Causal Inference with Recurrent Data via Propensity Score Methods

by

Haodi Liang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Propensity score methods are increasingly being used to reduce estimation bias of treatment effects for observational studies. Previous research has shown that propensity score methods consistently estimate the marginal hazard ratio for time to event data. However, recurrent data frequently arise in the biomedical literature and there is a paucity of research into the use of propensity score methods when data are recurrent in nature. The objective of my thesis is to extend the existing propensity score methods to recurrent data setting. We review current propensity score methods for estimating treatment effects when the outcome is a single time to event. Then we propose a new class of inverse probability treatment weighting (IPTW) estimators to estimate treatment effects for recurrent data. We illustrate our methods through both estimating equation theory and a series of Monte Carlo simulations. The simulation results indicate that when there is no censoring, the newly proposed IPTW estimators allow us to consistently estimate the marginal hazard ratio for each event. Under administrative censoring regime, the stabilized IPTW estimator consistently estimates the marginal hazard ratio while the conventional IPTW estimator yields significant bias, especially when the proportion of subjects being censored is high. For variance estimation, we incorporate the robust variance estimator and the bootstrap variance estimator to deal with the within-subject correlation induced by weighting. In addition, we apply our methods to a real life example. We note that although the Cox proportional hazards model we used for estimating the marginal hazard ratio may be subject to misspecification, the estimate still converges and has meaningful interpretations.

Acknowledgements

First I would like to express my deepest gratitude to my supervisor, Professor Cecilia Cotton, who is always there to help when I have questions. Her insightful thoughts and patient guidance help me overcome numerous difficulties I encountered throughout my research. Besides that, Cecilia offered to help revise my statement of purpose and CV for my Ph.D applications. Cecilia has my best interest at heart and I can see how invested she is in me doing well.

I would also like to thank Professor Yeying Zhu for offering me the first opportunity to do research in statistics during my undergraduate. Thank you to Professor Richard Cook for your insightful advice and your course survival analysis, which opened a gate to me in the field of biostatistics.

A special thank goes to my friend Ian Waudby-Smith, a Waterloo graduate and now an incoming biostatistics Ph.D student at the University of Washington. Ian gave me precious advice for my career path and offered tremendous support for my Ph.D applications.

In addition, I would like to thank Professor Pengfei Li and Professor Stephen New for their outstanding teaching. Through their courses I gained the statistics background and mathematical maturity that prove to be useful in my research.

Finally, I extend my thanks to my parents and friends for their support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1

Literature Review

Causal inference is an emerging field in statistics. In medical research, we are often interested in understanding the effect of treatment on an outcome. The gold standard is to conduct an experimental study where the treatment is randomized. This guarantees that the covariate distributions of the treatment group and control group do not differ systematically, in which case valid causal inference can be drawn by directly comparing the two groups. However, in reality, experimental studies are often impossible or unethical so we have to consider the alternative: observational studies. Observational studies differ from experimental studies in that treatment assignment is often dependent on other covariates, and we refer this as selection bias. As a result, the characteristics of the treatment and control group may be systematically different, which leads to biased estimation of the treatment effect. In order to make valid causal inference, adjustments must be made to balance the covariates between the two groups. Over the past several decades, different methods for reducing bias in observational studies have been developed and there has been an increasing interest in using the propensity score methods. In this section, we review the past research on causal inference based on the propensity score methods, specifically the use of inverse probability treatment weighting in reducing treatment effect estimation bias.

1.1 Historical Development on Propensity Score Methods

The propensity score is defined as the probability of treatment assignment conditional on measured baseline covariates [1]. There are four propensity score methods that are used most often in the biomedical literature: matching, stratification, inverse probability weighting and covariate adjustment. These methods allow us to reconstruct a pseudo-sample which mimics an experimental data setting, thus reducing or eliminating bias in estimating the treatment effect.

In the 1980s, researchers mainly focused on bias reduction on a linear scale. Rosenbaum and Rubin (1983) demonstrated that by dividing the sample into five mutually exclusive equal-sized strata based on the propensity score would result in an over 90% bias reduction [1, 2]. It was also proven that the inverse probability treatment weighting method using the propensity score gives consistent estimates of linear treatment effects [1]. Matching also yielded a similar performance by forming matched sets with similar values of propensity score between the treatment group and control group [3]. All three propensity score methods resulted in unbiased estimation of the treatment effect on a linear scale (i.e. when treatment effect is a difference in mean outcome) when there was no unmeasured confounding. However, little attention had been focused on other measures of treatment effects at that time.

1.2 Propensity Score Methods with Time to Event Data

It was not until recent years that propensity score methods for non-linear measures of treatment effects received attention. Some applied researchers used propensity score methods

to estimate non-linear treatment effects for time to event data before but the degree of bias they incurred had not been extensively studied [4]. To this end, Austin (2007) performed a series of Monte Carlo simulation studies to examine the degree of bias when treatment effects are measured using a hazard ratio, odds ratio and rate ratio [4]. The simulation results indicated that conditional on the propensity score, matching, stratification and inverse probability weighting all resulted in biased estimates of the true conditional hazard ratio and odds ratio, while regression adjustment yielded unbiased estimates of both the true conditional hazard ratio and odds ratio. Interestingly the rate ratio was consistently estimated for all propensity score methods. This is because conditional on the propensity score, we estimate the marginal treatment effect instead of the conditional treatment effect [5]. A conditional effect refers to the average effect at the individual level, of removing a subject from treated to untreated, while a marginal effect is the average effect at the population level, of moving the whole population from treated to untreated [5]. Austin (2007) [4] concluded that the marginal treatment effect coincides with the conditional treatment effect when the measure of treatment effect is a difference or rate ratio while the two effects do not coincide in the odds ratio and hazard ratio settings.

1.3 Review of Variance Estimation Methods

When dealing with clustered data, the use of naive variance estimator often results in biased estimation of the standard errors and poor coverage rates for confidence intervals [6]. By weighting we artificially induce a within-subject correlation by creating a cluster for each subject. Moreover, the fact that we inflate the sample size by weighting also leads to underestimated standard errors [6]. Lin and Wei (1989) argued that by using the robust variance estimator, one can eliminate the within-subject correlation induced by weighting [7]. However, the behavior of different variance estimators had not been extensively studied. Austin (2016) performed a series of Monte Carlo simulations using the naive variance estimator, the robust variance estimator proposed by Lin and Wei

and the bootstrap variance estimator [6]. The simulation results suggested that both the robust variance estimator and the bootstrap variance estimator significantly improved the accuracy of variance estimation with slightly better performance for the bootstrap variance estimator. This finding provided a helpful guideline of variance estimation to researchers in their research.

1.4 Recurrent Events

Methods for recurrent events analysis are covered extensively in Cook and Lawless (2007) [8]. Basic analysis methodologies include Poisson process and renewal process models where gap times are independent. The methods gained popularity because of its simplicity and well-established theoretical results. However, the independence assumption is often violated and to this end, cases where gap times are not independent to each other are discussed further. Additional modelling techniques for handling recurrent data include the Accelerated Failure Time model and the Cox proportional hazards model.

1.5 Discussion

Many observational data in real life are recurrent in nature. Though methods for making causal inference for time to event data have been developed, there is a paucity of research on making causal inference for recurrent data. Hence it would be desirable to extend the propensity score framework to two events and possibly multiple events. Therefore the objective of my thesis is to develop appropriate propensity score methods to estimate the treatment effect in the case of recurrent events.

Chapter 2

One Event Setting

The goal of this chapter is to draw causal inference in the case where the outcome of interest is a single time to event. The chapter focuses on the case where the treatment effect is measured using a hazard ratio. Baseline covariates, treatment assignment, outcome and some other related terminologies are defined at the beginning. Then, model assumptions and specifications for the treatment model and the outcome model are described. The consistency and asymptotic properties of the inverse probability treatment (IPTW) estimators are proven. Finally, a series of Monte Carlo simulations are conducted to extensively study the behaviour of the IPTW estimators under different parameter settings. A summary of the simulation results and trends is presented at the end of this chapter.

2.1 Notation and Model Set-up

We shall use the following notation throughout this chapter. Assume that we have a total of n subjects $i = 1, 2, ..., n$. We suppress the i notation in this section. Let Z be an indicator variable denoting treatment status, $\mathbf{X} = (1, x_1, x_2, ...x_{p-1})^T$ be a p-dimensional vector of covariates, and T be the observed survival time. Define T_0 to be the survival time under

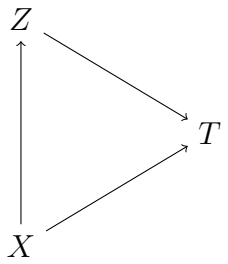

Figure 2.1: Causal graph for time to event data

control and T_1 to be the survival time under treatment. In an observational data setting, $\boldsymbol{X}$ is a confounder because it is both associated with treatment Z and is a risk factor for outcome T. Figure 2.1 illustrates the relationship among $\boldsymbol{X}$, Z and T. We define e to be the propensity score, which is given by

$$e(x) = P(Z = 1|\mathbf{X} = x)$$

The propensity score is the probability of receiving treatment conditional on observed covariates [1]. However, in most cases the propensity score is unknown and needs to be estimated. A common way of doing this is using a logistic regression model

$$log\Big(\frac{e}{1-e}\Big) = \mathbf{X}^T\boldsymbol{\alpha}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ...\alpha_{p-1})^T$ is a vector of regression coefficients. Let $\hat{\boldsymbol{\alpha}}$ denote the estimates of regression coefficients from the logistic regression model. Hence the estimated propensity score is

$$\hat{e}(x) = expit(\mathbf{X}^T\hat{\boldsymbol{\alpha}})$$

We define the conventional inverse probability treatment weights (IPTW) to be [9]

$$cw = \frac{Z}{e} + \frac{1-Z}{1-e}$$

Hence, each subject is weighted by the inverse of the probability of treatment the subject actually received. The IPTW weights sometimes induce extremely large weights for a few

subjects and as a result, these subjects will dominate the analysis [6]. To this end, we incorporate the stablized inverse probability weights, which is defined to be [10]

$$sw = \frac{pZ}{e} + \frac{(1-p)(1-Z)}{1-e}$$

where p is the treatment prevalence across the sample. That is, $p = P(Z = 1)$. Conventional weights often result in a few subjects having very large weights, thus yields unstable estimation of the marginal hazard ratio. The use of stabilized weights improves the stability of estimation by adjusting subjects with very large weights [6].

Assume there is no censoring. We define the marginal hazard ratio to be the hazard ratio of moving all population from treatment to control. For a given simulated dataset, to determine the true log marginal hazard ratio β^m, for each subject we simulate both potential outcomes under treatment and control conditions T_0 and T_1. Then we regress both potential outcomes on treatment indicator to obtain the log of the true marginal hazard ratio β^m [11]. To estimate the marginal treatment effect on the hazard of the occurrence of the outcome, the following assumptions need to be satisfied [1]:

1. There is no unmeasured confounding. i.e. $(T_0, T_1) \perp X|Z$
2. The probability of treatment is strictly positive i.e. $0 < P(Z = z|X) < 1$

We first regress treatment indicator on baseline covariates $\mathbf{X}$ through a logistic regression model to obtain the estimated propensity score $\hat{e}(x)$. Then we run a weighted Cox regression model of the form

$$h(t|x) = h_0(t)e^{\beta^m z}$$

where $h_0(t)$ is the baseline hazard and β^m is the log of the true marginal hazard ratio. We regress survival time T on treatment indicator Z, with both the conventional and stablized inverse probability weights defined above to get the estimate of $\hat{\beta}^m$. For the variance estimation of $\hat{\beta}^m$, we consider three different approaches. First we use the naive variance

estimator from the maximum partial likelihood estimator for the Cox proportional hazards model. Second, we use the robust variance estimator proposed by Lin [7]. The use of the IPTW artificially creates clusters with w_i copies of subject i for the i^{th} cluster. The covariance matrix for the i^{th} cluster is given by

$$Var(\boldsymbol{T}_i) = A_i^{\frac{1}{2}}(\beta) R_i(\alpha) A_i^{\frac{1}{2}}(\beta)$$

where $A_i(\beta) = \begin{bmatrix} Var(T_{i1}) & & & \\ & Var(T_{i2}) & & \\ & & \ddots & \\ & & & Var(T_{w_i}) \end{bmatrix}$ and $R_i(\alpha)$ is the working correlation matrix [12]. In the case where w_i are not integers, we can multiply the weights by some large constant to approximate the weights by integers. Since weighting induces a within-subject correlation, the naive variance estimator tends to incorrectly estimate the variance. Finally, we use the bootstrap variance estimator. We draw 200 bootstrap samples and for each simulation sample, and we repeat the same procedures described above to obtain the estimated log marginal hazard ratio. The standard deviation of the estimated log marginal hazard ratio from the 200 bootstrap samples is used as the bootstrap standard error [6].

2.2 Estimating Equation Theory

Estimating equations are a useful tool for semi-parametric models. In this section, we introduce some elementary theory regarding estimating equations and prove some key results. Let $\mathbf{X}$ denote a covariate vector and $\boldsymbol{\theta}$ denote a vector of unknown parameters. Let $\boldsymbol{\theta_0}$ be the true value of $\boldsymbol{\theta}$ and $U(\boldsymbol{X}; \boldsymbol{\theta})$ be a set of estimating equations. An unbiased estimating equation has expectation of 0 when evaluated at the true value $\boldsymbol{\theta_0}$. That is,

$$E[U(\mathbf{X}; \boldsymbol{\theta_0})] = 0$$

To estimate and make inference about $\boldsymbol{\theta_0}$ in the case where $dim(U) = dim(\boldsymbol{\theta})$, we solve the following equation to get an estimate $\hat{\boldsymbol{\theta}}$.

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i; \boldsymbol{\theta}) = 0$$

Note that the score function is an estimating equation where the model is fully specified.

Theorem 2.2.1. *Under certain regularity conditions, we have the following estimating equation properties:*

1. *Consistency:* $\hat{\theta} \xrightarrow{p} \theta_0$
2. *Asymptotic Normality:* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, E[\frac{\partial U(\theta_0)}{\partial \theta^T}]^{-1} Var(U(\theta_0)) E[\frac{\partial U(\theta_0)}{\partial \theta}]^{-1})$

The asymptotic normality property can be derived by applying Taylor expansion at the true value θ_0. We prove property 2 and the proof of property 1 can be done using the weak law of large number [13][14].

Proof.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\hat{\theta}) = 0$$

Apply Taylor expansion for $U_i(\theta)$ at θ_0:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Big[U_i(\theta_0) + \frac{\partial U_i(\tilde{\theta})}{\partial \theta^T}(\hat{\theta} - \theta_0)\Big] = 0$$

by the mean value theorem. $\tilde{\theta}$ is between θ_0 and $\hat{\theta}$. After some calculation, one can obtain:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial U_i(\tilde{\theta})}{\partial \theta^T}\sqrt{n}(\hat{\theta} - \theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\theta_0)$$

Using the following facts,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\theta_0) \xrightarrow{d} N(0, Var(U(\theta_0))$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial U_i(\tilde{\theta})}{\partial \theta^T} \xrightarrow{p} E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]$$

one can obtain

$$\sqrt{n}(\hat{\theta}-\theta_0) = -E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}\frac{1}{1+o_p(1)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\theta_0)$$

Since

$$\frac{1}{1+o_p(1)} = 1+o_p(1)$$

Then we have

$$\sqrt{n}(\hat{\theta}-\theta_0) = -E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\theta_0)+o_p(1)$$

Hence,

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N\Big(0, E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}Var(U(\theta_0))E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}\Big)$$

□

2.3 The IPTW Estimator

In this section, we construct the inverse probability treatment weighting estimator when the treatment effect is measured using a hazard ratio as well as explore the asymptotic properties.

Let $\hat{\theta}_{iptw}$ denote the conventional IPTW estimate and $\hat{\theta}_{siptw}$ denote the stabilized IPTW estimate. Let e_i and Z_i denote the propensity score and treatment status for individual i respectively. We obtain the estimate by solving the following weighted estimating equations:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{Z_i}{e_i}+\frac{1-Z_i}{1-e_i}\Big)U_i(\hat{\theta}_{iptw}) = 0$$

.

Assume $U(X;\theta)$ is an unbiased estimating equation. i.e. $E[U(X;\theta_0)] = 0$. The following theorems hold.

Theorem 2.3.1. *$\hat{\theta}_{iptw}$ is a consistent estimator of θ_0.*

Proof.

$$\begin{aligned} E\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)U(\theta_0)\Big] &= E\Big[E\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)U(\theta_0)|X\Big] \\ &= E\Big[\Big(\frac{E(Z|X)}{e}+\frac{1-E(Z|X)}{1-e}\Big)U(\theta_0)\Big] \\ &= E[2U(\theta_0)] \\ &= 0 \end{aligned}$$

□

Then the asymptotic normality property of the IPTW estimator can be easily obtained by applying Theorem 2.2.1.

Theorem 2.3.2. *$\sqrt{n}(\hat{\theta}_{iptw}-\theta_0) \xrightarrow{d} N(0, V_{iptw})$, where*

$$V_{iptw} = \frac{1}{4}E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1} E\Big[\Big(\frac{1}{e}+\frac{1}{1-e}\Big)U(\theta_0^{\otimes 2})\Big] E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}$$

Note: $U(\theta_0^{\otimes 2}) = U(\theta_0)U(\theta_0)^T$

Proof.

$$\begin{aligned} Var\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)U(\theta_0)\Big] &= E\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)^2 U(\theta_0^{\otimes 2})\Big] \\ &= E\Big[E\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)^2 U(\theta_0^{\otimes 2})|X\Big)\Big] \\ &= E\Big[\Big(\frac{E(Z|X)}{e^2}+\frac{1-E(Z|X)}{(1-e)^2}\Big)U(\theta_0^{\otimes 2})\Big] \\ &= E\Big[\Big(\frac{1}{e}+\frac{1}{1-e}\Big)U(\theta_0^{\otimes 2})\Big] \end{aligned}$$

$$
\begin{aligned}
E\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}\Big] &= E\Big[E\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}|X\Big] \\
&= E\Big[\Big(\frac{E(Z|X)}{e}+\frac{1-E(Z|X)}{1-e}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}\Big] \\
&= E\Big[2\frac{\partial U(\theta_0)}{\partial \theta^T}\Big] \\
&= 2E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]
\end{aligned}
$$

Hence, by Theorem 2.2.1, we have

$$
\begin{aligned}
V_{iptw} &= E\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1} Var\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)U(\theta_0)\Big] E\Big[\Big(\frac{Z}{e}+\frac{1-Z}{1-e}\Big)\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1} \\
&= \frac{1}{4}E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1} E\Big[\Big(\frac{1}{e}+\frac{1}{1-e}\Big)U(\theta_0^{\otimes 2})\Big] E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}
\end{aligned}
$$

□

We can derive properties of the stabilized inverse probability treatment weighting estimator using similar techniques. Hence, we have the following theorem.

Theorem 2.3.3. *Under certain regularity conditions, we have the following:*

1. $\hat{\theta}_{siptw}$ *is a consistent estimator of* θ_0.
2. $\sqrt{n}(\hat{\theta}_{siptw}-\theta_0) \xrightarrow{d} N(0, V_{siptw})$, *where*

$$
V_{siptw} = E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1} E\Big[\Big(\frac{p^2}{e}+\frac{(1-p)^2}{1-e}\Big)U(\theta_0^{\otimes 2})\Big] E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}
$$

The proof of this theorem follows the similar method as Theorem 2.3.1 and 2.3.2.

2.4 Estimating Equations for Cox Proportional Hazards Models

We define the following counting process notation. Let $Y(s) = I(s \leq w)$ denote the observation process and $\Delta N(s) = N(s + \Delta s) - N(s)$ denote the number of events in $[s, s + \Delta s)$. $dN(s) = I(s = w)$. The Cox proportional hazards model for estimating the marginal hazard ratio takes

$$h(w|x, z) = h_0(w)e^{\beta^m z}$$

where β^m is the log marginal hazard ratio. We consider semi-parametric regression where $h_0(w)$ is taken to be of no particular parametric form. The likelihood for subject i is of the form [15]:

$$L_i = h(w_i|x, z)exp(-\int_0^{w_i} h(t)dt)$$

The log-likelihood for l_i is:

$$\begin{aligned} l_i &= logh(w_i|x, z) - \int_0^{w_i} h(t|x, z)dt \\ &= \int_0^{\infty} dN_i(t)logh(t|x, z) - Y_i(t)h(t|x, z)dt \\ &= \int_0^{\infty} Y_i(t)[dN_i(t)(logh_0(t) + z_i\beta^m) - h_0(t)e^{z_i\beta^m}]dt \end{aligned}$$

Hence

$$l = \sum_{i=1}^{n} a_i l_i \tag{2.1}$$

$$= \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)a_i\Big[dN_i(t)(logh_0(t) + z_i\beta^m) - h_0(t)e^{z_i\beta^m}\Big] \tag{2.2}$$

where a_i are the IPTW weights.

Differentiate (2.2) with respect to $h_0(t)$ we get:

$$\frac{\partial l}{\partial h_0(t)} = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)a_i\Big[\frac{dN_i(t)}{h_0(t)} - h_0(t)z_i e^{z_i\beta^m}\Big]dt \tag{2.3}$$

Set (2.3) to be 0, we obtain:

$$\hat{h_0}(t) = \frac{\sum_{i=1}^{n} Y_i(t) a_i dN_i(t)}{\sum_{i=1}^{n} Y_i(t) a_i e^{z_i \beta^m}} \tag{2.4}$$

Differentiate l with respect to β^m we get:

$$\frac{\partial l}{\partial \beta^m} = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t) a_i [dN_i(t) z_i - h_0(t) z_i e^{z_i \beta^m}] dt \tag{2.5}$$

Plug $\hat{h_0}(t)$ into (2.5), we get:

$$U(\tilde{\beta}^m) = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t) a_i dN_i(t) \Big[z_i - \frac{\sum_{i=1}^{n} Y_i(t) a_i z_i e^{z_i \beta^m}}{\sum_{i=1}^{n} Y_i(t) a_i e^{z_i \beta^m}} \Big] dt$$

Define $S^{(0)}(\beta; t) = \sum_{i=1}^{n} Y_i(t) a_i e^{z_i \beta^m}$ and $S^{(1)}(\beta; t) = \sum_{i=1}^{n} Y_i(t) a_i z_i e^{z_i \beta^m}$

Then

$$U(\tilde{\beta}^m) = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t) dN_i(t) \Big[z_i - \frac{S^{(1)}(\beta; t)}{S^{(0)}(\beta; t)} \Big] dt$$

To obtain the estimate of the log marginal hazard ratio $\hat{\beta}^m$ we set $U(\tilde{\beta}^m) = 0$. We note that the estimating function theory deals with the possibility of misspecification.

2.5 Simulation Settings

2.5.1 Data Generation

We perform an extensive series of Monte Carlo simulations to examine the numerical performance of the IPTW estimator with both the conventional weights and the stablized weights. We simulate 10 independent baseline covariates $X_1, ..., X_{10}$. Of these 10 covariates, X_2, X_5 and X_9 follow a Bernoulli distribution with parameter equal to 0.5 and all other

covariates follow a standard normal distribution. The above simulation setting comes from Austin (2013) with slight modifications [11]. We assume X_1 - X_7 are associated with treatment assignment and X_4 - X_{10} are associated with outcome. Figure 2.2 summarizes the relationship among these variables. We allow covariates to have weak, moderate, strong and very strong effect on treatment assignment or outcome and denote the strength of association by α_w, α_m, α_s and α_{vs}, which is the log odds ratio per one unit increase in the corresponding covariate. We set the coefficients to be log(1.25), log(1.5), log(1.75) and log(2), respectively. We then use a logistic regression model to simulate treatment assignment probability for the i^{th} individual:

$$logit(p_i) = \alpha_0 + \alpha_w x_{1i} + \alpha_m x_{2i} + \alpha_s x_{3i} + \alpha_w x_{4i} + \alpha_m x_{5i} + \alpha_s x_{6i} + \alpha_{vs} x_{7i}$$

α_0 is determined by using a bisection approach suggested by Austin (2016) [6] to obtain the desired overall prevalence of treatment. We set α_0 to be -1.78 to achieve an overall prevalence of treatment of 25%. For each individual, we generate a treatment status from a Bernoulli distribution $Z_i \sim Beroulli(p_i)$ where p_i is the probability of treatment for the i^{th} individual generated as described above. Then, we simulate a linear predictor of the form:

$$LP_i = \beta^c Z_i + \alpha_w x_{4i} + \alpha_m x_{5i} + \alpha_s x_{6i} + \alpha_{vs} x_{7i} + \alpha_w x_{8i} + \alpha_m x_{9i} + \alpha_s x_{10i}$$

and denote it by LP_i. We generate a survival time for subject i from a Cox proportional hazard model with an exponential baseline hazard distribution with parameter $\lambda = 1$. That is, $h_0(t) = 1$. The generation algorithm is as follows: For each individual we generate an independent standard uniform random variable $u_i \sim U(0, 1)$. Then the survival time can be generated using the inverse CDF technique: $T_i = \frac{-log(u_i)}{e^{LP_i}}$[16]. As we can see from the data generation process, X_4 - X_7 both determine treatment assignment and are risk factors for outcome. Hence the treatment is confounded by X_4 - X_7 as we might expect in an observational data setting. The above data generation process is based on a conditional treatment effect (β^c). However, the IPTW estimator estimates the marginal effect [6]. When the treatment effect is measured using a hazard ratio, the conditional effect and

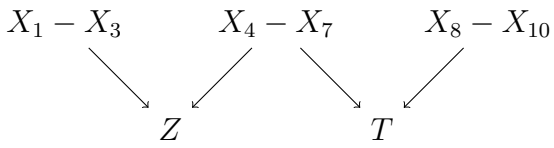

Figure 2.2: Causal Graph for simulation setting

the marginal effect do not coincide [4]. To this end, we employ a bisection approach [6] to determine the conditional log hazard ratio β^c that results in the desired true marginal hazard ratio. We simulate a dataset of size 10,000. For each subject, we simulate both potential outcomes under treatment and control conditions. Then we regress the survival outcome on treatment status to obtain the true marginal hazard ratio [11]. We note that this is a different simulation study than the one mentioned before and the reason why we do this is that the IPTW methods estimate the marginal treatment effect. Hence it is necessary to obtain the true marginal treatment effect in this setting.

We allow the true marginal hazard ratio to be 1, 1.5 and 2 (no treatment effect, weak treatment effect and strong treatment effect). For each true marginal hazard ratio, we estimate the marginal hazard ratio using a weighted Cox proportional hazard model with both the conventional weights and the stabilized weights. Hence, we examine a total of 6 scenarios. In each scenario, we simulate 1,000 datasets, each consisting of 10,000 subjects.

In each of the simulated datasets, we first regress the treatment status on X_1 - X_7 through a logistic regression model to estimate the propensity score. Then we regress the survival outcome on the treatment status using a weighted Cox proportional hazard model with both the conventional weights and the stabilized weights. We also incorporate three different variance estimators discussed in Section 2.1 to estimate the variance of the log marginal hazard ratio: the naive variance estimator, the robust variance estimator and the bootstrap variance estimator.

True log marginal HR β^m	True marginal HR e^{β^m}	True conditional HR β^c
0	1	0
0.4055	1.5	0.6894
0.6931	2	1.1711

HR: Hazard ratio

Table 2.1: Marginal and Conditional log hazard ratios used in simulation study

2.5.2 Simulation Results

As discussed before, our data generation process is based on a conditional treatment effect. However, the IPTW estimator estimates the marginal treatment effect. Table 2.1 gives the true conditional log hazard ratio generated in Section 2.5.1 for each of the three true marginal log hazard ratios.

In each of the 1,000 simulated datasets, we record the estimated log marginal hazard ratio $\hat{\beta}(j)$ and its naive standard error $\hat{\sigma}(j)$ from the Cox proportional hazard model output. We first calculate the mean of the log marginal hazard ratio across the 1000 iterations: $\overline{\hat{\beta}}_{iptw} = \frac{1}{1{,}000}\sum_{i=1}^{1{,}000}\hat{\beta}(i)$, so the estimated marginal hazard ratio is $e^{\overline{\hat{\beta}}_{iptw}}$. We define the average bias of the log marginal hazard ratio as: $\frac{\overline{\hat{\beta}}_{iptw}-\beta^m}{\beta^m}\cdot 100\%$ where β^m is the log true log marginal hazard ratio. Then we determine the average standard error of the log hazard ratio across the 1,000 iterations: $ASE = \hat{\sigma}_{iptw} = \frac{1}{1{,}000}\sum_{i=1}^{1{,}000}\hat{\sigma}(j)$. We also determine the empirical standard error of the 1,000 estimated log marginal hazard ratios: $ESE = sd(\hat{\beta}(j)) = \sqrt{\frac{\sum_{i=1}^{1{,}000}\left(\hat{\beta}(j)-\overline{\hat{\beta}}_{iptw}\right)^2}{1000-1}}$ [6]. If the variance of $\overline{\hat{\beta}}_{iptw}$ is correctly estimated, the average standard error should be close to the empirical standard error. That is, we expect the ratio $\frac{ASE}{ESE}$ to be close to 1. We summarize the simulation results with the conventional and stabilized IPTW weights in Table 2.2.

We observe that when treatment prevalence = 25%, both conventional IPTW and

True log cond HR β^c	True mar HR e^{β^m}	Estimated log mar HR $\overline{\hat{\beta}}_{iptw}$	Estimated mar HR $e^{\overline{\hat{\beta}}_{iptw}}$	Avg Bias	ASE	ESE	RSE	BSE
0	1	0.0011	1.0011	0.11%	0.0141	0.0377	0.0425	0.0393
0.6894	1.5	0.4085	1.5046	0.69%	0.0144	0.0443	0.0482	0.0451
1.1711	2	0.6967	2.0072	0.56%	0.0147	0.0507	0.0535	0.0505
0	1	0.0011	1.0011	0.11%	0.0232	0.0378	0.0426	0.0395
0.6894	1.5	0.4154	1.5150	2.39%	0.0234	0.0437	0.0482	0.0447
1.1711	2	0.7159	2.0460	3.33%	0.0237	0.0489	0.0530	0.0494

cond: Conditional
mar: Marginal
HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error
BSE: Average bootstrap standard error

Table 2.2: Simulation Results for Conventional and Stabilized IPTW Weights

stabilized IPTW result in approximately unbiased estimates of the marginal hazard ratio across all simulation scenarios. However, stabilized IPTW actually experiences greater bias. For variance estimation, the use of the naive variance estimator results in substantial bias in estimating the variance of log marginal hazard ratio. Although using the stabilized weights results in higher $\frac{ASE}{ESE}$ ratio, the improvement is not enough as the ratio is still far removed from one. We conjecture that this is because the inflated sample size induced by IPTW is taken into account while the within-subject correlation induced by IPTW is not. Both the robust variance estimator and the bootstrap variance estimator significantly reduce the bias in estimating the variance using both the conventional and the stabilized IPTW weights.

2.6 Discussion

The estimating equation theory shows that when the sample size goes to infinity, both the conventional IPTW estimator and the stabilized IPTW estimator lead to unbiased estimation of the marginal hazard ratio when the outcome is generated from a Cox proportional hazards model and correct propensity score model and outcome model are used. However, due to finite sample size of the simulated datasets, the IPTW method may yield biased estimation of the marginal hazard ratio. The simulation results indicate that the use of the conventional weights results in much lower bias in estimating the marginal hazard ratio than the use of the stabilized weights. Additionally, both the robust variance estimator and the bootstrap variance estimator provide a reasonable variance estimate. Therefore, to minimize estimation bias, we recommend researchers use the conventional IPTW with robust or bootstrap standard errors to estimate treatment effects in an observational data setting when the outcome is a single time to event.

Chapter 3

Two Events Setting

This chapter focuses on the development of propensity score methods in the setting of two events. This setting differs from the one event setting in that for each subject the two gap times may be correlated and as a result, the naive variance estimator often leads to incorrect variance estimates. To this end, methods for dealing with the within-subject correlation are discussed. Moreover, multiple propensity scores are estimated for a single subject if treatment changes and a new class of IPTW estimators are formulated. Three scenarios are discussed in this chapter: independent gap times, time-varying covariates with fixed treatment and time-varing covariates and treatment. For each scenario, model assumptions and specifications of the treatment model and the outcome model are given. A summary of the simulation results for all scenarios are presented at the end of the chapter.

3.1 Notation and Model Setup

We use the following notation throughout this chapter. Assume there are n subjects $i = 1, 2, ..., n$ although we suppress i notation in this chapter. Let $X(j)$ be a p-dimensional vector of covariates at the start of the j^{th} gap time and $Z(j)$ be treatment status at the

start of the j^{th} gap time. Let W_1 denote the first gap time and W_2 denote the second gap time. We define the propensity score for the first event e_1 to be

$$e_1 = P(Z(1) = 1|X(1) = x(1))$$

We define e_2 to be the propensity score for the second gap time i.e. the probability of treatment at the start of the second gap time conditional on all past covariate and treatment history. That is,

$$e_2 = P(Z(2) = 1|\overline{X}(2) = \overline{x}(2), \overline{Z}(1) = \overline{z}(1))$$

where $\overline{X}(j) = (X(1), X(2), ...X(j))$ is the covariate history through the start of the j^{th} gap time and $\overline{Z}(j) = (Z(1), Z(2), ...Z(j))$ is the treatment history through the start of the j^{th} gap time. Hence, the probability of treatment at the start of the first gap time and the start of the second gap time conditional on all past history is:

$$\begin{aligned} e_1e_2 &= P(Z(1) = 1, Z(2) = 1|\overline{X}(2) = \overline{x}(2), \overline{Z}(1) = \overline{z}(1)) \\ &= E(Z(1)Z(2)|\overline{X}(2) = \overline{x}(2), \overline{Z}(1) = \overline{z}(1)) \end{aligned}$$

Intuitively, the IPTW weights are defined to be the inverse of the probability of treatment path conditional on all past treatment and covariate history. We define the conventional inverse probability treatment weights for the first event to be

$$cw_1 = \frac{1}{P(Z(1) = z(1)|X(1) = x(1))} = \frac{Z(1)}{e_1} + \frac{1 - Z(1)}{1 - e_1}$$

and the stabilized inverse probability treatment weights to be

$$sw_1 = \frac{P(Z(1) = z(1))}{P(Z(1) = z(1)|X(1) = x(1))} = \frac{P(Z(1) = 1)Z(1)}{e_1} + \frac{P(Z(1) = 0)(1 - Z(1))}{1 - e_1}$$

as usual. We further define the conventional IPTW weights for the second event to be

$$\begin{aligned} cw_2 &= \frac{1}{(P(Z(1) = z(1)|X(1) = x(1))} \cdot \frac{1}{P(Z(2) = z(2)|\overline{X}(2) = \overline{x}(2), \overline{Z}(1) = z(1))} \\ &= \Big(\frac{Z(1)}{e_1} + \frac{1 - Z(1)}{1 - e_1}\Big)\Big(\frac{Z(2)}{e_2} + \frac{1 - Z(2)}{1 - e_2}\Big) \end{aligned}$$

The stabilized IPTW weights for the second event are defined to be

$$sw_2 = \frac{P(Z(1)=z(1))}{P(Z(1)=z(1),X(1)=x(1))} \cdot \frac{P(Z(2)=z(2)|Z(1)=z(1))}{P(Z(2)=z(2)|\overline{X}(2)=\overline{x}(2),\overline{Z}(1)=\overline{z}(1))}$$
$$= p_{11}\frac{Z(1)Z(2)}{e_1e_2} + p_{10}\frac{Z(1)(1-Z(2))}{e_1(1-e_2)} + p_{01}\frac{(1-Z(1))Z(2)}{(1-e_1)e_2} + p_{00}\frac{(1-Z(1))(1-Z(2))}{(1-e_1)(1-e_2)}$$

where $p_{ij} = P(Z(1)=i, Z(2)=j)$ [17].

The reason why we consider the stabilized weights is that the conventional weights sometimes result in extremely large weights for a few subjects. As a result, these subjects dominate the weighted analysis, and this results in unstable estimation of the marginal hazard ratio. The use of the conventional weights sometimes also leads to rather large variance for the conventional IPTW estimator [17].

3.2 Estimating Equation Theory for Recurrent Events

In this section we discuss asymptotic properties of IPTW estimators in a recurrent events data setting. Let $U(X;\theta)$ be a set of unbiased estimating equations for the second gap time i.e. $E(U(X;\theta_0)) = 0$. To obtain the IPTW estimate for the second gap time, we solve the following weighted estimating equations:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{Z_i(1)}{e_{1i}} + \frac{1-Z_i(1)}{1-e_{1i}}\Big)\Big(\frac{Z_i(2)}{e_{2i}} + \frac{1-Z_i(2)}{1-e_{2i}}\Big)U_i(\hat{\theta}) = 0$$

Denote the conventional IPTW estimate by $\hat{\theta}_{iptw}$. Under certain regularity conditions, we have the following theorems.

Theorem 3.2.1. *$\hat{\theta}_{iptw}$ is a consistent estimator of θ_0.*

Proof.

$$
\begin{aligned}
&E\Big[\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)U(\theta_0)\Big]\\
&=E\Big[E\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)U(\theta_0)|\overline{X}(2),Z(1)\Big]\\
&=E\Big[\frac{E[Z(1)Z(2)|\overline{X}(2),Z(1)]}{e_1e_2}U(\theta_0)\Big]+E\Big[\frac{E[Z(1)(1-Z(2))|\overline{X}(2),Z(1)]}{e_1(1-e_2)}U(\theta_0)\Big]\\
&+E\Big[\frac{E[(1-Z(1))Z(2)|\overline{X}(2),Z(1)]}{(1-e_1)e_2}U(\theta_0)\Big]+E\Big[\frac{E[(1-Z(1))(1-Z(2))|\overline{X}(2),Z(1)]}{(1-e_1)(1-e_2)}U(\theta_0)\Big]\\
&=4E[U(\theta_0)]\\
&=0
\end{aligned}
$$

□

Next we derive the asymptotic variance of the conventional IPTW estimator.

Theorem 3.2.2. *$\sqrt{n}(\hat{\theta}_{iptw}-\theta_0)\xrightarrow{d}N(0,V_{iptw})$, where*

$$V_{iptw}=\frac{1}{16}E\Big[\frac{\partial U(\theta_0)}{\partial\theta^T}\Big]^{-1}E\Big[\Big(\frac{1}{e_1e_2}+\frac{1}{e_1(1-e_2)}+\frac{1}{(1-e_1)e_2}+\frac{1}{(1-e_1)(1-e_2)}\Big)U(\theta_0^{\otimes 2})\Big]E\Big[\frac{\partial U(\theta_0)}{\partial\theta}\Big]^{-1}$$

Proof.

$$
\begin{aligned}
&Var\Big[\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)U(\theta_0)\Big]\\
&=E\Big[\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)^2\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)^2U(\theta_0^{\otimes 2})\Big]\\
&=E\Big[\Big(\frac{Z(1)Z(2)}{e_1^2e_2^2}+\frac{Z(1)(1-Z(2))}{e_1^2(1-e_2)^2}+\frac{(1-Z(1))Z(2)}{(1-e_1)^2e_2^2}+\frac{(1-Z(1))(1-Z(2))}{(1-e_1)^2(1-e_2)^2}\Big)U(\theta_0^{\otimes 2})\Big]\\
&=E\Big[\Big(\frac{E[Z(1)Z(2)|\overline{X}(2),Z(1)]}{e_1^2e_2^2}+\frac{E[Z(1)(1-Z(2))|\overline{X}(2),Z(1)]}{e_1^2(1-e_2)^2}\\
&+\frac{E[(1-Z(1))Z(2)|\overline{X}(2),Z(1)]}{(1-e_1)^2e_2^2}+\frac{E[(1-Z(1))(1-Z(2))|\overline{X}(2),Z(1)]}{(1-e_1)^2(1-e_2)^2}\Big)U(\theta_0^{\otimes 2})\Big]\\
&=E\Big[\Big(\frac{1}{e_1e_2}+\frac{1}{e_1(1-e_2)}+\frac{1}{(1-e_1)e_2}+\frac{1}{(1-e_1)(1-e_2)}\Big)U(\theta_0^{\otimes 2})\Big]
\end{aligned}
$$

$$
\begin{aligned}
&E\Big[\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]\\
&=E\Big[E\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}|\overline{X}(2),Z(1)\Big]\\
&=E\Big[\frac{E[Z(1)Z(2)|\overline{X}(2),Z(1)]}{e_1e_2}\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]+E\Big[\frac{E[Z(1)(1-Z(2))|\overline{X}(2),Z(1)]}{e_1(1-e_2)}\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]\\
&+E\Big[\frac{E[(1-Z(1))Z(2)|\overline{X}(2),Z(1)]}{(1-e_1)e_2}\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]+E\Big[\frac{E[(1-Z(1))(1-Z(2))|\overline{X}(2),Z(1)]}{(1-e_1)(1-e_2)}\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]\\
&=4E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]
\end{aligned}
$$

By Theorem 2.2.1, we have

$$
\begin{aligned}
V_{iptw}=&E\Big[\Big(\frac{Z(1)}{1-e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}Var\Big[\Big(\frac{Z(1)}{e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)U(\theta_0)\Big]\\
&E\Big[\Big(\frac{Z(1)}{1-e_1}+\frac{1-Z(1)}{1-e_1}\Big)\Big(\frac{Z(2)}{e_2}+\frac{1-Z(2)}{1-e_2}\Big)\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}\\
&=\frac{1}{16}E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}E\Big[\Big(\frac{1}{e_1e_2}+\frac{1}{e_1(1-e_2)}+\frac{1}{(1-e_1)e_2}+\frac{1}{(1-e_1)(1-e_2)}\Big)U(\theta_0^{\otimes 2})\Big]E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}
\end{aligned}
$$

□

The asymptotic distribution of the stabilized IPTW estimator $\hat{\theta}_{siptw}$ can be derived using the similar method. We give the asymptotic results and omit the proof.

Theorem 3.2.3. *Under certain regularity conditions,*

1. $\hat{\theta}_{siptw}$ *is a consistent estimator of* θ_0
2. $\sqrt{n}(\hat{\theta}_{siptw}-\theta_0)\xrightarrow{d}N(0,V_{siptw})$, *where*

$$
V_{siptw}=E\Big[\frac{\partial U(\theta_0)}{\partial \theta^T}\Big]^{-1}E\Big[\Big(\frac{p_{11}^2}{e_1e_2}+\frac{p_{10}^2}{e_1(1-e_2)}+\frac{p_{01}^2}{(1-e_1)e_2}+\frac{p_{00}^2}{(1-e_1)(1-e_2)}\Big)U(\theta_0^{\otimes 2})\Big]E\Big[\frac{\partial U(\theta_0)}{\partial \theta}\Big]^{-1}
$$

3.3 Time-Fixed Treatment and Covariates

We start with the simplest case where there are two independent gap times W_1 and W_2 for each subject. For simplicity we assume X is a 1-dimensional scalar. Figure 3.1 illustrates

the relationship among X, Z, W_1 and W_2. Our goal is to use propensity score methods to consistently estimate the marginal treatment effect. In this setting we assume treatment and covariates are fixed over time, so we use Z and X without the j notation. Define e to be the probability of treatment conditional on covariates. We regress treatment indicator Z on X to obtain the estimated propensity score $\hat{e}$:

$$\hat{e} = expit\big(\hat{\alpha}_0 + \hat{\alpha}_1 x\big)$$

To estimate the marginal treatment effect, we regress the survival outcomes W_1 and W_2 on the treatment status Z through a weighted Cox proportional hazards model with both the conventional and stabilized weights as defined in Section 3.1:

$$h_j(w|x,z) = h_0(w)e^{\beta^m z}$$

where $j = 1, 2$.

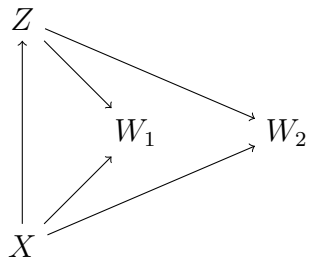

Figure 3.1: Causal graph for time-fixed treatment and covariate

We perform a simulation study to examine the numerical performance of the IPTW estimator with both the conventional and stabilized weights. For simplicity, we generate a standard normal covariate x. For each subject we generate a treatment probability through a logistic regression model:

$$\pi = expit\big(\alpha_0 + \alpha_1 x\big)$$

Then we generate a treatment status for each individual $z \sim Bernoulli(\pi)$. We set α_0 to be -1.1392 by a bisection approach to achieve an overall treatment prevalence of 25% [6].

Here α_1 represents the log odds ratio of treatment per unit increase in x and we set it to be log(1.5). For each subject we simulate two independent gap times W_1 and W_2 from a Cox proportional hazards model. We choose the baseline hazard to be an exponential distribution with $\lambda = 1$. Hence, the hazard takes the form:

$$h_j(w|x,z) = e^{\beta^c z + \beta_1 x}$$

where j =1 ,2. The association parameter between X and W_j is β_1, and is set to be log(1.5). The simulation algorithm for W_1 and W_2 is as follows [9]

- Simulate two independent standard uniform distribution u_1 and u_2
- Simulate $w_1 = \dfrac{-log(u_1)}{e^{\beta^c z + \beta_1 x}}$ and $w_2 = \dfrac{-log(u_2)}{e^{\beta^c z + \beta_1 x}}$

The above data generation method is based on a conditional hazard ratio e^{β^c}. However, the IPTW estimator estimates the marginal hazard ratio. To this end, we use a bisection approach to determine β^c that induces the specified marginal hazard ratio e^{β^m} [6].

To estimate β^m, first we obtain the estimated propensity score $\hat{e}$ through a logistic regression model. Then we calculate the conventional weights $cw_1 = \frac{Z}{e} + \frac{1-Z}{1-e}$ and the stabilized weights $sw_1 = \frac{P(Z=1)Z}{e} + \frac{P(Z=0)(1-Z)}{1-e}$. Finally we regress the gap times on the treatment indicator through a Cox proportional hazards model:

$$h_j(w|x,z) = h_0(w)e^{\beta^m z}$$

Doing this allows us to estimate the marginal treatment effect. Since weighting artificially creates a cluster for each subject, inducing a within-subject correlation, the naive variance estimator often fails to correctly estimate the variance of $\hat{\beta}^m$ [17]. To address this issue, we use the robust variance estimator proposed by Lin [7]. The robust variance estimator allows us to rewrite the dependence summations in estimating equations as independent, identical distributed summations, from which the asymptotic variance can be derived. To implement the robust variance estimator in R, we use the following formula:

$$coxph(Surv(W) \sim Z + cluster(id), weights = weight, data = dataset)$$

3.4 Time-Varying Covariates

Next we consider the case where covariates change over time while treatment remains the same. Assume $X(j)$ is a 1-dimensional scalar. The relationship among the variables is illustrated in Figure 3.2. Due to the change of covariates, the marginal hazard ratio may differ for the two gap times. The methodology for estimating the marginal hazard ratio is as follows: First we estimate the propensity score through a logistic regression model:

$$\hat{e} = expit\big(\hat{\alpha}_0 + \hat{\alpha}_1 x(1)\big)$$

Then to estimate the marginal hazard ratio for the first and second gap time, we run a weighted Cox proportional hazards model, whose hazard takes the form:

$$h_j(w|\overline{x}(j), \overline{z}(j)) = h_{0_j}(w)e^{\beta^{m_j} z}$$

where $j = 1, 2$.

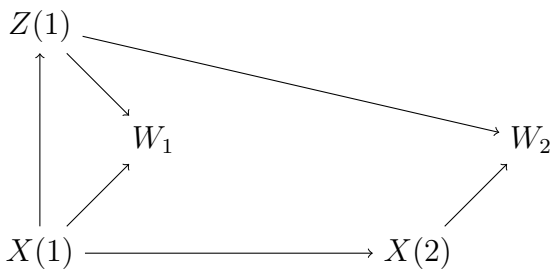

Figure 3.2: Causal graph for time-varying covariates

We perform the following simulation study to examine the numeric performance of the proposed IPTW estimator. We follow the same data generation methods described in Section 3.2 for the first gap time. Based on the first gap time W_1 and the covariate $X(1)$, we simulate the $X(2)$ covariate a second dependent gap time W_2 as follows [9]:

- Simulate a standard uniform random variable u_2

- Simulate a random variable $v \sim N(0, 16)$ independent of $x(1)$ and u_2
- Set $x(2) = x(1) + v$
- Simulate $w_2 = \dfrac{-log(u_2)}{e^{\beta^c z + \beta_1 x(2)}}$

The above data generation for the second gap time results in the same conditional hazard ratio e^{β^c} for both gap times. However, the marginal hazard ratio may not be the same for the two gap times. Given the log of the conditional hazard ratio β^c, we determine the true marginal hazard ratio for the second event using a similar method to that discussed in Section 3.1.

We obtain the estimated propensity score $\hat{e}$, along with the conventional weights and the stabilized weights for both gap times. Then, we regress the gap time W_j on treatment indicator $Z(j)$ through a weighted Cox proportional hazards model with both weights for both gap times to estimate the marginal hazard ratio.

$$h_j(w|\overline{x}(j), z) = h_{0_j}(w)e^{\beta^{m_j} z}$$

where $j = 1, 2$. Here β^{m_j} denotes the log marginal hazard ratio for the j^{th} gap time. Finally, we estimate the variance of $\hat{\beta}^{m_1}$ and $\hat{\beta}^{m_2}$ using both the naive variance estimator and the robust variance estimator.

3.5 Time-Varying Treatment and Covariates

We make further extensions by considering both time-varying treatment and covariates. Assume $X(j)$ is a 1-dimensional scalar. In such a setting, treatment status at the start of the second gap time, $Z(2)$, is dependent on treatment status at the beginning, $Z(1)$, and covariate value at the start of the second gap time $X(2)$. Figure 3.3 illustrates the relationship among these variables. In this setting the change of treatment results in a

different propensity score for the second gap time, and hence we need to estimate the IPTW weights for the second gap time as well. We estimate the propensity score for the first and second gap time through the following logistic regression models:

$$\hat{e}_1 = expit\big(\hat{\alpha}_0 + \hat{\alpha}_1 x(1)\big)$$

$$\hat{e}_2 = expit\big(\hat{\gamma}_0 + \hat{\gamma}_1 x(2) + \hat{\gamma}_2 z(1)\big)$$

Then we estimate the marginal hazard ratio for the first and second gap time through the following weighted Cox proportional hazards models:

$$h_j(w|\overline{x}(j), z(j)) = h_{0_j}(w)e^{\beta^{m_j} z(j)}$$

We illustrate our methodology for estimating the marginal treatment effect through a simulation study. We use the previously discussed data generation methods for the first gap time. We consider two dependence relationship between $X(1)$ and $X(2)$:

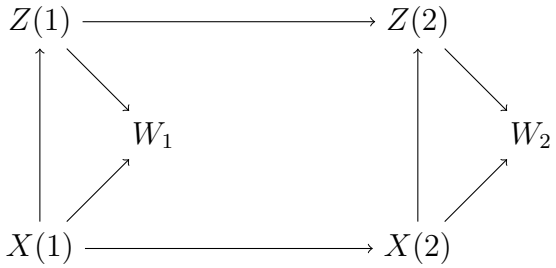

Figure 3.3: Causal graph for time-varying treatment and covariate

$$x(2) = x(1) + N(0, 16)$$

or

$$x(2) = x(1) + N(0, 1)$$

The correlation between $X(1)$ and $X(2)$ is approximately 0.24 for the first scenario and approximately 0.71 for the second scenario. Then, for each scenario, we simulate the

treatment status for the second gap time $Z(2)$ as follows. First we simulate a treatment probability through a logistic regression model:

$$logit(\pi_2) = \gamma_0 + \gamma_1 x(2) + \gamma_2 z(1)$$

We set γ_1 to be log(1.5) and allow γ_2 to be log(1.5) or log(0.25). Hence the log odds ratio of treatment when t = 2 is 1.5 per one unit increase in $X(2)$ keeping treatment at t = 1 the same. We set γ_0 to be 0.3338 and -0.1000 for the above two scenarios to achieve an overall treatment prevalence of 50% at t = 2 [6]. Having set all the parameters for the treatment model, we generate treatment status $Z(2) \sim Bernoulli(\pi_2)$. We simulate the first and second gap time W_1 and W_2 from a Cox proportional hazards model, whose hazard takes the form:

$$h_j(w|\overline{x}(j), \overline{z}(j)) = h_0(w)e^{\beta^c z(j) + \beta_1 x(j)}$$

The above generation technique results in the same conditional hazard ratio e^{β^c} for both gap times. However, the marginal hazard ratio may be different. We follow the similar method to that described in Section 3.1 to obtain the true marginal hazard ratio for both gap times.

To estimate the marginal hazard ratio for both gap times, first we obtain the estimated propensity score $\hat{e}_1$ and $\hat{e}_2$ through the following logistic regression models:

$$\hat{e}_1 = expit\big(\hat{\alpha_0} + \hat{\alpha_1} x(1)\big)$$

and

$$\hat{e}_2 = expit\big(\hat{\gamma_0} + \hat{\gamma_1} x(2) + \hat{\gamma_2} z(1)\big)$$

Then we regress the first gap time on treatment indicator $Z(1)$ through a Cox proportional hazards model using both the conventional weights cw_1 and the stabilized weights sw_1 from Section 3.1 to obtain the estimated marginal hazard ratio $e^{\hat{\beta}^{m_1}}$ for the first gap time. We run another weighted Cox proportional hazards model to regress the second gap time on treatment indicator $Z(2)$ using both the conventional weights cw_2 and the stabilized

weights sw_2 to obtain the estimated marginal hazard ratio $e^{\hat{\beta}^{m_2}}$ for the second gap time. The hazard takes the form:

$$h_j(w|\overline{x}(j), z(j)) = h_{0_j}(w)e^{\beta^{m_j} z(j)}$$

where $j = 1, 2$. Finally, we estimate the variance of $\hat{\beta}^{m_1}$ and $\hat{\beta}^{m_2}$ using both the naive variance estimator and the robust variance estimator.

3.6 Administrative Censoring

Often we have to deal with censored recurrent data where each subject has a different number of recurrent events. When censoring is a time-dependent confounder, the previous methods for estimating the marginal treatment effect without adjustments for censoring may yield biased results. To this end we incorporate weights for censoring to consistently estimate the marginal treatment effect. In this section we focus on the case where there is an administrative censoring time τ. We define the censoring indicator $\delta_1 = I(w_1 \leq \tau)$ and $\delta_2 = I(w_1 + w_2 \leq \tau)$. We can treat $(Z(i), \delta_i)$ as a treatment vector at the start of the i^{th} gap time. Thus, intuitively the IPTW weights are the inverse of the probability of treatment history the subject actually experienced.

The conventional censoring weights are defined as [17]

$$cw_1^{\dagger} = \frac{\delta_1}{P(\delta_1 = 1|x(1), z(1))}$$

$$cw_2^{\dagger} = \frac{\delta_1}{P(\delta_1 = 1|x(1), z(1))} \cdot \frac{\delta_2}{P(\delta_2 = 1|\delta_1 = 1, \overline{x}(2), \overline{z}(2))}$$

and the stabilized censoring weights are:

$$sw_1^{\dagger} = \frac{P(\delta_1 = 1)}{P(\delta_1 = 1|x(1), z(1))}$$

$$sw_2^{\dagger} = \frac{P(\delta_1 = 1)}{P(\delta_1 = 1|x(1), z(1))} \cdot \frac{P(\delta_2 = 1|\delta_1 = 1)}{P(\delta_2 = 1|\delta_1 = 1, \overline{x}(2), \overline{z}(2))}$$

To estimate the marginal hazard ratio for the first and second gap time, we solve the following weighted estimating equation

$$\sum_{i=1}^{n} a_i U_i(\tilde{\beta}^m) = 0$$

where $a_j = cw_j cw_j^{\dagger}$ for the conventional weights and $a_j = sw_j sw_j^{\dagger}$ for the stabilized weights where $j = 1, 2$.

We incorporate an administrative censoring time $\tau = 1$ in one of the time-varying treatment and covariates settings described in Section 3.4 with correlation of 0.24 between $X(1)$ and $X(2)$, and γ_2 is set to be $\log(1.5)$. The above setting results in approximately 30% of the subjects being censored for the first gap time and approximately 60% of the subjects being censored for the second gap time. We use both the conventional weights and the stabilized weights defined above to estimate the marginal hazard ratio and its standard errors for the first and second gap time. The simulation results are available in Table 3.8.

To further investigate the behaviour of the IPTW estimators, we increase the censoring proportion by incorporating another administrative censoring time $\tau = 0.25$. We keep other variables in the last setting the same. This results in approximately 70% of the subjects being censored for the first gap time and approximately 90% of the subjects being censored for the second gap time. We record the simulation results for this setting in Table 3.9.

3.7 Simulation Results

We allow the true marginal hazard ratio for the first gap time $e^{\beta^{m_1}}$ to be 1, 1.5 and 2. We determine the corresponding β^c that results in the specified marginal hazard ratios using a bisection approach [6]. For a given β^c, there is also a corresponding true marginal hazard ratio for the second gap time $e^{\beta^{m_2}}$. We summarize the relationship in Table 3.1.

True log marginal HR β^{m_1}	True marginal HR $e^{\beta^{m_1}}$	True log conditional HR β^c	True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$
0	1	0	0	1
0.4055	1.5	0.4599	0.2085	1.2318
0.6931	2	0.7830	0.3551	1.4263

HR: Hazard ratio

Table 3.1: Marginal and conditional log hazard ratios used in simulation study

For each of the three simulation settings, we simulate 1,000 datasets, each consisting of 10,000 subjects. In each of the 1,000 simulated datasets, we record the estimated log marginal hazard ratio for both gap times $\hat{\beta}_1(j)$ and $\hat{\beta}_2(j)$, along with its naive standard error $\hat{\sigma}_1(j)$ and $\hat{\sigma}_2(j)$. We record the average estimated log marginal hazard ratio $\overline{\hat{\beta}}^{m_k} = \sum_{j=1}^{1,000} \hat{\beta}_k(j)$ for $k = 1, 2$. We define the average bias of the log marginal hazard ratio as: $\frac{\overline{\hat{\beta}}^{m_j} - \beta^{m_j}}{\beta^{m_j}} \cdot 100\%$ where $j = 1, 2$. Then we determine the average standard error of the log hazard ratio across the 1,000 datasets: $ASE_k = \hat{\sigma}_k = \frac{1}{1,000} \sum_{j=1}^{1,000} \hat{\sigma}_k(j)$ where $k = 1, 2$. We also determine the empirical standard error of the 1,000 estimated log marginal hazard ratios for both gap times: $ESE_k = \sqrt{\frac{\sum_{j=1}^{1,000} \left(\hat{\beta}_k(j) - \beta^{m_k}\right)^2}{1,000-1}}$ where $k = 1, 2$ [6]. If the variance of $\hat{\beta}^{m_1}$ and $\hat{\beta}^{m_2}$ are correctly estimated, the average standard error should be close to the empirical standard error. For each of the three simulation settings, we record the average estimated log marginal hazard ratio, along with its naive average standard error, robust standard error and empirical standard error for both gap times. We summarize the simulation results for the second gap time in Table 3.2 for independent gap times with fixed treatment and covariates, Table 3.3 for time-varying covariates and Tables 3.4 - 3.7 for time-varying treatment and covariates. The upper half of the table is for the conventional weights and lower half for the stabilized weights.

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	-0.0004	0.9996	-0.04%	0.0100	0.0176	0.0196
0.4055	1.5	0.4054	1.4999	-0.01%	0.0102	0.0193	0.0209
0.6931	2	0.6939	2.0015	0.11%	0.0105	0.0211	0.0222
0	1	0.0002	1.0002	-0.02%	0.0163	0.0175	0.0196
0.4055	1.5	0.4091	1.5055	0.89%	0.0165	0.0191	0.0207
0.6931	2	0.7009	2.0156	1.13%	0.0168	0.0200	0.0216

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.2: Simulation results for independent gap times

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0001	1.0001	0.01%	0.0142	0.0249	0.0247
0.2085	1.2318	0.2085	1.2318	0.00%	0.0142	0.0245	0.0254
0.3551	1.4263	0.3554	1.4268	0.08%	0.0143	0.0265	0.0262
0	1	-0.0004	0.9996	-0.04%	0.0231	0.0244	0.0247
0.2085	1.2318	0.2104	1.2342	0.91%	0.0232	0.0255	0.0257
0.3551	1.4263	0.3619	1.4360	1.91%	0.0232	0.0265	0.0266

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.3: Simulation results for time-varying covariates

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0143	1.0144	1.43%	0.0101	0.0848	0.0673
0.2085	1.2318	0.2233	1.2502	7.10%	0.0101	0.0889	0.0724
0.3551	1.4263	0.3671	1.4435	3.38%	0.0102	0.0985	0.0777
0	1	0.0041	1.0041	0.41%	0.0201	0.0483	0.0493
0.2085	1.2318	0.2161	1.2412	3.64%	0.0201	0.0557	0.0534
0.3551	1.4263	0.3581	1.4306	0.84%	0.0203	0.0615	0.0582

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.4: Simulation results for time-varying treatment and covariates,$\gamma_2 = \log(0.25)$,Corr(X(1),X(2)) = 0.24

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0041	1.0041	0.41%	0.0100	0.0578	0.0556
0.2085	1.2318	0.2164	1.2416	3.79%	0.0101	0.0593	0.0593
0.3551	1.4263	0.3610	1.4348	1.66%	0.0101	0.0675	0.0637
0	1	0.0050	1.0050	0.50%	0.0201	0.0475	0.0501
0.2085	1.2318	0.2132	1.2376	2.25%	0.0201	0.0549	0.0547
0.3551	1.4263	0.3598	1.4330	1.32%	0.0203	0.0636	0.0586

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.5: Simulation results for time-varying treatment and covariates,$\gamma_2 = \log(1.5)$, Corr(X(1),X(2)) = 0.24

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0003	1.0003	0.03%	0.0100	0.0343	0.0346
0.3697	1.4473	0.3707	1.4487	0.27%	0.0101	0.0370	0.0372
0.6323	1.8819	0.6345	1.8861	0.35%	0.0104	0.0404	0.0394
0	1	0.0002	1.0002	0.02%	0.0200	0.0198	0.0222
0.3697	1.4473	0.3705	1.4485	0.22%	0.0203	0.0223	0.0238
0.6323	1.8819	0.6324	1.8821	0.02%	0.0207	0.0225	0.0252

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.6: Simulation results for time-varying treatment and covariates,$\gamma_2 = \log(0.25)$, Corr(X(1),X(2)) = 0.71

True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	-0.0004	0.9996	-0.04%	0.0100	0.0249	0.0273
0.3697	1.4473	0.3700	1.4477	0.08%	0.0101	0.0278	0.0291
0.6323	1.8819	0.6342	1.8857	0.30%	0.0101	0.0291	0.0310
0	1	0.0011	1.0011	0.11%	0.0200	0.0203	0.0224
0.3697	1.4473	0.3704	1.4482	0.19%	0.0203	0.0218	0.0240
0.6323	1.8819	0.6327	1.8827	0.06%	0.0208	0.0229	0.0256

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.7: Simulation results for time-varying treatment and covariates,$\gamma_2 = \log(1.5)$, Corr(X(1),X(2)) = 0.71

True log marginal HR β^{m_1}	True marginal HR $e^{\beta^{m_1}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_1}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_1}}$	Avg bias	ASE	ESE	RSE
0	1	-0.0006	0.9994	-0.06%	0.0178	0.0282	0.0305
0.4055	1.5	0.4258	1.5308	5.01%	0.0170	0.0280	0.0289
0.6931	2	0.7205	2.0555	3.95%	0.0167	0.0271	0.0287
0	1	-0.0002	0.0098	-0.02%	0.0231	0.0249	0.0259
0.4055	1.5	0.4087	1.5049	0.79%	0.0234	0.0252	0.0273
0.6931	2	0.7011	2.0160	1.15%	0.0238	0.0277	0.0286
True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0029	1.0029	0.29%	0.0181	0.0686	0.0723
0.2085	1.2318	0.2169	1.2422	4.03%	0.0164	0.0638	0.0689
0.3551	1.4263	0.3606	1.4341	1.55%	0.0154	0.0603	0.0671
0	1	0.0042	1.0042	0.42%	0.0201	0.0451	0.0494
0.2085	1.2318	0.2154	1.2404	3.31%	0.0202	0.0574	0.0529
0.3551	1.4263	0.3605	1.4340	1.52%	0.0203	0.0578	0.0575

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.8: Simulation results for time-varying treatment and covariates with administrative censoring time $\tau = 1$,$\gamma_2 = \log(1.5)$, Corr(X(1),X(2)) = 0.24

True log marginal HR β^{m_1}	True marginal HR $e^{\beta^{m_1}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_1}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_1}}$	Avg bias	ASE	ESE	RSE
0	1	-0.0027	0.9973	-0.27%	0.0293	0.0487	0.0485
0.4055	1.5	0.4491	1.5669	10.75%	0.0269	0.0422	0.0428
0.6931	2	0.7590	2.1361	9.51%	0.0258	0.0393	0.0399
0	1	-0.0004	0.0096	-0.04%	0.0231	0.0239	0.0259
0.4055	1.5	0.4082	1.5041	0.67%	0.0234	0.0262	0.0274
0.6931	2	0.7004	2.0146	1.05%	0.0238	0.0265	0.0286
True log marginal HR β^{m_2}	True marginal HR $e^{\beta^{m_2}}$	Estimated log marginal HR $\overline{\hat{\beta}}^{m_2}$	Estimated marginal HR $e^{\overline{\hat{\beta}}^{m_2}}$	Avg bias	ASE	ESE	RSE
0	1	0.0202	1.0204	2.04%	0.0428	0.1861	0.1797
0.2085	1.2318	0.2620	1.2995	25.66%	0.0362	0.1850	0.1730
0.3551	1.4263	0.4400	1.5527	23.91%	0.0325	0.1705	0.1618
0	1	0.0039	1.0039	0.39%	0.0201	0.0495	0.0498
0.2085	1.2318	0.2147	1.2395	2.97%	0.0201	0.0533	0.0534
0.3551	1.4263	0.3619	1.4361	1.91%	0.0203	0.0579	0.0568

HR: Hazard ratio
ASE: Average standard error
ESE: Empirical standard error
RSE: Average robust standard error

Table 3.9: Simulation results for time-varying treatment and covariates with administrative censoring time $\tau = 0.25$,$\gamma_2 = \log(1.5)$, Corr(X(1),X(2)) = 0.24

3.8 Discussion of Simulation Results

In this section we summarize main results from the above simulation studies. From the estimating equation theory, when there is no censoring the estimate $\hat{\beta}^{m_j}$ converges in probability to the log of the true log marginal hazard ratio β^{m_j} when the sample size n goes to infinity. For the above simulation scenarios with 10,000 subjects, the bias is negligible for time-fixed treatment and covariates and time-varying covariates. For the time-varying treatment and covariates scenarios, the use of the conventional weights tends to result in greater bias, whereas the bias is lower when using the stabilized weights. In the presence of censoring, the use of the conventional weights results in larger bias as the censoring proportion increases while the use of the stabilized weights seem to result in unbiased estimate of the marginal hazard ratio. For variance estimation, the naive variance estimator tends to either overestimate or underestimate the variance of the IPTW estimator across all simulation scenarios, while the robust variance estimator approximates the variance reasonably well across all scenarios. Therefore, we recommend researchers estimate the marginal treatment effects with the stabilized weights with the robust variance estimator for recurrent data.

3.9 Investigation of Proportional Hazards

If the Cox model for estimating the marginal hazard ratio is correctly specified, we would expect that the hazard ratio

$$\frac{h_j(w|\overline{X}(j) = \overline{x}(j), Z(j) = 1)}{h_j(w|\overline{X}(j) = \overline{x}(j), Z(j) = 0)} = e^{\beta^{m_j}}$$

To assess the proportional hazards assumption, we use the cox.zph function with identity link for both the first and second gap time with a critical p-value of 0.05. We test the following hypothesis [15].

$$H_0 : h_j(w|\overline{x}(j), z(j)) = h_{0_j}(w)e^{\beta^{m_j} z}$$

$$H_A : h_j(w|\overline{x}(j), z(j)) = h_{0_j}(w)e^{\beta^{m_j} z + \psi\beta^{m_j} w}$$

where $j = 1, 2$. We then are simply testing if $\psi = 0$. For time-varying treatment and covariates scenarios, the results show that for the first gap time the proportional hazards assumption is violated for about 50% of the simulation samples, while for the second gap time the proportional hazards assumption is violated for all simulation samples. The results are not surprising because the data are simulated using a conditional model including both treatment and covariates, whereas we fit the data using a marginal model including only the treatment indicator. Although this is a misspecified model, the estimate still has meaningful interpretations as the estimate converges to the log marginal hazard ratio instead of the log conditional hazard ratio[7].

3.10 Application

We apply our methods to the pulmonary exacerbations and rhDNase treatment example in Cook and Lawless (2007)[8]. We denote Z to be the rhDNase treatment, X be the forced expiratory volume (fev), δ to be the censoring indicator, W_1 to be the first gap time and W_2 to be the second gap time. To estimate the marginal hazard ratio for the first time and second gap time, we estimate the propensity score for subjects who experience the first and second gap time through the following two separate logistic regression models:

$$\hat{e}_1 = expit\big(\hat{\alpha}_0 + \hat{\alpha}_1 x\big)$$

$$\hat{e}_2 = expit\big(\hat{\gamma}_0 + \hat{\gamma}_1 x\big)$$

Then we regress the gap times W_1 and W_2 on the treatment status Z through a weighted Cox proportional hazard model with the stabilized weights:

$$h_j(w|x, z) = h_{0_j}(w)e^{\beta^{m_j} w}$$

where $j = 1, 2$. The results show that $\hat{\beta}^{m_1}$ = -0.362 ($p = 0.005$), which corresponds to a hazard ratio of 0.70 (95% CI: 0.54, 0.90) for subjects received treatment versus subjects

received placebo. This indicates strong positive treatment effect for the first gap time. The estimate of the log marginal hazard ratio for the second gap time $\hat{\beta}^{m_2}$ is 0.283 ($p = 0.2$). This suggests the treatment is not significant for the second gap time. Model checking can be carried out using the cox.zph function in R, which does not provide evidence of violation of the proportional hazard assumption for the first gap time ($p = 0.648$). However, there is evidence against the proportional hazard assumption for the second gap time ($p = 0.0366$).

3.11 Conclusion

In chapter 2, we showed that for observational data when the response is a time to event, the IPTW method consistently estimated the marginal hazard ratio when there was no unmeasured confouning [11]. We conducted a series Monte Carlo simulations to examine the performance of the IPTW estimator with both the conventional and stabilized weights. The results indicated that the conventional IPTW estimator resulted in lower bias than that of the stabilized IPTW estimator. Due to the within-subject correlation induced by weighting, the naive variance estimator failed to correctly estimate the variance of the IPTW estimator. Both the robust variance estimator and the bootstrap variance estimator accurately approximated the variance.

In chapter 3 we further considered settings where each subject experienced two events. For the time-fixed treatment and covariates scenarios, the IPTW estimator consistently estimated the overall marginal hazard ratio across two events. For the time-varying covariates and time-varying covariates and treatment scenarios, the IPTW estimator consistently estimated the marginal hazard ratio for each gap time. In the presence of censoring, we observed from the simulation results that the stabilized IPTW estimator generally resulted in unbiased estimation of the marginal hazard ratio. Whereas the conventional IPTW estimator resulted in biased estimate of the marginal hazard ratio. Moreover, the bias increases as the censoring proportion increases. We conjecture that the conventional IPTW

estimate converges to the log of another marginal hazard ratio for the first gap time, which is calculated given the administrative censoring time. For the second gap time it is not clear what the conventional IPTW estimate converges to. For the variance estimation, the robust variance estimator approximated the variance reasonably well with less than 10% bias across all scenarios. We note that we simulated the data from a conditional model with conditional hazard ratio, and we estimated the marginal hazard ratio using a marginal model. The misspecification of the model is what leads to the violation of the proportional hazards assumption. However, the estimate still converges and has meaningful interpretations [7].

To summarize, based on the simulation results, we recommend researchers use the conventional IPTW estimator to estimate the marginal hazard ratio with the robust variance estimator or the bootstrap variance estimator when the response is a single time to event, and use the stabilized IPTW estimator to estimate the marginal hazard ratio with the robust variance estimator for recurrent data. Further extensions can be made to settings with multiple events and the conventional weights and the stabilized weights can be formulated in the similar way to that described in Section 3.1 and 3.5 [17].

References

[1] Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

[2] Rosenbaum PR and Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79:516–24, 1984.

[3] Rubin DB. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29:185–203, 1973.

[4] Austin PC, Grootendorst P, Normand SL, and Anderson GM. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a monte carlo study. *Statistics in Medicine*, 26(4):754–768, 2007.

[5] Rosenbaum PR. *Propensity score*. John Wiley and Sons, Boston, 2005.

[6] Austin PC. Variance estimation when using inverse probability of treatment weighting (iptw) with survival analysis. *Statistics in Medicine*, 35(30):5642–5655, 2016.

[7] Lin DY and Wei LJ. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*, 84(408):1074–1078, 1989.

[8] Cook RJ and Lawless JF. *The Statistical Analysis of Recurrent Events*. Springer Science and Business Media, 2007.

[9] Lunceford JK and Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comprehensive study. *Statistics in Medicine*, 23:2937–2960, 2004.

[10] Cole SR and Hernan MA. Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine*, 75:45–49, 2004.

[11] Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, 32:2837–2849, 2013.

[12] Diggle PJ, Heagerty P, Liang KY, and Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press, 2002.

[13] Li P. *STAT 908: Statistical Inference, lecture notes*. Winter 2018.

[14] Godambe VP. *Estimating Functions*. The Clarendon Press Oxford University Press, 1991.

[15] Cook RJ. *STAT 935: Analysis of Survival Data, lecture notes*. Winter 2018.

[16] Bender R, Augustin T, and Blettner M. Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine*, 24:1713–1723, 2005.

[17] Robins JM, Hernan MA, and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.