L}_0) \\ \hat{\mu}_{OR,2} &= \mathbb{P}_n g_1^{(2)}(\bar{L}_0)\end{aligned}$$

4.2.4 Improved doubly-robust estimators under the simplified scenario

Based on the asymptotic properties from Murphy et al. (2001), our doubly robust estimator has the merit of maintaining consistency when either OR or PS model is correctly specified. When both model are correctly specified, its variance is minimized. However, when miss-specification occurs to either model, its variance may not be minimized any longer, especially when this miss-specification occurs to the OR model. To minimize the variance of our doubly-robust estimator, especially when OR model may be miss-specified, we adopted similar approach as Cao et al. (2009); Tsiatis et al. (2011) did by finding appropriate $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_b, \alpha_t)$ that satisfy i) the criteria for doubly-robustness and ii) the criteria of variance minimization as long as the PS model is correct simultaneously.

Regime 1

To minimize the variance of $\mu_{DR,1}$ as in (4.13), we want to minimize the quantity

$$Var\left[W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})\right].$$

As $E\{W_{\bar{p}_1}(A_1, \bar{L}_0)\} = 1$, we have

$$
\begin{aligned}
& E\left[W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] \\
= \; & E\{W_{\bar{p}_1}(A_1, \bar{L}_0)\} E\left[Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) \\
= \; & Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) = Y
\end{aligned}
$$

We also have

$$
\begin{aligned}
& Var\left[W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] \\
= \; & Var(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0))(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}))^2
\end{aligned}
$$

Consequently, based on $Var(\cdot) = E\{Var(\cdot|X, Y)\} + Var\{E(\cdot|X, Y)\}$,

$$
\begin{aligned}
& Var\left[W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})\right] \\
= \; & E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}))^2\right] + Var(Y)
\end{aligned}
$$

It is then equivalent to minimize $E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}))^2\right]$. We can achieve this by solving the equation:

$$
E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}))\frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})\right] = 0 \tag{4.14}
$$

On the other hand, the estimating equation to solve for $\boldsymbol{\alpha}$ in outcome regression models is

$$
\mathbb{P}_n I(C > t_1)\left[(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}))\frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})\right] = 0 \tag{4.15}
$$

Thus, an estimating equation to satisfy equation (4.14) and (4.15) simultaneously is

$$\mathbb{P}_n \frac{I(C > t_1)}{Pr(C > t_1|\bar{L}_0)} Var(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)[(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) \frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) = 0 \quad (4.16)$$

In this setting at $m = 1$, $W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0) = I(S_0 = 0) + I(U = t_1)\frac{1}{\lambda_1(\bar{X}_0)}$. Therefore,

$$\begin{aligned} Var\left[W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)\right] &= E\{W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)^2\} - E(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0))^2 \\ &= (1 - Pr(C > t_1|\bar{L}_0)) \cdot 1^2 + Pr(C > t_1|\bar{L}_0)\lambda_1(\bar{X}_0) \cdot \frac{1}{\lambda_1(\bar{X}_0)^2} - 1^2 \\ &= Pr(C > t_1|\bar{L}_0)\frac{1 - \lambda_1(\bar{X}_0)}{\lambda_1(\bar{X}_0)} \end{aligned}$$

Thus (4.16) could be further expressed as

$$\mathbf{P}_n I(C > t_1)\frac{1 - \lambda_1(\bar{X}_0)}{\lambda_1(\bar{X}_0)}[(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) \frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) = 0.$$

When the propensity score is correct but the outcome regression is or is not, the left hand side of (4.16) converges in probability to an expression of the form (4.14). Thus, the minimum variance is achieved even when outcome regression may be miss-specified. When the outcome regression is correct and the propensity score is not but has no unknown parameters, the left hand side of (4.16) converges to $E\left[(1 - \lambda_{1,0}^C(\bar{X}_0))\frac{1-\lambda_1(\bar{X}_0)}{\lambda_1(\bar{X}_0)}[(g_{1,0}^{(1)}(\bar{L}_0) - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) \frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})\right]$, which equals zero when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, so that $\hat{\boldsymbol{\alpha}}$ converges in probability to $\boldsymbol{\alpha}_0$. Note that here $\lambda_1(\bar{X}_0) = \lambda_{1,0}(\bar{X}_0)$ when propensity score model is correct and $\lambda_1^C(\bar{X}_0) = \lambda_{1,0}^C(\bar{X}_0)$ when outcome regression model is correct. $g_{1,0}^{(1)}(\bar{L}_0)$ denote the true outcome regression $E(Y|\bar{L}_0)$.

In practice, the model for the propensity score model is always parametrized. In order to maintain the above properties, especially when propensity score model is

miss-specified, an extra piece of the parameters in the propensity score model, say $\boldsymbol{\gamma}$ in $\lambda_1(\bar{X}_0, \boldsymbol{\gamma})$ need to be controlled. The steps needed is illustrated below:

Our likelihood for targeted treatment stop process is

$L = \prod \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)^{I(U=t_1)}(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))^{I(U>t_1)}$, where the discrete hazard function at t_1 is $\lambda_1(\boldsymbol{\gamma}, \bar{X}_0) = (1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}})^{-1}$.

Then we have the log likelihood

$$log(L) = \sum^{n} I(U=t_1) log\lambda_1(\boldsymbol{\gamma}, \bar{X}_0) + I(U>t_1) log(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))$$

and the score function

$$\begin{aligned} S_{\boldsymbol{\gamma}} &= \sum^{n} \left[\frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}} - \frac{I(U>t_1)}{1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}} \right] \\ &= \sum^{n} \left[\frac{I(U=t_1) - I(U=t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0) - I(U>t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))} \right] \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}} \\ &= \sum^{n} \left[\frac{I(U=t_1) - I(U \geq t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))} \right] \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}} \end{aligned}$$

Because

$$\begin{aligned} W_{\bar{p}_1}(A_1, \bar{L}_0) - 1 &= I(U<t_1) + \frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - 1 \\ &= \frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - (1 - I(U<t_1)) \\ &= \frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - I(U \geq t_1) \\ &= \frac{I(U=t_1) - I(U \geq t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \end{aligned}$$

We have

$$S_{\boldsymbol{\gamma}} = \frac{W_{\bar{p}_1}(A_1, \bar{L}_0) - 1}{1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \cdot \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}}$$

The influence function corresponding to estimator of form

$$W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) - \mu_1$$

when $\lambda_1(\bar{X}_0, \boldsymbol{\gamma})$ is correctly specified (i.e. $\lambda_1(\bar{X}_0, \boldsymbol{\gamma}_0) = \lambda_{1,0}(\bar{X}_0)$ for some $\boldsymbol{\gamma}_0$), while $g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha})$ may be miss-specified with $\hat{\boldsymbol{\alpha}}$ converge in probability to some $\boldsymbol{\alpha}^*$, have form

$$W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1)g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}^*) - \Gamma_0^T(\boldsymbol{\alpha}^*) \sum_{\gamma\gamma,0}^{-1} S_{\boldsymbol{\gamma}}(\bar{L}_0, \boldsymbol{\gamma}_0) - \mu_1$$

$$= W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1) \left[g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}^*) + \frac{\Gamma_0^T(\boldsymbol{\alpha}^*) \sum_{\gamma\gamma,0}^{-1}}{1 - \lambda_{1,0}(\bar{X}_0)} \cdot \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \right] - \mu_1$$

where

$$\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} = \frac{\partial \lambda_1(\bar{X}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}$$

$$\Gamma_0^T(\boldsymbol{\alpha}^*) = E \left[\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} (g_{1,0}^{(1)}(L_0) - g_1^{(1)}(L_0, \boldsymbol{\alpha}^*)) / \lambda_{1,0}(\bar{X}_0) \right]$$

$$\sum_{\gamma\gamma,0} = E \left[\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}}^T / \{\lambda_{1,0}(\bar{X}_0)(1 - \lambda_{1,0}(\bar{X}_0))\} \right]$$

To find $\hat{\boldsymbol{\alpha}}$ that converging to $\boldsymbol{\alpha}_{opt}^{**}$, which minimizes the variance of μ_1, we consider the following influence function

$$\begin{aligned} & W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1)g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}^*) - \mathbf{c}^{*T} S_{\boldsymbol{\gamma}}(\bar{L}_0, \boldsymbol{\gamma}_0) - \mu_1 \\ = \; & W_{\bar{p}_1}(A_1, \bar{L}_0)Y \\ & -(W_{\bar{p}_1}(A_1, \bar{L}_0) - 1) \left[g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}^*) + \frac{\mathbf{c}^{*T}}{1 - \lambda_{1,0}(\bar{X}_0)} \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \right] - \mu_1 \end{aligned} \quad (4.17)$$

for arbitrary $(\boldsymbol{\alpha}^*, \mathbf{c}^*)$. Taking the expression in braces above as a function of $(\boldsymbol{\alpha}^*, \mathbf{c}^*)$,

a solution $(\boldsymbol{\alpha}^{**}_{opt}, \mathbf{c}^{**}_{opt})$ to the following equation

$$E\left[Var(W_{\bar{p}_1}(A_1,\bar{L}_0))\{Y-g_1^{(1)}(L_0,\boldsymbol{\alpha}^*)-\mathbf{c}^{*T}\frac{\frac{\partial\lambda_{1,0}(\bar{X}_0)}{\partial\boldsymbol{\gamma}}}{1-\lambda_{1,0}(\bar{X}_0)}\}\left\{\begin{array}{c}\frac{\partial g_1^{(1)}(L_0,\boldsymbol{\alpha}^*)}{\partial\boldsymbol{\alpha}}\\ \frac{1}{1-\lambda_{1,0}(\bar{X}_0)}\cdot\frac{\partial\lambda_{1,0}(\bar{X}_0)}{\partial\boldsymbol{\gamma}}\end{array}\right\}\right]=0 \quad (4.18)$$

minimizes the variance of (4.17).

By doing an analogy to minimizing the variance of μ_1 when the discrete hazard function of treatment stoppage process is fully specified, the recommendation for estimating $\boldsymbol{\alpha}$ is to solve a weighted version of (4.18) as below:

$$\sum_{i=1}^{n}\{I(C_i>t_1)(\frac{1-\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}{\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})})\{Y_i-g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})-\mathbf{c}^T\frac{\frac{\partial\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}{\partial\boldsymbol{\gamma}}}{1-\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}\}\left\{\begin{array}{c}\frac{\partial g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\\ \frac{\frac{\partial\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}{\partial\boldsymbol{\gamma}}}{1-\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}\end{array}\right\}\}=0 \quad (4.19)$$

for $\boldsymbol{\alpha}$ and $\mathbf{c}$ simultaneously.

When the discrete hazard model for treatment stoppage process is correct and the outcome regression model may or may not be correct, the solution to (4.19), i.e.$\hat{\boldsymbol{\alpha}}$, converges in probability to $\boldsymbol{\alpha}^{**}_{opt}$. When outcome regression is correct but $\lambda_1(\bar{X}_0,\boldsymbol{\gamma})$ is not, assuming that $\hat{\boldsymbol{\gamma}}$ converges in probability to some $\boldsymbol{\gamma}^*$, the quantity to which the left hand side of (4.19) converges in probability equals zero when $(\boldsymbol{\alpha},\mathbf{c})=(\boldsymbol{\alpha}_0,0)$. Therefore, the solution to (4.19) yields an estimator for μ_1 that is not only doubly robust, but also achieves minimum asymptotic variance when the propensity model is correct.

To find the solution to (4.19) for $\boldsymbol{\alpha}$ and $\mathbf{c}$ simultaneously, we implemented Newton-Ralphson algorithm to solve the system of equations by the following steps:
Let

$$\begin{aligned}
S_a &= \sum_{i=1}^{n}\left[I(C_{,i}>t_1)(\frac{1-\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})})\{Y_i-g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})-\mathbf{c}^T\frac{\frac{\partial\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}}}{1-\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}\}\frac{\partial g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right]\\
S_c &= \sum_{i=1}^{n}\left[I(C_{,i}>t_1)(\frac{1}{\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})})\{Y_i-g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})-\mathbf{c}^T\frac{\frac{\partial\lambda_1(\hat{\boldsymbol{\gamma}},\bar{X}_{0,i})}{\partial\boldsymbol{\gamma}}}{1-\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}\}\frac{\partial\lambda_1(\bar{X}_{0,i}\hat{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}}\right],
\end{aligned}$$

$$
\begin{aligned}
-I_{aa} &= -\frac{\partial S_a}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^{n}\left[I(C_{,i}>t_1)\frac{1-\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}\cdot\frac{\partial g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\times\frac{\partial g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}^T\right]\\
-I_{ac} &= -\frac{\partial S_a}{\partial \mathbf{c}} = \sum_{i=1}^{n}\left[I(C_{,i}>t_1)\frac{1}{\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}\cdot\frac{\partial \lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}}\times\frac{\partial g_1^{(1)}(\bar{L}_{0,i},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}^T\right]\\
-I_{cc} &= -\frac{\partial S_c}{\partial \mathbf{c}} = \sum_{i=1}^{n}\left[I(C_{,i}>t_1)\frac{1}{\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})(1-\lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}}))}\cdot\frac{\partial \lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}}\times\frac{\partial \lambda_1(\bar{X}_{0,i},\hat{\boldsymbol{\gamma}})}{\partial\boldsymbol{\gamma}}^T\right]
\end{aligned}
$$

Then go through iterations by

$$
\begin{pmatrix}\boldsymbol{\alpha}\\ \mathbf{c}\end{pmatrix}_{j+1} = \begin{pmatrix}\boldsymbol{\alpha}\\ \mathbf{c}\end{pmatrix}_{j} + \begin{pmatrix}-I_{aa}, -I_{ac}\\ -I_{ac}^T, -I_{cc}\end{pmatrix}^{-1}\times\begin{pmatrix}S_a\\ S_c\end{pmatrix}
$$

$$
\text{until } \left|\begin{pmatrix}\boldsymbol{\alpha}\\ \mathbf{c}\end{pmatrix}_{j+1} - \begin{pmatrix}\boldsymbol{\alpha}\\ \mathbf{c}\end{pmatrix}_{j}\right| < 0.001
$$

Note that

$$
\begin{aligned}
\frac{\partial\lambda_1(\boldsymbol{\gamma},L_0)}{\partial\boldsymbol{\gamma}} &= \frac{\partial}{\partial\boldsymbol{\gamma}}\left(\frac{1}{1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}}}\right) = \frac{e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}}}{(1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}})^2}\cdot(1,X_b,X_{t_0})\\
\frac{\partial g_1^{(1)}(L_0,\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}} &= (1, I(C>t_1), X_b, X_{t_0})
\end{aligned}
$$

Regime 2

To minimize the variance of $\mu_{DR,2}$ as in (4.13), we want to minimize the quantity

$$
Var\left[W_{\bar{p}_1}(\bar{A}_1,\bar{L}_0)(Y-g_1^{(2)}(\bar{L}_0,\boldsymbol{\alpha}))+g_1^{(2)}(\bar{L}_0,\boldsymbol{\alpha})\right].
$$

As $E\{W_{\bar{p}_1}(A_1,\bar{L}_0)\}=1$, we have

$$
\begin{aligned}
& E\left[W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] \\
= \; & E\{W_{\bar{p}_1}(A_1, \bar{L}_0)\} E\left[Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) \\
= \; & Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) = Y
\end{aligned}
$$

We also have

$$
\begin{aligned}
& Var\left[W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) \mid X_0, X_{t_0}, Y\right] \\
= \; & Var(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0))(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))^2
\end{aligned}
$$

Consequently, based on $Var(\cdot) = E\{Var(\cdot|X, Y)\} + Var\{E(\cdot|X, Y)\}$,

$$
\begin{aligned}
& Var\left[W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})\right] \\
= \; & E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))^2\right] + Var(Y)
\end{aligned}
$$

It is then equivalent to minimize $E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))^2\right]$. We can achieve this by solving the equation:

$$
E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))\frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})\right] = 0 \tag{4.20}
$$

On the other hand, the estimating equation to solve for $\boldsymbol{\alpha}$ in outcome regression models is

$$
\mathbb{P}_n I(C > t_1)\left[(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))\frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})\right] = 0 \tag{4.21}
$$

Thus, an estimating equation to satisfy equation (4.20) and (4.21) simultaneously is

$$\mathbb{P}_n \frac{I(C > t_1)}{Pr(C > t_1|\bar{L}_0)} Var(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)[(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) \frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) = 0 \quad (4.22)$$

In this setting at $m = 2$, $W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0) = I(S_0 = 0) + I(U > t_1)\frac{1}{1-\lambda_1(\bar{X}_0)}$. Therefore,

$$\begin{aligned} Var\left[W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)\right] &= E\{W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)^2\} - E(W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0))^2 \\ &= (1 - Pr(C > t_1|\bar{L}_0)) \cdot 1^2 \\ &\quad + Pr(C > t_1|\bar{L}_0)(1 - \lambda_1(\bar{X}_0)) \cdot \frac{1}{(1 - \lambda_1(\bar{X}_0))^2} - 1^2 \\ &= Pr(C > t_1|\bar{L}_0) \frac{\lambda_1(\bar{X}_0)}{1 - \lambda_1(\bar{X}_0)} \end{aligned}$$

Thus (4.22) could be further expressed as

$$\mathbb{P}_n I(C > t_1) \frac{\lambda_1(\bar{X}_0)}{1 - \lambda_1(\bar{X}_0)} [(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) \frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(1)}(\bar{L}_0, \boldsymbol{\alpha}) = 0.$$

When the propensity score is correct but the outcome regression is or is not, the left hand side of (4.22) converges in probability to an expression of the form (4.20). Thus, the minimum variance is achieved even when outcome regression may be miss-specified. When the outcome regression is correct and the propensity score is not but has no unknown parameters, the left hand side of (4.22) converges to $E\left[(1 - \lambda_{1,0}^C(\bar{X}_0))\frac{\lambda_1(\bar{X}_0)}{1-\lambda_1(\bar{X}_0)}[(g_{1,0}^{(2)}(\bar{L}_0) - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}))\frac{\partial}{\partial \boldsymbol{\alpha}} g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})\right]$, which equals zero when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, so that $\hat{\boldsymbol{\alpha}}$ converges in probability to $\boldsymbol{\alpha}_0$. Note that here $\lambda_1(\bar{X}_0) = \lambda_{1,0}(\bar{X}_0)$ when propensity score model is correct and $\lambda_1^C(\bar{X}_0) = \lambda_{1,0}^C(\bar{X}_0)$ when outcome regression model is correct. $g_{1,0}^{(2)}(\bar{L}_0)$ denote the true outcome regression

$$E\{p(a_1|S_0)E(Y|\bar{A}_2, \bar{L}_1)\bar{L}_0\}$$

.

In practice, the model for the propensity score model is always parametrized. In order to maintain the above properties, especially when propensity score model is miss-specified, an extra piece of the parameters in the propensity score model, say $\boldsymbol{\gamma}$ in $\lambda_1(\bar{X}_0, \boldsymbol{\gamma})$ need to be controlled. The steps needed is illustrated below:

Our likelihood for targeted treatment stop process is

$L = \prod^n \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)^{I(U=t_1)}(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))^{I(t_1<U<t_2)}\{(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))\lambda_2(\boldsymbol{\gamma}, \bar{X}_1)\}^{I(U=t_2)}$, where the discrete hazard function at t_1 is $\lambda_1(\boldsymbol{\gamma}, \bar{X}_0) = (1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}})^{-1}$ and at t_2 is $\lambda_2(\boldsymbol{\gamma}, \bar{X}_1) = 1$.

Then we have the log likelihood
$log(L) = \sum^n I(U=t_1)log(\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)) + I(U>t_1)log(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))$ and the score function

$$\begin{aligned}
S_\gamma &= \sum^n \left[\frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}\frac{\partial\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial\boldsymbol{\gamma}} - \frac{I(U>t_1)}{1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}\frac{\partial\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial\boldsymbol{\gamma}}\right] \\
&= \sum^n \left[\frac{I(U=t_1)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - \frac{I(U>t_1)}{1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}\right]\frac{\partial\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial\boldsymbol{\gamma}} \\
&= \sum^n \left[\frac{I(U=t_1)(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)) - I(U>t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))}\right]\frac{\partial\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial\boldsymbol{\gamma}} \\
&= \sum^n \left[\frac{I(U=t_1) - I(U\geq t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)(1-\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))}\right]\frac{\partial\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial\boldsymbol{\gamma}}
\end{aligned}$$

Because

$$\begin{aligned} W_{\bar{p}_1}(A_1, \bar{L}_0) - 1 &= I(U < t_1) + \frac{I(U > t_1)}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - 1 \\ &= \frac{I(U > t_1)}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - (1 - I(U < t_1)) \\ &= \frac{I(U > t_1)}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} - I(U \geq t_1) \\ &= \frac{I(U > t_1) - I(U \geq t_1)(1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0))}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \\ &= \frac{I(U > t_1) - I(U \geq t_1) + I(U \geq t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \\ &= \frac{-I(U = t_1) + I(U \geq t_1)\lambda_1(\boldsymbol{\gamma}, \bar{X}_0))}{1 - \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \end{aligned}$$

Thus,

$$S_{\boldsymbol{\gamma}} = -\frac{W_{\bar{p}_1}(A_1, \bar{L}_0) - 1}{\lambda_1(\boldsymbol{\gamma}, \bar{X}_0)} \cdot \frac{\partial \lambda_1(\boldsymbol{\gamma}, \bar{X}_0)}{\partial \boldsymbol{\gamma}}$$

The influence function corresponding to estimator of form

$$\{W_{\bar{p}_1}(A_1, \bar{L}_0)(Y - g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})) + g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}) - \mu_2\}$$

when $\lambda_1(\bar{X}_0, \boldsymbol{\gamma})$ is correctly specified (i.e. $\lambda_1(\bar{X}_0, \boldsymbol{\gamma}_0) = \lambda_{1,0}(\bar{X}_0)$ for some $\boldsymbol{\gamma}_0$), while $g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha})$ may be miss-specified with $\hat{\boldsymbol{\alpha}}$ converge in probability to some $\boldsymbol{\alpha}^*$, have form

$$W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1)g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}^*) - \Gamma_0^T(\boldsymbol{\alpha}^*) \sum_{\gamma\gamma,0}^{-1} S_{\boldsymbol{\gamma}}(\bar{L}_0, \boldsymbol{\gamma}_0) - \mu_2$$

$$= W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1) \left[g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}^*) + \frac{\Gamma_0^T(\boldsymbol{\alpha}^*) \sum_{\gamma\gamma,0}^{-1}}{\lambda_{1,0}(\bar{X}_0)} \cdot \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \right] - \mu_2$$

where

$$\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} = \frac{\partial \lambda_1(\bar{X}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}$$
$$\Gamma_0^T(\boldsymbol{\alpha}^*) = E\left[-\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}}(g_{1,0}^{(2)}(L_0) - g_1^{(2)}(L_0, \boldsymbol{\alpha}^*))/(1 - \lambda_{1,0}(\bar{X}_0))\right]$$
$$\sum_{\gamma\gamma,0} = E\left[\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}}^T / \{\lambda_{1,0}(\bar{X}_0)(1 - \lambda_{1,0}(\bar{X}_0))\}\right]$$

To find $\hat{\boldsymbol{\alpha}}$ that converging to $\boldsymbol{\alpha}_{opt}^{**}$, which minimizes the variance of μ_2, we consider the following influence function

$$\{W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1)g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}^*) - \mathbf{c}^{*T} S_{\boldsymbol{\gamma}}(\bar{L}_0, \boldsymbol{\gamma}_0) - \mu_2\}$$
$$= \{W_{\bar{p}_1}(A_1, \bar{L}_0)Y - (W_{\bar{p}_1}(A_1, \bar{L}_0) - 1)\left[g_1^{(2)}(\bar{L}_0, \boldsymbol{\alpha}^*) + \frac{\mathbf{c}^{*T}}{\lambda_{1,0}(\bar{X}_0)} \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}}\right] - \mu_2\} \quad (4.23)$$

for arbitrary $(\boldsymbol{\alpha}^*, \mathbf{c}^*)$. Taking the expression in braces above as a function of $(\boldsymbol{\alpha}^*, \mathbf{c}^*)$, a solution $(\boldsymbol{\alpha}_{opt}^{**}, \mathbf{c}_{opt}^{**})$ to the following equation

$$E\left[Var(W_{\bar{p}_1}(A_1, \bar{L}_0))\{Y - g_1^{(2)}(L_0, \boldsymbol{\alpha}^*) - \mathbf{c}^{*T} \frac{\frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}}}{\lambda_{1,0}(\bar{X}_0)}\} \begin{Bmatrix} \frac{\partial g_1^{(2)}(L_0, \boldsymbol{\alpha}^*)}{\partial \boldsymbol{\alpha}} \\ \frac{1}{\lambda_{1,0}(\bar{X}_0)} \cdot \frac{\partial \lambda_{1,0}(\bar{X}_0)}{\partial \boldsymbol{\gamma}} \end{Bmatrix}\right] = 0 \quad (4.24)$$

minimizes the variance of (4.23).

By doing an analogy to minimizing the variance of μ_1 when the discrete hazard function of treatment stoppage process is fully specified, the recommendation for estimating $\boldsymbol{\alpha}$ is to solve a weighted version of (4.24) as below:

$$\sum_{i=1}^{n}\left[I(C_{,i} > t_1)(\frac{\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{1 - \lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})})\{Y_i - g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha}) - \mathbf{c}^T \frac{\frac{\partial \lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{\partial \boldsymbol{\gamma}}}{\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}\} \begin{Bmatrix} \frac{\partial g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ \frac{\frac{\partial \lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{\partial \boldsymbol{\gamma}}}{\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})} \end{Bmatrix}\right] = 0 \quad (4.25)$$

for $\boldsymbol{\alpha}$ and $\mathbf{c}$ simultaneously.

When the discrete hazard model for treatment stoppage process is correct and the outcome regression model may or may not be correct, the solution to (4.25), i.e.$\hat{\boldsymbol{\alpha}}$, converges in probability to $\boldsymbol{\alpha}_{opt}^{**}$. When outcome regression is correct but $\lambda_1(\bar{X}_0, \boldsymbol{\gamma})$ is

not, assuming that $\hat{\boldsymbol{\gamma}}$ converges in probability to some $\boldsymbol{\gamma}^*$, the quantity to which the left hand side of (4.25) converges in probability equals zero when $(\boldsymbol{\alpha}, \mathbf{c}) = (\boldsymbol{\alpha}_0, 0)$. Therefore, the solution to (4.25) yields an estimator for μ_2 that is not only doubly robust, but also achieves minimum asymptotic variance when the propensity model is correct.

To find the solution to (4.25) for $\boldsymbol{\alpha}$ and $\mathbf{c}$ simultaneously, we implemented Newton-Ralphson algorithm to solve the system of equations by the following steps:
Let

$$
\begin{aligned}
S_a &= \sum_{i=1}^{n} \left[I(C_{,i} > t_1)\left(\frac{\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{1-\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}\right)\{Y_i - g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha}) - \mathbf{c}^T \frac{\frac{\partial \lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}}}{\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}\} \frac{\partial g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right] \\
S_c &= \sum_{i=1}^{n} \left[I(C_{,i} > t_1)\left(\frac{1}{1-\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}\right)\{Y_i - g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha}) - \mathbf{c}^T \frac{\frac{\partial \lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{\partial \boldsymbol{\gamma}}}{\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}\} \frac{\partial \lambda_1(\bar{X}_{0,i} \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}} \right],
\end{aligned}
$$

$$
\begin{aligned}
-I_{aa} &= -\frac{\partial S_a}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^{n} \left[I(C_{,i} > t_1)\left(\frac{\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}{1-\lambda_1(\hat{\boldsymbol{\gamma}}, \bar{X}_{0,i})}\right) \cdot \frac{\partial g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \times \frac{\partial g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^T \right] \\
-I_{ac} &= -\frac{\partial S_a}{\partial \mathbf{c}} = \sum_{i=1}^{n} \left[I(C_{,i} > t_1)\frac{1}{1-\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})} \cdot \frac{\partial \lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}} \times \frac{\partial g_1^{(2)}(\bar{L}_{0,i}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^T \right] \\
-I_{cc} &= -\frac{\partial S_c}{\partial \mathbf{c}} = \sum_{i=1}^{n} \left[I(C_{,i} > t_1)\frac{1}{\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})(1-\lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}}))} \cdot \frac{\partial \lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}} \times \frac{\partial \lambda_1(\bar{X}_{0,i}, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}}^T \right]
\end{aligned}
$$

Then go through iterations by

$$
\begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{c} \end{pmatrix}_{j+1} = \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{c} \end{pmatrix}_{j} + \begin{pmatrix} -I_{aa}, -I_{ac} \\ -I_{ac}^T, -I_{cc} \end{pmatrix}^{-1} \times \begin{pmatrix} S_a \\ S_c \end{pmatrix}
$$

$$
\text{until} \left| \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{c} \end{pmatrix}_{j+1} - \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{c} \end{pmatrix}_{j} \right| < 0.001
$$

Note that

$$\frac{\partial \lambda_1(\boldsymbol{\gamma}, L_0)}{\partial \boldsymbol{\gamma}} = \frac{\partial}{\partial \boldsymbol{\gamma}}\left(\frac{1}{1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}}}\right) = \frac{e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}}}{(1+e^{-\gamma_0-\gamma_b X_b-\gamma_t X_{t_0}})^2}\cdot(1, X_b, X_{t_0})$$

$$\frac{\partial g_1^{(2)}(L_0, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = (1, I(C>t_1), I(C>t_1)Pr(C>t_2 \mid C>t_1, \bar{X}_0), X_b, X_{t_{last}})$$

where $X_{t_{last}} = X_{t_0}$ when $C < t_1$ and $X_{t_{last}} = E(X_{t_1} \mid X_{t_0})$ when $C > t_1$

4.3 Simulation Studies

4.3.1 Data Simulation

Simulation studies were conducted to assess the performance our proposed method. We take $K = 2$ time points $(t_1,\ t_2)$ and constructed a simulated dataset as follows:

1. Let $\boldsymbol{\gamma}^C = (\gamma_1^C, \gamma_2^C, \boldsymbol{\gamma}_b^C, \gamma_t^C) = (\gamma_1^C, \gamma_2^C, \gamma_{b1}^C, \gamma_{b2}^C, \gamma_t^C) = (\frac{1.3}{0.7}, \frac{2.6}{0.7}, -0.1, -0.1, -0.9)$; $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \boldsymbol{\gamma}_b, \gamma_t) = (\gamma_1, \gamma_2, \gamma_{b1}, \gamma_{b2}, \gamma_t) = (\frac{1}{0.7}, \frac{2}{0.7}, -0.2, -0.2, -0.8)$.

2. Simulate three independent normal random variables: $X_{b1} \sim N(0,1)$, $X_{b2} \sim N(0,1)$ and time-dependent effect X_t at baseline $X_{t_0} \sim N(4,1)$. Define $\mathbf{X}_b = (X_{b1}, X_{b2})$.

3. Generate C_1, a potential treatment terminating event among $[0, t_1)$ as a binomial random variable with $p = \lambda_1^C(\bar{X}_0, \boldsymbol{\gamma}^C) = \frac{1}{1+e^{-(\gamma_1^C+\gamma_b^C \mathbf{X}_b+\gamma_t^C X_{t_0})}}$. If $C_1 = 1$ then $U = t_1/2$, $\Delta = 0$. All the information afterward except the outcome are set to missing. (i.e. X_{t_1}, C_2, A_1, A_2) And the continuous outcome Y is generated as

$$Y = \alpha_0 + \alpha_{b1} X_{b1} + \alpha_{b2} X_{b2} + \alpha_t X_t + \epsilon,$$

where $\alpha_0 = 1.5$, $\alpha_{b1} = 1$, $\alpha_{b2} = 0.5$, $\alpha_t = 0.5$, X_t is the last observed time-varying covariate (i.e.X_{t_0} in this case) for each subject and ϵ is a standard

normal random variable.

4. If $C_1 = 0$, then generate A_1, the treatment decision at time t_1 as a binomial random variable with $p = \lambda_1(\bar{X}_0, \boldsymbol{\gamma}) = \frac{1}{1+e^{-(\gamma_1+\gamma_b \mathbf{X}_b+\gamma_t X_{t_0})}}$. If $A_1 = 1$, then $U = t_1$, $\Delta = 1$, and information after t_j will be set to missing (i.e. X_{t_1}, C_2, A_2). And Y is generated as

$$Y = \alpha_0 + \alpha_1 + \alpha_{b1} X_{b1} + \alpha_{b2} X_{b2} + \alpha_t X_t + \epsilon, \tag{4.26}$$

where $\alpha_1 = 0.5$, $X_t = X_{t_0}$.

5. Otherwise, when $A_j = C_j = 0$. Generate $X_{t_1} \sim N(X_{t_0}, 1)$

6. Generate C_2, a potential treatment terminating event among $[t_1, t_2)$ as a binomial random variable with $p = \lambda_2^C(\bar{X}_1, \boldsymbol{\gamma}) = \frac{1}{1+e^{-(\gamma_2^C+\gamma_b^C \mathbf{X}_b+\gamma_t^C X_{t_1})}}$. If $C_2 = 1$ then $U = (t_1 + t_2)/2$, $\Delta = 0$, and all information afterwards except the outcome are set to missing. Y is generated the same way as (4.26) except that $X_t = X_{t_1}$.

7. If $C_2 = 0$, then set $A_2 = 1$, $U = t_2$, $\Delta = 1$. The outcome Y is generated as

$$Y = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_{b1} X_{b1} + \alpha_{b2} X_{b2} + \alpha_t X_t + \epsilon,$$

where $\alpha_2 = 0.3$, $X_t = X_{t_1}$.

The population parameters of interest μ_j's, $j = 1, 2$. For each μ_j, the same algorithm above is followed except that the treatment decision is not simulated but forced to stop at t_j (i.e. $A_l = 0$, $l < j$; $A_j = 1$) if not already censored due to treatment terminating event. We generated the outcome Y 500,000 times and took the sample average as μ_j.

In addition to the scenario where both outcome regression and propensity score models are correctly specified, situations where either model is miss-specified are also

studied as follows:

- Propensity score (PS) model miss-specification: In the treatment decision event, where the probability to stop at t_1 follows a binomial distribution with $p = \lambda_1(\bar{X}_0, \boldsymbol{\gamma}) = \frac{1}{1+e^{-(\gamma_1+\gamma_b \mathbf{X}_b+\gamma_t X_{t_0})}}$. We replace $\gamma_t X_{t_0}$ with quadratic form of X_{t_0},i.e. $\gamma_m {X_{t_0}}^2$, where $\gamma_m = 0.2$.

- Outcome regression (OR) model miss-specification: ignoring the time-dependent covariate X_t when fitting OR model, only X_{b1} and X_{b2} are kept as covariates.

- Neither PS or OR model is correctly specified: implementing both settings above together.

Because the data simulation steps for OR incorrect case remain the same and the truth simulation steps does not involve modeling of treatment decision event, the simulated truth for these miss-specification scenarios maintains the same as that when both OR and PS models are correctly specified.

4.3.2 Simulation Results

The simulation results comparing IPW (inverse probability weighting estimator (Johnson and Tsiatis, 2004)), OR (outcome regression estimator using G computation) and three DR estimators (the ordinary DR estimator, the improved DR estimator with variance minimization and the improved DR estimator with variance minimization as well as adjusting for parametrization in PS model) using sample size 300 and 500 Monte Carlo datasets were displayed in Table 4.4. When both PS and OR models are correctly specified, all estimators perform well in this setting with nice coverage probabilities. The OR estimator has smaller variance than DR estimators and the IPW estimator has the largest variance of among the five. The three DR estimators achieve essentially the same consistency and efficiency.

Table 4.4: Simulation results for estimating the average potential outcomes

			Bias	SD	SEE	ECP
			PS correct, OR correct			
	IPW	t_1	-0.013	0.247	0.238	0.92
		t_2	0.008	0.103	0.104	0.964
	OR	t_1	-0.003	0.114	0.110	0.934
		t_2	-0.009	0.097	0.099	0.958
	Ordinary	t_1	0.003	0.192	0.193	0.962
		t_2	0.008	0.102	0.103	0.962
DR	Improved	t_1	0.005	0.194	0.194	0.956
		t_2	0.008	0.102	0.103	0.962
	Improved$_c$	t_1	0.006	0.193	0.196	0.954
		t_2	0.007	0.102	0.103	0.962
			PS incorrect, OR correct			
	IPW	t_1	-0.103	0.306	0.253	0.842
		t_2	-0.007	0.113	0.111	0.934
	OR	t_1	-0.014	0.114	0.111	0.946
		t_2	-0.016	0.102	0.099	0.942
	Ordinary	t_1	-0.005	0.206	0.209	0.946
		t_2	-0.002	0.109	0.107	0.944
DR	Improved	t_1	0.013	0.204	0.214	0.956
		t_2	-0.002	0.109	0.106	0.942
	Improved$_c$	t_1	0.017	0.207	0.220	0.960
		t_2	-0.002	0.108	0.105	0.938
			PS correct, OR incorrect			
	IPW	t_1	-0.012	0.247	0.238	0.920
		t_2	0.011	0.103	0.104	0.960
	OR	t_1	-0.263	0.111	0.110	0.360
		t_2	-0.006	0.097	0.099	0.958
	Ordinary	t_1	-0.001	0.207	0.216	0.936
		t_2	0.010	0.102	0.104	0.962
DR	Improved	t_1	0.002	0.195	0.198	0.950
		t_2	0.010	0.102	0.103	0.962
	Improved$_c$	t_1	0.005	0.196	0.200	0.956
		t_2	0.010	0.102	0.103	0.962
			PS incorrect, OR incorrect			
	IPW	t_1	-0.102	0.306	0.253	0.842
		t_2	-0.004	0.113	0.111	0.934
	OR	t_1	-0.295	0.114	0.110	0.248
		t_2	-0.014	0.102	0.099	0.944
	Ordinary	t_1	-0.096	0.264	0.241	0.850
		t_2	-0.004	0.110	0.109	0.940
DR	Improved	t_1	-0.003	0.209	0.229	0.960
		t_2	-0.002	0.109	0.106	0.936
	Improved$_c$	t_1	0.035	0.217	0.237	0.950
		t_2	0.001	0.108	0.105	0.940

Truth = (3.910, 4.032); SD, Monte Carlo standard deviation; SEE, standard error estimate based on bootstrap; ECP, empirical coverage probability for Wald-type 95% confidence interval.

For estimating μ_1, when PS model is incorrect but OR model is correct, the IPW estimator is no longer consistent, while the other four estimators perform well. The OR estimator has the smallest variance and the variance of the three DR estimators are very similar; when PS model is correct but OR model is incorrect, the OR estimator is no longer consistent, but the others demonstrate good performance. While variance of IPW estimator is the largest, the two improved DR estimators show appreciable smaller variance than the ordinary DR estimator; when neither PS nor OR is correct, the IPW, OR and ordinary DR are all inconsistent. Under this particular setting where PS model is mildly miss-specified, the two improved DR estimator demonstrate satisfactory coverage probabilities.

For estimating μ_2, the PS and OR models are always correct in spite of the miss-specification procedure we adopt. This is largely due to it is the last interval, so the propensity to stop at the last interval is always 1, regardless the miss-specification. Similar case was also found in Chapter 1, where the IPW estimator is not consistent in the first three time points, but maintains its consistency in the fourth time point under discrete time setting and PS miss-specification. The reason for OR estimator to remain consistent in spite of OR miss-specification is probably because of the special setting we implemented where the outcome is independent of treatment.

Note that in our studied settings, the improved DR estimator with adjustment for parametrization in PS model barely show any advantage over the improved DR estimator without the adjustment. This is probably due to in the special setting we adopted here, the parametrized PS model approximate the true model really well. It is conjectured that in more general settings, where the differences between the two models are more substantial, we would expect the improved DR with adjustment for parametrization will have better consistency.

4.4 Remarks

We proposed three doubly robust estimators for dynamic regime with time-varying effects. We showed both in theory and in simulation studies that they are consistent when either the PS model or OR model is correctly specified. We also demonstrate the improved doubly robust estimation that minimizes the variance when the OR model may or may not be correctly specified.

Chapter 5

Conclusions

In Chapter 2, we focused on direct outcome regression estimators that accommodate both continuous and discrete time and in Chapter 3, we proposed an outcome regression method to estimate both the distribution and the mean of the potential outcome for discrete time. In both two chapters, we focused on modeling the treatment-terminating process. They work well when the treatment terminating event is moderate, under which condition the IPW estimator may not perform well.

Both approaches proposed in the two chapters and their competing IPW estimators are good only if their postulated model is correct. To provide a more robust estimation, in Chapter 4, we proposed an ordinary doubly-robust estimator and two improved doubly-robust estimators for dynamic regime. Even though the improved doubly robust estimation has been implemented in several settings, it has never been incorporated in dynamic regime before.

The work in the three chapters are all associated with each other. The setting in Chapter 2 can be considered as an outcome regression for estimation at the first time point, while the outcome regression using G-computation algorithms (Chapter 3) estimate across all stages, which constitute an essential augmentation part of the doubly robust estimator in Chapter 4.

Bibliography

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). Statistical Models Based on Counting Processes. Springer-Verlag: New York.

Ash, R. and C. Doléans-Dade (1999). Probability & Measure Theory (2nd ed.). San Diego: Academic Press.

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. Biometrics 61(4), 962–973.

Boos, D. D. and L. A. Stefanski (2013). Essential Statistical Inference. Springer: New York.

Cao, W., A. A. Tsiatis, and M. Davidian (2009). Improving efficiency and robustness of the doubly robust. Biometrika 96, 723–734.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. Journal of the American Statistical Association 89, 81–87.

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

Daniel, R. M., B. L. De Stavola, and S. N. Cousens (2011). gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata Journal 11(4), 479–517(39).

Gulick, R. M., H. J. Ribaudo, S. Lustgarten, K. E. Squires, W. A. Meyer, E. P. Acosta, B. R. Schackman, C. D. Pilcher, R. L. Murphy, W. L. Maher, M. D. Witt, R. C. Reichman, S. Snyder, K. L. Klingman, and D. R. Kuritzkes (2004). Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. The New England Journal of Medicine 350, 1850–1861.

Hernán, M. Á., B. Brumback, and J. M. Robins (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. Epidemiology 11(5), 561–570.

Johnson, B., H. Ribaudo, R. Gulick, and J. Eron Jr. (2013). Modeling clinical endpoints as a function of time to switch to second-line art with incomplete data on switching times. Biometrics 69, 732–740.

Johnson, B. A. and A. A. Tsiatis (2004). Estimating mean response as a function of treatment duration in an observational study, where duration may be informatively censored. Biometrics 60, 315–23.

Johnson, B. A. and A. A. Tsiatis (2005). Semiparametric inference in observational duration-response studies, with duration possibly right-censored. Biometrika 92, 605–618.

Kalbfleisch, J. D. and R. L. Prentice (2002). The Statistical Analysis of Failure Time Data. John Wiley: New York.

Kang, D. Y. J. and J. L. Schafer (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion and rejoinder). Statist. Sci. 22, 52380.

Kosorok, M. (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer: New York.

Li, L., J. J. Eron, H. Ribaudo, R. M. Gulick, and B. A. Johnson (2012). Evaluating the effect of early versus late ARV regimen change if failure on an initial regimen: Results from the AIDS Clinical Trials Group study A5095. Journal of the American Statistical Association 107, 542–554.

Murphy, S. A., M. J. van der Laan, and J. M. Robins (2001). Marginal mean models for dynamic regimes. Journal of the American Statistical Association 96, 1410–23.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Statist. Sci. 5, 465–480.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods — application to control of the healthy worker survivor effect. Mathematical Modelling 7, 1393–1512.

Robins, J. M. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. in health service research methodology: a focus on AIDS (eds l. sechrest, h. freeman and a. mulley). US Public Health Service., 113–159.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In M. Berkane (Ed.), Latent Variable Modeling and Applications to Causality, Number 120 in Lecture Notes in Statistics, pp. 69–117. New York: Springer Verlag.

Robins, J. M., M. A. Hernan, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. Epidemiology 11(5), 550–560.

Robins, J. M., A. Rotnizky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association 89, 846–866.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Rosenbaum, P. R. and D. B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association 79, 516–524.

Rotnitzky, A., Q. Lei, M. Sued, and J. M. Robins (2012). Improved double-robust estimation in missing data and causal inference models. Biometrika 99(2), 439–456.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects. The Annals of Statistics 6, 34–58.

Scharfstein, D. O., A. Rotnizky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. Journal of the American Statistical Association 94, 1096–1120.

Snowden, J. M., S. Rose, and K. M. Mortimer (2011). Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. American Journal of Epidemiology 173(7), 731–738.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. Journal of the American Statistical Association 101, 1619–1637.

Tan, Z. (2010). Bounded, efficient, and doubly robust estimation with inverse weighting. Biometrika 97, 661–682.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. 72, 20–22.

Tsiatis, A. A. (2006). Semiparametric Theory and Missing Data. Springer.

Tsiatis, A. A., M. Davidian, and W. Cao (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. Biometrics 67(2), 536–545.

van der Laan, M. J. and D. Rubin (2006). Targeted maximum likelihood learning. The International Journal of Biostatistics 2(1).

van der Vaart, A. W. and J. A. Wellner (1996). Weak Convergence and Empirical Processes. New York: Springer.

Zhang, M., A. A. Tsiatis, M. Davidian, K. S. Pieper, and K. W. Mahaffey (2011). Inference on treatment effects from a randomize clincal trial in the presence of premature treatment discontinuation: the synergy trial. Biostatistics 12, 258–269.

Appendix A

Supplementary Material for Chapter 2

A.1 Details of Large Sample Properties

We assume that $Z_i = (Y_i, U_i, \Delta_i, X_i),\ i = 1, \ldots, n$, are independent and identically distributed random vectors. Define the true value

$$\mu_0(t) = \int_{\mathcal{X}} \left[m_1(t,x) G(t \mid x) - \int_{[0,t)} m_0(u,x)\, dG(u \mid x) \right] dH(x),$$

where, under the correctly specified models, we have $m_1(t,x) = m_1(t,x;\beta_0)$, $m_0(t,x) = m_0(t,x;\alpha_0)$, $G(t \mid x) = G(t \mid x;\gamma_0)$, and $\theta = (\alpha, \beta, \gamma)$. Let

$$\widehat{\mu}_n(t) = \mu(t;\widehat{\theta}_n) = \mathbb{P}_n B_1(Z;\widehat{\beta}_n,\widehat{\gamma}_n) - \mathbb{P}_n B_2(Z;\widehat{\alpha}_n,\widehat{\gamma}_n),$$

where

$$B_1(Z;\beta,\gamma) = m_1(t,X;\beta) G(t \mid X;\gamma), \quad B_2(Z;\alpha,\gamma) = \int_{[0,t)} m_0(u,X;\alpha)\, dG(u \mid X;\gamma),$$

$\widehat{\alpha}_n$ minimizes the negative log-likelihood

$$-\sum_{i=1}^{n}\left\{(1-\Delta_i)\log f_0(Y_i \mid U_i, X_i; \alpha)\right\}, \tag{A.1}$$

and $\widehat{\beta}_n$ minimizes the negative log-likelihood

$$-\sum_{i=1}^{n}\left\{\Delta_i \log f_1(Y_i \mid U_i, X_i; \beta)\right\}. \tag{A.2}$$

In Theorem 1, $\widehat{\gamma}_n$ minimizes the negative log-likelihood

$$-\sum_{i=1}^{n}\left[(1-\Delta_i)\log\left\{\lambda^{(0)}(U_i \mid X_i; \gamma)\right\} + \log\left\{G(U_i \mid X_i; \gamma)\right\}\right]. \tag{A.3}$$

The following conditions are stated in a style similar to Murphy et al. (2001, Appendix).

Conditions

C1. The integral $\int_{\mathcal{X}}\left[m_1(t,x)G(t \mid x) - \int_{[0,t)} m_0(u,x)\, dG(u \mid x)\right]^2 dH(x)$ exists and is finite. There exists a finite vector β_0 such that $0 = E\psi_\beta(Z;\beta_0)$, $\mathbb{P}_n\psi_\beta(Z;\beta)$ is the derivative of (A.2) with respect to β; there exists a finite vector α_0 such that $0 = E\psi_\alpha(Z;\alpha_0)$, $\mathbb{P}_n\psi_\alpha(Z;\alpha)$ is the derivative of (A.1) with respect to α; there exists a finite vector γ_0 such that $0 = E\psi_\gamma(Z;\gamma_0)$, $\mathbb{P}_n\psi_\gamma(Z;\gamma)$ is the derivative of (A.3) with respect to γ.

C2. Let $\mathbb{N}$ be an neighborhood of $\{\alpha_0, \beta_0, \gamma_0, \mu_0(t)\}$. The class of functions

$$\Big\{ \mathbb{1}(X \leq x), B_1(Z;\beta,\gamma), B_2(Z;\alpha,\gamma), B_1^2(Z;\beta,\gamma), B_2^2(Z;\alpha,\gamma),$$
$$\frac{\partial}{\partial\beta}B_1(Z;\beta,\gamma), \frac{\partial}{\partial\gamma}B_1(Z;\beta,\gamma), \frac{\partial}{\partial\alpha}B_2(Z;\alpha,\gamma), \frac{\partial}{\partial\gamma}B_2(Z;\alpha,\gamma),$$
$$\frac{\partial}{\partial\alpha}\psi_\alpha(Z;\alpha), \frac{\partial}{\partial\beta}\psi_\beta(Z;\beta), \frac{\partial}{\partial\gamma}\psi_\gamma(Z;\gamma), \psi_\alpha(Z;\alpha)^{\otimes 2}, \psi_\beta(Z;\beta)^{\otimes 2},$$
$$\psi_\gamma(Z;\gamma)^{\otimes 2}, (\alpha,\beta,\gamma,\mu_0(t)) \in \mathbb{N} \Big\}$$

is a Glivenko-Cantelli class.

C3. Assume that $\mathcal{I}_\alpha = E\{(\partial/\partial\alpha)\psi_\alpha(Z;\alpha)\}$, $\mathcal{I}_\beta = E\{(\partial/\partial\beta)\psi_\beta(Z;\beta)\}$, and $\mathcal{I}_\gamma = E\{(\partial/\partial\gamma)\psi_\gamma(Z;\gamma)\}$ are invertible at the true values α_0, β_0, and γ_0, respectively.

Theorem 1. *Under Conditions C1–C3, $\widehat{\mu}_n(t) = \mu(t, \widehat{\theta}_n)$ is a root-n consistent and asymptotically normal estimator for $\mu_0(t)$.*

Proof. To show consistency, note that $\widehat{\alpha}_n$, $\widehat{\beta}_n$, and $\widehat{\gamma}_n$ are all maximum likelihood estimators and consistency for their respective estimands follows from Conditions C1–C3 using standard arguments (cf. van der Vaart and Wellner, 1996, § 3.2, 3.4.1). To show the consistency of $\mu(t, \widehat{\theta}_n)$, we can construct a simple estimating function $\psi_\mu(Z;\mu(t),\theta) = B_1(Z;\beta,\gamma) - B_2(Z;\alpha,\gamma) - \mu(t)$, use Conditions C1–C3, and apply an ordinary theory of Z-estimation to the system of equations, $0 = \mathbb{P}_n(\psi_\mu, \psi_\alpha^{\mathrm{T}}, \psi_\beta^{\mathrm{T}}, \psi_\gamma^{\mathrm{T}})^{\mathrm{T}}$ (cf. van der Vaart and Wellner, 1996, Thm 3.3.1). The proofs of these results are omitted.

To show asymptotic normality, the Glivenko-Cantelli property and invertibility of $\mathcal{I}_\alpha$, $\mathcal{I}_\beta$, and $\mathcal{I}_\gamma$, lead to the results: $n^{1/2}(\widehat{\alpha}_n - \alpha_0) = n^{1/2}\mathbb{P}_n\varphi_\alpha(Z;\alpha_0) + o_p(1)$, $\varphi_\alpha(Z;\alpha) = \mathcal{I}_\alpha^{-1}\psi_\alpha(Z;\alpha)$, $n^{1/2}(\widehat{\beta}_n - \beta_0) = n^{1/2}\mathbb{P}_n\varphi_\beta(Z;\beta_0) + o_p(1)$, $\varphi_\beta(Z;\beta) = \mathcal{I}_\beta^{-1}\psi_\beta(Z;\beta)$, $n^{1/2}(\widehat{\gamma}_n - \gamma_0) = n^{1/2}\mathbb{P}_n\varphi_\gamma(Z;\gamma_0) + o_p(1)$, $\varphi_\gamma(Z;\gamma_0) = \mathcal{I}_\gamma^{-1}\psi_\gamma(Z;\gamma_0)$. Through first-

order Taylor-series approximations, we have that

$$\begin{aligned}
\mu(t;\widehat{\theta}_n) &= \mathbb{P}_n B_1(Z;\widehat{\beta}_n,\widehat{\gamma}_n) - \mathbb{P}_n B_2(Z;\widehat{\alpha}_n,\widehat{\gamma}_n) \\
&= \mathbb{P}_n B_1(Z;\beta_0,\gamma_0) - \mathbb{P}_n B_2(Z;\alpha_0,\gamma_0) \\
&\quad +\mathbb{P}_n \left\{ \frac{\partial}{\partial\alpha} B_2(Z;\alpha,\widehat{\gamma}_n)\Big|_{\alpha=\alpha^*} \right\} (\widehat{\alpha}_n - \alpha_0) \\
&\quad +\mathbb{P}_n \left\{ \frac{\partial}{\partial\beta} B_1(Z;\beta,\widehat{\gamma}_n)\Big|_{\beta=\beta^*} \right\} (\widehat{\beta}_n - \beta_0) \\
&\quad +\mathbb{P}_n \left[\frac{\partial}{\partial\gamma} \left\{ B_1(Z;\widehat{\beta}_n,\gamma) - B_2(Z;\widehat{\alpha}_n,\gamma) \right\} \Big|_{\gamma=\gamma^*} \right] (\widehat{\gamma}_n - \gamma_0)
\end{aligned}$$

where α^* is a vector of intermediate values between $\widehat{\alpha}_n$ and α_0, β^* is a vector of intermediate values between $\widehat{\beta}_n$ and β_0, and γ^* is a vector of intermediate values between $\widehat{\gamma}_n$ and γ_0. The first term in the expansion converges to $\mu_0(t)$ in probability by a law of large numbers. Apply the Glivenko-Cantelli property and the results above to yield the asymptotic expansion

$$n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\} = n^{1/2}\mathbb{P}_n\varphi_\mu(t,Z;\theta_0) + o_p(1),$$

where

$$\begin{aligned}
\varphi_\mu(t,Z;\theta_0) &= E\left\{ \frac{\partial}{\partial\alpha} B_2(Z;\alpha,\gamma_0)\Big|_{\alpha=\alpha_0} \right\} \varphi_\alpha(Z;\alpha_0) \qquad \text{(A.4)} \\
&\quad +E\left\{ \frac{\partial}{\partial\beta} B_1(Z;\beta,\gamma_0)\Big|_{\beta=\beta_0} \right\} \varphi_\beta(Z;\beta_0) \\
&\quad +E\left\{ \frac{\partial}{\partial\gamma} B_1(\beta_0,\gamma)\Big|_{\gamma=\gamma_0} - \frac{\partial}{\partial\gamma} B_2(\alpha_0,\gamma)\Big|_{\gamma=\gamma_0} \right\} \varphi_\gamma(Z;\gamma_0). \qquad \text{(A.5)}
\end{aligned}$$

Asymptotic normality then follows from the central limit theorem and Slutzky's Theorem. Thus, $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\}$ converges in distribution to a mean-zero normal random variable with variance $E(\varphi_\mu^{\otimes 2})$, with φ_μ defined in (A.5). □

Now, we introduce counting process notation for the treatment interruption data. Define the counting process $N_i^{(0)}(t) = I(U_i \leq t, \Delta_i = 0)$, $N_i(t) = I(U_i \leq t, \Delta_i = 1)$ the at-risk process $R_i(t) = I(U_i \geq t)$, and

$$M_i^{(0)}(t) = N_i^{(0)}(t) - R_i(t)\Lambda_0(t)\exp(X_i^{\mathrm{T}}\gamma_0),$$

a local martingale with respect to the filtration $\mathcal{F}_t = \sigma\{N_i(u), N_i^{(0)}(u), X_i, u \leq t,\ i = 1, \ldots, n\}$. We also define the sum of subject-specific martingales and counting processes as

$$\bar{M}^{(0)}(u) = \sum_{i=1}^{n} M_i^{(0)}(u), \qquad \bar{N}^{(0)}(u) = \sum_{i=1}^{n} N_i^{(0)}(u),$$

respectively. If $\widehat{\gamma}_n$ is the maximum partial likelihood estimator (MPLE), then it solves the estimating equations, $\mathbb{S}_n(\widehat{\gamma}_n) = 0$,

$$\mathbb{S}_n(\gamma) = \mathbb{P}_n\left[\int_0^\infty \left\{X - \widetilde{X}(u,\gamma)\right\} dN^{(0)}(u)\right],$$

where

$$S^{(k)}(u,\gamma) = \mathbb{P}_n\left\{X^{\otimes k}\exp(X^{\mathrm{T}}\gamma)R(u)\right\}, \quad s^{(k)}(u,\gamma) = \lim_{n\to\infty} S^{(k)}(u,\gamma),\ k = 0,1,2,$$
$$\widetilde{X}(u,\gamma) = \frac{S^{(1)}(u,\gamma)}{S^{(0)}(u,\gamma)}, \qquad \mu(u,\gamma) = \frac{s^{(1)}(u,\gamma)}{s^{(0)}(u,\gamma)},$$

where $v^{\otimes 2} = vv^{\mathrm{T}}$ for a vector v. Moreover, standard results from survival analysis yield the influence function for $\widehat{\gamma}_n$ as

$$\begin{aligned} n^{1/2}(\widehat{\gamma}_n - \gamma_0) &= n^{1/2}\mathbb{P}_n\{\varphi_\gamma(Z,\gamma_0)\} + o_p(1), \\ \varphi_\gamma(Z,\gamma_0) &= V_\gamma^{-1}\int_0^\infty \{X - \mu(u,\gamma_0)\}\, dM^{(0)}(u), \end{aligned}$$

where

$$V_\gamma = \lim_{n\to\infty} \mathbb{P}_n \left\{ \int_0^\infty \left[\frac{\sum_{j=1}^n \{X_j - \widetilde{X}(u,\gamma_0)\}^{\otimes 2} \exp(X_j^{\mathrm{T}}\gamma_0) R_j(u)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0) R_j(u)} \right] dN^{(0)}(u) \right\}.$$

Our proof of Theorem 2 relies on our ability to write the sum of stochastic integrals in $\widehat{\mu}_n(t)$ as the sum of independent and identically distributed random variables plus asymptotically negligible terms. To facilitate that objective, we state some of the key results in the following three lemmas. Lemma 1 provides the influence function for the predicted survival estimate at a fixed $X = x^*$, i.e. $\widehat{\mathrm{pr}}(C > t \mid x^*)$. This result is stated for completeness and whose proof is a standard result in advanced survival analysis texts. Hence, the proof of Lemma 1 is omitted. Lemma 2 derives the influence function for a stochastic integral with integrator equal to the Breslow estimator and is a key step in Lemma 3. Lemma 3 derives the influence function for a sample average of stochastic integrals, where the i-th integral has subject-specific integrator $G(t \mid X_i; \widehat{\gamma}_n)$.

Lemma 1. *For any $t \leq \tau$, τ the upper support of U, x^* lies in the interior of $\mathcal{X}$ and under Conditions VII.2.1–VII.2.2 from Andersen et al. (1993),*

$$n^{1/2}\{\widehat{G}_n(t \mid x^*; \widehat{\gamma}_n) - G(t \mid x^*)\} = n^{1/2}\mathbb{P}_n \{\varphi_G(Z, \gamma_0)\} + o_p(1),$$

where

$$\begin{aligned} \varphi_G(Z_i, \gamma_0) &= \exp\{(x^*)^{\mathrm{T}}\gamma_0\}\{\int_0^t \frac{dM_i^{(0)}(u)}{s^{(0)}(u,\gamma_0)} \\ &\quad + \{K(t)\}^{\mathrm{T}} V_\gamma^{-1} \int_0^\infty \{X_i - \mu(u,\gamma_0)\}\, dM_i^{(0)}(u)\}, \\ K(t) &= -\int_0^t \{x^* - \mu(u,\gamma_0)\}\, d\Lambda_0(u). \end{aligned}$$

Lemma 2. *Let $H(t)$ be a $\mathcal{F}_t$-adapted function, $\widehat{\Lambda}_0(t)$ is the Breslow estimator for the*

integrated baseline hazard function $\Lambda_0(t)$, and define

$$\widehat{\eta}_n = \int_a^b H(u)\, d\widehat{\Lambda}_0(u;\widehat{\gamma}_n), \quad \eta_0 = \int_a^b H(u)\, d\Lambda_0(u),$$

for $0 \leq a < b < \tau$, τ is the upper support point of U and is finite. Under Conditions VII.2.1–VII.2.2 from Andersen et al. (1993),

$$n^{1/2}(\widehat{\eta}_n - \eta_0) = n^{1/2}\mathbb{P}_n\left\{\varphi_{\eta,H}(Z,\gamma_0)\right\} + o_p(1),$$

where

$$\begin{aligned}
\varphi_{\eta,H}(Z_i,\gamma_0) &= \int_a^b \frac{H(t)\, dM_i^{(0)}(t)}{s^{(0)}(t,\gamma_0)} + \\
&\qquad \int_0^\infty \left\{K_{H(b)}(b) - K_{H(a)}(a) + J(a,b)\right\}^{\mathrm{T}} V_\gamma^{-1}\left\{X_i - \mu(t,\gamma_0)\right\} dM_i^{(0)}(t), \\
K_{H(b)}(b) &= -H(b)\int_0^b \mu(u,\gamma_0)\, d\Lambda_0(u), \\
K_{H(a)}(a) &= -H(a)\int_0^a \mu(u,\gamma_0)\, d\Lambda_0(u), \\
J_H(a,b) &= \int_a^b \left\{\int_0^u \mu(t,\gamma_0)\, d\Lambda_0(t)\right\} dH(u).
\end{aligned}$$

Proof. First, we expand the difference by adding and subtracting like quantities,

$$\begin{aligned}
n^{1/2}(\widehat{\eta}_n - \eta_0) &= n^{1/2}\int_a^b H(u)\left\{d\widehat{\Lambda}_0(u,\widehat{\gamma}_n) - d\widehat{\Lambda}_0(u,\gamma_0)\right\} && \text{(A.6)} \\
&\quad + n^{1/2}\int_a^b H(u)\left\{d\widehat{\Lambda}_0(u,\gamma_0) - d\Lambda_0(u)\right\} && \text{(A.7)}
\end{aligned}$$

Using integration by parts, we expand the first term as the following three terms

$$\begin{aligned}
\text{(A.6)} &= n^{1/2}\left[H(b)\{\widehat{\Lambda}_0(b;\widehat{\gamma}_n)-\widehat{\Lambda}_0(b;\gamma_0)\}\right] && \text{(A.8)}\\
&\quad -n^{1/2}\left[H(a)\{\widehat{\Lambda}_0(a;\widehat{\gamma}_n)-\widehat{\Lambda}_0(a;\gamma_0)\}\right] && \text{(A.9)}\\
&\quad -n^{1/2}\int_a^b\left\{\widehat{\Lambda}_0(u;\widehat{\gamma}_n)-\widehat{\Lambda}_0(u;\gamma_0)\right\}\,dH(u). && \text{(A.10)}
\end{aligned}$$

Through a first-order Taylor-series approximation, law of large numbers, and Lenglart's Inequality,

$$\text{(A.8)} = n^{-1/2}\sum_{i=1}^{n}\{K_{H(b)}(b)\}^{\mathrm{T}}V_\gamma^{-1}\int_0^\infty\{X_i-\mu(u,\gamma_0)\}\,dM_i(u)+o_p(1),$$

where $K_{H(b)}(b)$ is defined above. Similarly, we have that

$$\text{(A.9)} = -n^{-1/2}\sum_{i=1}^{n}\{K_{H(a)}(a)\}^{\mathrm{T}}V_\gamma^{-1}\int_0^\infty\{X_i-\mu(u,\gamma_0)\}\,dM_i(u)+o_p(1),$$

and $K_{H(a)}(a)$ is defined above. Lastly, we have that

$$\begin{aligned}
\text{(A.10)} &= -n^{1/2}\int_a^b\left\{\frac{\partial}{\partial\gamma}\int_0^u\frac{d\bar{N}^{(0)}(t)}{\sum_{j=1}^n\exp(X_j^{\mathrm{T}}\gamma)R_j(t)}\bigg|_{\gamma=\gamma^*}\right\}dH(u)\\
&= -n^{1/2}\int_a^b\left[\left\{-\int_0^u\mu(t,\gamma_0)\,d\Lambda_0(t)\right\}^{\mathrm{T}}(\widehat{\gamma}_n-\gamma_0)\,dH(u)\right]+o_p(1),\\
&= \left[\int_a^b\left\{\int_0^u\mu(t,\gamma_0)\,d\Lambda_0(t)\right\}dH(u)\right]^{\mathrm{T}}n^{1/2}(\widehat{\gamma}_n-\gamma_0)+o_p(1),\\
&= n^{-1/2}\sum_{i=1}^{n}\{J(a,b)\}^{\mathrm{T}}\int_0^\infty V_\gamma^{-1}\left\{X_i-\mu(u,\gamma_0)\right\}\,dM_i^{(0)}(u)+o_p(1),
\end{aligned}$$

where the second expression follows from a first-order Taylor-series approximation, law of large numbers, and Lenglart's Inequality, analogous to the preceding argu-

ments, and $J(a,b)$ is defined above. Hence,

$$\begin{aligned}(\text{A.6}) &= n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \left\{K_{H(b)}(b) - K_{H(a)}(a) + J(a,b)\right\}^{\mathrm{T}} V_\gamma^{-1}\left\{X_i - \mu(u,\gamma_0)\right\} \\ &\quad dM_i^{(0)}(u) + o_p(1).\end{aligned}$$

Similar to (A.6), we expand (A.7) into three terms using integration by parts, i.e.,

$$\begin{aligned}(\text{A.7}) &= n^{1/2}\left[H(b)\{\widehat{\Lambda}_0(b;\gamma_0) - \Lambda_0(b)\}\right] && (\text{A.11})\\ &\quad -n^{1/2}\left[H(a)\{\widehat{\Lambda}_0(a;\gamma_0) - \Lambda_0(a)\}\right] && (\text{A.12})\\ &\quad -n^{1/2}\int_a^b \left\{\widehat{\Lambda}_0(u;\gamma_0) - \Lambda_0(u)\right\} dH(u). && (\text{A.13})\end{aligned}$$

By definition, we have

$$\begin{aligned}(\text{A.11}) &= n^{1/2}H(b)\left\{\int_0^b \frac{d\bar{N}^{(0)}(u)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(u)} - d\Lambda_0(u)\right\} \\ &= n^{1/2}H(b)\int_0^b \frac{d\bar{M}^{(0)}(u)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(u)} \\ &= n^{-1/2}\sum_{i=1}^{n} H(b)\int_0^b \frac{dM_i^{(0)}(u)}{s^{(0)}(u,\gamma_0)} + o_p(1),\end{aligned}$$

where the third line follows by a law of large numbers. Similarly,

$$(\text{A.12}) = -n^{-1/2}\sum_{i=1}^{n} H(a)\int_0^a \frac{dM_i^{(0)}(u)}{s^{(0)}(u,\gamma_0)} + o_p(1).$$

Finally,

$$\begin{aligned}(\text{A.13}) &= -n^{1/2}\int_a^b \left\{\widehat{\Lambda}_0(u,\gamma_0) - \Lambda_0(u)\right\} dH(u), \\ &= -n^{1/2}\int_a^b \left\{\int_0^u \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)}\right\} dH(u). && (\text{A.14})\end{aligned}$$

The next step is to change the order of integration. Note, the area of integration is

regular in the sense that it is contiguous; however, the region is trapezoidal which results in the sum of two regions after the change. There are two equivalent expressions for (A.14):

$$\begin{aligned}(\text{A.14}) &= -n^{1/2}\Bigg[\int_0^b \left\{\int_t^b dH(u)\right\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)} \\ &\qquad - \int_0^a \left\{\int_t^a dH(u)\right\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)}\Bigg] \qquad (\text{A.15}) \\ &= -n^{1/2}\Bigg[\int_0^a \left\{\int_a^b dH(u)\right\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)} \\ &\qquad + \int_a^b \left\{\int_t^b dH(u)\right\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)}\Bigg].\end{aligned}$$

Note, that if $a = 0$, the two expression are identical. By the Fundamental Theorem of Calculus, we can evaluate the inner integral at the upper and lower limits of integration: $\int_t^b dH(u) = H(b) - H(t)$ and $\int_t^a dH(u) = H(a) - H(t)$. Then,

$$\begin{aligned}(\text{A.15}) &= -n^{1/2}\int_0^b \{H(b) - H(t)\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)} \qquad (\text{A.16}) \\ &\quad +n^{1/2}\int_0^a \{H(a) - H(t)\} \frac{d\bar{M}^{(0)}(t)}{\sum_{j=1}^n \exp(X_j^{\mathrm{T}}\gamma_0)R_j(t)} \\ &= n^{-1/2}\sum_{i=1}^n \int_0^b \frac{\{H(t) - H(b) + [H(a) - H(t)]I(0 \le t \le a)\}\, dM_i^{(0)}(t)}{s^{(0)}(t,\gamma_0)} + o_p(1),\end{aligned}$$

where the last line follows by an ordinary law of large numbers. Putting (A.11)–(A.13) together,

$$\begin{aligned}(\text{A.7}) &= (\text{A.11}) + (\text{A.12}) + (\text{A.13}) \\ &= n^{-1/2}\sum_{i=1}^n \int_0^b \frac{\{H(t) - H(t)I(0 \le t \le a)\}\, dM_i^{(0)}(t)}{s^{(0)}(t,\gamma_0)} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^n \int_a^b \frac{H(t) dM_i^{(0)}(t)}{s^{(0)}(t,\gamma_0)} + o_p(1).\end{aligned}$$

Thus,

$$
\begin{aligned}
&\text{(A.6)} + \text{(A.7)} = \\
&\quad n^{-1/2} \sum_{i=1}^{n} \left[\int_a^b \frac{H(t)\, dM_i^{(0)}(t)}{s^{(0)}(t, \gamma_0)} \right. \\
&\quad \left. + \int_0^\infty \{K_{H(b)}(b) - K_{H(a)}(a) + J(a,b)\}^{\mathrm{T}} V_\gamma^{-1} \{X_i - \mu(t, \gamma_0)\}\, dM_i^{(0)}(t) \right] + o_p(1).
\end{aligned}
$$

□

Lemma 3. *Let $H_i(t)$ be a $\mathcal{F}_t$-adapted function and define*

$$
\widehat{\zeta}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t H_i(u)\, d\widehat{G}(u \mid X_i; \widehat{\gamma}_n), \quad \zeta_0(t) = E\left\{ \int_0^t H_i(u)\, dG(u, X_1) \right\}.
$$

Under Conditions VII.2.1–VII.2.2 from Andersen et al. (1993),

$$
n^{1/2}\{\widehat{\zeta}_n(t) - \zeta_0(t)\} = n^{1/2} \mathbb{P}_n \left\{ \int_0^\infty \varphi_{\zeta,H}(u, Z_i, \gamma_0)\, dM_i^{(0)}(u) \right\} + o_p(1),
$$

where

$$
\begin{aligned}
\varphi_{\zeta,H}(u, Z_i, \gamma_0) &= \{J(\gamma_0) + K(\gamma_0)\} V_\gamma^{-1} \{X_i - \mu(u, \gamma_0)\} + \frac{E\{-H_1(u) G(u, X_1) e^{X_1^{\mathrm{T}} \gamma_0}\}}{s^{(0)}(u, \gamma_0)} \\
J(\gamma_0) &= E\{ \int_0^\infty \{-H_1(u) \exp(X_1^{\mathrm{T}} \gamma_0) X_1 I(0 < u < t) \{G(u \mid X_1) d\Lambda_0(u) \\
&\quad + \Lambda_0(u)\, dG(u \mid X_1)\}\}\} \\
K(\gamma_0) &= \int_0^\infty E\{-H_1(u) G(u \mid X_1) \exp(X^{\mathrm{T}} \gamma_0)\} I(0 < u < t)\, d\Lambda_0(u).
\end{aligned}
$$

Proof. We begin by noting $G(t \mid X_i) = \exp\{-\Lambda(t, X_i)\}$ implies the identity

$$
dG(t \mid X_i) = d\{e^{-\Lambda(t, X_i)}\} = G(t \mid X_i) d\{-\Lambda(t, X_i)\} = -G(t \mid X_i) d\Lambda(t, X_i).
$$

As a result of the functional delta method and the uniform consistency of

$\widehat{G}(u \mid X_i)$ for $G(u \mid X_i)$ for $0 \leq u \leq t$, we have that

$$n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} H_i(u)\, d\{\widehat{G}(u \mid X_i; \widehat{\gamma}_n) - G(u \mid X_i)\} =$$
$$-n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} H_i(u) G(u \mid X_i) d\{\widehat{\Lambda}(u, X_i; \widehat{\gamma}_n) - \Lambda(u, X_i)\} + o_p(1). \quad \text{(A.17)}$$

Expression (A.17) is expanded as the following two terms

$$\begin{aligned}
\text{(A.17)} \quad = \quad & -n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} H_i(u) G(u \mid X_i) && \text{(A.18)} \\
& d[\{\widehat{\Lambda}_0(u, \widehat{\gamma}_n) e^{X_i^{\mathrm{T}} \widehat{\gamma}_n}\} - \{\Lambda_0(u) e^{X_i^{\mathrm{T}} \widehat{\gamma}_n}\}] && \text{(A.19)} \\
& -n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} H_i(u) G(u \mid X_i) d[\{\Lambda_0(u) e^{X_i^{\mathrm{T}} \widehat{\gamma}_n}\} - \{\Lambda_0(u) e^{X_i^{\mathrm{T}} \gamma}\}] && \text{(A.20)} \\
& + o_p(1). && \text{(A.21)}
\end{aligned}$$

Using the consistency of $\widehat{\gamma}_n$ and the continuity of $\exp(X_i^{\mathrm{T}} \gamma)$,

$$\text{(A.19)} = -n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} H_i(u) G(u \mid X_i) e^{X_i^{\mathrm{T}} \gamma_0}\, d\{\widehat{\Lambda}_0(u, \widehat{\gamma}_n) - \Lambda_0(u)\} + o_p(1). \quad \text{(A.22)}$$

Now, we see that (A.22) is $n^{1/2}$ multiplied by a sample average of stochastic integrals in Lemma 2 plus asymptotically negligible terms. Let $H(u) = H_i(u) G(u \mid X_i) \exp(X_i^{\mathrm{T}} \gamma_0)$, $a = 0$, and $b = t$ in Lemma 2, then

$$\text{(A.22)} = -n^{1/2} \sum_{i=1}^{n} \varphi_{\zeta, H_i}(Z_i; \gamma_0) + o_p(1),$$

where

$$\begin{aligned}
\varphi_{\zeta,H_i}(Z_i;\gamma_0) &= \int_0^t \frac{E\{H_1(u)G(u \mid X_1)\exp(X_1^{\mathrm{T}}\gamma_0)\}\, dM_i(u)}{s^{(0)}(u,\gamma_0)} \\
&\quad +\{K_{\zeta,H_i}(t) - J_{\zeta,H_i}(t)\}^{\mathrm{T}} \int_0^\infty V_\gamma^{-1}\{X_i - \mu(u,\gamma_0)\}\, dM_i^{(0)}(u), \\
K_{\zeta,H_i}(t) &= -\int_0^t E\left\{H_1(t)G(u \mid X_1)\exp(X_1^{\mathrm{T}}\gamma_0)\right\}\mu(u,\gamma_0)\, d\Lambda_0(u) \\
J_{\zeta,H_i}(t) &= E\left[\int_0^t \left\{\int_0^u \mu(s,\gamma_0)d\Lambda_0(s)\right\}\exp(X_1^{\mathrm{T}}\gamma_0)d\{H_1(u)G(u \mid X_1)\}\right].
\end{aligned}$$

The second term in (A.17) is

$$\begin{aligned}
\text{(A.21)} &= -n^{-1/2}\sum_{i=1}^n \int_{[0,t)} H_i(u)G(u \mid X_i)(e^{X_i^{\mathrm{T}}\widehat{\gamma}_n} - e^{X_i^{\mathrm{T}}\gamma_0})\, d\Lambda_0(u) \\
&= -n^{1/2}\left[\int_{[0,t)}\left\{n^{-1}\sum_{i=1}^n H_i(u)G(u \mid X_i)e^{X_i^{\mathrm{T}}\gamma^*}X_i\right\} d\Lambda_0(u)\right](\widehat{\gamma}_n - \gamma_0) \\
&= -n^{-1/2}\sum_{i=1}^n \int_0^\infty \left[\{\Omega_2(t)\}^{\mathrm{T}}V_\gamma^{-1}\{X_i - \mu(u,\gamma_0)\}\right] dM_i(u) + o_p(1),
\end{aligned}$$

where

$$\Omega_2(t) = \int_{[0,t)} E\{H_1(u)G(u \mid X_1)e^{X_1^{\mathrm{T}}\gamma_0}X_1\}\, d\Lambda_0(u).$$

Putting these three terms together, we have

$$\begin{aligned}
\text{(A.17)} &= \text{(A.19)} + \text{(A.21)} \\
&= n^{-1/2}\sum_{i=1}^n (-1)\varphi_{\zeta,H_i}(Z_i;\gamma_0) + \Omega_2(t)V_\gamma^{-1}\int_0^\infty \{X_i - \mu(u,\gamma_0)\}\, dM_i(u) + o_p(1),
\end{aligned}$$

□

For consistency in presentation, we state the following conditions for proving Theorem 2 in a format similar to those for Theorem 1 but accommodate the more complex partial likelihood estimator for γ and Breslow estimator for $\Lambda_0(t)$. Note that Conditions V11.2.1–VII.2.2 from Andersen et al. (1993) are implied by Conditions C1*–C3*

below so that the conclusions from Lemmas 1–3 all hold.

Conditions

C1*. The integral $\int_{\mathcal{X}} \left[m_1(t,x)G(t \mid x) - \int_{[0,t)} m_0(u,x)\, dG(u \mid x) \right]^2 dH(x)$ exists and is finite. There exists a finite vector β_0 such that $0 = E\psi_\beta(Z;\beta_0)$, $\mathbb{P}_n\psi_\beta(Z;\beta)$ is the derivative of (A.2) with respect to β; there exists a finite vector α_0 such that $0 = E\psi_\alpha(Z;\alpha_0)$, $\mathbb{P}_n\psi_\alpha(Z;\alpha)$ is the derivative of (A.1) with respect to α; there exists a finite vector γ_0 such that $0 = E\psi_\gamma(Z;\gamma_0)$, $\mathbb{S}_n(\gamma) = \mathbb{P}_n\psi_\gamma(Z;\gamma)$ is Cox's score function; τ, the upper support point of U, is finite.

C2*. Let $\mathbb{N}$ be an neighborhood of $\{\alpha_0, \beta_0, \gamma_0, \mu_0(t)\}$. The class of functions

$$\Big\{ \mathbb{1}(X \le x), B_1(Z;\beta,\gamma), B_2(Z;\alpha,\gamma), B_1^2(Z;\beta,\gamma), B_2^2(Z;\alpha,\gamma),$$
$$\frac{\partial}{\partial\beta}B_1(Z;\beta,\gamma), \frac{\partial}{\partial\gamma}B_1(Z;\beta,\gamma), \frac{\partial}{\partial\alpha}B_2(Z;\alpha,\gamma), \frac{\partial}{\partial\gamma}B_2(Z;\alpha,\gamma),$$
$$\frac{\partial}{\partial\alpha}\psi_\alpha(Z;\alpha), \frac{\partial}{\partial\beta}\psi_\beta(Z;\beta), \psi_\alpha(Z;\alpha)^{\otimes 2}, \psi_\beta(Z;\beta)^{\otimes 2},$$
$$N^{(0)}(u), XN^{(0)}(u), X^{\otimes 2}N^{(0)}(u), R(u)e^{X^{\mathrm{T}}\gamma}, XR(u)e^{X^{\mathrm{T}}\gamma}, X^{\otimes 2}R(u)e^{X^{\mathrm{T}}\gamma},$$
$$u \in [0,\tau], (\alpha,\beta,\gamma,\mu_0(t)) \in \mathbb{N} \Big\}$$

is a Glivenko-Cantelli class.

C3*. Assume that $\mathcal{I}_\alpha = E\{(\partial/\partial\alpha)\psi_\alpha(Z;\alpha)\}$, $\mathcal{I}_\beta = E\{(\partial/\partial\beta)\psi_\beta(Z;\beta)\}$, and V_γ are invertible at the true values α_0, β_0, and γ_0, respectively.

Theorem 2. *Under Conditions C1*–C3*, $\widehat{\mu}_n(t)$ is a root-n consistent estimator for $\mu_0(t)$ and $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\}$ converges in distribution to a mean-zero normal random variable with covariance $E(\varphi_\mu^{\otimes 2})$, where φ_μ is given in* (A.35).

Proof. The consistency of $\widehat{\mu}_n(t)$ follows similarly to Theorem 1 and is omitted. We focus on deriving the analytic form of the variance for the limiting distribution. To this end, we will show that $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\} = n^{-1/2}\sum_{i=1}^n \varphi_\mu(t, Z_i; \vartheta_0) + o_p(1)$,

which will imply the conclusion by an ordinary central limit theorem and Slutzky's theorem. We begin by separating the estimator and estimand into two parts:

$$
\begin{aligned}
& n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\} = \\
& \quad n^{-1/2}\sum_{i=1}^{n}\left[m_1(t, X_i; \widehat{\beta}_n)\widehat{G}(t \mid X_i; \widehat{\gamma}_n) - m_1(t, X_i; \beta_0)G(t \mid X_i)\right] && \text{(A.23)} \\
& \quad -n^{-1/2}\sum_{i=1}^{n}\{\int_{[0,t)} m_0(u, X_i; \widehat{\alpha}_n)\, d\widehat{G}(u \mid X_i; \widehat{\gamma}_n) && \text{(A.24)} \\
& \quad -\int_{[0,t)} m_0(u, X_i; \alpha_0)\, dG(u \mid X_i)\}. && \text{(A.25)}
\end{aligned}
$$

The first term on the right-hand side of the above expression is

$$
\begin{aligned}
\text{(A.23)} \;=\; & n^{-1/2}\sum_{i=1}^{n}\left[m_1(t, X_i; \widehat{\beta}_n)\left\{\widehat{G}(t \mid X_i; \widehat{\gamma}_n) - G(t \mid X_i)\right\}\right] && \text{(A.26)} \\
& +n^{1/2}\sum_{i=1}^{n}\{\{m_1(t, X_i; \widehat{\beta}_n) && \text{(A.27)} \\
& -m_1(t, X_i; \beta_0)\}G(t \mid X_i)\} && \text{(A.28)}
\end{aligned}
$$

Note that $\widehat{G}(t \mid X_i; \widehat{\gamma}_n)$ is the standard predicted survival function at $X = X_i$ whose asymptotic expansion is given in Lemma 1 and leads to the following expression:

$$
\begin{aligned}
\text{(A.26)} \;=\; & n^{-1/2}\sum_{i=1}^{n}\left[E\{-w_1(t; Z_1, \theta_0)\}\int_{[0,t)} \frac{dM_i^{(0)}(u)}{s^{(0)}(u, \gamma_0)} + \right. \\
& \left. \{\Gamma_1(t)\}^{\mathrm{T}} V_\gamma^{-1}\int_{[0,\infty)} \{X_i - \mu(u, \gamma_0)\}\, dM_i^{(0)}(u)\right] + o_p(1), \quad \text{(A.29)}
\end{aligned}
$$

where

$$
\begin{aligned}
w_1(t, Z_i; \theta_0) &= m_1(t, X_i; \beta_0)G(t \mid X_i; \gamma_0)e^{X_i^{\mathrm{T}}\gamma_0}, \\
\Gamma_1(t) &= -E\left[w_1(t, Z_1; \theta_0)\int_{[0,t)} \{X_1 - \mu(u, \gamma_0)\}\, d\Lambda_0(u)\right].
\end{aligned}
$$

A first-order Taylor series approximation of $m_1(t, X_i; \widehat{\beta}_n)$ about β_0 in (A.28) leads to

the asymptotic expansion

$$\begin{aligned}(\text{A.28}) &= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\beta}m_1(t,X_i;\beta)\Big|_{\beta=\beta^*}\right\}G(t\mid X_i)(\widehat{\beta}_n-\beta_0),\\ &= n^{-1/2}\sum_{i=1}^{n}\{L(t)\}^{\mathrm{T}}\varphi_\beta(Z_i;\beta_0)+o_p(1), \qquad (\text{A.30})\end{aligned}$$

where β^* is a vector of intermediate points on the line segment between $\widehat{\beta}$ and β_0 and $L(t) = E[\{(\partial/\partial\beta)m_1(t,X_1;\beta_0)\}G(t\mid X_1)]$. Combining (A.29) and (A.30) leads to the expression

$$\begin{aligned}(\text{A.23}) &= n^{-1/2}\sum_{i=1}^{n}\left[\{L(t)\}^{\mathrm{T}}\varphi_\beta(Z_i;\beta_0)+E\{-w_1(t;Z_1,\theta_0)\}\int_{[0,t)}\frac{dM_i^{(0)}(u)}{s^{(0)}(u,\gamma_0)}+\right.\\ &\qquad \left.\{\Gamma_1(t)\}^{\mathrm{T}}V_\gamma^{-1}\int_{[0,\infty)}\{X_i-\mu(u,\gamma_0)\}\,dM_i^{(0)}(u)\right]+o_p(1). \qquad (\text{A.31})\end{aligned}$$

Similarly, we can split (A.25) into two parts by adding and subtracting like terms:

$$\begin{aligned}-1\times(\text{A.25}) &= n^{-1/2}\sum_{i=1}^{n}\int_{[0,t)}\{m_0(u,X_i;\widehat{\alpha}_n)-m_0(u,X_i;\alpha_0)\}\,dG(u\mid X_i) \qquad (\text{A.32})\\ &\quad +n^{1/2}\sum_{i=1}^{n}\int_{[0,t)}m_0(u,X_i;\widehat{\alpha}_n)\,d\left\{\widehat{G}(u\mid X_i;\widehat{\gamma}_n)-G(u\mid X_i)\right\} \qquad (\text{A.33})\end{aligned}$$

Applying a Taylor-series approximation to the function $m_0(u,X_i;\widehat{\alpha}_n)$ about α_0, followed by Condition C1 and a law of large numbers leads to

$$(\text{A.32}) = n^{-1/2}\sum_{i=1}^{n}\{J_\alpha(t)\}^{\mathrm{T}}\varphi_\alpha(Z_i;\alpha_0)+o_p(1), \qquad (\text{A.34})$$

where

$$J_\alpha(t) = \int_0^t(\partial/\partial\alpha)m_0(u,X_1;\alpha_0)\,dG(u\mid X_1).$$

Under Condition C1 and the root-n consistency of $\widehat{\alpha}_n$, a Taylor-series approximation

yields

$$(\text{A.33}) = n^{-1/2} \sum_{i=1}^{n} \int_{[0,t)} m_0(u, X_i; \alpha_0)\, d\{\widehat{G}(u \mid X_i; \widehat{\gamma}_n) - G(u \mid X_i)\} + o_p(1).$$

Using Lemma 3 with $H_i(u) = m_0(u, X_i; \alpha_0)$, we have that

$$\begin{aligned}
(\text{A.33}) &= n^{-1/2} \sum_{i=1}^{n} (-1) \varphi_{\zeta, m_0}(Z_i; \gamma_0) \\
&\quad + \{\Omega_{2,m_0}(t)\}^{\mathrm{T}} V_\gamma^{-1} \int_0^\infty \{X_i - \mu(u, \gamma_0)\}\, dM_i^{(0)}(u) + o_p(1) \\
\varphi_{\zeta, m_0}(Z_i; \gamma_0) &= \int_{[0,t)} \frac{E\{m_0(u, X_1; \alpha_0) G(u \mid X_1) \exp(X_1^{\mathrm{T}} \gamma_0)\} dM_i^{(0)}(u)}{s^{(0)}(u, \gamma_0)} \\
&\quad \{K_{\zeta, m_0}(t) - J_{\zeta, m_0}(t)\}^{\mathrm{T}} V_\gamma^{-1} \int_0^\infty \{X_i - \mu(u, \gamma_0)\}\, dM_i(u), \\
K_{\zeta, m_0}(t) &= -\int_{[0,t)} E\{m_0(u, X_1; \alpha_0) G(u \mid X_1) \exp(X_1^{\mathrm{T}} \gamma_0)\} \mu(u, \gamma_0)\, d\Lambda_0(u) \\
J_{\zeta, m_0}(t) &= E\left[\int_{[0,t)} \left\{ \int_0^u \mu(s, \gamma_0)\, d\Lambda_0(s) \right\} \exp(X_1^{\mathrm{T}} \gamma_0)\, d\{m_0(u, X_1; \alpha_0) G(u \mid X_1)\} \right] \\
\Omega_{2,m_0}(t) &= \int_{[0,t)} E\left[m_0(u, X_1; \alpha_0) G(u \mid X_1) e^{X_1^{\mathrm{T}} \gamma_0} X_1 \right] d\Lambda_0(u).
\end{aligned}$$

Therefore,

$$(\text{A.23}) + (\text{A.25}) = n^{-1/2} \sum_{i=1}^{n} \varphi_\mu(t, Z_i; \vartheta_0) + o_p(1),$$

where

$$\begin{aligned}
\varphi_\mu(t, Z_i; \vartheta_0) &= \{L(t)\}^{\mathrm{T}} \varphi_\beta(Z_i; \beta_0) + E\{-w_1(t; Z_1, \theta_0)\} \int_{[0,t)} \frac{dM_i^{(0)}(u)}{s^{(0)}(u, \gamma_0)} \quad (\text{A.35}) \\
&\quad + \{\Gamma_1(t)\}^{\mathrm{T}} V_\gamma^{-1} \int_{[0,\infty)} \{X_i - \mu(u, \gamma_0)\}\, dM_i^{(0)}(u) \\
&\quad - \{J_\alpha(t)\}^{\mathrm{T}} \varphi_\alpha(Z_i; \alpha_0) + \varphi_{\zeta, m_0}(Z_i; \gamma_0) \\
&\quad - \{\Omega_{2,m_0}(t)\}^{\mathrm{T}} V_\gamma^{-1} \int_{[0,\infty)} \{X_i - \mu(u, \gamma_0)\}\, dM_i^{(0)}(u).
\end{aligned}$$

Therefore, by the central limit theorem and Slutzky's Theorem, $n^{1/2}\{\widehat{\mu}_n(t) - \mu_0(t)\}$ converges in distribution to a mean-zero normal random variable with covariance

$E(\varphi_\mu^{\otimes 2})$. □

A.2 Details of Binning Strategies for Inverse Probability Weighting Estimator

Table A.1: Simulation results for IPW estimator with varying binning strategy at continuous time setting

	Original				Centered by natural scale				Centered by log scale			
	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP	Bias	SD	SEE	ECP
						n=100						
t_1	-0.111	0.467	0.370	0.852	0.035	0.473	0.374	0.856	0.034	0.579	0.435	0.829
t_2	0.018	0.289	0.265	0.932	0.053	0.326	0.295	0.919	0.129	0.296	0.267	0.895
t_3	0.037	0.213	0.197	0.914	0.055	0.202	0.195	0.920	0.221	0.207	0.197	0.770
						n=200						
t_1	-0.114	0.346	0.312	0.894	0.029	0.346	0.309	0.914	0.029	0.417	0.384	0.918
t_2	0.025	0.204	0.196	0.946	0.056	0.237	0.225	0.934	0.134	0.205	0.197	0.880
t_3	0.036	0.152	0.142	0.922	0.054	0.141	0.141	0.920	0.220	0.146	0.141	0.639
						n=300						
t_1	-0.071	0.299	0.282	0.902	0.055	0.279	0.265	0.926	0.072	0.361	0.351	0.909
t_2	0.026	0.170	0.163	0.929	0.051	0.194	0.186	0.925	0.133	0.171	0.163	0.856
t_3	0.044	0.124	0.116	0.912	0.062	0.116	0.116	0.908	0.229	0.122	0.116	0.488

ECP, empirical coverage probability for Wald-type 95% confidence interval

To examine the effects of preparation strategy on the performance of inverse probability weighting estimator, we compared Johnson and Tsiatis (2004) estimator with three binning strategies respectively in simulation studies (Table A.1).

Under the same continuous time setting as in §2.4.2. Under 1000 Monte Carlo simulation iterations, three binning strategies were compared: 1) (*Original*) Use the mid point of two adjacent targeted time points on the log scale as the middle cut off points. The cut-off points on the two extreme are close to e^{-10} and e^{10}. This way, all continuous time points were classified into their closest targeted discrete time points. However, the bins created by this policy is not centered around their targeted time points, which may affect the consistency of the estimator. To investigate this

centering effect, another two binning strategies were adopted. 2) (*Centered on natural scale*) This strategy uses bins that are centered around its targeted time points on the natural scale of time. The width of bins on difference time points may be different since the spacing of the targeted time points are not identical. Some observations with observed time length very small or very large may be excluded from the analysis. 3)(*Centered on log scale*) this approach is essentially the same as last one except that the bins are centered around its targeted time points on the log scale of time instead of natural scale. Using our targeted treatment length as in § 2.4.2,There were only a tiny proportion of observations excluded from the analysis using strategy 2 and 3.

The simulation shows that strategy 2 is most appropriate for the IPW estimators, with good coverage probabilities for all targeted treatment lengths. The strategy 1 is also okay, although the consistency for the targeted treatment lengths on the lower or upper end could be improved. Strategy 3 works poor in this setting.

A.3 Details of ESPRIT Infusion Trial Data Analysis Results

The parameter estimates for outcome regression on adverse event within 30 days (for censored data and observed data respectively) and Cox proportional hazard regression models are listed in Table A.2. Statistically significant association is also found among patient demographics, infusion time and the outcome.

The proposed estimators constructed under Weibull and Cox proportional hazard models as well as the inverse probability weighting estimator with two binning strategies are illustrated in Figure 2.1 and Table A.3. While the point estimates of the four estimators are generally similar, our proposed estimators are substantially more efficient than IPW estimators overall. The results suggests that longer infusion lengths may corresponds to higher probability of encounter an adverse event within

Table A.2: Estimates of coefficients (and their standard errors) for outcome logistic regression on adverse event within 30 days and Cox proportional hazard regression with respect to infusion time

	α	β	γ
Infusion time	-0.293 (0.212)	2.226 (1.087)	NA
Diabetes	-0.555 (0.712)	-0.296 (0.368)	0.165 (0.245)
PTCA	0.901 (0.595)	0.530 (0.297)	-0.132 (0.247)
Angina	0.043 (0.593)	-0.402 (0.303)	-0.269 (0.218)
Heparin	0.368 (0.675)	0.258 (0.375)	0.173 (0.272)
Weight	-0.011 (0.282)	-0.016 (0.141)	-0.208 (0.106)

PTCA, percutaneous transluminal coronary angioplasty

Table A.3: Estimated event proportion (and their standard errors) for the ESPRIT trial.

	Direct		IPW	
Hours	PH	Extreme Value	strategy 1	strategy 2
16	0.053(0.010)	0.064(0.012)	0.040(0.016)	0.029(0.020)
18	0.064(0.008)	0.074(0.011)	0.066(0.010)	0.055(0.010)
20	0.076(0.008)	0.086(0.011)	0.078(0.017)	0.068(0.018)
22	0.089(0.012)	0.101(0.014)	0.071(0.024)	0.065(0.028)
24	0.102(0.019)	0.113(0.021)	0.121(0.035)	0.093(0.002)

IPW, inverse probability weighting

Strategy 1, *Original* in Table A.1

Strategy 2, *Centered by natural scale* in Table A.1

30 days.

A.4 Details of ACTG Data Analysis Results

The parameter estimates for outcome regression (OR) on CD4 cell count (for censored data and observed data respectively) and Cox proportional hazard (Cox PH) regression models are listed in Table A.4. This demonstrates that not only the time between virologic failure to switch is closely related to patients clinical atributes (eg. time to virologic failure), but also the CD4 cell count is closely associated with patients clinical characteristics (baseline CD4 cell count, time to virologic failure, etc.). This justifies the necessity of adjustment on those covariates in the estimation.

Table A.4: Parameter estimates for outcome regression on CD4 cell count and Cox proportional hazard regression with respect to time to switch

	OR(censored)	OR(observed)	Cox PH
Time to switch	0.03 (0.03)	-0.10 (0.04)	NA
HIV-1 RNA	-0.01 (0.03)	0.05 (0.05)	0.09 (0.18)
HIV-1 RNA at virologic failure	-0.04 (0.02)	-0.11 (0.04)	0.03 (0.11)
Time to failure	0.08 (0.02)	0.08 (0.03)	1.05 (0.14)
Baseline CD4 cell count	0.04 (0.00)	0.05 (0.01)	0.01 (0.03)
Baseline CD8 cell count	-0.01 (0.01)	0.03 (0.04)	-0.11 (0.09)
Body Weight	-0.01 (0.02)	-0.01 (0.04)	0.11 (0.14)
Age	0.00 (0.00)	0.00 (0.00)	-0.01 (0.01)
Sex	0.01 (0.05)	-0.02 (0.09)	-0.05 (0.31)
Drug History	-0.01 (0.06)	-0.13 (0.08)	0.64 (0.34)
Race			
White	-	-	-
Black	-0.05 (0.04)	-0.10 (0.08)	0.17 (0.26)
Hispanic or other	-0.02 (0.05)	-0.04 (0.10)	0.04 (0.29)

Table A.5 displays the data analysis results for CD4 outcome using a variety of estimators. For the outcome regression, we adopted the linear regression models (MLR), the spline models (Spline) and the generalized additive model (GAM). For spline models, we used cubic spline regression with respect to the time effect. Meanwhile, the generalized additive model was built by implementing a smoothing

Table A.5: ACTG data analysis with CD4 outcome

	Direct					
		PH		Extreme Value	IPW	
Hours	MLR	Spline	GAM	MLR	Strategy 1	Strategy 4
1	2.551(0.039)	2.694(0.139)	2.556(0.064)	2.550(0.042)	2.511(0.053)	2.533(0.085)
2	2.479(0.039)	2.675(0.117)	2.515(0.048)	2.479(0.038)	2.509(0.061)	2.365(0.130)
4	2.413(0.056)	2.607(0.074)	2.476(0.048)	2.409(0.048)	2.355(0.098)	1.975(0.044)
6	2.387(0.062)	2.555(0.050)	2.456(0.051)	2.379(0.053)		
8	2.375(0.065)	2.539(0.045)	2.443(0.052)	2.372(0.052)		
10	2.366(0.068)	2.523(0.038)	2.434(0.047)	2.377(0.048)		
12	2.371(0.064)	2.512(0.035)	2.428(0.045)	2.392(0.041)		
14	2.395(0.052)	2.499(0.030)	2.423(0.044)	2.411(0.035)		
16	2.418(0.041)	2.488(0.028)	2.421(0.044)	2.426(0.029)	2.451(0.035)	2.492(0.032)
18	2.443(0.033)	2.478(0.027)	2.419(0.043)	2.438(0.025)		
20	2.456(0.030)	2.475(0.027)	2.419(0.043)	2.449(0.023)		
22	2.466(0.028)	2.474(0.027)	2.418(0.043)	2.458(0.022)		
24	2.473(0.027)	2.474(0.028)	2.418(0.043)	2.465(0.023)		

MLR, multiple linear regression;GAM, generalized additive model.

function on time variable. For IPW estimator, we adopted the binning strategy 1 as mentioned earlier in Table A.1, and the binning strategy 4, which the same as strategy 2 except that the bins are not only symmetric around the targeted time point, but also have the same width. Observations that fall outside the bins are excluded from the analysis. For strategy 4, the number of events in the four intervals are 18, 9, 1, 2 respectively. This explains why the estimate at the third time interval is a little off. Out of 182 subjects, 127 were included by binning strategy 4. This shows that the performance of IPW estimator varies with different binning strategy and it is not easy to find an appropriate binning strategy. On the other hand, the results using our direct estimators are more precise, could be implemented relatively easily and estimate more time points. As illustrated in Figure A.1, the direct estimators based on spline models can detect more curvature than that based on linear regression. And they are both more flexible to detect different trends across time than the direct estimator based on the general additive models.

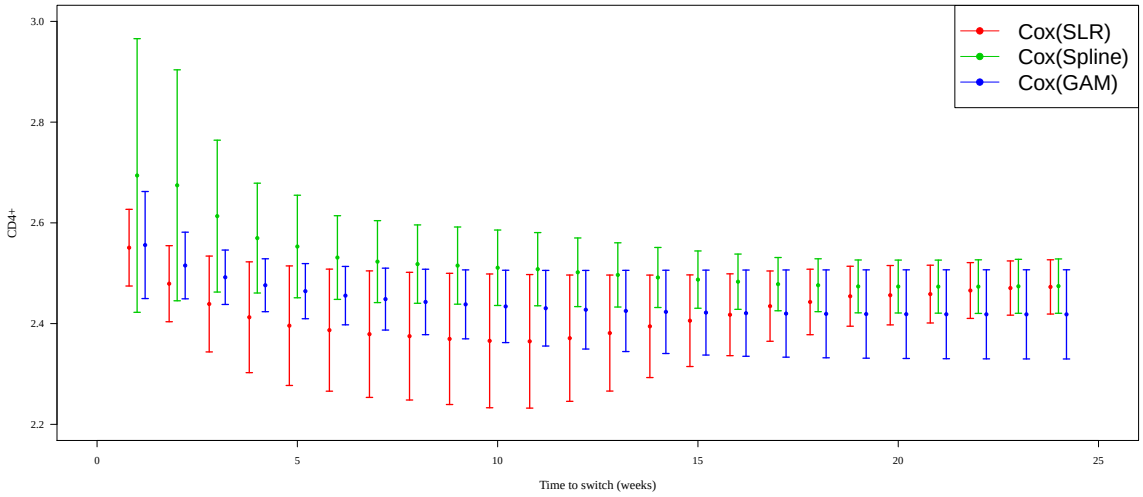

Figure A.1: Estimated CD4+ counts and their 95% confidence intervals across switch time using our direct estimator based on Cox PH models and linear regression (red), Cox PH models and spline regression (green), and Cox PH models and generalized additive models (blue).

Appendix B

Supplementary Material for Chapter 4

B.0.1 Derivation of doubly robust estimators

For estimating μ_2, we have the estimating equation for JT2004 estimator:

$$\begin{aligned}\Psi(X) &= \left[\frac{I(U=t_2,\Delta=1)}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}+\frac{I(t_1<U<t_2,\Delta=0)}{1-\lambda_1(\bar{X}_0)}+I(U<t_1,\Delta=0)\right](Y-\mu_2)\\ &= \left[\frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}+\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}+(1-S_0)\right](Y-\mu_2)\end{aligned}$$

Then the augmentation term is

$$\sum_{j=0}^{1}\{E_{obs}\left[\Psi(X)\mid\bar{A}_{j+1},\bar{L}_j\right]-E_{obs}\left[\Psi(X)\mid\bar{A}_j,\bar{L}_j\right]\}$$

Now we derive these condition expectations as follows:

For $j=1$ and for one observation,

$$
\begin{aligned}
& E_{obs}\left[\Psi(X) \mid \bar{A}_2, \bar{L}_1\right] \\
= \; & E\left[\frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}(Y-\mu_2) \mid A_2=1, C>t_2, \bar{X}_1\right] \\
& +E\left[\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}(Y-\mu_2) \mid A_1=0, t_1<C<t_2, \bar{X}_1\right] \\
& +E\left[(Y-\mu_2) \mid C<t_1, \bar{X}_1\right] \\
= \; & \frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}\left[E(Y \mid A_2=1, C>t_2, \bar{X}_1)-\mu_2\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid A_1=0, t_1<C<t_2, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid C<t_1, \bar{X}_1)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}\left[E(Y \mid U=t_2, \Delta=1, \bar{X}_1)-\mu_2\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right]
\end{aligned}
$$

Meanwhile,

$$
\begin{aligned}
& E_{obs}\left[\Psi(X) \mid \bar{A}_1, \bar{L}_1\right] \\
= \; & E\left[\frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}(Y-\mu_2) \mid A_1=0, C>t_2, \bar{X}_1\right] \\
& +E\left[\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}(Y-\mu_2) \mid A_1=0, t_1<C<t_2, \bar{X}_1\right] \\
& +E\left[(1-S_0)(Y-\mu_2) \mid C<t_1, \bar{X}_0\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}E\left[A_2(Y-\mu_2) \mid A_1=0, C>t_2, \bar{X}_1\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}E\left[(Y-\mu_2) \mid A_1=0, t_1<C<t_2, \bar{X}_1\right] \\
& +(1-S_0)\left[E(Y \mid C<t_1, \bar{X}_0)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}E\left[E\{A_2(Y-\mu_2) \mid A_2=1, C>t_2, \bar{X}_1\} \mid A_1=0, C>t_2, \bar{X}_1\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}E\left[Y-\mu_2 \mid A_1=0, t_1<C<t_2, \bar{X}_1\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}E\left[A_2E\{(Y-\mu_2) \mid A_2=1, C>t_2, \bar{X}_1\} \mid A_1=0, C>t_2, \bar{X}_1\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid A_1=0, t_1<C<t_2, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}Pr(U=t_2, \mid U \geq t_2, \bar{X}_1)\left[E\{E(Y \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_1\}-\mu_2\right] \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}\lambda_2(\bar{X}_1)\{E(Y \mid A_2=1, C>t_2, \bar{X}_1)-\mu_2\} \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right] \\
= \; & \frac{S_0(1-A_1)S_1}{(1-\lambda_1(\bar{X}_0))}\{E(Y \mid U=t_2, \Delta=1, \bar{X}_1)-\mu_2\} \\
& +\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}\left[E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)-\mu_2\right] \\
& +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_1)-\mu_2\right]
\end{aligned}
$$

Therefore,

$$\begin{aligned} & E_{obs}\left[\Psi(X) \mid \bar{A}_2, \bar{L}_1\right] - E_{obs}\left[\Psi(X) \mid \bar{A}_1, \bar{L}_1\right] \\ = & \frac{I(U = t_2, \Delta = 1) - \lambda_2(\bar{X}_1) I(U \geq t_2)}{(1 - \lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)} \{E(Y \mid U = t_2, \Delta = 1, \bar{X}_1) - \mu_2\} \end{aligned}$$

which equals 0 due to $\lambda_2(\bar{X}_1) = 1$ and $I(U = t_2, \Delta = 1) = I(U \geq t_2)$.

Note that the quantities for $(1 - S_0)$ and $S_0(1 - A_1)(1 - S_1)$ is the same in two conditional expectations respectively, so they are canceled out in the augmentation term.

The above equation could also be written as Murphy's format in (A.2) as

$$\left[I(A_2 = a_2 = 1) - Pr(a_2 = 1 \mid \bar{A}_1, \bar{L}_1)\right] \cdot \frac{I(U \geq t_2)}{(1 - \lambda_1 \bar{X}_0)\lambda_2(\bar{X}_1)} \cdot \{g_2^{(2)}(a_2 = 1, \bar{A}_1, \bar{L}_1)) - \mu_2\}$$

where $\frac{I(U \geq t_2)}{(1 - \lambda_1 \bar{X}_0)\ \lambda_2(\bar{X}_1)} = W_{\bar{p}_2}(a_2 = 1, \bar{A}_1, \bar{L}_1)$

Note that a_2 could not be 0 since everyone must stop at t_2 if not earlier.

For $j = 0$,

$$\begin{aligned} & E_{obs}\left[\Psi(X) \mid \bar{A}_1, \bar{L}_0\right] \\ = & E\left[\frac{S_0(1 - A_1) S_1 A_2}{(1 - \lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)} (Y - \mu_2) \mid A_1 = 0, C > t_1, \bar{X}_0\right] \\ & + E\left[\frac{S_0(1 - A_1)(1 - S_1), \Delta = 0)}{1 - \lambda_1(\bar{X}_0)} (Y - \mu_2) \mid A_1 = 0, C > t_1, \bar{X}_0\right] \\ & + E\left[(1 - S_0)(Y - \mu_2) \mid C < t_1, \bar{X}_0\right] \\ = & \frac{S_0(1 - A_1)}{(1 - \lambda_1(\bar{X}_0))} E\left[\frac{S_1 A_2}{\lambda_2(\bar{X}_1)} (Y - \mu_2) \mid A_1 = 0, C > t_1, \bar{X}_0\right] \\ & + \frac{S_0(1 - A_1)}{1 - \lambda_1(\bar{X}_0)} E\left[(1 - S_1)(Y - \mu_2) \mid A_1 = 0, C > t_1, \bar{X}_0\right] \\ & + (1 - S_0)\left[E(Y \mid C < t_1, \bar{X}_0) - \mu_2\right] \end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} E\left[E(\frac{S_1A_2}{\lambda_2(\bar{X}_1)}\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + \frac{S_0(1-A_1)}{1-\lambda_1(\bar{X}_0)} E\left[E((1-S_1)\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} E\left[\frac{S_1A_2}{\lambda_2(\bar{X}_1)} E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + \frac{S_0(1-A_1)}{1-\lambda_1(\bar{X}_0)} E\left[(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} \frac{Pr(U = t_2, \Delta = 1 \mid U \geq t_2, \bar{X}_0)}{E\{\lambda_2(\bar{X}_1) \mid \bar{X}_0\}} E\left[S_1E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + \frac{S_0(1-A_1)}{1-\lambda_1(\bar{X}_0)} E\left[(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} E\left[S_1E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + \frac{S_0(1-A_1)}{1-\lambda_1(\bar{X}_0)} E\left[(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} E\left[E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} \left[E\{E(Y \mid \bar{A}_2, \bar{L}_1) \mid \bar{A}_1, \bar{L}_0\} - \mu_2\right] \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right] \\
= \quad & \frac{S_0(1-A_1)}{(1-\lambda_1(\bar{X}_0))} (E\{S_1E(Y \mid U = t_2, \Delta = 1, \bar{X}_1) \\
& + (1-S_1)E(Y \mid t_1 < U < t_2, \Delta = 0, \bar{X}_1)\} - \mu_2) \\
& + (1-S_0)\left[E(Y \mid U < t_1, \Delta = 0, \bar{X}_0) - \mu_2\right]
\end{aligned}
$$

$$
\begin{aligned}
& E_{obs}\left[\Psi(X) \mid \bar{L}_0\right] \\
= \; & E\left[\frac{S_0(1-A_1)S_1A_2}{(1-\lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)}(Y-\mu_2) \mid \bar{L}_0\right] \\
& +E\left[\frac{S_0(1-A_1)(1-S_1)}{1-\lambda_1(\bar{X}_0)}(Y-\mu_2) \mid \bar{L}_0\right] \\
& +E\left[(1-S_0)(Y-\mu_2) \mid \bar{L}_0\right] \\
= \; & \frac{S_0}{(1-\lambda_1(\bar{X}_0))}E\left[\frac{(1-A_1)S_1A_2}{\lambda_2(\bar{X}_1)}(Y-\mu_2) \mid \bar{L}_0\right] \\
& +\frac{S_0}{1-\lambda_1(\bar{X}_0)}E\left[(1-A_1)(1-S_1)(Y-\mu_2) \mid \bar{L}_0\right] \\
& +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
= \; & \frac{S_0}{(1-\lambda_1(\bar{X}_0))}E\left[E(\frac{(1-A_1)S_1A_2}{\lambda_2(\bar{X}_1)}\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +\frac{S_0}{1-\lambda_1(\bar{X}_0)}E\left[E((1-A_1)(1-S_1)\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
= \; & \frac{S_0}{(1-\lambda_1(\bar{X}_0))}E\left[\frac{(1-A_1)S_1A_2}{\lambda_2(\bar{X}_1)}E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +\frac{S_0}{1-\lambda_1(\bar{X}_0)}E\left[(1-A_1)(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
= \; & \frac{S_0}{(1-\lambda_1(\bar{X}_0))}\cdot\frac{(1-Pr(U=t_1 \mid U\geq t_1,\bar{L}_0))Pr(U=t_2 \mid U\geq t_2,\bar{L}_0)}{E\{\lambda_2(\bar{X}_1) \mid \bar{L}_0\}} \\
& \cdot E\left[S_1E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +\frac{S_0}{1-\lambda_1(\bar{X}_0)}(1-Pr(U=t_1 \mid U\geq t_1,\bar{L}_0))E\left[(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
= \; & \frac{S_0}{(1-\lambda_1(\bar{X}_0))}(1-\lambda_1(\bar{X}_0))E\left[S_1E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +\frac{S_0}{1-\lambda_1(\bar{X}_0)}(1-\lambda_1(\bar{X}_0))E\left[(1-S_1)E(\{Y-\mu_2\} \mid \bar{A}_2,\bar{L}_1) \mid \bar{L}_0\right] \\
& +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{S_0}{(1-\lambda_1(\bar{X}_0))}(1-\lambda_1(\bar{X}_0))E\left[E(\{Y-\mu_2\} \mid \bar{A}_2, \bar{L}_1) \mid \bar{L}_0\right] \\
&\quad +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
&= S_0\left[E\{E(Y \mid \bar{A}_2, \bar{L}_1) \mid \bar{L}_0\}-\mu_2\right] \\
&\quad +(1-S_0)\left[E(Y \mid \bar{L}_0)-\mu_2\right] \\
&= S_0(E\{S_1E(Y \mid U=t_2, \Delta=1, \bar{X}_1) \\
&\quad +(1-S_1)E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)\}-\mu_2) \\
&\quad +(1-S_0)\left[E(Y \mid U<t_1, \Delta=0, \bar{X}_0)-\mu_2\right]
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& E_{obs}\left[\Psi(X) \mid \bar{A}_1, \bar{L}_0\right]-E_{obs}\left[\Psi(X) \mid \bar{L}_0\right] \\
= & \frac{\{(1-A_1)-(1-\lambda_1(\bar{X}_0)\}S_0}{1-\lambda_1(\bar{X}_0)}\left[E\{E(Y \mid U>t_1, \bar{X}_1) \mid \bar{L}_0\}-\mu_2\right] \\
= & \frac{\{I(U>t_1)-(1-\lambda_1(\bar{X}_0)\}I(U \geq t_1)}{1-\lambda_1(\bar{X}_0)}\{E\{S_1E(Y \mid U=t_2, \Delta=1, \bar{X}_1) \\
+ & (1-S_1)E(Y \mid t_1<U<t_2, \Delta=0, \bar{X}_1)\}-\mu_2\}
\end{aligned}
$$

This is equivalent to

$$
\left[I(A_1=a_1=0)-Pr(a_1=0 \mid \bar{L}_0)\right] \cdot \frac{I(U \geq t_1)}{(1-\lambda_1(\bar{X}_0))} \cdot \{g_1^{(2)}(a_1=0, \bar{L}_0)-\mu_2\}
$$

where $\frac{I(U \geq t_1)}{(1-\lambda_1(\bar{X}_0))}=W_{\bar{p}_1}(a_1=0, \bar{L}_0)$

Note that when $a_1=1$, $W_{\bar{p}_1}(a_1, \bar{L}_0)=0$, so

$$
\left[I(A_1=a_1=0)-Pr(a_1=0 \mid \bar{L}_0)\right] W_{\bar{p}_1}(a_1=0, \bar{L}_0)\{g_1^{(2)}(a_1=1, \bar{L}_0)-\mu_2\}=0.
$$

To sum up, the augmentation term is

$$\begin{aligned}
&\left[I(A_1 = a_1 = 0) - Pr(a_1 = 0 \mid \bar{L}_0)\right] \cdot \frac{I(U \geq t_1)}{(1 - \lambda_1(\bar{X}_0))} \cdot \{g_1^{(2)}(a_1 = 0, \bar{L}_0) - \mu_2\} \\
&+ \left[I(A_2 = a_2 = 1) - Pr(a_2 = 1 \mid \bar{A}_1, \bar{L}_1)\right] \\
&\cdot \frac{I(U \geq t_2)}{(1 - \lambda_1 \bar{X}_0)\lambda_2(\bar{X}_1)} \cdot \{g_2^{(2)}(a_2 = 1, \bar{A}_1, \bar{L}_1) - \mu_2\}
\end{aligned}$$

in Murphy's format

and is

$$\begin{aligned}
&\frac{I(U > t_1) - (1 - \lambda_1(\bar{X}_0))I(U \geq t_1)}{1 - \lambda_1(\bar{X}_0)} \{g_1^{(2)}(a_1 = 0, \bar{L}_0) - \mu_2\} \\
&+ \frac{I(U = t_2, \Delta = 1) - \lambda_2(\bar{X}_1)I(U \geq t_2)}{(1 - \lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)} \{g_2^{(2)}(a_2 = 1, \bar{A}_1, \bar{L}_1) - \mu_2\} \\
= \; &- \frac{I(U = t_1, \Delta = 1) - \lambda_1(\bar{X}_0)I(U \geq t_1)}{1 - \lambda_1(\bar{X}_0)} \{g_1^{(2)}(a_1 = 0, \bar{L}_0) - \mu_2\} \\
&+ \frac{I(U = t_2, \Delta = 1) - \lambda_2(\bar{X}_1)I(U \geq t_2)}{(1 - \lambda_1(\bar{X}_0))\lambda_2(\bar{X}_1)} \{g_2^{(2)}(a_2 = 1, \bar{A}_1, \bar{L}_1) - \mu_2\}
\end{aligned}$$

in Tsiatis's format.

And is equivalent to

$$\begin{aligned}
&\{W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0) - 1\} \left[g_1^{(2)}(a_1 = 0, \bar{L}_0) - \mu_2\right] \\
&+ \{W_{\bar{p}_2}(\bar{A}_2, \bar{L}_1) - W_{\bar{p}_1}(\bar{A}_1, \bar{L}_0)\} \left[g_2^{(2)}(a_2 = 1, \bar{A}_1, \bar{L}_0) - \mu_2\right]
\end{aligned}$$

Note that when $\lambda_2(\bar{X}_1) = 1$, the latter half of the augmentation term is zero.

For estimating μ_1, we have the estimating equation for JT2004 estimator:

$$\begin{aligned}\Psi(X) &= \left[\frac{I(U=t_1,\Delta=1)}{\lambda_1(\bar{X}_0)}+I(U<t_1,\Delta=0)\right](Y-\mu_1)\\ &= \left[\frac{S_0A_1}{\lambda_1(\bar{X}_0)}+(1-S_0)\right](Y-\mu_1)\end{aligned}$$

Following similar arguments, we can get the augmentation term

$$\begin{aligned}&\frac{(A_1-\lambda_1(\bar{X}_0))S_0}{\lambda_1(\bar{X}_0)}\{E(Y\mid U=t_1,\Delta=1,\bar{X}_0)-\mu_1\}\\ =\;&\frac{I(U=t_1,\Delta=1)-\lambda_1(\bar{X}_0)I(U\geq t_1)}{\lambda_1(\bar{X}_0)}\{E(Y\mid U=t_1,\Delta=1,\bar{X}_0)-\mu_1\}\end{aligned}$$

which is equivalent to

$$\left[I(A_1=a_1=1)-Pr(a_1=1\mid\bar{L}_0)\right]\cdot\frac{I(U\geq t_1)}{\lambda_1(\bar{X}_0)}\cdot\{g_1^{(1)}(a_1=1,\bar{L}_0)-\mu_1\}$$

And also equivalent to

$$\{W_{\bar{p}_1}(\bar{A}_1,\bar{L}_0)-1\}\left[g_1^{(1)}(\bar{L}_0)-\mu_1\right]$$

B.0.2 Connection of 4.2.3 to Chapter 1

Under regime 1 in § 4.2.3 , (4.8) could be further written as

$$\begin{aligned}\mu_1 &= E(Y|U,\Delta,X_0)\\ &= E\left[E(Y|U=t_1,\Delta=1,X_0)I(C>t_1)+E(Y|U\leq t_1,\Delta=0,X_0)I(C\leq t_1)\right].\end{aligned}$$

Because

$$E(Y|U \le t_1, \Delta = 0, X_0) = \frac{1}{P(C \le t_1|X_0)} \int_0^{t_1} E(Y|U = u, \Delta = 0, X_0)\, dP(C \le u|X_0)$$
$$E\left[E(Y|U = t_1, \Delta = 1, X_0)I(C > t_1)\right] = E\left[E(Y|U = t_1, \Delta = 1, X_0)P(C > t_1|X_0)\right],$$

we have the equivalent expression that

$$\mu_1 = E_{X_0}\left[m_1(t_1, X_0)P(C > t_1|X_0) + \int_0^{t_1} m_0(u, X_0)\, dP(C \le u|X_0)\right], \qquad \text{(B.1)}$$

where $m_1(t, X_0) = E(Y|U = t, \Delta = 1, X_0)$ and $m_0(t, X_0) = E(Y|U = t, \Delta = 0, X_0)$ as defined in our Chapter 1.

Appendix C

R codes

C.1 R codes for Chapter 3

```
#program for illustrating the time-dependent covariates
#scenario in notes 3/14/2014
# revised 10/29/2014
# note that U= midpoint of L intervals in observation
#censoring during the interval.


lucy.dat <- function(n=5,p,L,p0=2,eff=1,diff=0.5){


prob0=prob=x=C=T=d=U=lucy.y=matrix(0,n,p)


x0=matrix(rnorm(n*p0),n,p0)   # fixed covariates x0
b0=matrix(0.1,p0,1)
b=matrix(0.2,p0,1)
```

```
a0= 1.3*seq(1:p)/eff
a= seq(1:p)/eff

#-----------------------------------------------------
# baseline info
#-----------------------------------------------------

x[,1]=4+rnorm(n)

prob0[,1]=1/(exp(-a0[1]+0.9*x[,1]+x0%*%b0)+1)
C[,1]=rbinom(n,1,prob0[,1])

prob[,1]=1/(exp(-a[1]+0.8*x[,1]+x0%*%b)+1)
T[,1]=rbinom(n,1,prob[,1])
T[,1]=ifelse(C[,1]==0,T[,1],NA)

tmp1=L[1]/2
U[,1]=ifelse(C[,1]==1,tmp1, L[1])

xt=cbind(0,x[,1])
#lucy.y[,1]=-2.5+0.5*x0[,1]+0.5*x0[,2]-5*x[,1]+2*(1-d)
#---------------------------------
# follow-up
#---------------------------------
for (i in 2:p){
x[,i]=diff+rnorm(n)+x[,i-1]
x[,i]=ifelse(T[,i-1]==0,x[,i],NA)
```

```
      prob0[,i]=1/(exp(-a0[i]+0.9*x[,i]+x0%*%b0)+1)
C[,i]=rbinom(n,1,prob0[,i])

prob[,i]=1/(exp(-a[i]+0.8*x[,i]+x0%*%b)+1)
T[,i]=rbinom(n,1,prob[,i])
T[,i]=ifelse(C[,i]==0,T[,i],NA)



tmp2=(L[i-1]+L[i])/2
U[,i]=ifelse(C[,i]==1,tmp2,L[i])
xtmp=cbind(L[i-1],x[,i])
xt=rbind(xt,xtmp)
}

Lmid=(L+c(0,L[-length(L)]))/2
L1=rep(L,each=n)
L2=rep(Lmid,each=n)
U2=apply(U,1,function(x) max(x,na.rm=TRUE))




T[,p]=ifelse(U2==L[p],1,NA)
d=rowSums(T,na.rm=TRUE)
colnames(xt)=c("t",paste("xt",seq(1:(dim(xt)[2]-1)),sep=""))
id=seq(1:n)

tmpx=x  # processed below to represent most recently xt
```

```
for (j in 1:(p)){
if(j==1){tmpx[,j]=x[,1]}
else{tmpx[,j]=apply(x[,1:j], 1,function(x) x[length(x[!is.na(x)])])}
}


#identify the last non-missing x
xx=apply(x,1,function(x) x[length(x[!is.na(x)])])

y=1.5+x0[,1]+0.5*x0[,2]-2*d+log(U2)+0.5*xx+rnorm(n)

dat.brent=data.frame(id=id,L=L1,U=U2,d=d,x0=x0,xt=xt,Lmid=L2,y=y,xx=xx)
dat.brent$yb=as.numeric(dat.brent$U == L & dat.brent$d == 1)

return(list(U=U2,d=d,T=T,C=C,x0=x0,xt=xt,tmpx=tmpx,y=y,x=x,
xx=xx,id=id,df=dat.brent, prob0=prob0, prob=prob))
}
##########################################################
#-----------Simulated Truth----------------------------------
p.sim <-length(L)
n.sim=500000

sim.x0 <- matrix(NA,p.sim,n.sim)
sim.lambda <- matrix(NA,p.sim,n.sim)
sim.Cens <- sim.U<-sim.Delta<-matrix(NA, p.sim, n.sim)


sim.x1 <- rnorm(n.sim,0,1)
sim.x2 <- rnorm(n.sim,0,1)
```

```
Lmid=(L+c(0,L[-p.sim]))/2

for (k in 1){

sim.x0[k,] <- rnorm(n.sim,(3.5+0.5*k),1)
sim.lambda[k,] <- 1/(1+exp(-1.3*k/eff+0.9*sim.x0[k,]+0.1*sim.x1+0.1*sim.x2))
sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.U[k,]=ifelse(sim.Cens[k,]==1,Lmid[k], L[k])
sim.Delta[k,]= ifelse(sim.Cens[k,]==0,1,0)

}

for (k in 2:p.sim){

sim.x0[k,] <- diff+rnorm(n.sim)+sim.x0[k-1,]
sim.lambda[k,] <- 1/(1+exp(-1.3*k/eff+0.9*sim.x0[k,]+0.1*sim.x1+0.1*sim.x2))

sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.Cens[k,]=ifelse(sim.Cens[k-1,]==0,sim.Cens[k,],NA)

sim.Delta[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),0,1)

sim.U[k,]=ifelse(is.na(sim.Cens[k,]),sim.U[k-1,],Lmid[k])
sim.U[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),sim.U[k,],L[k])
}
x1rep=matrix(sim.x1,p.sim,n.sim,byrow=TRUE)
x2rep=matrix(sim.x2,p.sim,n.sim,byrow=TRUE)
```

```
sim.Y <- 1.5 +x1rep+0.5*x2rep-2*sim.Delta+0.5*sim.x0 +log(sim.U)

#sim.Y <- 1.5 +x1rep+0.5*x2rep+0.5*sim.x0

trumu<-rowMeans(sim.Y)


t1=matrix(sim.Y,ncol=1)

t2=rep(L,n.sim)

t3=data.frame(Y=t1,L=t2)


ft.t=lm(Y~L,data=t3)

trumu2=predict(ft.t,newdata=list(L=L))

trubeta=ft.t$coeff


#-----------------------------------------------------------------------

# 5_11_2014

#-----------------------------------------------------------------------


Lucy.gcomp<- function(x0,time.x,x,u,d,y,L,Lx){


n=length(u)

#------------estimation of parameters

ft3=lucy.discrete.haz.dep.cens(u=u,d=1-d,L=L,x0=x0,time.x=time.x,Lx=Lx)

alpha=ft3$beta[1:4]

bxt=ft3$beta[7]

b0=ft3$beta[5:6]


mu.x=colMeans(x0)

s.x=apply(x0,2,sd)


xt10=x[,1]
```

```
mu.x0=mean(xt10)
s.x0=sd(xt10)

t20=!is.na(x[,2])
xt10.20=xt10[t20]
xt20.20=x[,2][t20]

t30=!is.na(x[,3])
xt20.30=x[,2][t30]
xt30.30=x[,3][t30]

t40=!is.na(x[,4])
xt30.40=x[,3][t40]
xt40.40=x[,4][t40]

x.mk=c(xt10.20,xt20.30,xt30.40)
y.mk=c(xt20.20,xt30.30,xt40.40)
ft.mk=lm(y.mk~x.mk)
mk.a=ft.mk$coeff[1]
mk.b=ft.mk$coeff[2]

m1=mu.x0
m2=m1*mk.b+mk.a
m3=m2*mk.b+mk.a
m4=m3*mk.b+mk.a
mt=c(m1,m2,m3,m4)

#identify the last non-missing x
last.x=apply(x,1,function(x) x[length(x[!is.na(x)])])
```

```
ftlm=lm(y ~ x0+d+last.x+log(u))
cf=ftlm$coef

#_________________MC integration_________________

p.sim <-length(L)
n.sim=1000

sim.x0 <- matrix(NA,p.sim,n.sim)
sim.lambda <- matrix(NA,p.sim,n.sim)
sim.Cens <- sim.U<-sim.Delta<-matrix(NA, p.sim, n.sim)



sim.x1 <- rnorm(n.sim,mu.x[1],s.x[1])
sim.x2 <- rnorm(n.sim,mu.x[2],s.x[2])

Lmid=(L+c(0,L[-p.sim]))/2

for (k in 1){

sim.x0[k,] <- rnorm(n.sim,mt[k],s.x0)
sim.lambda[k,] <- 1/(1+exp(-alpha[k]-sim.x0[k,]*bxt-sim.x1*b0[1]-sim.x2*b0[2]))
sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.U[k,]=ifelse(sim.Cens[k,]==1,Lmid[k], L[k])
sim.Delta[k,]= ifelse(sim.Cens[k,]==0,1,0)
}

for (k in 2:p.sim){
```

```
sim.x0[k,] <- rnorm(n.sim,mk.a,s.x0)+sim.x0[k-1,]*mk.b
sim.lambda[k,] <- 1/(1+exp(-alpha[k]-sim.x0[k,]*bxt-sim.x1*b0[1]
-sim.x2*b0[2]))

sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.Cens[k,]=ifelse(sim.Cens[k-1,]==0,sim.Cens[k,],NA)

sim.Delta[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),0,1)

sim.U[k,]=ifelse(is.na(sim.Cens[k,]),sim.U[k-1,],Lmid[k])
sim.U[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),sim.U[k,],L[k])
}
x1rep=matrix(sim.x1,p.sim,n.sim,byrow=TRUE)
x2rep=matrix(sim.x2,p.sim,n.sim,byrow=TRUE)

sim.Y <- cf[1] +cf[2]*x1rep+cf[3]*x2rep+ cf[4]*sim.Delta+cf[5]*sim.x0
+cf[6]*log(sim.U)



mu_hat <- rowMeans(sim.Y)

t1=matrix(sim.Y,ncol=1)
t2=rep(L,n.sim)
t3=data.frame(Y=t1,L=t2)

ft.t=lm(Y~L,data=t3)
beta_hat=ft.t$coeff
mu_hat_model=predict(ft.t,newdata=list(L=L))
```

```
return(list(mu=mu_hat,beta=ft3$beta, mu2=mu_hat_model,b=beta_hat))

}




#-------------------------------------------------------------------
# 8_15_2014
#-------------------------------------------------------------------

Lucy.gcomp.pred<- function(x0,time.x,x,u,d,y,L,Lx){

n=length(u)
#------------estimation of parameters
ft3=lucy.discrete.haz.dep.cens(u=u,d=1-d,L=L,x0=x0,time.x=time.x,Lx=Lx)
alpha=ft3$beta[1:4]
bxt=ft3$beta[7]
b0=ft3$beta[5:6]

mu.x=colMeans(x0)
s.x=apply(x0,2,sd)

xt10=x[,1]
mu.x0=mean(xt10)
s.x0=sd(xt10)

t20=!is.na(x[,2])
```

```
xt10.20=xt10[t20]
xt20.20=x[,2][t20]

t30=!is.na(x[,3])
xt20.30=x[,2][t30]
xt30.30=x[,3][t30]

t40=!is.na(x[,4])
xt30.40=x[,3][t40]
xt40.40=x[,4][t40]

x.mk=c(xt10.20,xt20.30,xt30.40)
y.mk=c(xt20.20,xt30.30,xt40.40)
ft.mk=lm(y.mk~x.mk)
mk.a=ft.mk$coeff[1]
mk.b=ft.mk$coeff[2]

m1=mu.x0
m2=m1*mk.b+mk.a
m3=m2*mk.b+mk.a
m4=m3*mk.b+mk.a
mt=c(m1,m2,m3,m4)

#identify the last non-missing x
last.x=apply(x,1,function(x) x[length(x[!is.na(x)])])

ftlm=lm(y ~ x0+d+last.x+log(u))
cf=ftlm$coef
```

```
#_________________prediction_________________

predY=pred.lambda=pred.prob=matrix(NA,length(L),n)
tmpd=matrix(d,n,length(L))
tmpu=matrix(u,n,length(L))
tmpL=matrix(L,n,length(L),byrow=TRUE)

tmpp=tmpu<=tmpL
tmpd=ifelse(tmpp,tmpd,1)
tmpu=ifelse(tmpp,tmpu,tmpL)



myx2=x[,1:2]
tmpx2=apply(myx2,1,function(x) x[length(x[!is.na(x)])])
myx3=x[,1:3]
tmpx3=apply(myx3,1,function(x) x[length(x[!is.na(x)])])
myx4=x[,1:4]
tmpx4=apply(myx4,1,function(x) x[length(x[!is.na(x)])])

pred.lambda[1,] <- 1/(1+exp(-alpha[1]-x[,1]*bxt-x0[,1]*b0[1]-x0[,2]*b0[2]))
pred.lambda[2,] <- 1/(1+exp(-alpha[2]-tmpx2*bxt-x0[,1]*b0[1]-x0[,2]*b0[2]))
pred.lambda[3,] <- 1/(1+exp(-alpha[3]-tmpx3*bxt-x0[,1]*b0[1]-x0[,2]*b0[2]))
pred.lambda[4,] <- 1/(1+exp(-alpha[4]-tmpx4*bxt-x0[,1]*b0[1]-x0[,2]*b0[2]))



p1.1=(1-pred.lambda[1,])
p1.0=pred.lambda[1,]

part1=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[4]+cf[5]*x[,1] +cf[6]*log(L[1])
```

```
part2=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+cf[5]*x[,1] +cf[6]*log(Lmid[1])

predY[1,]=part1*p1.1+part2*p1.0


p2.1=(1-pred.lambda[1,])*(1-pred.lambda[2,])
p2.0a=pred.lambda[1,]
p2.0b=(1-pred.lambda[1,])*pred.lambda[2,]


m2.1=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[4]+cf[5]*tmpx2 +cf[6]*log(L[2])
m2.0a=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx2 +cf[6]*log(Lmid[1])
m2.0b=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx2 +cf[6]*log(Lmid[2])

predY[2,]=p2.1*m2.1+p2.0a*m2.0a+p2.0b*m2.0b


p3.1=(1-pred.lambda[1,])*(1-pred.lambda[2,])*(1-pred.lambda[3,])
p3.0a=pred.lambda[1,]
p3.0b=(1-pred.lambda[1,])*pred.lambda[2,]
p3.0c=(1-pred.lambda[1,])*(1-pred.lambda[2,])*pred.lambda[3,]

m3.1=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[4]+cf[5]*tmpx3 +cf[6]*log(L[3])
m3.0a=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx3 +cf[6]*log(Lmid[1])
m3.0b=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx3 +cf[6]*log(Lmid[2])
m3.0c=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx3 +cf[6]*log(Lmid[3])

predY[3,]=p3.1*m3.1+p3.0a*m3.0a+p3.0b*m3.0b+p3.0c*m3.0c
```

```
p4.1=(1-pred.lambda[1,])*(1-pred.lambda[2,])*(1-pred.lambda[3,])*(
1-pred.lambda[4,])
p4.0a=pred.lambda[1,]
p4.0b=(1-pred.lambda[1,])*pred.lambda[2,]
p4.0c=(1-pred.lambda[1,])*(1-pred.lambda[2,])*pred.lambda[3,]
p4.0d=(1-pred.lambda[1,])*(1-pred.lambda[2,])*(1-pred.lambda[3,]
)*pred.lambda[4,]

m4.1=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[4]+cf[5]*tmpx4 +cf[6]*log(L[4])
m4.0a=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx4 +cf[6]*log(Lmid[1])
m4.0b=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx4 +cf[6]*log(Lmid[2])
m4.0c=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx4 +cf[6]*log(Lmid[3])
m4.0d=cf[1] +cf[2]*x0[,1]+cf[3]*x0[,2]+ cf[5]*tmpx4 +cf[6]*log(Lmid[4])

predY[4,]=p4.1*m4.1+p4.0a*m4.0a+p4.0b*m4.0b+p4.0c*m4.0c+p4.0d*m4.0d

mu_hat <- rowMeans(predY)



t1=matrix(predY,ncol=1)
t2=rep(L,n)
t3=data.frame(Y=t1,L=t2)

ft.t=lm(Y~L,data=t3)
beta_hat=ft.t$coeff
mu_hat_model=predict(ft.t,newdata=list(L=L))

return(list(mu=mu_hat,beta=ft3$beta, mu2=mu_hat_model,b=beta_hat))
```

```
}
```

```
####################################
# wrote 4_18_2014
# add time dependent time.x

#first column of time.x is Lx, it is the time when time-dep
# covariates are measured
# .cens is when considering censoring rather than event.
#Number of time points in the models are different

lucy.discrete.haz.dep.cens <- function (x0=NULL,time.x=NULL,u,d,L,Lx){

K=length(L)
a0=rep(NA,K)
n=length(u)

if(is.null(x0)){
p0=0
}else{
x0=as.matrix(x0)
p0= dim(x0)[2]
}

if(is.null(time.x)){
```

```
pt=0
}else{
time.x=as.matrix(time.x)
pt0= (dim(time.x)[2]-1)
pt= pt0
}



#----get initial value for intercept-----------
LL=c(0,L[1:(K-1)])
midpt=(L+LL)/2

ind1= matrix(u,n,K)==matrix(midpt,n,K,byrow=TRUE) & matrix(d,n,K)==1
ind2= matrix(u,n,K)>=matrix(LL,n,K,byrow=TRUE)

p1= colSums(ind1)/colSums(ind2)
p1=ifelse(p1==0,0.000001,p1)
#print(p1)
a0= log(p1/(1-p1))
#print(a0)
xnew0=as.matrix(cbind(1,x0))

p=p0+pt
theta2=matrix(c(a0,rep(0,p)),p+K,1)



theta=0
iter <- 0
itermax=150
```

```
tolerance=0.001

#---Newton Ralphson--------------------------------
while (sum(abs(theta - theta2)) >= tolerance) {
#print(abs(theta - theta2)[4])
iter=iter+1
#print(paste("iter =",iter))



Ha=Sa=matrix(NA,1,K)
Sb=Hab=matrix(NA,K,p)
Hb=0

theta=theta2
xdf=NULL

if(!is.null(x0)& !is.null(time.x)){

for (j in 1:K){
#print(paste("J =",j))

bnew=theta[c(j,(K+1):(K+p))]

Ltmp=max(Lx[which(Lx<L[j])])

xt=time.x[which(time.x[,1]==Ltmp),-1]
pos=which(!is.na(xt))

xnew=as.matrix(cbind(1,x0,xt))[pos,]
```

```
x=as.matrix(cbind(x0,xt))[pos,]

xb=xnew%*%bnew

lp=exp(-xb)
Flp=lp/((1+lp)^2)
Ha[j]=sum(ind2[pos,j]*Flp)
Sa[j]=sum(ind1[pos,j])-sum(ind2[pos,j]/(lp+1))

Hb=Hb+t(x)%*%(matrix(ind2[pos,j]*Flp,length(pos),p)*x)
Sb[j,]=t(ind1[pos,j])%*%x-t(as.matrix(ind2[pos,j]/(lp+1)))%*%x
Hab[j,]=t(ind2[pos,j]*Flp)%*%x
xtmp=cbind(L=L[j],xnew)
xdf=rbind(xdf,xtmp)
}

S=rbind(t(Sa),as.matrix(colSums(Sb)))
H1=cbind(diag(as.vector(Ha)),Hab)
H2=cbind(t(Hab),Hb)
H=rbind(H1,H2)
}

theta2=theta+solve(H)%*%S
beta=t(theta2)
colnames(beta)=c( paste("a",seq(1:K),sep=""),
paste("x0",seq(1:p0),sep=""),
paste("xt",seq(1:pt0),sep="") )
V=solve(H)
colnames(V)=rownames(V)=colnames(beta)
```

```
#print(beta)


}
return(list(beta=beta,V=V,xdf=xdf))
}



#____________________with respect to event

lucy.discrete.haz.dep <- function (x0=NULL,time.x=NULL,u,d,L,Lx){

K=length(L)-1
a0=rep(NA,K)
n=length(u)

if(is.null(x0)){
p0=0
}else{
x0=as.matrix(x0)
p0= dim(x0)[2]
}

if(is.null(time.x)){
pt=0
}else{
time.x=as.matrix(time.x)
pt0= (dim(time.x)[2]-1)
pt= pt0
```

```
}


#----get initial value for intercept-----------
ind1= matrix(u,n,K)== matrix(L[1:K],n,K,byrow=TRUE) & matrix(d,n,K)==1
ind2= matrix(u,n,K)>=matrix(L[1:K],n,K,byrow=TRUE)

p1= colSums(ind1)/colSums(ind2)
p1=ifelse(p1==0,0.000001,p1)
a0= log(p1/(1-p1))
xnew0=as.matrix(cbind(1,x0))

p=p0+pt
theta2=matrix(c(a0,rep(0,p)),p+K,1)



theta=0
iter <- 0
itermax=150
tolerance=0.00000000001

#---Newton Ralphson-------------------------------
while (sum(abs(theta - theta2)) >= tolerance) {
iter=iter+1
# print(paste("iter =",iter))



Ha=Sa=matrix(NA,1,K)
Sb=Hab=matrix(NA,K,p)
```

```
Hb=0

theta=theta2
xdf=NULL

if(!is.null(x0)& !is.null(time.x)){

for (j in 1:K){
#print(paste("J =",j))

bnew=theta[c(j,(K+1):(K+p))]

Ltmp=max(Lx[which(Lx<L[j])])

xt=time.x[which(time.x[,1]==Ltmp),-1]
pos=which(!is.na(xt))

xnew=as.matrix(cbind(1,x0,xt))[pos,]
x=as.matrix(cbind(x0,xt))[pos,]

xb=xnew%*%bnew

lp=exp(-xb)
Flp=lp/((1+lp)^2)
Ha[j]=sum(ind2[pos,j]*Flp)
Sa[j]=sum(ind1[pos,j])-sum(ind2[pos,j]/(lp+1))

Hb=Hb+t(x)%*%(matrix(ind2[pos,j]*Flp,length(pos),p)*x)
Sb[j,]=t(ind1[pos,j])%*%x-t(as.matrix(ind2[pos,j]/(lp+1)))%*%x
```

```
Hab[j,]=t(ind2[pos,j]*Flp)%*%x
xtmp=cbind(L=L[j],xnew)
xdf=rbind(xdf,xtmp)
}

S=rbind(t(Sa),as.matrix(colSums(Sb)))
if (length(Ha)==1){
H1=cbind(Ha,Hab)
}
if (length(Ha)>1){
H1=cbind(diag(as.vector(Ha)),Hab)
}
H2=cbind(t(Hab),Hb)
H=rbind(H1,H2)
}

theta2=theta+solve(H)%*%S
beta=t(theta2)
colnames(beta)=c( paste("a",seq(1:K),sep=""),
paste("x0",seq(1:p0),sep=""),
paste("xt",seq(1:pt0),sep="") )
V=solve(H)
colnames(V)=rownames(V)=colnames(beta)

# print(beta)

}
return(list(beta=beta,V=V,xdf=xdf))
}
```

C.2 R codes for Chapter 4

```
#####################################################################
# 2 stage  y is indep of U,  Dec,2014
#####################################################################
lucy.dat2 <- function(n=5,p=2,L,p0=2,eff=1,diff=0){


prob0=prob=x=C=T=d=U=lucy.y=matrix(0,n,p)


x0=matrix(rnorm(n*p0),n,p0)   # fixed covariates x0
b0=matrix(0.1,p0,1)
b=matrix(0.2,p0,1)



a0= 1.3*seq(1:p)/eff
a= seq(1:p)/eff


#-------------------------------------------------------
# baseline info
#-------------------------------------------------------


x[,1]=4+rnorm(n)


prob0[,1]=1/(exp(-a0[1]+0.9*x[,1]+x0%*%b0)+1)
C[,1]=rbinom(n,1,prob0[,1])
y1= 1.5+x0[,1]+0.5*x0[,2]+0.5*x[,1]+rnorm(n)
y1=ifelse(C[,1]==1,y1,NA)
```

```
prob[,1]=1/(exp(-a[1]+0.8*x[,1]+x0%*%b)+1)
T[,1]=rbinom(n,1,prob[,1])
T[,1]=ifelse(C[,1]==0,T[,1],NA)
yt=1.5+x0[,1]+0.5*x0[,2]+0.5*x[,1]+0.5+rnorm(n)
y=ifelse(T[,1]==1 & !is.na(T[,1]),yt,y1)

tmp1=L[1]/2
U[,1]=ifelse(C[,1]==1,tmp1, L[1])

xt=cbind(0,x[,1])

#----------------------------------
# follow-up
#----------------------------------
for (i in 2:p){
x[,i]=diff+rnorm(n)+x[,i-1]
x[,i]=ifelse(T[,i-1]==0,x[,i],NA)


      prob0[,i]=1/(exp(-a0[i]+0.9*x[,i]+x0%*%b0)+1)
C[,i]=rbinom(n,1,prob0[,i])
yt= 1.5+x0[,1]+0.5*x0[,2]+0.5*x[,i]+0.5+rnorm(n)
y=ifelse(C[,i]==1& !is.na(C[,i]),yt,y)

prob[,i]=1/(exp(-a[i]+0.8*x[,i]+x0%*%b)+1)
T[,i]=rbinom(n,1,prob[,i])
T[,i]=ifelse(C[,i]==0,T[,i],NA)
yt= 1.5+x0[,1]+0.5*x0[,2]+0.5*x[,i]+0.5+0.3+rnorm(n)
```

```
y=ifelse(T[,i]==1& !is.na(T[,i]),yt,y)

tmp2=(L[i-1]+L[i])/2
U[,i]=ifelse(C[,i]==1,tmp2,L[i])
xtmp=cbind(L[i-1],x[,i])
xt=rbind(xt,xtmp)
}

Lmid=(L+c(0,L[-length(L)]))/2
L1=rep(L,each=n)
L2=rep(Lmid,each=n)
U2=apply(U,1,function(x) max(x,na.rm=TRUE))



T[,p]=ifelse(U2==L[p],1,NA)
yt= 1.5+x0[,1]+0.5*x0[,2]+0.5*x[,p]+0.5+0.3+rnorm(n)
y=ifelse(T[,p]==1& !is.na(T[,p]),yt,y)

d=rowSums(T,na.rm=TRUE)
colnames(xt)=c("t",paste("xt",seq(1:(dim(xt)[2]-1)),sep=""))
id=seq(1:n)

tmpx=x  # processed below to represent most recently xt

for (j in 1:(p)){
if(j==1){tmpx[,j]=x[,1]}
else{tmpx[,j]=apply(x[,1:j], 1,function(x) x[length(x[!is.na(x)])])}
}
```

```
#identify the last non-missing x
xx=apply(x,1,function(x) x[length(x[!is.na(x)])])

dat.brent=data.frame(id=id,L=L1,U=U2,d=d,x0=x0,xt=xt,Lmid=L2,y=y,xx=xx)
dat.brent$yb=as.numeric(dat.brent$U == L & dat.brent$d == 1)

return(list(U=U2,d=d,T=T,C=C,x0=x0,xt=xt,tmpx=tmpx,y=y,x=x,xx=xx,
id=id,df=dat.brent, prob0=prob0, prob=prob))
}

##########################################
#simulated truth for lucy.dat2
##########################################

my.simtru<-function(L,eff=2,diff=0){

p.sim <-length(L)
n.sim=500000

sim.x0 <- matrix(NA,p.sim,n.sim)
sim.lambda <- matrix(NA,p.sim,n.sim)
sim.Y<-sim.Cens <- sim.U<-sim.Delta<-matrix(NA, p.sim, n.sim)

sim.x1 <- rnorm(n.sim,0,1)
sim.x2 <- rnorm(n.sim,0,1)

Lmid=(L+c(0,L[-p.sim]))/2
```

```
for (k in 1){

sim.x0[k,] <- rnorm(n.sim,(3.5+0.5*k),1)
sim.lambda[k,] <- 1/(1+exp(-1.3*k/eff+0.9*sim.x0[k,]+0.1*sim.x1+0.1*sim.x2))
sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.U[k,]=ifelse(sim.Cens[k,]==1,Lmid[k], L[k])
sim.Delta[k,]= ifelse(sim.Cens[k,]==0,1,0)
sim.y0= 1.5+sim.x1+0.5*sim.x2+0.5*sim.x0[k,]
sim.yt=1.5+sim.x1+0.5*sim.x2+0.5*sim.x0[k,]+0.5
sim.Y[k,]=ifelse(sim.Cens[k,]==1,sim.y0,sim.yt)
}

for (k in 2:p.sim){

sim.x0[k,] <- diff+rnorm(n.sim)+sim.x0[k-1,]
sim.lambda[k,] <- 1/(1+exp(-1.3*k/eff+0.9*sim.x0[k,]+0.1*sim.x1+0.1*sim.x2))

sim.Cens[k,]<-rbinom(n.sim,1,sim.lambda[k,])
sim.Cens[k,]=ifelse(sim.Cens[k-1,]==0,sim.Cens[k,],NA)

sim.Delta[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),0,1)

sim.U[k,]=ifelse(is.na(sim.Cens[k,]),sim.U[k-1,],Lmid[k])
sim.U[k,]=ifelse(sim.Cens[k,]==1|is.na(sim.Cens[k,]),sim.U[k,],L[k])

sim.y02= 1.5+sim.x1+0.5*sim.x2+0.5*sim.x0[k,]+0.5
sim.yt=1.5+sim.x1+0.5*sim.x2+0.5*sim.x0[k,]+0.5+0.3
```

```
sim.Y[k,]=ifelse(sim.Cens[1,]==1,sim.y0,0)
sim.Y[k,]=ifelse(sim.Cens[k,]==1& !is.na(sim.Cens[k,]),sim.y02,
sim.Y[k,])
sim.Y[k,]=ifelse(sim.Cens[k,]==0 & !is.na(sim.Cens[k,]),sim.yt,
sim.Y[k,])



}

trumu<-rowMeans(sim.Y)
return(trumu)
}

#4_9_2015
# add the c parameter of control variance when ps is missspecified
#5_2_2015



lucy.DR.m2 <- function(y,U,L,x0,d,time.x,x,tmpx,xx){

ind1=ifelse(U<L[1],0,1) #I(C>t1)
ind2=ifelse(U==L[2],1,0) # I(C>t2)
ind3=ifelse(U==L[1],1,0) # I(U=t1)

p0=dim(x0)[2]
n=length(y)
#------------estimation of parameters-------------------------

ft3=lucy.discrete.haz.dep.cens(u=U,d=1-d,L=L,x0=x0,time.x=time.x,Lx=L-10)
```

```
alpha=ft3$beta[1:p]

bxt=ft3$beta[p+p0+1]

b0=ft3$beta[(p+1):(p+p0)]


ft4=lucy.discrete.haz.dep(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10)

a=ft4$beta[1:(p-1)]

bxt4=ft4$beta[p+p0]

b=ft4$beta[(p):(p+p0-1)]

ftm=lm(x[,2]~x[,1])


# predict lambda's and xt1

pred.xt1=ftm$coeff[1]+ftm$coeff[2]*x[,1]


pred.lambda0.0=1/(1+exp(-alpha[1]-x[,1]*bxt-x0%*%b0))

pred.lambda0.1=1/(1+exp(-alpha[2]-pred.xt1*bxt-x0%*%b0))


#pred.lambda0.0=1/(1+exp(-alpha[1]-x0%*%b0))

#pred.lambda0.1=1/(1+exp(-alpha[2]-x0%*%b0))


pred.lambda.1<- 1/(1+exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2]))



# estimate w wt

ft1=calc.w(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10,M=1,tmpx=tmpx)

ft2=calc.w(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10,M=2,tmpx=tmpx)



##############################

# solving alpha_g for m=1
```

```
################################
# a%*%x=b
wt1.1=1
p1d=cbind(1,ind1,x0,x[,1])
MxA=(t(p1d)*matrix(wt1.1,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))%*%p1d/n
MxB=t(p1d)%*%as.matrix(y*wt1.1)/n
sol1=solve(MxA,MxB)


# variance minimization
wt.i1=ind1*(1-1/pred.lambda.1)


p1d=cbind(1,ind1,x0,x[,1])
MxA.i=(t(p1d)*matrix(wt.i1,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))%*%p1d/n
MxB.i=t(p1d)%*%as.matrix(y*wt.i1)/n
sol1.t=solve(MxA.i[2:5,2:5],MxB.i[2:5])
sol1.i=c(0,sol1.t)

#cT adjustment
#derivative of g1.1
dg11=cbind(ind1,x0,x[,1])

#d(lambda.1)/d(gamma)=lambda.1^2*exp(-gammaX)  .X
dlam.nox=pred.lambda.1^2*exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2])
dx=cbind(1,x[,1],x0)

theta2=c(sol1.t,rep(0.1, dim(dx)[2]))
iter <- 0
```

```
#itermax=150
tolerance=0.000001
theta=0
#---Newton Ralphson-------------------------------
while (sum(abs(theta - theta2)) >= tolerance) {
#print(abs(theta - theta2)[4])
iter=iter+1
#print(paste("iter =",iter))

theta=theta2
c=theta[(length(sol1.t)+1):length(theta)]
ag=theta[1:length(sol1.t)]
g11=dg11%*%ag
l1=(dlam.nox/(1-pred.lambda.1)*(dx%*%c))

Sa=t(ind1*(1/pred.lambda.1-1)*(y-g11-l1))%*% dg11
Sb=t(ind1/pred.lambda.1*(y-g11-l1)*dlam.nox)%*%dx
S=cbind(Sa,Sb)

#-I(theta)
Iaa=t(matrix(ind1*(1/pred.lambda.1-1),ncol=dim(dg11)[2],nrow=n)*dg11)%*%dg11
Ica=t(matrix(ind1/pred.lambda.1*dlam.nox,ncol=dim(dx)[2],nrow=n)*dx)%*%dg11
Icc=t(matrix(ind1*dlam.nox^2/((1-pred.lambda.1)*pred.lambda.1),
ncol=dim(dx)[2],nrow=n)*(dx))%*%dx
I=rbind(cbind(Iaa,t(Ica)),cbind(Ica, Icc))

theta2=theta+solve(I)%*%t(S)

#print(theta2)
```

```
}
```

```
################################
# solving alpha_g for m=2
################################
wt1=1
wt2=1#1-ind3

q1d=cbind(1,ind1,ind2,x0,xx) # 1st derivative for g2^(2)
prc2=(1-pred.lambda0.1)

xlast=pred.xt1*ind1+ x[,1]*(1-ind1)
q2d=cbind(1,ind1,ind1*prc2,x0,xlast)  # 1st derivative for g1^(2)
# 1st derivative for sum p(a|s)*g2^(2)-g1^(2)
q2p=cbind(alpha0=0,alpha1=0,alpha2=ind1*(ind2-prc2),alphab1=0,alphab2=0,
alphat=ind1*(xx-pred.xt1))

Mxb=t(q1d)%*%as.matrix(y*wt1)/n
Mxa=((t(q1d)*matrix(wt1,ncol=n,nrow=dim(x0)[2]+4,byrow=TRUE))
%*%q1d-(t(q2d)*matrix(wt2,ncol=n,nrow=dim(x0)[2]+4,byrow=TRUE))%*%q2p)/n

sol2=solve(Mxa,Mxb)
```

```
# variance minimization
wt.i2=ind1*(pred.lambda.1/(1-pred.lambda.1))



Mxb.i=t(q1d)%*%as.matrix(y*wt.i2)/n
Mxa.i=((t(q1d)*matrix(wt.i2,ncol=n,nrow=dim(x0)[2]+4,byrow=TRUE))
%*%q1d-(t(q2d)*matrix(0,ncol=n,nrow=dim(x0)[2]+4,byrow=TRUE))%*%q2p)/n

sol2t=solve(Mxa.i[2:6,2:6],Mxb.i[2:6])
sol2.i=c(0,sol2t)

#cT adjustment
#derivative of g1.2
#dg120=cbind(ind1,x0,xlast,ind1*prc2)
dg12=cbind(ind1,ind1*prc2,x0,xlast)
# dg12=cbind(ind1,ind2,x0,xx)
#d(lambda.1)/d(gamma)=lambda.1^2*exp(-gammaX)  .X
dlam.nox2=(pred.lambda.1)^2*exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2])
dx=cbind(1,x[,1],x0)

theta4=c(sol2t,rep(-0.1, dim(dx)[2]))
iter <- 0
#itermax=150
tolerance=0.000001
theta3=0
#---Newton Ralphson-------------------------------
while (sum(abs(theta3 - theta4)) >= tolerance) {
#print(abs(theta3 - theta4)[4])
iter=iter+1
```

```
#print(paste("iter =",iter))

theta3=theta4
c2=theta3[(length(sol2t)+1):length(theta3)]
ag2=theta3[1:length(sol2t)]
g12=dg12%*%ag2
l2=(dlam.nox2/(pred.lambda.1)*(dx%*%c2))

Sa2=t(ind1*(pred.lambda.1/((1-pred.lambda.1)))*(y-g12-l2))%*% dg12
Sb2=t(ind1/((1-pred.lambda.1))*(y-g12-l2)*dlam.nox2)%*%dx
S2=cbind(Sa2,Sb2)

#-I(theta)
Iaa2=t(matrix(ind1*(pred.lambda.1/((1-pred.lambda.1))),
ncol=dim(dg12)[2],nrow=n)*dg12)%*%dg12
Ica2=t(matrix(ind1/((1-pred.lambda.1))*dlam.nox2,ncol=dim(dx)[2],
nrow=n)*dx)%*%dg12
Icc2=t(matrix(ind1*dlam.nox2^2/((1-pred.lambda.1)*pred.lambda.1),
ncol=dim(dx)[2],nrow=n)*(dx))%*%dx
I2=rbind(cbind(Iaa2,t(Ica2)),cbind(Ica2, Icc2))

theta4=theta3+solve(I2)%*%t(S2)

#print(theta4)

}
```

```
####### fit in g funtions#########################


g1.1=p1d%*%as.matrix(sol2[-3])

g2.2=q1d%*%as.matrix(sol2)

g1.2=q2d%*%as.matrix(sol2)

g1.1i=p1d%*%as.matrix(sol1.i)

g1.2i=q2d%*%as.matrix(sol2.i)


g1.1c=dg11%*%as.matrix(theta2[1:length(sol1.t)])

g1.2c=cbind(ind1,ind1*prc2,x0,xlast)%*%as.matrix(theta4[1:length(sol2t)])




mu.dr=mu.jt=mu.or=mu.or.i=mu.dr.i=mu.dr.c=rep(NA,length(L))

#------------DR m=1-------------------------

mu.dr[1]=sum(ft1$W*y-ft1$W*g1.1+g1.1)/n

mu.jt[1]=sum(ft1$W*y)/sum(ft1$W)

mu.or[1]=mean(g1.1)

mu.or.i[1]=mean(g1.1i)

mu.dr.i[1]=sum(ft1$W*y-ft1$W*g1.1i+g1.1i)/n

mu.dr.c[1]=sum(ft1$W*y-ft1$W*g1.1c+g1.1c)/n



#------------DR m=2-------------------------

t.num.d=sum((ft2$W*y-(ft2$W-1)*g1.2))

t.num.di=sum((ft2$W*y-(ft2$W-1)*g1.2i)) #(1-ind3)*
```

```
t.num.dc=sum((ft2$W*y-(ft2$W-1)*g1.2c))

t.den=n#sum(1-ind3)
mu.jt[2]=sum(ft2$W*y)/sum(ft2$W)
mu.or[2]=mean(g1.2)
mu.dr[2]=t.num.d/t.den
mu.dr.i[2]=t.num.di/t.den
mu.or.i[2]=mean(g1.2i)
mu.dr.c[2]=t.num.dc/t.den

return(list(mu.dr=mu.dr,mu.jt=mu.jt,mu.or=mu.or,
mu.or.i=mu.or.i, mu.dr.i=mu.dr.i,mu.dr.c=mu.dr.c))
}

#################### omit the time dependent x##############



lucy.DR.m2.or <- function(y=dat$y,U=dat$U,L=L,x0=dat$x0,d=dat$d,
time.x=dat$xt,x=dat$x,tmpx=dat$tmpx,xx=dat$xx){

ind1=ifelse(U<L[1],0,1) #I(C>t1)
ind2=ifelse(U==L[2],1,0) # I(C>t2)
ind3=ifelse(U==L[1],1,0) # I(U=t1)

p0=dim(x0)[2]
n=length(y)
#------------estimation of parameters-------------------------
```

```
ft3=lucy.discrete.haz.dep.cens(u=U,d=1-d,L=L,x0=x0,time.x=time.x,Lx=L-10)
alpha=ft3$beta[1:p]
bxt=ft3$beta[p+p0+1]
b0=ft3$beta[(p+1):(p+p0)]

ft4=lucy.discrete.haz.dep(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10)
a=ft4$beta[1:(p-1)]
bxt4=ft4$beta[p+p0]
b=ft4$beta[(p):(p+p0-1)]
ftm=lm(x[,2]~x[,1])

# predict lambda's and xt1
pred.xt1=ftm$coeff[1]+ftm$coeff[2]*x[,1]

pred.lambda0.0=1/(1+exp(-alpha[1]-x[,1]*bxt-x0%*%b0))
pred.lambda0.1=1/(1+exp(-alpha[2]-pred.xt1*bxt-x0%*%b0))
pred.lambda.1<- 1/(1+exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2]))



# estimate w wt
ft1=calc.w(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10,M=1,tmpx=tmpx)
ft2=calc.w(u=U,d=d,L=L,x0=x0,time.x=time.x,Lx=L-10,M=2,tmpx=tmpx)



##############################
# solving alpha_g for m=1
################################
# a%*%x=b
wt1.1=1
```

```
p1d=cbind(1,ind1,x0)
MxA=(t(p1d)*matrix(wt1.1,ncol=n,nrow=dim(x0)[2]+2,byrow=TRUE))%*%p1d/n
MxB=t(p1d)%*%as.matrix(y*wt1.1)/n
sol1=solve(MxA,MxB)



# variance minimization
wt.i1=ind1*(1-1/pred.lambda.1)



p1d=cbind(1,ind1,x0)
MxA.i=(t(p1d)*matrix(wt.i1,ncol=n,nrow=dim(x0)[2]+2,byrow=TRUE))%*%p1d/n
MxB.i=t(p1d)%*%as.matrix(y*wt.i1)/n
sol1.t=solve(MxA.i[2:4,2:4],MxB.i[2:4])
sol1.i=c(0,sol1.t)

#cT adjustment
#derivative of g1.1
dg11=cbind(ind1,x0)

#d(lambda.1)/d(gamma)=lambda.1^2*exp(-gammaX)  .X
dlam.nox=pred.lambda.1^2*exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2])
dx=cbind(1,x[,1],x0)

theta2=c(sol1.t,rep(0.1, dim(dx)[2]))
iter <- 0
#itermax=150
tolerance=0.000001
theta=0
```

```
#---Newton Ralphson--------------------------------
while (sum(abs(theta - theta2)) >= tolerance) {
#print(abs(theta - theta2)[4])
iter=iter+1
#print(paste("iter =",iter))

theta=theta2
c=theta[(length(sol1.t)+1):length(theta)]
ag=theta[1:length(sol1.t)]
g11=dg11%*%ag
l1=(dlam.nox/(1-pred.lambda.1)*(dx%*%c))

Sa=t(ind1*(1/pred.lambda.1-1)*(y-g11-l1))%*% dg11
Sb=t(ind1/pred.lambda.1*(y-g11-l1)*dlam.nox)%*%dx
S=cbind(Sa,Sb)

#-I(theta)
Iaa=t(matrix(ind1*(1/pred.lambda.1-1),ncol=dim(dg11)[2],nrow=n)*dg11)%*%dg11
Ica=t(matrix(ind1/pred.lambda.1*dlam.nox,ncol=dim(dx)[2],nrow=n)*dx)%*%dg11
Icc=t(matrix(ind1*dlam.nox^2/((1-pred.lambda.1)*pred.lambda.1),
ncol=dim(dx)[2],nrow=n)*(dx))%*%dx
I=rbind(cbind(Iaa,t(Ica)),cbind(Ica, Icc))

theta2=theta+solve(I)%*%t(S)

#print(theta2)

}
```

```
##############################
# solving alpha_g for m=2
################################
wt1=1
wt2=1#1-ind3

q1d=cbind(1,ind1,ind2,x0) # 1st derivative for g2^(2)
prc2=(1-pred.lambda0.1)

xlast=pred.xt1*ind1+ x[,1]*(1-ind1)
q2d=cbind(1,ind1,ind1*prc2,x0)  # 1st derivative for g1^(2)
# 1st derivative for sum p(a|s)*g2^(2)-g1^(2)
q2p=cbind(alpha0=0,alpha1=0,alpha2=ind1*(ind2-prc2),alphab1=0,alphab2=0)

Mxb=t(q1d)%*%as.matrix(y*wt1)/n
Mxa=((t(q1d)*matrix(wt1,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))
%*%q1d-(t(q2d)*matrix(wt2,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))%*%q2p)/n

sol2=solve(Mxa,Mxb)



# variance minimization
wt.i2=ind1*(pred.lambda.1/(1-pred.lambda.1))
```

```
Mxb.i=t(q1d)%*%as.matrix(y*wt.i2)/n
Mxa.i=((t(q1d)*matrix(wt.i2,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))
%*%q1d-(t(q2d)*matrix(0,ncol=n,nrow=dim(x0)[2]+3,byrow=TRUE))%*%q2p)/n

sol2t=solve(Mxa.i[2:5,2:5],Mxb.i[2:5])
sol2.i=c(0,sol2t)

#cT adjustment
#derivae of g1.2
#dg120=cbind(ind1,x0,xlast,ind1*prc2)
dg12=cbind(ind1,ind1*prc2,x0)
# dg12=cbind(ind1,ind2,x0,xx)
#d(lambda.1)/d(gamma)=lambda.1^2*exp(-gammaX)  .X
dlam.nox2=(pred.lambda.1)^2*exp(-a[1]-x[,1]*bxt4-x0[,1]*b[1]-x0[,2]*b[2])
dx=cbind(1,x[,1],x0)

theta4=c(sol2t,rep(-0.1, dim(dx)[2]))
iter <- 0
#itermax=150
tolerance=0.000001
theta3=0
#---Newton Ralphson-------------------------------
while (sum(abs(theta3 - theta4)) >= tolerance) {
#print(abs(theta3 - theta4)[4])
iter=iter+1
#print(paste("iter =",iter))

theta3=theta4
c2=theta3[(length(sol2t)+1):length(theta3)]
```

```
ag2=theta3[1:length(sol2t)]
g12=dg12%*%ag2
l2=(dlam.nox2/(pred.lambda.1)*(dx%*%c2))

Sa2=t(ind1*(pred.lambda.1/((1-pred.lambda.1)))*(y-g12-l2))%*% dg12
Sb2=t(ind1/((1-pred.lambda.1))*(y-g12-l2)*dlam.nox2)%*%dx
S2=cbind(Sa2,Sb2)

#-I(theta)
Iaa2=t(matrix(ind1*(pred.lambda.1/((1-pred.lambda.1))),
ncol=dim(dg12)[2],nrow=n)*dg12)%*%dg12
Ica2=t(matrix(ind1/((1-pred.lambda.1))*dlam.nox2,ncol=dim(dx)[2],
nrow=n)*dx)%*%dg12
Icc2=t(matrix(ind1*dlam.nox2^2/((1-pred.lambda.1)*pred.lambda.1),
ncol=dim(dx)[2],nrow=n)*(dx))%*%dx
I2=rbind(cbind(Iaa2,t(Ica2)),cbind(Ica2, Icc2))

theta4=theta3+solve(I2)%*%t(S2)

#print(theta4)

}




####### fit in g funtions#########################
```

```
g1.1=p1d%*%as.matrix(sol2[-3])

g2.2=q1d%*%as.matrix(sol2)

g1.2=q2d%*%as.matrix(sol2)

g1.1i=p1d%*%as.matrix(sol1.i)

g1.2i=q2d%*%as.matrix(sol2.i)


g1.1c=dg11%*%as.matrix(theta2[1:length(sol1.t)])

g1.2c=cbind(ind1,ind1*prc2,x0)%*%as.matrix(theta4[1:length(sol2t)])




mu.dr=mu.jt=mu.or=mu.or.i=mu.dr.i=mu.dr.c=rep(NA,length(L))

#------------DR m=1-------------------------

mu.dr[1]=sum(ft1$W*y-ft1$W*g1.1+g1.1)/n

mu.jt[1]=sum(ft1$W*y)/sum(ft1$W)

mu.or[1]=mean(g1.1)

mu.or.i[1]=mean(g1.1i)

mu.dr.i[1]=sum(ft1$W*y-ft1$W*g1.1i+g1.1i)/n

mu.dr.c[1]=sum(ft1$W*y-ft1$W*g1.1c+g1.1c)/n



#------------DR m=2-------------------------

t.num.d=sum((ft2$W*y-(ft2$W-1)*g1.2))

t.num.di=sum((ft2$W*y-(ft2$W-1)*g1.2i)) #(1-ind3)*

t.num.dc=sum((ft2$W*y-(ft2$W-1)*g1.2c))


t.den=n#sum(1-ind3)

mu.jt[2]=sum(ft2$W*y)/sum(ft2$W)
```

```
mu.or[2]=mean(g1.2)

mu.dr[2]=t.num.d/t.den

mu.dr.i[2]=t.num.di/t.den

mu.or.i[2]=mean(g1.2i)

mu.dr.c[2]=t.num.dc/t.den


return(list(mu.dr=mu.dr,mu.jt=mu.jt,mu.or=mu.or,

mu.or.i=mu.or.i, mu.dr.i=mu.dr.i,mu.dr.c=mu.dr.c))

}


#revised 5/2/2015

# calculate weight Wpk for the IPW estimator as well as

#the intermediate weight Wpt, t<k=m


calc.w <- function (u,d,L,M,x0,time.x,Lx,tmpx){


p=length(L)

n=length(u)

p0=dim(x0)[2]

#------------estimation of parameters-------------------------

ft4=lucy.discrete.haz.dep(u=u,d=d,L=L,x0=x0,time.x=time.x,Lx=Lx)

a=ft4$beta[1:(p-1)]

bxt4=ft4$beta[p+p0]

b=ft4$beta[(p):(p+p0-1)]


# ---calculate Lambda from discrete hazard function with

#respect to physician stop-----

pred.lambda=matrix(NA,p,n)

pred.lambda2=matrix(0,p,n)
```

```
for (j in 1:(p)){
pred.lambda[j,]<- 1/(1+exp(-a[j]-tmpx[,j]*bxt4-x0[,1]*b[1]-x0[,2]*b[2]))
}

# lambda4=1 since it is the end of the study.
pred.lambda[p,]<-1

K=p
#---calculate denominator of W-----------------------------------------
ind1= matrix(u,n,K)==matrix(L,n,K,byrow=TRUE) & matrix(d,n,K)==1
#I(U=tj, d=1)
ind2= matrix(u,n,K)> matrix(L,n,K,byrow=TRUE)  #I(U>tj)

# denominator for each element by individual
den=(t(1-pred.lambda))^ind2*(t(pred.lambda)^ind1)

prodden=apply(den,1,prod) #denominator for W by individual

#-------calculate numerator of W-----------------------
num=matrix(NA,n,K)
num= (matrix(L,n,K,byrow=TRUE)==L[M])^ind1*(matrix(u,n,K)<=L[M])^ind2
# tj=tm or u<=tm

prodnum=apply(num,1,prod) #numerator for W by individual

##########-W-##############
W=prodnum/prodden
```

```
#-----------------calc Wt---------------------

tmpden=den

tmpnum=num


for (j in 2:p){

tmpden[,j]=apply(den[,1:j],1,prod)

tmpnum[,j]=apply(num[,1:j],1, prod)

}

Wt=tmpnum/tmpden


######calc W{t-1}##########

Wt1=matrix(1,n,p)

Wt1[,-1]=Wt[,-p]



return(list(W=W,Wt=Wt,Wt1=Wt1,num=num, den=den, tmpnum=tmpnum, tmpden=tmpden))


}
```