ABSTRACT

BAI, XIAOFEI. Doubly-robust Estimators in Observational Studies with and without a Stratified Sub-sample. (Under the direction of Anastasios A. Tsiatis.)

In this dissertation, we study semiparametric estimators in observational survival analysis. Observational studies are frequently conducted to compare the effects of two treatments on survival. For such studies we must be concerned about confounding; that is, there are covariates that affect both the treatment assignment and the survival distribution. With confounding the usual treatment-specific Kaplan-Meier estimator might be a biased estimator of the underlying treatment-specific survival distribution and the ordinary log rank test might lead to biased result.

This dissertation has three parts. In Chapter 1, we use semiparametric theory to derive doubly robust estimators (and their asymptotic variance) of the treatment-specific survival distributions and the difference in the treatment-specific survival distributions in the case when the collected covariates are enough to adjust for all confounding in the main study. In Chapter 2, we develop treatment-specific survival estimator in the case when we further conduct a stratified sampling substudy to collect additional potential confounding covariates. In Chapter 3, we develop log rank type test statistics to test for the difference in overall treatment curves, in both case of main study and substudy. All developed methods are applied to a real data example along with different simulation scenarios.

Doubly-robust Estimators in Observational Studies with and without a Stratified Sub-sample

by
Xiaofei Bai

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2014

APPROVED BY:

Marie Davidian	Daowen Zhang
Wenbin Lu	Anastasios A. Tsiatis Chair of Advisory Committee

DEDICATION

To my family.

BIOGRAPHY

The author was born in Beijing, China. She entered Nankai University and obtained the Bachelor's degree in Statistics from the College of Mathematical Sciences in 2009. She began her graduate study in the statistics department at North Carolina State University and got the Master's degree in 2011 on the path of pursuing a Ph.D. degree. With the valuable instruction and guidance from her adviser Dr. Anastasios (Butch) Tsiatis, she focused working on developing semiparametric estimators in observational survival analysis. She will complete the Ph.D. study in May, 2014.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude and appreciation to my adviser, Dr. Anastasios (Butch) Tsiatis, who provided me insightful instruction and generous support throughout my research and my dissertation. He has always been encouraging and understanding and showed great patience and kindness to me. I learned a lot from Dr. Tsiatis, not only on the expertise knowledge in statistics but also on how to conduct research work. I feel very lucky to have him as my adviser. Without his help and guidance, I would not be able to accomplish this work.

I am greatly grateful to my committee members, Dr. Marie Davidian, Dr. Daowen Zhang and Dr. Wenbin Lu, for reviewing my dissertation and sharing their insightful comments and suggestions to my research. I would like extend my appreciation to all the faculty members of the Department of Statistics at North Carolina State University, who offered a variety of interesting and useful lectures. I also want to thank all the staff members, who make our department feel like a big warm family.

Last but not the least, I wish to thank my family for their unconditional support and love. They are the source of my strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Doubly-robust Estimators of Survival Distributions in Observational Studies

1.1 Introduction

In observational survival studies, treatments are often determined according to clinician/patient preference and thus may be subject to confounding; that is, patients receiving the two treatments may be prognostically different. Under these conditions, the Kaplan-Meier estimator (Kaplan and Meier, 1958) for a treatment-specific survival distribution will in general be an inconsistent estimator for the true treatment-specific survival distribution (that is, the survival distribution for the underlying population of patients with disease had everyone in the population received a specific treatment.) This is formally defined with the use of potential outcomes in Section 1.2. Estimators that take account of confounding have been proposed by, e.g., Makuch (1982) and Lee et al. (1992) using a form of regression-based adjustment and Cole and Hernán (2004) using propensity score adjustment. The former methods require that the analyst posit a model for the survival distribution as a function of potential confounding variables while the latter depends on a model for the propensity score, the probability of receiving a treatment given covariates (confounders). Although these estimators lead to valid inference if these models are correctly specified under the assumption that all potential confounding variables are recorded in the database (no unmeasured confounders), if the models do not capture the true relationships between outcome or treatment assignment and covariates, respectively, the estimators may be inconsistent.

To address this limitation and to achieve improved efficiency, we propose to use the theory of Robins and Rotnitzky (1992) and Hubbard et al. (2000) for deriving locally efficient augmented

inverse probability weighted complete case (AIPWCC) estimators for treatment-specific distributions in the observational data context. Under the no unmeasured confounders assumption, such estimators will not only yield greater efficiency, we will show they are doubly robust in that they will be consistent for the true survival distribution if either of the required postulated model for the survival distribution as a function of covariates or those for the propensity score and censoring distribution are correctly specified, even if the other is not. Thus, the proposed estimators will also offer protection against model misspecification to some extent.

1.2 Notations and Assumptions

We are interested in estimating treatment-specific survival distributions and their difference in an observational study. We assume here that there are two treatments under consideration which we label 1 and 0. We define potential outcomes for the N individuals in our study as $(X_i, T_i^{(0)}, T_i^{(1)}, C_i^{(0)}, C_i^{(1)}), i = 1, \ldots, N$, assumed identically and independently distributed, where for individual i in our sample X_i denotes a vector of baseline covariates, $T_i^{(j)}$ denotes the potential survival time and $C_i^{(j)}$ denotes the potential censoring time if individual i were given treatment (possibly contrary to fact) j, for $j = 0, 1$. The treatment-specific survival distribution is then defined as $S^{(j)}(u) = P(T^{(j)} \geq u)$ for $j = 0, 1$ and will be the main focus of inference in this chapter.

In some well-designed trials, all patients will be followed from the date they entered the study until the time of data analysis and their survival status will be fully ascertained during this period. Consequently, the potential censoring time was the time from entry into study until time of analysis which would be the same under both treatments. Therefore it would be reasonable to assume $C_i^{(1)} = C_i^{(0)} = C_i$ for $i = 1, \ldots, N$.

We also denote by Z_i the binary treatment assignment for patient i, where $Z_i = 1, 0$ and make the strong ignorability assumption (Rubin, 1978), or no unmeasured confounders assumption, that Z is conditionally independent of $T^{(j)}$ given X, denoted by $Z \perp\!\!\!\perp T^{(j)}|X$. In a randomized study, we can reasonably assume that treatment assignment Z is independent of baseline covariates X as well as potential outcome $T^{(j)}, j = 0, 1$. However, this may no longer be a reasonable assumption in an observational study. If, however, we collected sufficient covariates X that are related to survival and that we believe were used in the treatment decision, then the assumption of no unmeasured confounders may be reasonable.

We also make the usual assumption of non-informative censoring; namely, that $C \perp\!\!\!\perp T^{(j)}|X, Z$. Together these assumptions imply that

$$(Z, C) \perp\!\!\!\perp T^{(j)}|X. \tag{1.1}$$

We use this assumption in the remainder of this chapter and refer to it as assumption (1.1) .

In contrast to the potential outcomes, some of which may not be observed, the data that are observable can be summarized as $(Z_i, X_i, U_i, \Delta_i), i = 1, \ldots, N$, where, in addition to (Z_i, X_i), already defined, we also observe for individual i their time to death or censoring $U_i = \min(T_i, C_i)$ and the failure indicator $\Delta_i = I(T_i \leq C_i)$ where $T_i = Z_i T_i^{(1)} + (1 - Z_i)T_i^{(0)}$; i.e., the time to death on the assigned treatment.

1.3 Inference for Treatment-specific Survival Distributions

In this section, we mainly focus on estimating the underlying survival distribution on treatment 1 while obvious analogs could be used for treatment 0. Thus, the primary goal is to estimate the function $S^{(1)}(u) = P(T^{(1)} \geq u), u \geq 0$ using the observed data $(Z_i, X_i, U_i, \Delta_i), i = 1, \ldots, N$. Of course, we only get to observe $T^{(1)}$ in the data when $Z = 1$ and $\Delta = 1$ $(C > T)$; otherwise $T^{(1)}$ is missing or coarsened. As indicated by Tsiatis (2006), section 8.6, this can be viewed as monotone coarsening, where we only get to observe the data on treatment 1 when $Z = 1$ and only data that were observed through censoring time C and the variable of interest $T^{(1)}$ is observed (complete-case) when $Z = 1, C \geq T^{(1)}$. Because $(Z, C) \perp\!\!\!\perp T^{(1)}|X$, this implies that we have monotone coarsening at random and thus the probability of not being coarsened; i.e., observing a complete case, is given by $\pi(X)K_c^{(1)}(T^{(1)}, X)$, where $\pi(X) = P(Z = 1|X)$ is the so called propensity score, and $K_c^{(1)}(r, X) = P(C \geq r|X, Z = 1)$ is the conditional survival function of the treatment specific censoring time given X.

If the model for the coarsening, defined through the propensity score and the treatment specific censoring distribution given the covariates are correctly specified, then without making any additional assumption on the survival distribution, as shown by Robins and Rotnitzky (1992) and Tsiatis (2006), that all regular asymptotic linear estimator of the survival distribution can be written as AIPWCC estimators with corresponding estimating functions:

$$\frac{Z\Delta}{\pi(X)K_c^{(1)}(T^{(1)}, X)}\left\{I(T^{(1)} \geq u) - S^{(1)}(u)\right\}$$
$$-\left\{\frac{Z - \pi(X)}{\pi(X)}\right\}h_1(u, X) + \int_0^\infty \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c^{(1)}(r, X)}{K_c^{(1)}(r, X)}h_2(u, r, X), \tag{1.2}$$

where $\mathrm{d}M_c^{(1)}(r, X) = \mathrm{d}N_c(r) - \lambda_c^{(1)}(r, X)Y(r)$ is the martingale increment for the censoring distribution, $N_c(r) = I(U \leq r, \Delta = 0)$, $Y(r) = I(U \geq r)$ and $\lambda_c^{(1)}(r, X) = \frac{-\mathrm{d}\log K_c^{(1)}(r,X)}{\mathrm{d}r}$ is the conditional hazard function for C given $Z = 1$ and X, $h_1(u, X)$ and $h_2(u, r, X)$ are arbitrary functions. We note that $\{I(T^{(1)} \geq u) - S^{(1)}(u)\}$ denotes the full-data estimating function that would be used if $T^{(1)}$ were observed on all subjects. The first term in (1.2) is the inverse

probability weighted complete case term, and the latter two terms are augmentation terms.

For monotonely coarsened data, we can use the results from Theorem 10.4 of Tsiatis (2006) to show that the optimal choice for $h_1(u, X)$ and $h_2(u, r, X)$, i.e., the estimator with the smallest asymptotic variance is given by

$$h_1(X) = E\left\{I(T^{(1)} \geq u) - S^{(1)}(u)|X\right\},$$

and

$$h_2(r, X) = E\left\{I(T^{(1)} \geq u) - S^{(1)}(u)|T^{(1)} \geq r, X\right\},$$

respectively.

Let us define $P(T^{(1)} \geq r|X) = H^{(1)}(r, X)$. We also note that $P(T \geq r|Z = 1, X) = P(T^{(1)} \geq r|Z = 1, X) = P(T^{(1)} \geq r|X)$, the last equality following from assumption (1.1). Consequently, the optimal estimating function is given as

$$\frac{Z\Delta}{\pi(X)K_c^{(1)}(U, X)}\left\{I(U \geq u) - S^{(1)}(u)\right\} - \left\{\frac{Z - \pi(X)}{\pi(X)}\right\}\left\{H^{(1)}(u, X) - S^{(1)}(u)\right\}$$
$$+ \int_0^\infty \frac{Z}{\pi(X)} \frac{\mathrm{d}M_c^{(1)}(r, X)}{K_c^{(1)}(r, X)}\left[I(r < u)\left\{\frac{H^{(1)}(u, X)}{H^{(1)}(r, X)} - S^{(1)}(u)\right\} + I(r \geq u)\left\{1 - S^{(1)}(u)\right\}\right].$$

As shown in Appendix A, the above estimating function can be further simplified as

$$\frac{ZI(U \geq u)}{\pi(X)K_c^{(1)}(u, X)} - \left\{\frac{Z - \pi(X)}{\pi(X)}\right\}H^{(1)}(u, X) + \int_0^u \frac{Z}{\pi(X)} \frac{\mathrm{d}M_c^{(1)}(r, X)}{K_c^{(1)}(r, X)}\left\{\frac{H^{(1)}(u, X)}{H^{(1)}(r, X)}\right\} - S^{(1)}(u). \tag{1.3}$$

If we knew the propensity score $\pi(X)$, the conditional censoring distribution $K_c^{(1)}(r, X)$, and the conditional survival distribution $H^{(1)}(r, X)$, then this estimating function could be used as a basis to obtain the optimal AIPWCC estimator for $S^{(1)}(u)$; namely,

$$\hat{S}^{(1)}(u) = N^{-1}\sum_{i=1}^{N}\left[\frac{Z_i I(U_i \geq u)}{\pi(X_i)K_c^{(1)}(u, X_i)}\right.$$
$$\left. - \left\{\frac{Z_i - \pi(X_i)}{\pi(X_i)}\right\}H^{(1)}(u, X_i) + \int_0^u \frac{Z_i}{\pi(X_i)} \frac{\mathrm{d}M_{c,i}^{(1)}(r, X_i)}{K_c^{(1)}(r, X_i)} \frac{H^{(1)}(u, X_i)}{H^{(1)}(r, X_i)}\right], \tag{1.4}$$

where

$$
\begin{aligned}
& \int_0^u \frac{Z_i}{\pi(X_i)} \frac{\mathrm{d}M_{c,i}^{(1)}(r, X_i)}{K_c^{(1)}(r, X_i)} \frac{H^{(1)}(u, X_i)}{H^{(1)}(r, X_i)} \\
= \; & H^{(1)}(u, X_i) \left[\int_0^u \frac{Z_i}{\pi(X_i)} \frac{\mathrm{d}N_{c,i}(r, X_i)}{K_c^{(1)}(r, X_i) H^{(1)}(r, X_i)} - \int_0^u \frac{Z_i}{\pi(X_i)} \frac{\lambda_c^{(1)}(r, X_i) I(U_i \geq r)}{K_c^{(1)}(r, X_i) H^{(1)}(r, X_i)} \right] \\
= \; & \frac{Z_i H^{(1)}(u, X_i)}{\pi(X_i)} \left[\frac{I(U_i \leq u)(1 - \Delta_i)}{K_c^{(1)}(U_i, X_i) H^{(1)}(U_i, X_i)} - \int_0^{\min(u, U_i)} \frac{\lambda_c^{(1)}(r, X_i)}{K_c^{(1)}(r, X_i) H^{(1)}(r, X_i)} \right],
\end{aligned}
$$

and the variance of the estimator for $\hat{S}^{(1)}(u)$ can be estimated by using the standard sandwich variance estimator; namely,

$$
\begin{aligned}
N^{-1} \sum_{i=1}^{N} \Bigg[& \frac{Z_i I(U_i \geq u)}{\pi(X_i) K_c^{(1)}(u, X_i)} - \left\{ \frac{Z_i - \pi(X_i)}{\pi(X_i)} \right\} H^{(1)}(u, X_i) \\
& + \int_0^u \frac{Z_i}{\pi(X_i)} \frac{\mathrm{d}M_{c,i}^{(1)}(r, X_i)}{K_c^{(1)}(r, X_i)} \left\{ \frac{H^{(1)}(u, X_i)}{H^{(1)}(r, X_i)} \right\} - \hat{S}^{(1)}(u) \Bigg]^2 . \qquad (1.5)
\end{aligned}
$$

The estimator given above is similar to the estimator proposed by Hubbard et al. (2000), using the same principle of developing the locally efficient estimator, however, we demonstrate in Appendix B that this estimator is doubly-robust; that is, if either the estimator for the coarsening defined through the estimator for $P(Z = 1|X) = \pi(X)$ and the estimator for $P(C \geq r|Z = 1, X) = K_c^{(1)}(r, X)$ are both consistent or the estimator for $P(T^{(1)} \geq r|X) = H^{(1)}(r, X)$ is consistent, the estimator (1.4) will be consistent. Moreover, we use this estimator as a springboard for the AIPWCC estimator in the case of stratified sampling in Chapter 2.

Of course, we do not know $\pi(X)$, $K_c^{(1)}(r, X)$, or $H^{(1)}(r, X)$; hence, each of these must be estimated from the data. A logistic regression model is commonly used to estimate the propensity score $P(Z = 1|X)$ with data $\{(Z_i, X_i), i = 1, \ldots, n\}$. For $K_c^{(1)}(r, X)$ a treatment-specific proportional hazards model $\lambda_c^{(1)}(r, X) = \lambda_c^{(1)}(r) \exp\{\beta_c^T X\}$ is often used with data $\{(U_i, (1 - \Delta_i), X_i), i : Z_i = 1\}$. For $H^{(1)}(r, X)$ a treatment-specific proportional hazards model $\lambda_t^{(1)}(r, X) = \lambda_t^{(1)}(r) \exp\{\beta_t^T X\}$ with data $\{(U_i, \Delta_i, X_i), i : Z_i = 1\}$ is often used.

We must however be careful with regards to the variance estimator (1.5) because there may be an effect on the variance due to estimating the functions $\pi(X)$, $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$ and their possible misspecification. What we know from the theory of AIPWCC estimators is that if all these functions were consistently estimated, then the asymptotic variance of the corresponding estimator would be unaffected in which case we could estimate the variance of $\hat{S}^{(1)}(u)$ consistently by substituting the estimators $\hat{\pi}(X)$, $\hat{K}_c^{(1)}(r, X)$ and $\hat{H}^{(1)}(r, X)$ into (1.5). If, however, the propensity score $\pi(X)$ and the conditional censoring distribution $K_c^{(1)}(r, X)$ were

consistently estimated whereas, the conditional survival distribution $H^{(1)}(r, X)$ was not, then the variance estimator would be conservative. In practice it has been shown that a reasonable attempt for estimating $H^{(1)}(r, X)$ will often result in negligible conservativeness even with misspecification for this model. This will be investigated further in our simulation studies in Section 1.5. When the estimator for $H^{(1)}(r, X)$ is consistent and either of the estimators for $\pi(X)$ or $K_c^{(1)}(r, X)$ are not, then even though the resulting estimator for $S^{(1)}(u)$ is consistent (part of the double robustness property), the resulting estimator for the variance may be biased and there is no theory that we know of that suggests the direction or extent of the bias. In such cases, we may want to use a bootstrap estimator of the asymptotic variance. This too will be investigated in the simulation studies of Section 1.5.

1.4 Inference on Difference of Treatments

As well as estimating the individual treatment specific survival distributions, we are also interested in testing the null hypothesis on whether the survival distributions of the two treatments are the same. That is, estimating the difference in the treatment-specific survival distribution $S_{diff}(u) = S^{(1)}(u) - S^{(0)}(u) = P(T^{(1)} \geq u) - P(T^{(0)} \geq u), u \geq 0$. In Section 1.3, we developed the estimating function for $S^{(1)}(u)$ and by analogy we can easily obtain the the estimating function for $S^{(0)}(u)$. It is then straightforward that the estimating function for $S_{diff}(u)$ is just the difference between the estimating function of $S^{(1)}(u)$ and of $S^{(0)}(u)$. For ease of notation we denote for patient i and treatment $j = 0, 1$

$$\begin{aligned}\phi_j^*(V_i) &= \left[\frac{(Z_i)^j(1-Z_i)^{1-j}I(U_i \geq u)}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j}K_c^{(j)}(u, X_i)} - \frac{(2j-1)\{Z_i - \pi(X_i)\}}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j}}H^{(j)}(u, X_i)\right. \\ &\quad \left. + \int_0^u \frac{(Z_i)^j(1-Z_i)^{1-j}}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j}} \frac{\mathrm{d}M_c^{(j)}(r, X_i)}{K_c^{(j)}(r, X_i)} \frac{H^{(j)}(u, X_i)}{H^{(j)}(r, X_i)}\right], \qquad (1.6)\end{aligned}$$

where $V_i = (Z_i, X_i, U_i, \Delta_i)$ are the data observed for the i-th individual, $K_c^{(j)}(r, X) = P(C \geq r|X, Z = j)$, and $H^{(j)}(r, X) = P(T \geq r|X, Z = j)$. We also denote by $\hat{\phi}_j^*(V_i)$ the same as in (1.6) after substituting the estimators for $\pi(X_i)$, $K_c^{(j)}(r, X_i)$, and $H^{(j)}(r, X_i)$. Hence, the doubly robust estimator for $S_{diff}(u)$ is given by

$$\hat{S}_{diff}(u) = N^{-1}\sum_{i=1}^{N}\left\{\hat{\phi}_1^*(V_i) - \hat{\phi}_0^*(V_i)\right\},$$

and for the variance estimator of $\hat{S}_{diff}(u)$ we propose using the sandwich variance

$$\widehat{\text{Var}}\{\hat{S}_{diff}(u)\} = N^{-1}\sum_{i=1}^{N}\left\{\hat{\phi}_1^*(V_i) - \hat{\phi}_0^*(V_i) - \hat{S}_{diff}(u)\right\}^2.$$

The same comments regarding the appropriateness of this estimator when some of the models are misspecified apply as in the estimation of $S^{(1)}(u)$ discussed previously in Section 1.3.

1.5 Simulations

Several simulation studies have been carried out to demonstrate the performance of the proposed estimators in Section1.3 and Section 1.4. Results from the main simulation study are shown here and results from other scenarios can be found in the Appendices. The primary goal of this simulation study is to demonstrate the double robustness property of the proposed estimator for $S^{(1)}(u)$ using the observed data $V_i = (Z_i, X_i, U_i, \Delta_i)$ assuming that all potential confounders were available.

We generated 500 replications and within each replicate, $N = 2,000$ observations are generated as follows: $B \sim Bernoulli(0.5), W \sim N(0,1)$ and $X_2 \sim N(0,1)$, mutually independent. Here B is a strata variable with two levels, mimicking the real data example (ASCERT data) where patients have either two- or three-vessel disease. More details about the ASCERT data can be found in Section 2.6. We denote by $X_1 = (B, W, W^2)^T$, and $X = (X_1^T, X_2)^T$. The treatment assignment propensity is generated by $P(Z = 1|X) = \frac{\exp(0.1B+0.1W+0.5W^2+0.5X_2)}{1+\exp(0.1B+0.1W+0.5W^2+0.5X_2)}$, and the treatment-specific hazard functions are given by $\lambda_{Z=1}(t|X) = \exp(-1 + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$ and $\lambda_{Z=0}(t|X) = \exp(-0.5 + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$, respectively. We also generated an independent censoring variable C as Uniform$(0, 4)$.

We considered four scenarios: the first one uses all the covariates X to fit all three models $\pi(X)$, $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$. This scenario is the "correct-correct" model, where $\pi(X)$, $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$ are estimated consistently. In the second scenario, we leave out the quadratic term W^2 and use only $(B, W, X_2)^T$ to fit $\pi(X), K_c^{(1)}(r, X)$ and use X to fit $H^{(1)}(r, X)$. We refer to this as the "wrong propensity, correct regression" scenario. The third one, which is referred as the "correct propensity, wrong regression" scenario, is where we use X to fit $\pi(X), K_c^{(1)}(r, X)$ and $(B, W, X_2)^T$ to fit $H^{(1)}(r, X)$. Finally, in the last scenario, we use only $(B, W, X_2)^T$ to fit $\pi(X)$, $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$ and this is the "wrong propensity, wrong regression" scenario. Note that the last scenario would also correspond to the situation where the assumption of no unmeasured confounders would not hold in the main study and reflects the possible bias that may result. Because the distribution of the censoring variable is independent of

the covariates, the misspecification of the coarsening model is only through $\pi(X)$. In Appendix C, we explore the scenario where misspecification of the coarsening model occurs because of misspecifying $K_c^{(1)}(r, X)$ as well.

Table 1.1: Simulation 1, AIPWCC estimator for $\hat{S}^{(1)}(u)$ at selected time points when there is no stratified sampling. Sample size N=2000. Scenario 1 corresponds to "correct-correct" case; scenario 2 corresponds to"wrong propensity, correct regression" case; scenario 3 corresponds to"correct propensity, wrong regression" case; scenario 4 corresponds to"wrong propensity, wrong regression" case.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate
Scenario 1	0.5	0.693	0.694	0.001	0.013	0.013	0.96
	1.0	0.523	0.524	0.001	0.015	0.015	0.95
	1.5	0.408	0.408	0.000	0.016	0.016	0.94
	2.0	0.324	0.325	0.001	0.017	0.016	0.94
	2.5	0.262	0.262	0.000	0.017	0.017	0.95
	3.0	0.214	0.214	0.000	0.017	0.018	0.96
Scenario 2	0.5	0.693	0.694	0.001	0.013	0.013	0.96
	1.0	0.523	0.524	0.001	0.015	0.015	0.93
	1.5	0.408	0.408	0.000	0.016	0.015	0.94
	2.0	0.324	0.325	0.001	0.016	0.015	0.93
	2.5	0.262	0.261	-0.001	0.016	0.015	0.93
	3.0	0.214	0.214	0.000	0.016	0.016	0.94
Scenario 3	0.5	0.693	0.694	0.001	0.013	0.013	0.96
	1.0	0.523	0.524	0.001	0.015	0.016	0.95
	1.5	0.408	0.408	0.000	0.016	0.016	0.95
	2.0	0.324	0.325	0.001	0.017	0.017	0.95
	2.5	0.262	0.261	-0.001	0.017	0.017	0.95
	3.0	0.214	0.214	0.000	0.017	0.018	0.96
Scenario 4	0.5	0.693	0.661	-0.032	0.014	0.014	0.35
	1.0	0.523	0.489	-0.035	0.015	0.015	0.36
	1.5	0.408	0.374	-0.034	0.016	0.016	0.41
	2.0	0.324	0.294	-0.031	0.016	0.016	0.49
	2.5	0.262	0.234	-0.028	0.016	0.016	0.57
	3.0	0.214	0.190	-0.025	0.016	0.017	0.67

In Table 1.1, we present results for estimating the treatment-specific survival distribution for treatment 1, $S^{(1)}(u)$. Results regarding the difference between two treatments $S_{diff}(u)$ are similar and shown in Table 1.2. We can see in scenarios 1, 2, and 3 that the estimator defined

in (1.4) is indeed consistent as desired. For scenarios 2 and 3, even though parts of the models (propensity models/regression model) are not correctly specified, the resulting estimator is still consistent. This illustrates the doubly-robustness property. In scenario 4, the estimator is biased as is expected because the models $\pi(X)$ and $H^{(1)}(u, X)$ are misspecified. The coverage rate is calculated using the 95% confidence interval derived as the estimator plus/minus 1.96 estimated standard errors using the sandwich variance formulas for the four scenarios. Because in scenario 4, the estimator $\hat{S}^{(1)}(u)$ is not consistent, the coverage rate is far below 0.95. In theory, the estimated variance in scenario 3 may be conservative however, as expected, the coverage rate was close to the nominal rate of 0.95.

Note that the coverage rate in scenario 2 did not achieve the 0.95 nominal rate. This indeed is the scenario discussed at the end of Section 1.3 where the sandwich variance estimator is not expected to be consistent. For this scenario we also estimated the asymptotic variance and the coverage rate of the confidence interval using a nonparametric bootstrap sampling. We used 100 bootstrap replicates and compared the bootstrap standard error with the Monte-Carlo standard error. As shown in Table 1.3, the bootstrap standard error provides an accurate estimator of the standard error and the corresponding confidence interval computed as the estimator plus/minus 1.96 bootstrap standard error attains the nominal coverage. This suggests that, despite of computation intensity, bootstrap variance estimator is close to the Monte-Carlo variance and may be used in the cases when coarsening probabilities could not be estimated consistently.

For the interest of smaller sample size performance, a simulation study with sample size $N = 200$ is conducted. The data generating mechanism and parameter values remain the same. The AIPWCC estimator is computed in all four scenarios as before. Table 1.4 gives the summary results of biasedness and standard error estimates at several time points, suggesting that the treatment-specific survival distribution $S^{(1)}(u)$ can still be estimated consistently even with smaller sample size. The proposed standard error estimator, on the other hand, tends to be unstable especially at the tail of the curve. The bootstrap standard error estimator performs better and hence is preferable for such a smaller sample size.

Other competing estimators were also considered: the inverse probability weighted estimator, the outcome regression estimator and the treatment-specific Kaplan-Meier estimator. The numerical results can be found in Appendix D. When correctly specified, the first two estimators are consistent. The proposed AIPWCC estimator is more efficient than the inverse probability weighted estimator and slightly less efficient than the outcome regression estimator. On the other hand, both of these competing estimators are biased if the models are misspecified. The unadjusted treatment-specific Kaplan-Meier estimator is always biased as it does not account for confounding.

1.6 Conclusion and Discussion

Viewing censored survival data from an observational study as monotonely coarsened versions of treatment-specific potential survival times, we were able to use the powerful semiparametric theory for missing data to derive doubly robust estimators for the underlying treatment-specific survival distributions as well as the difference in these treatment-specific distributions. We also derived an estimator for the asymptotic variance of these estimators. Using simulation studies we demonstrated that the large sample properties of these estimators and corresponding confidence intervals worked well in most cases where we expected consistent estimators.

Table 1.2: Simulation 1, AIPWCC estimator for $S_{diff}(u)$ at selected time points when there is no stratified sampling. Sample size N=2000. Scenario 1 corresponds to "correct-correct" case; scenario 2 corresponds to"wrong propensity, correct regression" case; scenario 3 corresponds to"correct propensity, wrong regression" case; scenario 4 corresponds to"wrong propensity, wrong regression" case.

	u	$S_{diff}(u)$	$\hat{S}_{diff}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate
Scenario 1	0.5	0.118	0.119	0.001	0.022	0.022	0.94
	1.0	0.143	0.144	0.001	0.021	0.022	0.96
	1.5	0.143	0.142	0.000	0.023	0.022	0.94
	2.0	0.133	0.133	0.000	0.022	0.021	0.95
	2.5	0.121	0.120	-0.001	0.022	0.021	0.93
	3.0	0.108	0.107	-0.001	0.021	0.028	0.96
Scenario 2	0.5	0.118	0.119	0.001	0.022	0.022	0.94
	1.0	0.143	0.144	0.001	0.021	0.023	0.97
	1.5	0.143	0.142	-0.001	0.022	0.022	0.96
	2.0	0.133	0.134	0.000	0.022	0.022	0.96
	2.5	0.121	0.119	-0.001	0.022	0.022	0.94
	3.0	0.108	0.107	-0.001	0.021	0.022	0.97
Scenario 3	0.5	0.118	0.118	0.000	0.027	0.029	0.96
	1.0	0.143	0.143	0.000	0.023	0.026	0.98
	1.5	0.143	0.142	-0.001	0.024	0.025	0.96
	2.0	0.133	0.133	-0.001	0.022	0.023	0.96
	2.5	0.121	0.119	-0.001	0.022	0.023	0.95
	3.0	0.108	0.107	-0.001	0.021	0.023	0.97
Scenario 4	0.5	0.118	0.034	-0.084	0.023	0.023	0.04
	1.0	0.143	0.063	-0.080	0.023	0.024	0.09
	1.5	0.143	0.073	-0.070	0.024	0.024	0.19
	2.0	0.133	0.075	-0.058	0.023	0.023	0.28
	2.5	0.121	0.071	-0.050	0.023	0.022	0.38
	3.0	0.108	0.066	-0.042	0.022	0.023	0.54

Table 1.3: Simulation 1, AIPWCC estimator for $S^{(1)}(u)$ at selected time points when there is no stratified sampling. Sample size N=2000. Bootstrap inference for scenario 2, that is, the "wrong propensity, correct regression" case where variance estimator (1.5) is no longer valid. Bootstrap replicate=100.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}(u)$	Bias	Monte-Carlo SE	Bootstrap SE	Coverage Rate
Scenario 2	0.5	0.693	0.694	0.001	0.013	0.013	0.96
	1.0	0.523	0.524	0.001	0.015	0.015	0.94
	1.5	0.408	0.408	0.000	0.016	0.016	0.95
	2.0	0.324	0.325	0.000	0.016	0.016	0.95
	2.5	0.262	0.261	-0.001	0.016	0.016	0.94
	3.0	0.214	0.214	0.000	0.016	0.017	0.95

Table 1.4: Simulation 1, AIPWCC estimator for $\hat{S}^{(1)}(u)$ at selected time points when there is no stratified sampling. Sample size N=200. Scenario 1 corresponds to "correct-correct" case; scenario 2 corresponds to"wrong propensity, correct regression" case; scenario 3 corresponds to"correct propensity, wrong regression" case; scenario 4 corresponds to"wrong propensity, wrong regression" case. Bootstrap replicate=500.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}(u)$	Bias	Monte-Carlo SE	Estimated SE	Bootstrap SE	CR (Proposed)	CR (Bootstrap)
Scenario 1	0.5	0.693	0.696	0.003	0.039	0.041	0.042	0.95	0.96
	1.0	0.523	0.524	0.001	0.049	0.048	0.049	0.94	0.94
	1.5	0.408	0.410	0.002	0.050	0.051	0.052	0.95	0.95
	2.0	0.324	0.326	0.001	0.051	0.052	0.054	0.95	0.95
	2.5	0.262	0.259	-0.003	0.054	0.053	0.056	0.92	0.93
	3.0	0.214	0.203	-0.012	0.061	0.054	0.067	0.89	0.95
Scenario 2	0.5	0.693	0.697	0.003	0.039	0.042	0.041	0.96	0.96
	1.0	0.523	0.524	0.001	0.048	0.046	0.048	0.93	0.94
	1.5	0.408	0.411	0.003	0.048	0.048	0.051	0.94	0.95
	2.0	0.324	0.326	0.002	0.050	0.048	0.052	0.94	0.95
	2.5	0.262	0.260	-0.002	0.052	0.048	0.054	0.91	0.93
	3.0	0.214	0.205	-0.009	0.059	0.049	0.060	0.86	0.94
Scenario 3	0.5	0.693	0.696	0.003	0.039	0.042	0.042	0.96	0.95
	1.0	0.523	0.524	0.000	0.049	0.049	0.049	0.95	0.94
	1.5	0.408	0.410	0.002	0.050	0.052	0.052	0.96	0.95
	2.0	0.324	0.326	0.001	0.051	0.053	0.054	0.95	0.96
	2.5	0.262	0.259	-0.003	0.054	0.054	0.056	0.93	0.94
	3.0	0.214	0.203	-0.011	0.062	0.055	0.067	0.89	0.94
Scenario 4	0.5	0.693	0.664	-0.030	0.042	0.044	0.044	0.94	0.94
	1.0	0.523	0.488	-0.035	0.050	0.049	0.050	0.89	0.89
	1.5	0.408	0.376	-0.032	0.050	0.050	0.051	0.90	0.90
	2.0	0.324	0.295	-0.030	0.050	0.050	0.052	0.89	0.90
	2.5	0.262	0.232	-0.030	0.051	0.050	0.053	0.88	0.91
	3.0	0.214	0.180	-0.034	0.058	0.051	0.059	0.84	0.90

Chapter 2

Doubly-robust Estimators of Survival Distributions in Stratify Sampled Observational Studies with Additional Information from a Subsample

2.1 Introduction

In Chapter 1, we derive doubly robust estimators (and their asymptotic variance) of the treatment-specific survival distributions and the difference in the treatment-specific survival distributions in the case when the collected covariates are enough to adjust for all confounding in the main study by using the semiparametric theory for missing data. However, it is important to keep in mind that the consistency of previously derived estimator depends on the crucial no unmeasured confounder assumption that all potential confounders are captured, which in reality may not true.

This is indeed the case in a data example known as the ASCERT study, an observational study funded by the National Heart Lung and Blood Institute to compare two widely used treatment options for patients with coronary artery disease (Weintraub et al., 2012), where the investigators had concerns about the possibility of residual confoundings. The ASCERT study was a retrospective analysis patients who had either two-vessel or three-vessel coronary artery disease and were treated either by surgical revascularization (coronary artery bypass grafting; CABG) or catheter-based revascularization (percutaneous coronary intervention; PCI). Not all

patients were followed until the endpoint of interest (all-cause mortality); accordingly, survival outcomes for these patients were censored. Moreover the investigators recognized that assessment of coronary anatomy was limited to a few relatively crude variables. For example, patients with certain high-risk coronary features may have been preferentially selected for CABG over PCI. To address this limitation, a sub-study was conducted in which detailed data on coronary anatomy were collected on a stratified random sample of patients at 54 hospitals in the main ASCERT study. According to the protocol, approximately 500 patient records were to be randomly selected without replacement from each of four strata formed by all combinations of the two treatments and whether or not the patients had two- or three-vessel disease. For each randomly selected patient, X-ray films of coronary angiograms taken prior to the patient's revascularization procedure were retrieved from storage and sent to a central laboratory for expert interpretation and analysis. More details of this study are given in Section 2.6.

The challenge is that of acknowledging this design in the comparison of treatment-specific survival distributions. There is a vast literature on methods for incorporating data from a subsample into such analyses (e.g., Breslow and Cain, 1988; Flanders and Greenland, 1991; Zhao and Lipsitz, 1992; Mark and Katki, 2006; Stürmer et al., 2007, and references therein). However, to our knowledge, there is no work applicable to a setting like that of ASCERT to develop estimators for survival distributions that take full advantage of the data collected on the entire cohort to achieve the greatest efficiency possible and that possess the double robustness property to protect against model misspecification. Currently, theory for AIPWCC estimators is based on the assumption that observations across subjects are independent and identically distributed (iid), which is clearly not the case for data collected according to a stratified sampling scheme.

In this chapter, we develop estimators for survival distribution in the case when a stratified sampling substudy is further conducted in order to collect additional potential confounding covariates.

2.2 Notations and Assumptions

Let $V_i = (U_i, \Delta_i, Z_i, X_{1i}), i = 1, \ldots, N$ denote the data available on all subjects in the main study, where now X_1 represents covariates collected on these subjects. The key requirement for using the previously proposed method is the strong ignorability assumption, which here would take the form $Z \perp\!\!\!\perp T^{(j)} | X_1$. If, however, there is concern that this assumption is untenable based only on X_1, then a subset of patients may be sampled from the main study to collect additional covariates X_2. As in the ASCERT study the subsample may be obtained using a stratified sampling design in which K strata are identified based on (Z, X_1), and letting ϵ denote the stratum indicator taking values $1, \ldots, K$ such that $N_k = \sum_{i=1}^{N} I(\epsilon_i = k)$ is the number of

subjects in stratum $k, k = 1, \ldots, K$, then from stratum k, a fixed (by design) number of subjects n_k are sampled from the N_k subjects at random without replacement. For each subject so included in the substudy, additional covariates X_2 are collected, where, letting $X = (X_1^T, X_2^T)^T$, it is believed that the strong ignorability assumption $Z \perp\!\!\!\perp T^{(j)}|X$ holds. Let $R_i = 1$ if subject $i = 1, \ldots, N$, was selected into the subsample and 0 otherwise. The goal is to estimate $S^{(1)}(u)$ (and $S_{diff}(u)$) using the observed data $O_i = (U_i, \Delta_i, Z_i, X_{1i}, R_i, R_i X_{2i})$

2.3 Inference for Treatment-specific Survival Distributions

Again, here we use the missing data analogy. We identify the "full data" that would ideally have been collected on all N subjects as $F_i = (U_i, \Delta_i, Z_i, X_{1i}, X_{2i}), i = 1, \ldots, N$, recognizing that X_2 is collected only on subjects in the substudy. Were there full data available, a consistent, asymptotically normal, doubly robust estimator for $S^{(1)}(u)$ could be found by solving

$$\sum_{i=1}^{N} \phi\{F_i, S^{(1)}(u)\} = 0$$

, where $\phi\{F, S^{(1)}(u)\}$ is the estimating function (1.3). We may identify the observed data as full data with X_2 missing for subjects not selected in the subsample. Accordingly, following Tsiatis (2006), we propose the class of AIPWCC estimators defined as the solution to

$$\sum_{i=1}^{N} \left[\frac{R_i \phi\{F_i, S^{(1)}(u)\}}{\eta(\epsilon_i)} - \left\{ \frac{R_i - \eta(\epsilon_i)}{\eta(\epsilon_i)} \right\} h(V_i) \right] = 0, \tag{2.1}$$

where $\eta(\epsilon_i) = P(R_i = 1|\epsilon_i)$, taking values n_k/N_k when $\epsilon_i = k$, the probability of selection into the subsample as a function of the stratum to which subject i belongs; and $h(V_i)$ is an arbitrary function of the data collected on all subjects. Here, if we set $h(V_i) = 0$, then the resulting estimator is a simple inverse probability weighted estimator, which we refer to as the weighted estimator. However, the choice of $h(V)$ will affect the asymptotic variance of the resulting estimator.

The usual theory for AIPWCC estimators applies when the observed data $O_i, i = 1, \ldots, N$, are iid. We briefly outline our proposed modification to account for the stratified sampling. Rearranging (2.1), we write this estimating equation as $T\{\mathbf{O}, S^{(1)}(u)\} = 0$, where $\mathbf{O} = (O_1, \ldots, O_n)$ (all the observed data) and

$$T\{\mathbf{O}, S^{(1)}(u)\} = \sum_{k=1}^{K} \frac{N_k}{N} \left(\frac{\sum_{i:\epsilon_i=k} R_i[\phi\{F_i, S^{(1)}(u)\} - h(V_i)]}{n_k} + \frac{\sum_{i:\epsilon_i=k} h(V_i)}{N_k} \right). \tag{2.2}$$

Letting $\mu_k(\phi) = E[\phi\{F, S^{(1)}(u)\}|\epsilon = k]$ and $\mu_k(h) = E[h(V)|\epsilon = k]$, it is straightforward to observe that

$$\begin{aligned} E[T\{\mathbf{O}, S^{(1)}(u)\}] &= E(E[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]) \\ &= E[\sum_{k=1}^{K}(N_k/N)\{\mu_k(\phi) - \mu_k(h) + \mu_k(h)\}] \\ &= \sum_{k=1}^{K} p_k \mu_k(\phi) \\ &= E[\phi\{F, S^{(1)}(u)\}], \end{aligned}$$

where $p_k = P(\epsilon = k)$. This implies that $T\{\mathbf{O}, S^{(1)}(u)\}$ is an unbiased estimating function. Consequently, the solution to (2.1) will yield an estimator for $S^{(1)}(u)$ which will converge in limit to the same estimand as the solution to the full data estimating equation given by (1.3) with $X = (X_1^T, X_2^T)^T$. In particular, this means that the proposed AIPWCC estimator will have the double robustness property as described in the previous section. This result is true regardless of the choice for $h(V)$.

To derive the optimal choice of $h(V)$ and to support large sample inference, we must derive the asymptotic variance of estimators solving (2.1) and an estimator for the asymptotic variance of $T\{\mathbf{O}, S^{(1)}(u)\}$. Note that

$$\text{Var}[T\{\mathbf{O}, S^{(1)}(u)\}] = E(\text{Var}[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]) + \text{Var}(E[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]).$$

We showed above that

$$E[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]) = \sum_{k=1}^{K} \frac{N_k}{N} \mu_k(\phi),$$

which, because $(N_1, \ldots, N_K)$ is multinomial $(N; p_1, \ldots, p_K)$, implies

$$\text{Var}(E[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]) = N^{-1}\left[\sum_{k=1}^{K} p_k \mu_k^2(\phi) - \left\{\sum_{k=1}^{K} p_k \mu_k(\phi)\right\}^2\right]. \tag{2.3}$$

Because the data from the n_k and $N_k - n_k$ individuals in stratum k who were selected or not in the subsample are independent realizations from the population in stratum k, and because the data from each stratum are independent, after some algebra (see Appendix E), we obtain

that $\mathrm{Var}[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K]$ equals

$$\sum_{k=1}^{K} \left(\frac{N_k}{N}\right)^2 \left[\{\mathrm{Var}_k(\phi) - 2\mathrm{Cov}_k(\phi, h) + \mathrm{Var}_k(h)\} \left(\frac{1}{n_k} - \frac{1}{N_k}\right) + \mathrm{Var}_k(\phi) \left(\frac{1}{N_k}\right)\right], \qquad (2.4)$$

where $\mathrm{Var}_k(\phi) = \mathrm{Var}[\phi\{F, S^{(1)}(u)\}|\epsilon = k], \mathrm{Var}_k(h) = \mathrm{Var}\{h(V)|\epsilon = k\}$, and $\mathrm{Cov}_k(\phi, h) = \mathrm{Cov}[\phi\{F, S^{(1)}(u)\}, h(V)|\epsilon = k]$. Thus $\mathrm{Var}[T\{\mathbf{O}, S^{(1)}(u)\}]$ is given by (3.5) plus the expectation of (2.4), and the function $h(V)$ that minimizes this variance is obtained by minimizing (2.4). Because

$$\{\mathrm{Var}_k(\phi) - 2\mathrm{Cov}_k(\phi, h) + \mathrm{Var}_k(h)\} = \mathrm{Var}[\phi\{F, S^{(1)}(u)\} - h(V)|\epsilon = k],$$

it is straightforward to show that the optimal choice is

$$h^{opt}(V) = E[\phi\{F, S^{(1)}(u)\}|V],$$

which, although ϵ is a function of V, we write as

$$h^{opt}(V) = E[\phi\{F, S^{(1)}(u)\}|V, \epsilon] = \sum_{k=1}^{K} I(\epsilon = k) E[\phi\{F, S^{(1)}(u)\}|V, \epsilon = k]$$

to emphasize that the best choice for $h(V)$ in each summand of (2.2) is the conditional expectation of $\phi\{F, S^{(1)}(u)\}$ given V within each stratum.

If we write $\phi\{F, S^{(1)}(u)\}$ of equation (1.3) as $\phi_1^*(F) - S^{(1)}(u)$, where $\phi_1^*(F)$ equals

$$\frac{ZI(U \geq u)}{\pi(X)K_c^{(1)}(u, X)} - \left\{\frac{Z - \pi(X)}{\pi(X)}\right\} H^{(1)}(u, X) + \int_0^u \frac{Z}{\pi(X)} \frac{\mathrm{d}M_c^{(1)}(r, X)}{K_c^{(1)}(r, X)} \left\{\frac{H^{(1)}(u, X)}{H^{(1)}(r, X)}\right\}, \qquad (2.5)$$

then after some algebra, we can show that the optimal estimator for $S^{(1)}(u)$ is given by

$$\hat{S}^{(1)}_{strat}(u) = N^{-1} \sum_{i=1}^{N} \left[\frac{R_i \phi_1^*(F_i)}{\eta(\epsilon_i)} - \left\{\frac{R_i - \eta(\epsilon_i)}{\eta(\epsilon_i)}\right\} E\{\phi_1^*(F_i)|V_i, \epsilon_i\}\right], \qquad (2.6)$$

where $E\{\phi_1^*(F_i)|V_i, \epsilon_i\} = \sum_{k=1}^{K} I(\epsilon_i = k) E\{\phi_1^*(F_i)|V_i, \epsilon_i = k\}$.

As the conditional expectation $E\{\phi_1^*(F)|V, \epsilon = k\}$ are not known, we would propose models $w_k(V, \psi_k^{(1)})$ for them in terms of parameters $\psi_k^{(1)}$. The $\psi_k^{(1)}$ may be estimated via least squares as follows using the data on subjects in the subsample from stratum k (where both F and V are collected). For each subject i in the subsample, form an estimator $\hat{\phi}_{1i}^*$, say, for $\phi_1^*(F_i)$ as given by (2.5) by substituting appropriate estimators for $K_c^{(1)}(r, X_i), H^{(1)}(r, X_i)$ and $\pi(X_i)$. The

estimator $\hat{\psi}_k^{(1)}$ is derived by minimizing $\sum_{i:R_i=1,\epsilon_i=k}\{\hat{\phi}_{1i}^* - w_k(V_i, \psi_k^{(1)})\}^2$ in $\psi_k^{(1)}$. The estimator for $S^{(1)}$ is the solution to equation (2.6) with $E\{\phi_1^*(F_i)|V_i, \epsilon_i\}$ replaced by $w_k(V_i, \hat{\psi}_k^{(1)})$ when $\epsilon_i = k$.

Care must be taken in estimating $\pi(X)$, $H^{(1)}(r, X)$ and $K_c^{(1)}(r, X)$, as all variables in X are collected only on subjects in the subsample. Because $H^{(1)}(r, X)$ and $K_c^{(1)}(r, X)$ are probabilities conditional on X and $Z = 1$, and because the strata are defined through X_1 (the subset of X observed for all subjects) and Z, consistent estimators for these functions may be derived by modeling and fitting based on the subsample in a manner analogous to that based the entire sample when there is no stratified sampling. However, this is not the case for the propensity score $\pi(X)$. For the sake of clarity, suppose as in ASCERT that the $K = 2M$ strata are defined by all combinations of M categories derived through components of X_1 and the two treatment levels 0 and 1. Denote the number of subjects in the stratum corresponding to the m-th category for treatments 1 and 0 as N_{1m} and N_{0m} and the corresponding numbers in the subsample as n_{1m} and n_{0m}, respectively, $m = 1, \ldots, M$. Then, using redundant notation similar to that above and letting Cat be the indicator for category, it is straightforward to show that if

$$\text{logit}\{P(Z = 1|X, Cat = m)\} = \zeta(X, m),$$

then

$$\text{logit}\{P(Z = 1|X, Cat = m, R = 1)\} = \zeta(X, m) + \log(n_{1m}/N_{1m}) - \log(n_{0m}/N_{0m}),$$

where $\text{logit}(\theta) = \log\{\theta/(1 - \theta)\}$. Consequently, writing

$$\pi(X) = \sum_{m=1}^{M} I(Cat = m)\pi(X, m),$$

where $\pi(X, m) = P(Z = 1|X, Cat = m)$, we propose estimating $\pi(X)$ by fitting postulated models for each $\pi(X, m)$. This may be based on fitting logistic regression models to the subsample data $(Z_i, X_i), i : R_i = 1$ of the form

$$\text{logit}\{P(Z = 1|X, Cat = m, R = 1)\} = \zeta^*(X, m),$$

so that the estimator for $\pi(X, m)$ is of the form

$$\text{expit}\{\zeta^*(X, m) - \log(n_{1m}/N_{1m}) + \log(n_{0m}/N_{0m})\},$$

where $\text{expit}(\theta) = e^\theta/(1 + e^\theta)$.

A simple but effective suggestion for modeling $w_k(V, \psi_k^{(1)})$ is based on the following: If the

additional covariate information X_2 was not predictive for $\pi(X)$, $K_c^{(1)}(r,X)$ or $H^{(1)}(r,X)$, then $E\{\phi_1^*(F)|V,\epsilon=k\}=\phi_1^*(V)$. Of course, this is generally not the case but this suggests that we might consider a model where we assume that $E\{\phi_1^*(F)|V,\epsilon=k]=\psi_{0k}^{(1)}+\psi_{1k}^{(1)}\phi_1^*(V)$. Keep in mind that we will obtain a doubly robust estimator for $S^{(1)}(u)$ regardless of whether or not the model for $w_k(V,\psi_k^{(1)})$ is correctly specified. However, a good approximation for $E\{\phi_1^*(F)|V,\epsilon=k\}$ should result in gains in efficiency. Consequently, the suggested strategy is to derive an estimated $\tilde{\phi}_1^*(V_i)$ for individual i exactly as was suggested in Section 2 for estimating (2.5) using all the data from the main study with only X_1. The parameter $\psi_k^{(1)}$ is estimated using least squares by minimizing $\sum_{i:R_i=1,\epsilon_i=k}\{\hat{\phi}_{1i}^*-\psi_{0k}^{(1)}-\psi_{1k}^{(1)}\tilde{\phi}_1^*(V_i)\}^2$ in $\psi_k^{(1)}$. The resulting AIPWCC stratified estimator is given by

$$\hat{S}_{strat}^{(1)}(u)=N^{-1}\sum_{i=1}^{N}\left[\frac{R_i\hat{\phi}_{1i}^*}{\eta(\epsilon_i)}-\left\{\frac{R_i-\eta(\epsilon_i)}{\eta(\epsilon_i)}\right\}\{\hat{\psi}_{0k}^{(1)}+\hat{\psi}_{1k}^{(1)}\tilde{\phi}_1^*(V_i)\}\right]. \tag{2.7}$$

As we will demonstrate in the simulations studies (Section 2.5), this strategy leads to an estimator that has good performance.

The variance estimator for the estimator of $S^{(1)}(u)$ would then use (3.5) plus (2.4), substituting the natural estimators for $p_k,\mu_k(\phi),\mu_k(h),\mathrm{Var}_k(\phi),\mathrm{Var}_k(h)$ and $\mathrm{Cov}_k(\phi,h)$, where here we take

$$h(V)=\sum_{k=1}^{K}I(\epsilon=k)\{\psi_{0k}^{(1)}+\psi_{1k}^{(1)}\phi_1^*(V)-S^{(1)}(u)\}.$$

As in the case when there is no stratified sampling, the double robustness property as well as the adequacy of the proposed variance estimator will depend on whether or not the models for $\pi(X)$ (used for the propensity score), $K_c^{(1)}(r,X)$ and $H^{(1)}(r,X)$ are correctly specified.

2.4 Inference on Difference of Treatments

Since we are also interested in making inference on $S_{diff}(u)$, we also provide an estimator for $S_{diff}(u)$ and its variance estimator when a substudy is conducted. In Section 2.3, we have shown that (2.7) is a consistent estimator of $S^{(1)}(u)$. Similarly, we can estimate the difference function $S_{diff}(u)$ by

$$\hat{S}_{diff,strat}(u)=N^{-1}\sum_{i=1}^{N}\left(\frac{R_i\hat{\phi}_{diff,i}^*}{\eta(\epsilon_i)}-\left\{\frac{R_i-\eta(\epsilon_i)}{\eta(\epsilon_i)}\right\}\hat{h}_{diff}^*(V_i)\right),$$

where $\hat{\phi}_{diff,i}^*=\hat{\phi}_{1i}^*-\hat{\phi}_{0i}^*$, for treatment $j=(0,1)$, $\hat{\phi}_{j_i}^*$ is the estimator for $\phi_j^*(F_i)$ defined by equation (1.6) with $X_i=(X_{1i}^T,X_{2i}^T)^T$, and $\hat{h}_{diff}^*(V_i)=\{\hat{\psi}_{0k}^{(1)}+\hat{\psi}_{1k}^{(1)}\tilde{\phi}_1^*(V_i)\}-\{\hat{\psi}_{0k}^{(0)}+\hat{\psi}_{1k}^{(0)}\tilde{\phi}_0^*(V_i)\}$,

where $\hat{\psi}_k^{(j)}$ is obtained using least squares by minimizing $\sum_{i:R_i=1,\epsilon_i=k}\{\hat{\phi}_{j\,i}^* - \psi_{0k}^{(j)} - \psi_{1k}^{(j)}\tilde{\phi}_j^*(V_i)\}^2$, $j = 0, 1$. The asymptotic variance can be estimated by substituting the above $\hat{\phi}_{diff}^*$ and $\hat{h}_{diff}(V)$ into (3.5) and (2.4).

2.5 Simulations

In this simulation we consider the case where we conduct a substudy using stratified sampling and consider the performance of our stratified AIPWCC estimators. Using the same data generating mechanism as in Section 1.5, we collected X_1 on the main study of $N = 5,000$ observations and randomly sampled $n_{11} = n_{12} = n_{01} = n_{02} = 300$ observations from each of $K = 4$ stratum combination of (Z, B) without replacement and collected additional covariates X_2 on these subsamples. In these simulations we used the correct models for $\pi(X)$, $K_c^{(j)}(r, X)$, and $H^{(j)}(r, X)$, $j = 0, 1$.

We compared the consistency and efficiency of three estimators for $S^{(1)}(u)$ under this setting. The first estimator is the estimator (2.7). We refer to this estimator as the AIPWCC stratified estimator. The second one used the estimator resulting from (2.1), where the augmentation term $h(V_i) = 0$, referred to as the weighted estimator. We also considered the infeasible estimator (1.4) with covariates X_1, X_2 on the whole dataset as the gold standard.

As expected, all three estimators along with their estimated standard errors are consistent and the confidence intervals have coverage rate close to the nominal rate of 0.95.

The relative efficiency is computed by taking the ratio of the Monte-Carlo variance of each estimator over the Monte-Carlo variance of the gold standard estimator. Here we see the large gains in efficiency using the AIPWCC stratified estimator compared to the weighted estimator with little loss of efficiency compared to the gold standard. Recall that the weighted estimator only uses data from the substudy whereas the AIPWCC stratified estimator uses all the observed data.

This illustrates that choosing the augmentation term appropriately can recover a great deal of information from the main study and result in large gains in efficiency.

We also compared the performance of estimator for $S_{diff}(u)$ (see Table 2.2) and the results are similar to the above estimators of $S^{(1)}(u)$. In addition, we conducted simulation studies where the models for either the propensity score $\pi(X)$ or the conditional survival distribution $H^{(j)}(r, X)$ were misspecified, and the results were as expected. The estimators were doubly robust and the estimated variance and coverage behaved as it did for the first set of simulations where X was collected on all the patients in the main study (see Appendix F).

Table 2.1: Simulation 2, estimators for $\hat{S}^{(1)}_{strat}(u)$ at selected time points when there is subsample to collect additional covariates. Sample size N=5000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$ samples randomly selected from K=4 strata. The relative efficiency (RE) is the ratio of Monte-Carlo variance of each estimator over the Monte-Carlo variance of the gold standard estimator.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}_{strat}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate	Relative Efficiency
AIPWCC	0.5	0.693	0.694	0.000	0.009	0.009	0.96	1.20
	1.0	0.523	0.523	0.000	0.011	0.011	0.95	1.31
	1.5	0.408	0.408	0.000	0.012	0.012	0.95	1.36
	2.0	0.324	0.324	0.000	0.012	0.012	0.95	1.34
	2.5	0.262	0.262	0.000	0.013	0.013	0.95	1.37
	3.0	0.214	0.213	-0.001	0.013	0.013	0.94	1.32
Weighted	0.5	0.693	0.694	0.001	0.019	0.019	0.95	5.19
	1.0	0.523	0.524	0.001	0.021	0.021	0.96	5.10
	1.5	0.408	0.410	0.002	0.023	0.023	0.94	5.56
	2.0	0.324	0.325	0.001	0.022	0.023	0.96	4.76
	2.5	0.262	0.262	0.000	0.023	0.024	0.96	4.46
	3.0	0.214	0.214	0.000	0.024	0.025	0.96	4.26
Gold	0.5	0.693	0.693	0.000	0.008	0.008	0.95	1.00
	1.0	0.523	0.523	0.000	0.009	0.010	0.95	1.00
	1.5	0.408	0.407	0.001	0.010	0.010	0.94	1.00
	2.0	0.324	0.324	0.000	0.010	0.010	0.96	1.00
	2.5	0.262	0.261	-0.001	0.011	0.011	0.94	1.00
	3.0	0.214	0.214	0.000	0.012	0.011	0.94	1.00

2.6 Analysis of the ASCERT Study

In this section, we apply the proposed AIPWCC estimator to data from the ASCERT study. The goal of the analysis was to estimate the AIPWCC-adjusted 4-year survival distribution for PCI and CABG using baseline covariate data from the main ASCERT study database augmented with information on coronary anatomy from a subsample of patients in the ASCERT angiographic companion study. The sampling frame consisted of records from $9,800$ patients in the ASCERT database who underwent a coronary revascularization procedure (CABG or PCI) at one of 54 hospitals participating in the ASCERT study that agreed to be part of the companion study. For the purpose of this analysis we considered the patients from the 54 hospitals of the companion study to be the main focus of inference. Eighteen covariates were used on the full sample which included demographics (e.g., age, sex), risk factors (e.g., body mass index, smoking), symptoms and history of cardiovascular disease (e.g., chest pain, congestive heart failure), and comorbidities (e.g., diabetes). The subsample includes records

Table 2.2: Simulation 2, estimators for $\hat{S}_{diff,strat}(u)$ at selected time points when there is subsample to collect additional covariates. Sample size N=5000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$ samples randomly selected from K=4 strata. The relative efficiency (RE) is the ratio of Monte-Carlo variance of each estimator over the Monte-Carlo variance of the gold standard estimator.

	u	$S_{diff}(u)$	$\hat{S}_{diff,strat}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate	Relative Efficiency
AIPWCC	0.5	0.118	0.119	0.001	0.017	0.016	0.94	1.53
	1.0	0.143	0.143	0.000	0.016	0.017	0.96	1.56
	1.5	0.143	0.143	0.000	0.0165	0.017	0.95	1.48
	2.0	0.133	0.133	0.000	0.016	0.016	0.97	1.39
	2.5	0.121	0.121	0.000	0.017	0.016	0.95	1.39
	3.0	0.108	0.107	0.000	0.017	0.017	0.94	1.35
Weighted	0.5	0.118	0.119	0.001	0.029	0.027	0.92	4.48
	1.0	0.143	0.144	0.001	0.028	0.028	0.96	4.52
	1.5	0.143	0.145	0.003	0.028	0.028	0.96	4.32
	2.0	0.133	0.134	0.000	0.028	0.028	0.95	4.30
	2.5	0.121	0.120	-0.001	0.029	0.028	0.94	4.13
	3.0	0.108	0.108	0.000	0.029	0.029	0.94	3.86
Gold	0.5	0.118	0.119	0.001	0.029	0.027	0.95	1.00
	1.0	0.143	0.143	0.000	0.028	0.028	0.97	1.00
	1.5	0.143	0.142	0.000	0.028	0.028	0.96	1.00
	2.0	0.133	0.133	0.000	0.028	0.028	0.95	1.00
	2.5	0.121	0.120	0.000	0.029	0.028	0.94	1.00
	3.0	0.108	0.107	0.000	0.029	0.029	0.93	1.00

from approximately $2,000$ patients chosen by design (roughly 500 in each of the four strata determined by all combinations of the two treatments and whether or not the patients had two- or three-vessel disease). Information collected on the subsample includes features of the patient's coronary anatomy (e.g., left-side dominance) and features of each individual blockage (e.g., lesion length, tortuosity, calcification, degree of stenosis).

The original sampling frame comprised 9,800 patients. Subsequently it was determined that some patients were ineligible for analysis. Consequently, the main study participants consisted of $7,393$ eligible patients among the $9,800$ patients in the 54 hospitals of the companion study and the subsample consisted of $1,554$ eligible patients among the $2,049$ originally sampled. Table 2.3 provides details on the sample size before and after eligibility was determined.

Figure 2.1 presents plots of the estimated AIPWCC-adjusted survival distributions with accompanying 95% pointwise confidence intervals for PCI, CABG and their difference as estimated by the proposed AIPWCC estimator. Overall, results are highly consistent with the published primary ASCERT analysis (Weintraub et al., 2012). Short-term risk appears to be lower with PCI but long-term risk appears to be lower with CABG.

Table 2.3: Number of Patients in the Sampling Frame and Sample.

		Original		Subset meeting NEJM inclusion/exclusion	
		sampling frame	sample	sampling frame	sample
2-vessel	CABG	973	554	718	406
	PCI	3938	486	3681	453
3-vessel	CABG	3357	527	1659	260
	PCI	1532	482	1335	435
Total		9800	2049	7939	1554

Table 2.4 presents the estimated AIPWCC-adjusted cumulative incidence (one minus the survival distribution) of mortality at two time points (30 days and 4 years) and compares two versions of the adjusted analysis together with the unadjusted treatment-specific Kaplan-Meier estimators. The first AIPWCC estimator adjusts only for covariates available in the full sample and includes all $7,393$ patients. The second analysis used the AIPWCC stratified estimator and includes the additional covariates on coronary anatomy collected on the subsample and corresponds to the analysis shown in Figure 2.1. In each case, the 30-day estimate favors PCI whereas the 4-year estimate favors CABG. Both adjusted analyses gave similar results with possibly CABG having slightly better performance compared to PCI in the unadjusted analysis.

Overall all three analyses gave similar results suggesting that the confounding, either with covariates used in the main study or those collected in the substudy, was not substantial.

2.7 Conclusion and Discussion

The AIPWCC estimator derived in chapter 1 provide valid inference if the key assumption of no unmeasured confounder holds. In the case when additional covariate information is necessary to make the assumption of no unmeasured confounders tenable, we could conduct a substudy using a stratified sampling design to collect such additional covariates and we have proposed a method for obtaining doubly robust estimators with such a design which uses the data from the main study to gain efficiency. The corresponding estimator and standard error estimator was shown in simulation studies to perform well.

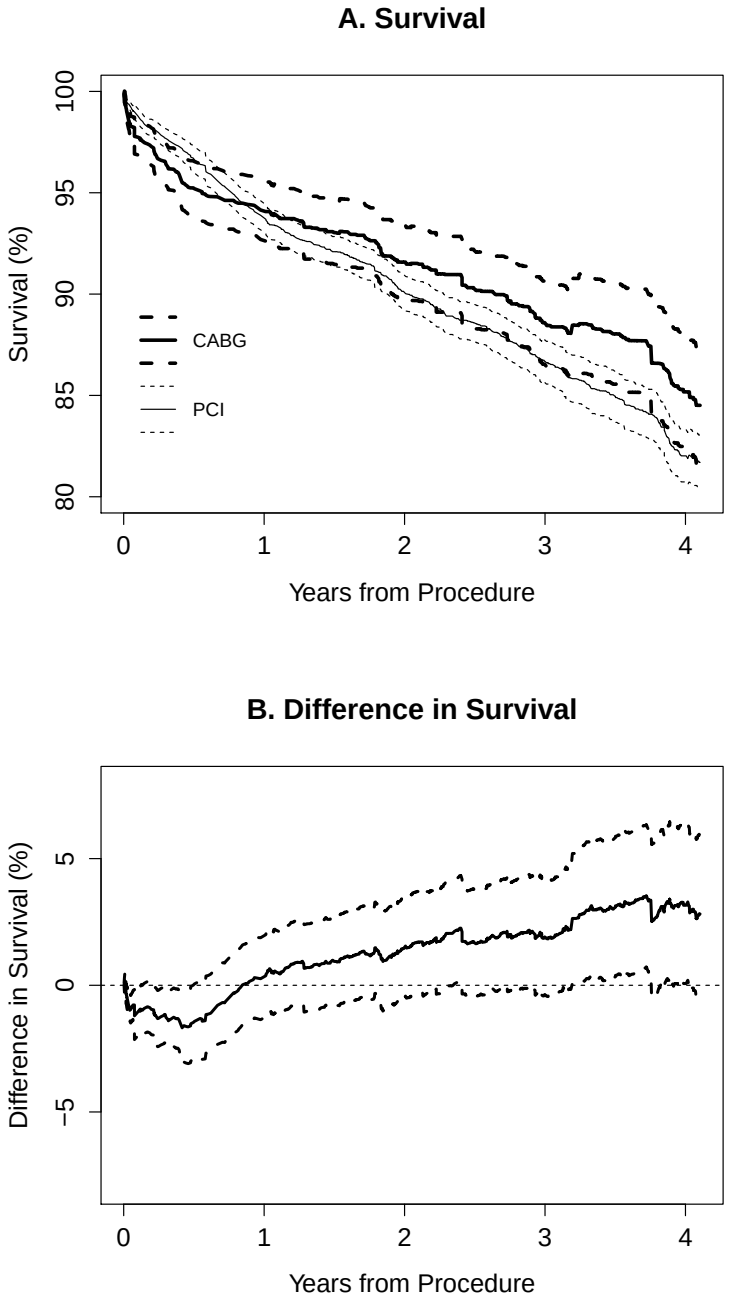

Figure 2.1: Estimated AIPWCC Stratified Survival Estimator for CABG and PCI. The confidence interval is computed as the treatment-specific survival estimator plus/minus 1.96 times the proposed standard error estimator.

Table 2.4: Estimated AIPWCC-Adjusted Cumulative Incidence of Mortality at 30 Days and 4 Years by Treatment Group, Expressed as Percentage With 95% Confidence Interval. The proposed standard error estimator is used to compute the confidence interval.

		30-day (%)	4-year (%)
Unadjusted Kaplan Meier	CABG	1.77 (1.24, 2.30)	14.69 (12.90, 16.49)
	PCI	1.14 (0.84, 1.43)	18.64 (17.25, 20.03)
	Difference	-0.64 (-1.24, -0.03)	3.95 (1.68, 6.22)
AIPWCC (Full)	CABG	2.17 (1.50, 2.83)	15.93 (13.98, 17.88)
	PCI	1.08 (0.79, 1.38)	18.31 (16.95, 19.67)
	Difference	-1.08 (-1.81, -0.36)	2.38 (0.04, 4.73)
AIPWCC (Stratified)	CABG	2.26 (1.39, 3.13)	14.83 (12.00, 17.65)
	PCI	1.21 (0.85, 1.57)	17.98 (16.70, 19.27)
	Difference	-1.05 (-1.99, -0.11)	3.16 (0.07, 6.25)

Chapter 3

A Doubly-Robust Log Rank Type Test for Comparing Survival Distributions in an Observational Study With and Without an Auxiliary Substudy

3.1 Introduction

Besides treatment-specific survival curve estimates discussed in the previous two chapters, the overall comparison of treatment-specific survival distributions is also an important issue in survival analysis. In a randomized controlled clinical trial where covariate distribution is balanced among different treatment groups, the log rank test is most commonly used to test the null hypothesis that there is no significant difference between treatments. In an observational study, however, the traditional log rank test is no longer valid due to possible confounding.

Xie and Liu (2005) proposed an inverse probability weighted log rank test to adjust for such confounding. They computed a weighted version of log rank statistics by substituting inverse probability weighted number of subject at risk and number of subject that die in each group. This estimator would provide valid inference if the propensity model is consistently estimated. Zhang and Schaubel (2012) compared treatment groups in term of the difference in restricted mean lifetime, which they referred as the average causal effect. This is a semiparametric estimator with a doubly-robust property. However, its consistency relies on the underlying assumption that the censoring time and survival time are independent conditional on the treatment. If the

covariates also have effect on censoring which may likely be the case in observational studies, this method might be subject to bias.

In this chapter, we propose a log rank type test statistics to compare treatment groups. With a nonparametric bootstrap estimator of the denominator, the resulting test statistics will be doubly-robust. Moreover, we generalize the statistics into the case where a substudy is to be conducted to collect additional covariates as discussed in Chapter 2 to account for residual confounding.

3.2 Test Statistic for Main Study when all Potential Confounders are Captured

In this section, we develop the log rank type statistic when the covariates collected in the main study are believed to capture all potential confounding. The assumptions are the same as those described in Chapter 1 where we developed the doubly-robust estimator for treatment-specific distributions $S^{(j)}(u) = P(T^{(1)} \geq u), u \geq 0, j = 0, 1$ as

$$\begin{aligned}\hat{S}^{(j)}(u) &= N^{-1} \sum_{i=1}^{N} \phi_j^*(u; V_i) \\ &= N^{-1} \sum_{i=1}^{N} \left[\frac{(Z_i)^j (1-Z_i)^{1-j} I(U_i \geq u)}{\{\pi(X_i)\}^j \{1-\pi(X_i)\}^{1-j} K_c^{(j)}(u, X_i)} - \frac{(2j-1)\{Z_i - \pi(X_i)\}}{\{\pi(X_i)\}^j \{1-\pi(X_i)\}^{1-j}} H^{(j)}(u, X_i) \right. \\ &\qquad \left. + \int_0^u \frac{(Z_i)^j (1-Z_i)^{1-j}}{\{\pi(X_i)\}^j \{1-\pi(X_i)\}^{1-j}} \frac{\mathrm{d}M_c^{(j)}(r, X_i)}{K_c^{(j)}(r, X_i)} \frac{H^{(j)}(u, X_i)}{H^{(j)}(r, X_i)} \right], \end{aligned} \tag{3.1}$$

where $\pi(X) = P(Z = 1|X)$ is the propensity score, $K_c^{(j)}(r, X) = P(C \geq r|X, Z = j)$ is the conditional survival function of the treatment specific censoring time given X, $\mathrm{d}M_c^{(j)}(r, X)$ is the martingale increment for the censoring distribution, namely, $\mathrm{d}M_c^{(j)}(r, X) = \mathrm{d}N_c(r) - \lambda_c^{(j)}(r, X)Y(r)$, $N_c(r) = I(U \leq r, \Delta = 0)$, $Y(r) = I(U \geq r)$ and $\lambda_c^{(j)}(r, X) = \frac{-\mathrm{d}\log K_c^{(j)}(r,X)}{\mathrm{d}r}$ is the conditional hazard function for C given $Z = j$ and X, and $H^{(j)}(r, X) = P(T^{(j)} \geq r|X)$. In Chapter 1 we also derived the estimator $\hat{S}^{(1)}(u) - \hat{S}^{(0)}(u)$ and its variance estimator allowing us to compare the survival distributions at a fixed time point u.

The treatment comparison of the overall survival distributions is often of interest and is commonly based on estimating the treatment-specific hazard ratio assuming a proportional hazards model. That is, we assume $\lambda^{(1)}(u) = \lambda^{(0)}(u)\exp(\alpha)$, where $\lambda^{(j)}(u) = \lim_{h \to 0} h^{-1} P(u \leq T^{(j)} < u + h|T^{(j)} \geq u)$ is the hazard function of $T^{(j)}$ at time u, $j = 0, 1$. The log hazard ratio α is then of interest.

In a randomized study with data $(Z_i, U_i, \Delta_i), i = 1, \ldots, N$, the maximum likelihood estima-

tor for α solves

$$\sum_{i=1}^{N} \int \{Z_i - \bar{Z}(u,\alpha)\} \mathrm{d}N_i(u) = 0,$$

where $N_i(u) = I(U_i \leq u, \Delta_i = 1)$ and $\bar{Z}(u,\alpha) = \sum_{i=1}^{N} Z_i \exp(\alpha Z_i) Y_i(u) / \sum_{i=1}^{N} \exp(\alpha Z_i) Y_i(u)$. After some algebra, this estimator can be written as the solution to

$$\int W_N(u,\alpha) \{\mathrm{d}\hat{\Lambda}_1(u) \exp(-\alpha) - \mathrm{d}\hat{\Lambda}_0(u)\} = 0,$$

where $Y_1(u) = \sum_{i=1}^{N} Z_i Y_i(u)$, $Y_0(u) = \sum_{i=1}^{N} (1 - Z_i) Y_i(u)$, $N_1(u) = \sum_{i=1}^{N} Z_i N_i(u)$, $N_0(u) = \sum_{i=1}^{N} (1 - Z_i) N_i(u)$, $\mathrm{d}\hat{\Lambda}_j(u) = \mathrm{d}N_j(u)/Y_j(u)$ is the usual treatment specific Nelson-Aalen estimator for the hazard function for the survival distribution for treatment $j = 0, 1$ and $W_N(u,\alpha) = \frac{Y_1(u)\exp(\alpha) Y_0(u)}{Y_1(u)\exp(\alpha) + Y_0(u)}$.

Motivated by the representation of this estimating equation in a randomized study case, we suggest estimating α in the observational study as the solution to the estimating equation

$$\int W_N(u,\alpha) \{\mathrm{d}\hat{\Lambda}^{(1)}(u) \exp(-\alpha) - \mathrm{d}\hat{\Lambda}^{(0)}(u)\} = 0, \tag{3.2}$$

where $W_N(u,\alpha)$ is a weight function and $\mathrm{d}\hat{\Lambda}^{(j)}(u) = -\mathrm{d}\hat{S}^{(j)}(u)/\hat{S}^{(j)}(u), j = 0, 1$. Here the estimator for the treatment-specific survival distributions $\hat{S}^{(j)}(u)$ use the doubly robust estimator (3.1). Subject to regularity conditions, any choice for $W_N(u,\alpha)$, as long as $N^{-1} W_N(u,\alpha)$ converges to a deterministic function $w(u,\alpha)$, will lead to a consistent, asymptotically normal, doubly robust estimator for α.

Similarly, in a randomized study, $\int W_N(u,0) \{\mathrm{d}\hat{\Lambda}_1(u) - \mathrm{d}\hat{\Lambda}_0(u)\}$ is used to compute the logrank test statistic. By analogy, we propose to use $\int W_N(u) \{\mathrm{d}\hat{\Lambda}^{(1)}(u) - \mathrm{d}\hat{\Lambda}^{(0)}(u)\}$ to develop a log rank type test of the null hypothesis of treatment equality in observation studies. In particular, we use $W_N(u) = \frac{1}{N} \frac{Y_1(u) Y_0(u)}{Y_1(u) + Y_0(u)}$, define

$$T_N = \int W_N(u) \{\mathrm{d}\hat{\Lambda}^{(1)}(u) - \mathrm{d}\hat{\Lambda}^{(0)}(u)\} = \int W_N(u) \left\{ \frac{\mathrm{d}\hat{S}^{(0)}(u)}{\hat{S}^{(0)}(u)} - \frac{\mathrm{d}\hat{S}^{(1)}(u)}{\hat{S}^{(1)}(u)} \right\}, \tag{3.3}$$

and

$$T_N^* = \int W_N(u) \left[\left\{ \frac{\mathrm{d}\hat{S}^{(0)}(u)}{\hat{S}^{(0)}(u)} - \frac{\mathrm{d}S^{(0)}(u)}{S^{(0)}(u)} \right\} - \left\{ \frac{\mathrm{d}\hat{S}^{(1)}(u)}{\hat{S}^{(1)}(u)} - \frac{\mathrm{d}S^{(1)}(u)}{S^{(1)}(u)} \right\} \right].$$

Under the null hypothesis of $\mathrm{H}_0 : S^{(1)}(u) = S^{(0)}(u)$, we have $T_N = T_N^*$, hence $\mathrm{Var}(T_N) =$

$\mathrm{Var}(T_N^*)$. Moreover,

$$\begin{aligned} T_N^* &= \int W_N(u) \left[\left\{ \frac{\mathrm{d}\hat{S}^{(0)}(u)}{\hat{S}^{(0)}(u)} - \frac{\mathrm{d}S^{(0)}(u)}{\hat{S}^{(0)}(u)} + \frac{\mathrm{d}S^{(0)}(u)}{\hat{S}^{(0)}(u)} - \frac{\mathrm{d}S^{(0)}(u)}{S^{(0)}(u)} \right\} \right. \\ &\quad \left. - \left\{ \frac{\mathrm{d}\hat{S}^{(1)}(u)}{\hat{S}^{(1)}(u)} - \frac{\mathrm{d}S^{(1)}(u)}{\hat{S}^{(1)}(u)} + \frac{\mathrm{d}S^{(1)}(u)}{\hat{S}^{(1)}(u)} - \frac{\mathrm{d}S^{(1)}(u)}{S^{(1)}(u)} \right\} \right] \\ &= \int \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d} \left\{ \hat{S}^{(0)}(u) - S^{(0)}(u) \right\} - \frac{W_N(u) \mathrm{d}S^{(0)}(u)}{\hat{S}^{(0)}(u) S^{(0)}(u)} \left\{ \hat{S}^{(0)}(u) - S^{(0)}(u) \right\} \right] \\ &\quad - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d} \left\{ \hat{S}^{(1)}(u) - S^{(1)}(u) \right\} - \frac{W_N(u) \mathrm{d}S^{(1)}(u)}{\hat{S}_1(u) S^{(1)}(u)} \left\{ \hat{S}^{(1)}(u) - S^{(1)}(u) \right\} \right]. \end{aligned}$$

Since $\hat{S}^{(j)}(u) - S^{(j)}(u) = \frac{1}{N} \sum_{i=1}^{N} \phi_j(u; V_i)$, where $\phi_j(u; V_i) = \phi_j^*(u; V_i) - S^{(j)}(u)$, then

$$\begin{aligned} \hat{T}_N^* = \frac{1}{N} \sum_{i=1}^{N} \int & \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d}\phi_0(u; V_i) - \frac{W_N(u) \mathrm{d}\hat{S}^{(0)}(u)}{\{\hat{S}^{(0)}(u)\}^2} \phi_0(u; V_i) \right] \\ & - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d}\phi_1(u; V_i) - \frac{W_N(u) \mathrm{d}\hat{S}^{(1)}(u)}{\{\hat{S}^{(1)}(u)\}^2} \phi_1(u; V_i) \right]. \end{aligned}$$

Hence

$$\begin{aligned} \widehat{\mathrm{Var}}(\hat{T}_N^*) = \frac{1}{N} \mathrm{Var} \left(\int \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d}\phi_0(u; V_i) - \frac{W_N(u) \mathrm{d}\hat{S}^{(0)}(u)}{\{\hat{S}^{(0)}(u)\}^2} \phi_0(u; V_i) \right] \right. \\ \left. - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d}\phi_1(u; V_i) - \frac{W_N(u) \mathrm{d}\hat{S}^{(1)}(u)}{\{\hat{S}^{(1)}(u)\}^2} \phi_1(u; V_i) \right] \right). \end{aligned}$$

The log rank type statistic takes the form

$$G = \frac{T_N}{\sqrt{\widehat{\mathrm{Var}}(\hat{T}_N^*)}}, \tag{3.4}$$

which asymptotically follows the standard normal distribution $N(0, 1)$ under the null hypothesis of $\mathrm{H}_0 : S^{(1)}(u) = S^{(0)}(u)$. Hence, null hypothesis is rejected at the α level when $|G| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of standard normal distribution.

In Chapter 1, we showed that $\hat{S}^{(j)}(u), j = 0, 1$ is doubly-robust; that is, if either the estimator for the coarsening defined through the estimator for $P(Z = 1|X) = \pi(X)$ and the estimator for $P(C \geq r|Z = j, X) = K_c^{(j)}(r, X)$ are both consistent or the estimator for $P(T^{(j)} \geq r|X) = H^{(j)}(r, X)$ is consistent, then the estimator (1.4) will be consistent. Hence, the numerator T_N is doubly-robust (with expectation mean zero) if either $\pi(X)$ and $K_c^{(j)}(r, X)$

are both consistently estimated or $H^{(j)}(r, X)$ is consistently estimated. The denominator estimator $\sqrt{\widehat{\text{Var}}(\hat{T}_N^*)}$, however, is not guaranteed to be valid under model misspecification. If the propensity score $\pi(X)$ and the conditional censoring distribution $K_c^{(j)}(r, X)$ were consistently estimated whereas, the conditional survival distribution $H^{(j)}(r, X)$ was not, then the variance estimator would be conservative. The resulting test statistic $\frac{T_N}{\sqrt{\widehat{\text{Var}}(\hat{T}_N^*)}}$ would be a normal distribution with mean 0 and variance less than 1. On the other hand, when the estimator for $H^{(j)}(r, X)$ is consistent and either of the estimators for $\pi(X)$ or $K_c^{(j)}(r, X)$ are not, then the resulting denominator $\sqrt{\widehat{\text{Var}}(\hat{T}_N^*)}$ may be biased and there is no theoretical result on the direction or extent of the bias. In such cases, it is suggested to use a bootstrap estimator of the asymptotic variance. To be more specific, we resample the dataset $(Z_i, X_i, U_i, \Delta_i), i = 1, \ldots, n$, with replacement B times. In each bootstrap replicate, we compute the numerator $T_N^{(b)}, b = 1, \ldots, B$. We then compute the sample standard deviation of $T_N^{(b)}, b = 1, \ldots, B$, and denote that by $\sqrt{\text{Var}\{T_N^{(b)}\}}$. We propose the bootstrap test statistics as $G_{bootstrap} = \frac{T_N}{\sqrt{\text{Var}\{T_N^{(b)}\}}}$. As shown later in our simulation studies, the bootstrap version of these test statistics have good asymptotic properties under varying model misspecification.

Similar to the ordinary log rank test, under local alternatives, these test statistics follow a normal distribution with a non-centrality parameter and variance 1. We compare the performance of this statistic with other competing methods in terms of power under various alternatives in the simulation studies in Section 3.4.

One thing worth pointing out is that both the numerator and denominator of (3.4) involve the reciprocal of $\hat{S}^{(j)}(u), j = 0, 1$, which tends to be unstable at the tail of survival curve. Therefore, in order to lessen this problem one may truncate the range of integration of our test statistic. The validity of of the distribution of this truncated test statistic will still hold under the null hypothesis. However, with truncation there is potential loss of efficiency by not using all the survival data. Thus, we need to evaluate the balance between increased stability versus loss of efficiency in the choice of the truncation point. This issue will also be examined in our simulation studies.

3.3 Test Statistics in Stratified Sampling with Additional Information from a Subsample

As discussed in Chapter 2, the inference derived in Section 3.2 will be valid only when the critical no unmeasured confounders assumption $Z \perp\!\!\!\perp T^{(j)}|X$ holds for covariates collected in the main study. If this assumption is questionable then one may decide to collect additional covariate information in a substudy in order to make this assumption tenable. This indeed

was the case for the ASCERT study discussed in Chapter 2 where additional covariates were collected using a stratified sampling design. We now show how one can construct a valid log rank type test for comparing treatment-specific survival distributions where a substudy using a stratified sampling design to collect such additional covariates was used.

Using the same notation as in Chapter 2, X_1 denotes the covariate collected in the main study and X_2 denotes the additional covariates collected on subsample. Use $F_i = (U_i, \Delta_i, Z_i, X_{1i}, X_{2i}), i = 1, \ldots, N$ to denote the "full data" and

$$\begin{aligned}\phi_j^*(u; F_i) &= \left[\frac{(Z_i)^j(1-Z_i)^{1-j} I(U_i \geq u)}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j} K_c^{(j)}(u, X_i)} - \frac{(2j-1)\{Z_i - \pi(X_i)\}}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j}} H^{(j)}(u, X_i)\right. \\ &\quad \left. + \int_0^u \frac{(Z_i)^j(1-Z_i)^{1-j}}{\{\pi(X_i)\}^j\{1-\pi(X_i)\}^{1-j}} \frac{\mathrm{d}M_c^{(j)}(r, X_i)}{K_c^{(j)}(r, X_i)} \frac{H^{(j)}(u, X_i)}{H^{(j)}(r, X_i)}\right]\end{aligned}$$

to denote the i-th element of estimating equation that would have been used to estimate $S^{(j)}(u)$ if we had access to all the full data in the main study. This is similar to the estimating equation (3.1) but using both X_1 and X_2. As shown in Chapter 2, the doubly-robust estimator for the treatment-specific survival probability at time point u is:

$$\hat{S}^{(j)}(u) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{R_i}{\eta(\epsilon_i)}\hat{\phi}_j^*(u; F_i) - \left\{\frac{R_i - \eta(\epsilon_i)}{\eta(\epsilon_i)}\right\} h^{(j)}(u; V_i)\right],$$

and its variance can be estimated by the summation of

$$N^{-1}\left[\sum_{k=1}^{K} p_k \mu_k^2(\phi_j) - \left\{\sum_{k=1}^{K} p_k \mu_k(\phi_j)\right\}^2\right] \tag{3.5}$$

and

$$\sum_{k=1}^{K}\left(\frac{N_k}{N}\right)^2\left[\left\{\mathrm{Var}_k(\phi_j) - 2\mathrm{Cov}_k(\phi_j, h^{(j)}) + \mathrm{Var}_k(h^{(j)})\right\}\left(\frac{1}{n_k} - \frac{1}{N_k}\right) + \mathrm{Var}_k(\phi_j)\left(\frac{1}{N_k}\right)\right], \tag{3.6}$$

where R_i is the subsample indicator, $\eta(\epsilon_i) = P(R_i = 1|\epsilon_i)$, $h^{(j)}(u; V_i) = E\{\phi_j^*(F_i)|V_i, \epsilon_i\}$, $p_k = P(\epsilon = k)$, $\phi_j = \phi_j^* - S^{(j)}(u)$, $\mu_k(\phi_j) = E[\phi_j\{F, S^{(j)}(u)\}|\epsilon = k]$, $\mu_k(h^{(j)}) = E[h^{(j)}(V)|\epsilon = k]$, $\mathrm{Var}_k(\phi_j) = \mathrm{Var}[\phi_j\{F, S^{(j)}(u)\}|\epsilon = k]$, $\mathrm{Var}_k(h^{(j)}) = \mathrm{Var}\{h^{(j)}(V)|\epsilon = k\}$, and $\mathrm{Cov}_k(\phi_j, h^{(j)}) = \mathrm{Cov}[\phi_j\{F, S^{(j)}(u)\}, h^{(j)}(V)|\epsilon = k]$.

As in the case of the main study, the numerator of the test statistic is given by

$$T_N = \int W_N(u)\left\{\frac{\mathrm{d}\hat{S}^{(0)}(u)}{\hat{S}^{(0)}(u)} - \frac{\mathrm{d}\hat{S}^{(1)}(u)}{\hat{S}^{(1)}(u)}\right\}.$$

Under the null hypothesis, we have

$$\begin{aligned} T_N = T_N^* &= \int \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d}\left\{ \hat{S}^{(0)}(u) - S^{(0)}(u) \right\} - \frac{W_N(u)\mathrm{d}S^{(0)}(u)}{\hat{S}^{(0)}(u)S^{(0)}(u)} \left\{ \hat{S}^{(0)}(u) - S^{(0)}(u) \right\} \right] \\ &\quad - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d}\left\{ \hat{S}^{(1)}(u) - S^{(1)}(u) \right\} - \frac{W_N(u)\mathrm{d}S^{(1)}(u)}{\hat{S}_1(u)S^{(1)}(u)} \left\{ \hat{S}^{(1)}(u) - S^{(1)}(u) \right\} \right]. \end{aligned}$$

Hence,

$$\begin{aligned} \hat{T}_N^* &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{R_i}{\eta(\epsilon_i)} \int \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d}\phi_0(u; F_i) - \frac{W_N(u)\mathrm{d}\hat{S}^{(0)}(u)}{\{\hat{S}^{(0)}(u)\}^2} \phi_0(u; F_i) \right] \right. \\ &\qquad - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d}\phi_1(u; F_i) - \frac{W_N(u)\mathrm{d}\hat{S}^{(1)}(u)}{\{\hat{S}^{(1)}(u)\}^2} \phi_1(u; F_i) \right] \\ &\quad - \frac{R_i - \eta(\epsilon_i)}{\eta(\epsilon_i)} \int \left[\frac{W_N(u)}{\hat{S}^{(0)}(u)} \mathrm{d}h^{(0)}(u; V_i) - \frac{W_N(u)\mathrm{d}\hat{S}^{(0)}(u)}{\{\hat{S}^{(0)}(u)\}^2} h^{(0)}(u; V_i) \right] \\ &\qquad \left. - \left[\frac{W_N(u)}{\hat{S}^{(1)}(u)} \mathrm{d}h^{(1)}(u; V_i) - \frac{W_N(u)\mathrm{d}\hat{S}^{(1)}(u)}{\{\hat{S}^{(1)}(u)\}^2} h^{(1)}(u; V_i) \right] \right). \end{aligned}$$

Notice that $\hat{T}_N^*$ is of the same form as the AIPWCC stratified estimator (2.7), its variance can also be estimated by adding the estimators of the variance of conditional expectation terms like (3.5) and the estimators of the expectation of conditional variance term like (3.6). The final log rank type test statistic in the stratified sampling framework is given by

$$G_{strat} = \frac{T_N}{\sqrt{\widehat{\mathrm{Var}}(\hat{T}_N^*)}}. \tag{3.7}$$

The discussion at the end of Section 3.2 also applies in the case stratified sampling: If all models are correctly specified, the test statistics follow a standard normal distribution under the null hypothesis and follow a noncentral normal distribution with variance 1 under the alternative hypothesis. Misspecification of the regression model $H^{(j)}(r, X)$ would result in a test statistic with variance less than 1; misspecification of the propensity model $\pi(X)$ and/or $K_c^{(j)}(r, X)$ would affect the variance of the test statistic in an unknown direction. Similar to the case of the main study, the nonparametric bootstrap procedure might provide a more accurate estimator of the asymptotic variance. One way to conduct the bootstrap sampling is to randomly sample with replacement the main study subjects and substudy subjects separately. That is, given a bootstrap number B, for each bootstrap replicate $b = 1, \ldots, B$, a total of $\sum_{i=1}^{N} R_i$ subjects are sampled with replacement among all subject with $R_i = 1$ and a total of $\sum_{i=1}^{N}(1 - R_i)$

subjects are sampled with replacement among all subject with $R_i = 0$. Combining the new N subjects sampled from the b-th bootstrap replicate, we compute $T^{(b)}_{N,strat}$. The bootstrap test statistics in stratified sampling take would take the form $G_{strat,bootstrap} = \frac{T_N}{\sqrt{\text{Var}\{T^{(b)}_{N,strat}\}}}$, where $\sqrt{\text{Var}\{T^{(b)}_{N,strat}\}}$ is the sample standard deviation of $T^{(b)}_{N,strat}, b = 1, \ldots, B$. As shown in the simulation study, the bootstrap version of test performs better compared to (3.7) in the case of misspecification. Also, appropriate truncation could result in more stable performance with possible slight efficiency loss.

3.4 Simulations

3.4.1 Simulation 1: Main Study Only

In this section, we conduct a simulation study on the performance of test statistics (3.4) under both the null hypothesis and alternative hypothesis. Different type of misspecification scenarios, including misspecification of propensity models and of regression models, are also being considered.

In this simulation, we generated 500 replications and within each replicate, $N = 2000$ observations are generated as follows: $B \sim Bernoulli(0.5), W \sim N(0,1)$ and $X_2 \sim N(0,1)$, mutually independent. We denote by $X_1 = (B, W, W^2)^T$, and $X = (X_1^T, X_2)^T$. The treatment assignment propensity is generated by $P(Z = 1|X) = \frac{\exp(0.1B+0.1W+0.5W^2+0.5X_2)}{1+\exp(0.1B+0.1W+0.5W^2+0.5X_2)}$, and the treatment-specific hazard functions are given by $\lambda_{Z=1}(t|X) = \lambda_{Z=0}(t|X) = \exp(-0.5 + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$. We also generated an independent censoring variable C as Uniform$(0, 4)$. This is a scenario under the null hypothesis and we expect the test statistics (3.4) to follow a standard normal distribution.

Table 3.1: Simulation 1, under the null hypothesis, the performance of test statistics (3.4) when there is no stratified sampling. Sample size N=2000. Correct model specification.

truncation	Pr(stat $\geq$ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| $\geq$ 1.96)
no	0.026	0.024	0.050
at 3.50	0.022	0.028	0.050
at 3.25	0.022	0.028	0.050
at 3.00	0.022	0.026	0.048
at 2.75	0.022	0.030	0.052
at 2.50	0.024	0.026	0.050

With all models, $\pi(X)$, $K_c^{(j)}(r, X)$ and $H^{(j)}(r, X)$, being correctly specified, the performance of the log rank type test statistics (3.4) is summarized in Table 3.1. It is expected that the test statistics follow the standard normal distribution and the type I error rate is indeed close to 0.05, with or without truncation. We also computed the ordinary log rank test and because of confounding, this test does not provide valid inference and rejects the null hypothesis with probability 1. The mean of ordinary log rank test is 7.07 and the standard deviation is 0.99.

Table 3.2: Simulation 1, under the null hypothesis, the performance of test statistics (3.4) when there is no stratified sampling. Sample size N=2000. Correct regression models and wrong propensity models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.024	0.022	0.046
at 3.50	0.022	0.026	0.048
at 3.25	0.024	0.024	0.048
at 3.00	0.024	0.020	0.044
at 2.75	0.024	0.024	0.048
at 2.50	0.028	0.016	0.044

Table 3.3: Simulation 1, under the null hypothesis, the performance of bootstrap test statistics $G_{bootstrap}$ when there is no stratified sampling. Sample size N=2000. Correct regression models and wrong propensity models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.020	0.032	0.052
at 3.50	0.024	0.040	0.064
at 3.25	0.026	0.040	0.066
at 3.00	0.026	0.034	0.060
at 2.75	0.024	0.030	0.054
at 2.50	0.026	0.028	0.054

If we misspecify the propensity model by leaving out the W^2 term in $\pi(X)$ and keeping the other two models $K_c^{(j)}(r, X)$ and $H^{(j)}(r, X)$ correctly specified, the performance of log rank type test statistics (3.4) is summarized in Table 3.2. Theoretically, in this scenario, the test

statistics should still follow the normal distribution with mean 0 but not necessarily a variance of one. In this scenario the variance estimator was close to one and hence the nominal level of significance was close to 0.05, however, this is not guaranteed, at least theoretically, to occur in general. Table 3.3 presents the bootstrap version of test statistics $G_{bootstrap}$ with $B = 100$ bootstrap resampling under wrong propensity model. We see after appropriate truncation, the bootstrap test provides nominal level of significance.

Table 3.4: Simulation 1, under the null hypothesis, the performance of test statistics (3.4) when there is no stratified sampling. Sample size N=2000. Correct propensity models and wrong regression models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.008	0.010	0.018
at 3.50	0.010	0.012	0.022
at 3.25	0.012	0.012	0.024
at 3.00	0.010	0.012	0.022
at 2.75	0.008	0.008	0.016
at 2.50	0.006	0.012	0.018

Table 3.5: Simulation 1, under the null hypothesis, the performance of Bootstrap test statistics $G_{boostrap}$ when there is no stratified sampling. Sample size N=2000. Correct propensity models and wrong regression models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.024	0.028	0.052
at 3.50	0.018	0.026	0.044
at 3.25	0.018	0.026	0.044
at 3.00	0.022	0.024	0.046
at 2.75	0.020	0.026	0.046
at 2.50	0.022	0.024	0.046

We also considered misspecification of the regression models by leaving out the W^2 term in $H^{(j)}(r, X)$ and keeping the propensity models $\pi(X)$ and $K_c^{(j)}(r, X)$ correctly specified, the performance of log rank type test statistics (3.4) is summarized in Table 3.4. Theoretically,

in this scenario, the test statistics should still follow the normal distribution with mean 0 and variance less than 1, which is indeed supported by our simulation results. The bootstrap version of test statistics $G_{bootstrap}$ with $B = 100$ bootstrap resampling under wrong regression model is presented in Table 3.5. It showed that, by bootstrap on the variance, the log rank test statistic $G_{bootstrap}$ provides nominal significance level close to 0.05, even when the regression model is misspecified.

Table 3.6: Simulation 1, under the alternative hypothesis, the power of test statistics (3.4) when there is no stratified sampling. Sample size N=2000. Correct model specification.

truncation	Power	
	$\beta_0^{(1)} = 0.6$	$\beta_0^{(1)} = 0.7$
no	0.354	0.894
at 3.50	0.372	0.902
at 3.25	0.358	0.900
at 3.00	0.360	0.900
at 2.75	0.352	0.892
at 2.50	0.332	0.884

Under the alternative hypothesis of $\lambda_{Z=1}(t|X) = \exp(-\beta_0^{(1)}+0.1B+0.1W+0.5W^2+0.5X_2)$, and $\lambda_{Z=0}(t|X) = \exp(-0.5 + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$, the power performance of the test statistics (3.4) is summarized in Table 3.6. For comparison, we fit a Cox proportional hazards model with covariate X and Z to the data and used the coefficient of Z divided by its estimated standard deviation as the gold standard test statistic. The power of this gold standard test statistic is 0.376 when $\beta_0^{(1)} = 0.6$ and 0.914 when $\beta_0^{(1)} = 0.7$. This indicates that (3.4) is comparable to the optimal model based gold standard test.

3.4.2 Simulation 2: With Substudy

In this section, we consider the performance of test statistics (3.7) under both null and alternative hypotheses. Similar to the previous study, we also compute the test under different misspecification scenarios, including misspecification of the propensity and regression models.

The data generating mechanism is the same as the previous studies, where $X_1 = (B, W, W^2)^T$ is collected on $N = 2000$ observations in the main study and X_2 is only observable for a random subsample of size $n_{11} = n_{12} = n_{01} = n_{02} = 300$, where n_{jk} denotes the number of subjects in stratum $Z = j$ and $B = k - 1$, for $j = 0, 1, k = 1, 2$. Besides (3.7), which will be referred to as

AIPWCC stratified test statistics, we compute similar test statistics with no augmentation terms. That is, we compute (3.7) by substituting $h^{(j)}(u;V) = 0$, $j = 0, 1$ in both the numerator and denominator. This estimator, referred to as weighted test statistics, only makes use of information collected on the main study and is hence expected to be less efficient than the AIPWCC stratified test statistics. Similar to $G_{strat,bootstrap}$, a bootstrap version of weighted test statistics, denoted by $G^{weighted}_{strat,bootstrap}$, could be computed by using the sample standard deviation as the denominator. The performance of $G_{strat,bootstrap}$ and $G^{weighted}_{strat,bootstrap}$ will be compared in the case of model misspecification.

Table 3.7: Simulation 2, under the null hypothesis, the performance of AIPWCC stratified test statistics and weighted test statistics with stratified sampling. In each cell, the left column corresponds to AIPWCC stratified test statistics and the right column corresponds to weighted test statistics. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct model specification.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.028—0.026	0.024—0.028	0.052—0.054
at 3.50	0.028—0.026	0.026—0.032	0.054—0.058
at 3.25	0.022—0.026	0.030—0.032	0.052—0.058
at 3.00	0.024—0.022	0.030—0.032	0.054—0.054
at 2.75	0.026—0.020	0.028—0.032	0.054—0.052
at 2.50	0.026—0.022	0.026—0.032	0.052—0.054

When all models, $\pi(X)$, $K_c^{(j)}(r, X)$ and $H^{(j)}(r, X)$, are correctly specified, the performance of AIPWCC stratified test statistics and weighted test statistics are summarized in Table 3.7. Both test statistics follows the standard normal distribution and the type I error rate is close to the nominal 0.05 level, regardless of the degree of truncation. Similar to the case of main study only, the ordinary log rank test statistic does not account for the confounding appropriately and leads to biased result with mean 7.00 and standard deviation 1.01.

When the propensity model is misspecified by leaving out the W^2 term in $\pi(X)$ and the other two models $K_c^{(j)}(r, X)$ and $H^{(j)}(r, X)$ are correctly specified, the performance of the AIPWCC stratified test statistics and weighted test statistics are summarized in Table 3.8. Theoretically, in this scenario, the test statistics should still follow a normal distribution with mean 0 but with variance that may not equal one; however, our simulation results show that the variance of both test statistics are close to 1 in this setting leading to nominal significance levels close to 0.05. Table 3.9 presents the result of the bootstrap test statistics $G_{strat,bootstrap}$ and $G^{weighted}_{strat,bootstrap}$. Both provide valid inference with nominal significance level.

Table 3.8: Simulation 2, under the null hypothesis, the performance of AIPWCC stratified test statistics and weighted test statistics with stratified sampling. In each cell, the left column corresponds to AIPWCC stratified test statistics and the right column corresponds to weighted test statistics. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct regression models and wrong propensity models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.026—0.026	0.016—0.032	0.042—0.058
at 3.50	0.034—0.028	0.020—0.032	0.054—0.060
at 3.25	0.034—0.026	0.020—0.030	0.054—0.056
at 3.00	0.034—0.022	0.020—0.030	0.054—0.052
at 2.75	0.032—0.022	0.020—0.030	0.052—0.052
at 2.50	0.026—0.022	0.024—0.028	0.050—0.050

If we misspecify the regression models by leaving out the W^2 term in $H^{(j)}(r, X)$ and keep the propensity models $\pi(X)$ and $K_c^{(j)}(r, X)$ correctly specified, the performance of the AIPWCC stratified test statistics and weighted test statistics are summarized in Table 3.10. As showed, in this scenario, both test statistics should still follow the normal distribution with mean 0 and variance less than 1. In particular, the variance of the AIPWCC stratified test statistics are larger than the weighted test statistics. This indeed is what we would expect. The bootstrap test statistics $G_{strat,bootstrap}$ and $G^{weighted}_{strat,bootstrap}$ are summarized in Table 3.11. Compared with Table 3.10, we see the bootstrap version of tests are more accurate. Hence, it would be suggested to perform the bootstrap test when there exists the possibility of misspecification.

Under the alternative hypothesis of $\lambda_{Z=1}(t|X) = \exp(-\beta_0^{(1)} + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$, and $\lambda_{Z=0}(t|X) = \exp(-0.5 + 0.1B + 0.1W + 0.5W^2 + 0.5X_2)$, the power performance of the test statistics (3.7) is summarized in Table 3.12. As expected, the AIPWCC stratified test statistics are more powerful than weighted test statistics. Again, we computed the gold standard test by fitting a proportional hazard Cox model with covariate X and treatment assignment Z on the subsample observations. We use the coefficient of Z divided by its estimated standard deviation as the gold standard test statistic. Note that for those subject not selected in the subsample, their information will not be used to compute the gold standard test statistics. The power of gold standard test statistics is 0.282 when $\beta_0^{(1)} = 0.6$ and 0.780 when $\beta_0^{(1)} = 0.7$. The simulation results indicate that the AIPWCC stratified test statistics are more powerful than the gold standard test. The efficiency gain comes from the use of the main study observations that are not selected in the subsample.

Table 3.9: Simulation 2, under the null hypothesis, the performance of AIPWCC bootstrap stratified test statistics $G_{strat,bootstrap}$ and bootstrap weighted test statistics $G^{weighted}_{strat,bootstrap}$ with stratified sampling. In each cell, the left column corresponds to $G_{strat,bootstrap}$ and the right column corresponds to $G^{weighted}_{strat,bootstrap}$. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct regression models and wrong propensity models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.014—0.024	0.020—0.028	0.034—0.052
at 3.50	0.026—0.022	0.024—0.034	0.050—0.056
at 3.25	0.030—0.024	0.022—0.032	0.052—0.056
at 3.00	0.028—0.030	0.022—0.032	0.050—0.062
at 2.75	0.032—0.028	0.024—0.032	0.056—0.060
at 2.50	0.030—0.022	0.022—0.034	0.052—0.056

3.5 Analysis on ASCERT Data

In this section, we apply the proposed log rank test statistics to data from the ASCERT study. The goal is to test for the null hypothesis that there is no significant difference between PCI and CABG using baseline covariate data from the main ASCERT study database augmented with information on coronary anatomy from a subsample of patients in the ASCERT angiographic companion study. The main study consisted of records from $9,800$ patients in the ASCERT database who underwent a coronary revascularization procedure (CABG or PCI) at one of 54 hospitals participating in the ASCERT study that agreed to be part of the companion study. Subsequently it was determined that some patients were ineligible for analysis. Consequently, the main study participants consisted of $7,393$ eligible patients among the $9,800$ patients in the 54 hospitals. For the purpose of this analysis we considered the patients from the 54 hospitals of the companion study to be the main focus of inference. Twenty-eight covariates were used on the full sample which included demographics (e.g., age, sex), risk factors (e.g., body mass index, smoking), symptoms and history of cardiovascular disease (e.g., chest pain, congestive heart failure), and comorbidities (e.g., diabetes). The subsample includes records from approximately $2,000$ patients chosen by design (roughly 500 in each of the four strata determined by all combinations of the two treatments and whether or not the patients had two- or three-vessel disease). After consideration of eligibility issue, $1,554$ eligible patients remained in the substudy for analysis. Information collected on the subsample includes features of the patient's coronary anatomy (e.g., left-side dominance) and features of each individual blockage (e.g., lesion length, tortuosity, calcification, degree of stenosis).

We are interested in testing the treatment effect between PCI and CABG up to time point 1

Table 3.10: Simulation 2, under the null hypothesis, the performance of AIPWCC stratified test statistics and weighted test statistics with stratified sampling. In each cell, the left column corresponds to AIPWCC stratified test statistics and the right column corresponds to weighted test statistics. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct propensity models and wrong regression models.

truncation	$\Pr(\text{stat} \geq 1.96)$	$\Pr(\text{stat} \leq -1.96)$	$\Pr(\lvert\text{stat}\rvert \geq 1.96)$
no	0.022—0.016	0.012—0.010	0.034—0.026
at 3.50	0.026—0.014	0.014—0.010	0.040—0.024
at 3.25	0.026—0.014	0.016—0.012	0.042—0.026
at 3.00	0.026—0.014	0.014—0.010	0.040—0.024
at 2.75	0.028—0.016	0.014—0.010	0.042—0.026
at 2.50	0.022—0.010	0.012—0.010	0.034—0.020

month and 4 years. Figure 3.1 displays the estimated treatment-specific curves using AIPWCC estimator on the main study, the stratified AIPWCC estimator with $h = 0$ and with $h = h^{opt}$, respectively. The computed curves are quite similar: PCI has higher short-term survival probability while CABG has higher long-term survival probability. This result suggests the confounding of additional covariate collected on the substudy is not very influential.

Table 3.13 shows the log-rank test statistics on the main study, weighted test statistics, AIPWCC stratified test statistics and ordinary log rank test statistics truncated at 1 month and 4 years. Consistent with the treatment-specific curves, log rank test statistics in the main study suggests PCI is significantly better than CABG up to 1 month and contrary at 4 years. The weighted test statistics only use data from subsample and hence lack power. Meanwhile, the AIPWCC stratified test statistics is able to recover more information by making use of subjects in the main study. The ordinary log rank test suggests that there is not much confounding due to main study covariates either.

3.6 Discussion

In this chapter, we used semiparametric theory to develop log rank type statistics in observational survival studies. When there exists confounding covariates in the main study, the developed test statistics is efficient and doubly-robust. If the confounding is negligible, the proposed test does not lose too much power compared to the ordinary log rank test. When the issue of confounding is still in question within the main study, a substudy could be conducted to collect additional potential confounding variables. We proposed AIPWCC stratified test statistics in such a scenario, which maintains the doubly-robust property. If in fact the additional

Table 3.11: Simulation 2, under the null hypothesis, the performance of AIPWCC bootstrap stratified test statistics $G_{strat,bootstrap}$ and bootstrap weighted test statistics $G^{weighted}_{strat,bootstrap}$ with stratified sampling. In each cell, the left column corresponds to $G_{strat,bootstrap}$ and the right column corresponds to $G^{weighted}_{strat,bootstrap}$. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct propensity models and wrong regression models.

truncation	Pr(stat ≥ 1.96)	Pr(stat ≤ -1.96)	Pr(|stat| ≥ 1.96)
no	0.022—0.020	0.014—0.018	0.036—0.038
at 3.50	0.030—0.022	0.016—0.020	0.046—0.042
at 3.25	0.032—0.020	0.016—0.020	0.048—0.040
at 3.00	0.032—0.022	0.014—0.018	0.046—0.040
at 2.75	0.032—0.022	0.014—0.016	0.046—0.040
at 2.50	0.030—0.024	0.010—0.016	0.040—0.040

covariates are not confounders, the AIPWCC stratified test statistics are able to recover much of the information from the main study subjects and resulting in good power.

Table 3.12: Simulation 2, under the alternative hypothesis, the power of AIPWCC stratified test statistics and weighted test statistics with stratified sampling. In each cell, the left column corresponds to AIPWCC stratified test statistics and the right column corresponds to weighted test statistics. Sample size N=2000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$. Correct models specification.

truncation	Power	
	$\beta_0^{(1)} = 0.6$	$\beta_0^{(1)} = 0.7$
no	0.294 – 0.250	0.748 – 0.714
at 3.50	0.320 – 0.258	0.864 – 0.734
at 3.25	0.326 – 0.254	0.868 – 0.740
at 3.00	0.332 – 0.258	0.874 – 0.736
at 2.75	0.326 – 0.256	0.874 – 0.732
at 2.50	0.328 – 0.260	0.868 – 0.720

Table 3.13: Log rank tests applied on the ASCERT data, including log rank test statistics on main study, weighted test statistics, AIPWCC stratified test statistics and ordinary log-rank test statistics. All tests are truncated at time point of interest: 1 month and 4 years.

Tuncation	Main Study	Weighted	AIPWCC Stratified	Ordinary
1 month	2.87	0.39	2.47	2.31
4 years	-2.34	-0.66	-1.91	-3.97

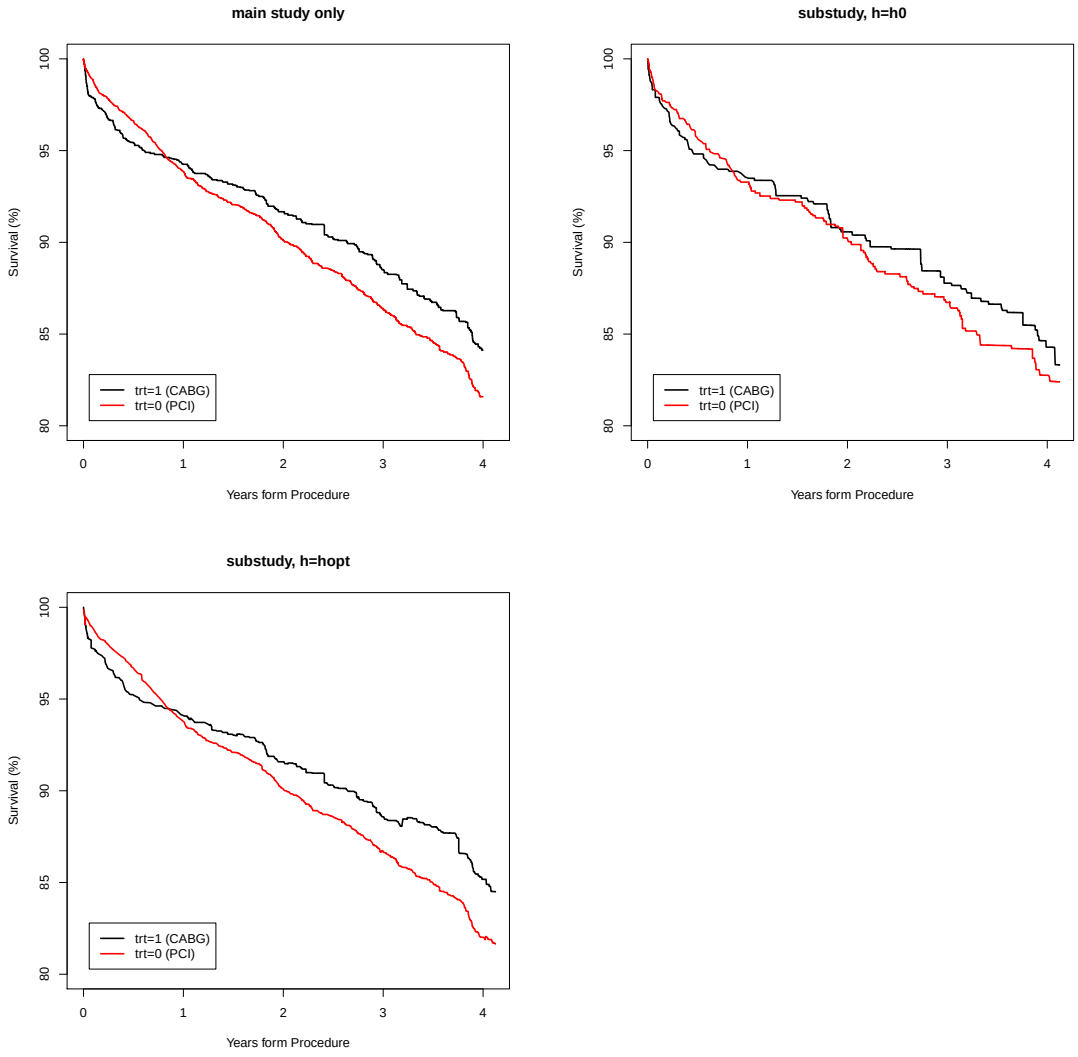

Figure 3.1: Estimated AIPWCC and AIPWCC Stratified Survival Estimator for CABG and PCI.

REFERENCES

Breslow, N. and Cain, K.C. (1988). Logistic regression for two-stage case-control data. *Biometrika* **75**, 11–20.

Cole, S.R. and Hernán, M.A. (2004). Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine* **75**, 45-49.

Hubbard, A.E., van der Laan, M.J., and Robins, J.M. (1999). Nonparametric locally efficient estimation of the treatment specific survival distributions with right censored data and covariates in observational studies. In *Statistical Models in Epidemiology: The Environment and Clinical Trials* (E. Halloran, D. Berry, eds). New York: Springer, pp. 134–178.

Flanders, W.D. and Greenland, S. (1991). Analytic methods for two-stage case control studies and other stratified designs. *Statistics in Medicine* **10**, 739–747.

Kaplan, E.L. and Meier, Paul. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

Lee, J., Yoshizawa, C. Wilkens, L. and Lee, H.P. (1992). Covariance adjustment of survival curves based on Cox's proportional hazards regression model. *Computer Applications in the Biosciences* **8**, 23-27.

Makuch, R.W. (1982). Adjusted survival-curve estimation using covariates. *Journal of Chronic Diseases* **35**, 437–443.

Mark, S.D. and Katki, H.A. (2006). Specifying and implementing nonparametric and semiparametric survival estimators in two-stage (nested) cohort studies with missing case data. *Journal of the American Statistical Association* **101**, 460–471.

Robins, J.M., and Rotnitzky, A. (1992). Recovery information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology, Methodological Issues* (N. Jewell, K. Dietz, V. Farewell, eds.). Boston: Birkhauser, pp. 297–331.

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

Stürmer, T., Glynn, R.J., Rothman, K.J., Avorn, J., and Schneeweiss, S. (2007). Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Medical Care* **45**, S157–S165.

Tsiatis, A.A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Weintraub, W.S., Grau-Sepulveda, M.V., Weiss, J.M., O'Brien, S.M, Peterson, E.D., Kolm, P., Zhang, Z., Klein, L.W., Shaw, R.E., McKay, C., Ritzenthaler, L.L., Popma, J.J., Messenger, J.C., Shahian, D.M., Grover, F.L., Mayer, J.E., Shewan, C.M., Garratt, K.N., Moussa, I.D., Dangas, G.D., and Edwards, F.H. (2012). Comparative effectiveness of revascularization strategies. *New England Journal of Medicine* **366**, 1467–1476.

Xie, J. and Liu, C. (2005). Adjusted Kaplan-Meier Estimator and Log-rank Test with Inverse Probability of Treatment Weighting for Survival Data. *Statistics in Medicine* **24**,3089–3110.

Zhang, M. and Schaubel, D.E. (2012). Contrasting treatment-specific survival using double-robust estimators. *Statistics in Medicine* **31**,4255–4268.

Zhao, L.P. and Lipsitz, S. (1992). Design and analysis of two-stage studies. *Statistics in Medicine* **11**, 769–782.

APPENDICES

Appendix A

Simplification of the Estimating Function

This appendix demonstrates how the estimating function is simplified to equation (1.3) of the main paper in this section. For simplicity, we suppress the superscript (1) of $S^{(1)}(u), K_c^{(1)}(u,X)$, $H^{(1)}(u,X)$ and $M_c^{(1)}(u,X)$. Hence, the estimating function we want to simplify is

$$\frac{Z\Delta}{\pi(X)K_c(U,X)}\left\{I(U\geq u)-S^{(1)}(u)\right\} \tag{A.1}$$

$$-\left\{\frac{Z-\pi(X)}{\pi(X)}\right\}\left\{H(u,X)-S^{(1)}(u)\right\} \tag{A.2}$$

$$+\int_0^\infty \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)}\left[I(r<u)\left\{\frac{H(u,X)}{H(r,X)}-S^{(1)}(u)\right\}+I(r\geq u)\left\{1-S^{(1)}(u)\right\}\right]. \tag{A.3}$$

Note that (A.3) can be expressed as

$$\int_0^u \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)}\left\{\frac{H(u,X)}{H(r,X)}\right\} \tag{A.4}$$

$$+\int_u^\infty \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)} \tag{A.5}$$

$$-S(u)\int_0^\infty \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)}. \tag{A.6}$$

We first compute (A.6)=

$$-\frac{S(u)Z}{\pi(X)}\left\{\int_0^\infty \frac{\mathrm{d}N_c(r,X)}{K_c(r,X)}-\int_0^\infty \frac{\lambda_c(r,X)I(U\geq r)}{K_c(r,X)}\right\}. \tag{A.7}$$

Because $\lambda(r,X) = \frac{-\mathrm{d}\log K_c(r,X)}{\mathrm{d}r}$, we obtain $\int \frac{\lambda_c(r,X)}{K_c(r,X)} = \frac{1}{K_c(r,X)}$. Hence (A.7) equals

$$-\frac{S(u)Z}{\pi(X)}\left\{\frac{1-\Delta}{K_c(U,X)} - \frac{1}{K_c(r,X)}\Big|_0^U\right\} = -\frac{S(u)Z}{\pi(X)}\left\{1 - \frac{\Delta}{K_c(U,X)}\right\}. \tag{A.8}$$

We now compute (A.5):

$$\begin{aligned}\int_u^\infty \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)} &= \frac{Z}{\pi(X)}\left\{\int_u^\infty \frac{\mathrm{d}N_c(r,X)}{K_c(r,X)} - \int_u^\infty \frac{\lambda_c(r,X)I(U\geq r)}{K_c(r,X)}\right\}\\ &= \frac{Z}{\pi(X)}I(U\geq u)\left\{\frac{1-\Delta}{K_c(U,X)} - \int_u^U \frac{\lambda_c(r,X)}{K_c(r,X)}\right\}\\ &= \frac{Z}{\pi(X)}I(U\geq u)\left[\frac{1}{K_c(u,X)} - \frac{\Delta}{K_c(U,X)}\right]. \end{aligned}\tag{A.9}$$

Combining (A.1) with (A.6)=(A.8) and (A.5)=(A.9), we obtain

$$\begin{aligned}&\frac{Z\Delta I(U\geq u)}{\pi(X)K_c(U,X)} - \frac{Z\Delta S(u)}{\pi(X)K_c(U,X)} + \frac{ZI(U\geq u)}{\pi(X)K_c(u,X)} - \frac{Z\Delta I(U\geq u)}{\pi(X)K_c(U,X)} - \frac{ZS(u)}{\pi(X)} + \frac{Z\Delta S(u)}{\pi(X)K_c(U,X)}\\ &\qquad = \frac{ZI(U\geq u)}{\pi(X)K_c(u,X)} - \frac{ZS(u)}{\pi(X)}.\end{aligned}\tag{A.10}$$

So the estimating function now becomes (A.10)+(A.2)+(A.4)

$$\frac{ZI(U\geq u)}{\pi(X)K_c(u,X)} - \frac{ZS(u)}{\pi(X)} - \left\{\frac{Z-\pi(X)}{\pi(X)}\right\}\{H(u,X)-S(u)\} + \int_0^u \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)}\left\{\frac{H(u,X)}{H(r,X)}\right\}.$$

Also note that $-S(u)\frac{Z}{\pi(X)} + S(u)\frac{Z-\pi(X)}{\pi(X)} = -S(u)$. Therefore, the estimating function is simplified to be

$$\frac{ZI(U\geq u)}{\pi(X)K_c(u,X)} - \left\{\frac{Z-\pi(X)}{\pi(X)}\right\}H(u,X) + \int_0^u \frac{Z}{\pi(X)}\frac{\mathrm{d}M_c(r,X)}{K_c(r,X)}\left\{\frac{H(u,X)}{H(r,X)}\right\} - S(u). \tag{A.11}$$

This is identical to the simplified estimating equation (1.3).

Appendix B

Demonstration of the Double Robustness Property

This appendix demonstrates the double robustness property of estimator (1.4) of the main paper. By analogy, it is straightforward to show double robustness of $\hat{S}^{(0)}(u)$ and $\hat{S}_{diff}(u)$. Again, we suppress the superscript (1) of $K_c^{(1)}(u,X), H^{(1)}(u,X)$ and $M_c^{(1)}(u,X)$. Suppose that our estimator for $\pi(X)$ converges to $\pi^*(X)$, and the estimator for $K_c(u,X)$ converges to $K_c^*(u,X)$ and the estimator for $H(u,X)$ converges to $H^*(u,X)$. In that case the estimator for $S(u)$ will be consistent if the estimating function (A.11) has mean zero when we use $\pi^*(X)$, $K_c^*(u,X)$ and $H^*(u,X)$ substituted for $\pi(X)$, $K_c(u,X)$ and $H(u,X)$ respectively.

Toward that end, consider the integral

$$\int_0^u \frac{Z}{\pi^*(X)} \frac{\mathrm{d}M_c^*(r,X)}{K_c^*(r,X)} = \int_0^u \frac{Z}{\pi^*(X)} \frac{I(T^{(1)} \geq r)\left\{\mathrm{d}\widetilde{N}(r) - \lambda_c^*(r,X)I(C \geq r)\right\}}{K_c^*(r,X)},$$

where $\widetilde{N}(r) = I(C \leq r)$. Now let us compute

$$\begin{aligned}
&\int_0^u \frac{\mathrm{d}\widetilde{N}(r) - \lambda_c^*(r,X)I(C \geq r)}{K_c^*(r,X)} \\
&= \frac{I(C \leq u)}{K_c^*(C,X)} - \int_0^{\min(u,C)} \frac{\lambda_c^*(r,X)}{K_c^*(r,X)} \\
&= \frac{I(C \leq u)}{K_c^*(C,X)} - I(C \leq u)\left\{\frac{1}{K_c^*(C,X)} - 1\right\} - I(C \geq u)\left\{\frac{1}{K_c^*(u,X)} - 1\right\} \\
&= 1 - \frac{I(C \geq u)}{K_c^*(u,X)}.
\end{aligned}$$

Hence we have shown that $\frac{I(C \geq u)}{K_c^*(u,X)} = 1 - \int_0^u \frac{\mathrm{d}\widetilde{N}(r) - \lambda_c^*(r,X)I(C \geq r)}{K_c^*(r,X)}$. Going back to the estimating

function (1.3), note that

$$\begin{aligned}\frac{ZI(U \geq u)}{\pi^*(X)K_c^*(u,X)} &= \frac{ZI(T \geq u)I(C \geq u)}{\pi^*(X)K_c^*(u,X)} \\ &= \frac{ZI(T^{(1)} \geq u)}{\pi^*(X)} \left\{1 - \int_0^u \frac{\mathrm{d}\widetilde{M}^*(r)}{K_c^*(r,X)}\right\} \\ &= \frac{ZI(T^{(1)} \geq u)}{\pi^*(X)} - \frac{Z}{\pi^*(X)} \int_0^u \frac{I(T^{(1)} \geq u)\mathrm{d}\widetilde{M}^*(r)}{K_c^*(r,X)}. \end{aligned} \tag{B.1}$$

Combining (B.1) with the second term of (A.11) and $-S(u)$ we obtain

$$\begin{aligned}&\frac{ZI(T^{(1)} \geq u)}{\pi^*(X)} - \frac{Z}{\pi^*(X)} \int_0^u \frac{I(T^{(1)} \geq u)\mathrm{d}\widetilde{M}^*(r)}{K_c^*(r,X)} - \left\{\frac{Z - \pi^*(X)}{\pi^*(X)}\right\} H^*(u,X) - S(u) \\ &= I(T^{(1)} \geq u) + \left\{\frac{Z - \pi(X)}{\pi(X)}\right\} I(T^{(1)} \geq u) - \frac{Z}{\pi^*(X)} \int_0^u \frac{I(T^{(1)} \geq u)\mathrm{d}\widetilde{M}^*(r)}{K_c^*(r,X)} \\ &\quad - \left\{\frac{Z - \pi^*(X)}{\pi^*(X)}\right\} H^*(u,X) - S(u) \\ &= \Big\{I(T^{(1)} \geq u) - S(u)\Big\} + \left\{\frac{Z - \pi^*(X)}{\pi^*(X)}\right\} \Big\{I(T^{(1)} \geq u) - H^*(u,X)\Big\} \\ &\quad - \frac{Z}{\pi^*(X)} \int_0^u \frac{I(T^{(1)} \geq u)\mathrm{d}\widetilde{M}^*(r)}{K_c^*(r,X)}. \end{aligned} \tag{B.2}$$

Now combining the last term of (B.2) with the third term of (A.11), we obtain

$$\frac{Z}{\pi^*(X)} \int_0^u \frac{\mathrm{d}\widetilde{M}^*(r,X)}{K_c^*(r,X)} \left\{\frac{I(T^{(1)} \geq r)H^*(u,X)}{H^*(r,X)} - I(T^{(1)} \geq u)\right\}.$$

Thus the estimating function equals

$$I(T^{(1)} \geq u) - S(u) \tag{B.3}$$

$$+ \left\{\frac{Z - \pi^*(X)}{\pi^*(X)}\right\} \Big\{I(T^{(1)} \geq u) - H^*(u,X)\Big\} \tag{B.4}$$

$$+ \frac{Z}{\pi^*(X)} \int_0^u \frac{\mathrm{d}\widetilde{M}^*(r,X)}{K_c^*(r,X)} \left\{\frac{I(T^{(1)} \geq r)H^*(u,X)}{H^*(r,X)} - I(T^{(1)} \geq u)\right\}. \tag{B.5}$$

Note that E(B.3)$= 0$ by definition of $S(u) = P(T^{(1)} \geq u)$. Because $Z \perp\!\!\!\perp T^{(1)}|X$, E(B.4)$= 0$ if either $\pi^*(X) = P(Z = 1|X)$ or $H^*(u, X_i) = P(T^{(1)} \geq u|X_i)$. If $P(C \geq u|Z = 1, X_i) = P(C \geq u|Z = 1, T^{(1)}, X_i) = K_c^*(u, X_i)$, then $\mathrm{d}\widetilde{M}(r,X) = \mathrm{d}\widetilde{N}(r) - \lambda^*(r,X)I(C \geq r)$ is a martingale increment and hence (B.5) has mean zero. If $H^*(u, X_i) = P(T^{(1)} \geq u|X_i)$, then E(B.5) can be written as $\int E\left[\frac{Z}{\pi^*(X)} \frac{\mathrm{d}\widetilde{M}^*(r,X)}{K_c^*(r,X)} \left\{\frac{I(T^{(1)} \geq r)H^*(u,X)}{H^*(r,X)} - I(T^{(1)} \geq u)\right\}\right]$. Conditioning on C, Z and X,

it equals

$$\int E\left(\frac{Z}{\pi^*(X)}\frac{\mathrm{d}\widetilde{M}^*(r,X)}{K_c^*(r,X)}E\left[\left\{\frac{I(T^{(1)}\geq r)H^*(u,X)}{H^*(r,X)}-I(T^{(1)}\geq u)\right\}|C,Z,X\right]\right).$$

By assumption (1.1),

$$\begin{aligned}
&E\left[\left\{\frac{I(T^{(1)}\geq r)H^*(u,X)}{H^*(r,X)}-I(T^{(1)}\geq u)\right\}|C,Z,X\right]\\
&=\frac{P(T^{(1)}\geq r|X)H(u,X)}{H(r,X)}-P(T^{(1)}\geq u|X)\\
&=\frac{H(r,X)H(u,X)}{H(r,X)}-H(u,X)=0.
\end{aligned}$$

Hence we have a double robustness property in that our estimator will be consistent if the estimator for $P(Z=1|X)=\pi(X)$ and for $P(C\geq u|Z=1,X)=K_c(u,X)$ are both consistent or the estimator for $P(T^{(1)}\geq u|X)=H(u,X)$ is consistent.

Appendix C

Simulation Study with Misspecified Censoring Distribution

In the simulation of main paper (Chapter 1 and Chapter 2), the censoring variable follows the uniform distribution hence free from potential misspecification. In this appendix, we consider the case when the distribution of the censoring variable depends on the covariates and check the double robustness property of AIPWCC estimator (4). The data generating mechanism for $(X, Z, T^{(1)}, T^{(0)})$ is exactly same as in Section 4, with the censoring variable C having hazard function $\lambda_C(t|X) = \exp(0.1B + 0.1W - 0.5W^2 + 0.5X_2)$. A total of 500 replicates are generated, and the sample size within each replicate is $N = 2,000$ observations.

Similar to the main paper, four more scenarios are considered: scenario 5 estimates $\pi(X)$ and $H^{(1)}(r, X)$ consistently using X to fit the models, and estimates $K_c^{(1)}(r, X)$ inconsistently, using only $(B, W, X_2)^T$ to fit the model. Scenario 6 is the "wrong π, wrong K, correct H" case, when $(B, W, X_2)^T$ is used to fit $\pi(X)$ and $K_c^{(1)}(r, X)$, and X is used to fit $H^{(1)}(r, X)$. In scenario 7, $(B, W, X_2)^T$ is used to fit $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$, while X is used to fit $\pi(X)$. And finally, only $(B, W, X_2)^T$ is used to fit $\pi(X)$, $K_c^{(1)}(r, X)$ and $H^{(1)}(r, X)$ in scenario 8.

Table C.1 provides the summary results at several time points. In scenario 5 and scenario 6, the regression function $H^{(1)}(r, X)$ is correctly estimated in both cases. The resulting estimator $\hat{S}^{(1)}(u)$ is indeed consistent illustrating the doubly-robust property. Similar to the scenario 2 in the main paper, the standard error estimator is biased and hence the coverage rate of the computed confidence interval is far from the nominal 0.95. In scenario 7 and scenario 8, neither the estimator for $\pi(X)$ and $K_c^{(1)}(r, X)$ are both consistent, nor the estimator for $H^{(1)}(r, X)$ is consistent, the double robustness property no longer holds and we obtain biased estimators $\hat{S}^{(1)}(u)$ as expected.

Table C.1: Simulation 1 with misspecified censoring distribution, AIPWCC estimator for $\hat{S}^{(1)}(u)$ at selected time points when there is no stratified sampling. Sample size N=2000. Both scenario 5 and scenario 6 corresponds to"wrong propensity, correct regression" case; both scenario 7 and scenario 8 corresponds to"wrong propensity, wrong regression" case.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate
Scenario 5	0.5	0.693	0.694	0.001	0.014	0.014	0.96
	1.0	0.523	0.525	0.002	0.018	0.017	0.96
	1.5	0.408	0.409	0.001	0.021	0.019	0.94
	2.0	0.324	0.324	0.000	0.024	0.021	0.92
	2.5	0.262	0.262	0.000	0.025	0.022	0.90
	3.0	0.214	0.214	-0.001	0.026	0.024	0.92
Scenario 6	0.5	0.693	0.694	0.001	0.014	0.014	0.97
	1.0	0.523	0.525	0.001	0.017	0.017	0.95
	1.5	0.408	0.409	0.001	0.021	0.019	0.92
	2.0	0.324	0.324	0.000	0.023	0.020	0.91
	2.5	0.262	0.262	0.000	0.024	0.021	0.89
	3.0	0.210	0.214	-0.001	0.025	0.023	0.92
Scenario 7	0.5	0.693	0.683	-0.010	0.014	0.014	0.89
	1.0	0.523	0.503	-0.020	0.018	0.018	0.80
	1.5	0.408	0.381	-0.026	0.021	0.020	0.72
	2.0	0.324	0.294	-0.030	0.023	0.021	0.67
	2.5	0.262	0.231	-0.031	0.024	0.023	0.69
	3.0	0.214	0.183	-0.031	0.025	0.024	0.69
Scenario 8	0.5	0.693	0.649	-0.044	0.015	0.015	0.17
	1.0	0.523	0.466	-0.058	0.018	0.018	0.12
	1.5	0.408	0.345	-0.062	0.020	0.020	0.13
	2.0	0.324	0.261	-0.063	0.022	0.021	0.14
	2.5	0.262	0.202	-0.060	0.022	0.022	0.21
	3.0	0.214	0.158	-0.056	0.023	0.023	0.30

Appendix D

Computation with Competing Methods

In this Appendix, we study the performance of three competing estimators: the ordinary inverse probability weighted (IPW) estimator, the outcome regression (OR) estimator and the unadjusted treatment-specific Kaplan Meier estimator in the case of no stratified sampling.

The IPW estimator takes the form

$$\hat{S}^{(1)}_{IPW}(u) = N^{-1} \sum_{i=1}^{N} \frac{Z_i I(U_i \geq u)}{\pi(X_i) K_c^{(1)}(u, X_i)},$$

which is essentially the first part of estimator (4). As before, the functions $\pi(X)$ and $K_c^{(1)}(u, X)$ will be estimated using a logistic regression model and treatment specific proportional hazards regression model, respectively.

For the OR estimator, we propose to fit a treatment-specific proportional hazards regression model using the data $\{(U_i, \Delta_i, X_i), i : Z_i = 1\}$ to estimate $H^{(1)}(u, X) = P(T^{(1)} \geq u|X)$. Then the OR estimator takes the form:

$$\hat{S}^{(1)}_{OR}(u) = N^{-1} \sum_{i=1}^{N} \hat{H}^{(1)}(u, X_i).$$

We used the same data as generated in the first set of simulations, and computed the IPW and OR estimator with both correctly and incorrectly specified models. To be more specific, in the "IPW (correct)" case, X is used to fit both $\pi(X)$ and $K_c^{(1)}(u, X)$. In the "IPW (incorrect)" case, only (B, W, X_2) is used to fit both $\pi(X)$ and $K_c^{(1)}(u, X)$. Similarly, $H^{(1)}(u, X)$ is fitted by using X in the "OR (correct)" case and by using (B, W, X_2) in the "OR (incorrect)" case. The biasedness and standard error estimates are shown in Table D.1.

Both the IPW and the OR estimator are consistent if the corresponding models are correctly specified. When compared with our proposed AIPWCC estimator (4), the OR estimator is more efficient, while the IPW estimator is less efficient. If the models are not consistently specified, however, both IPW and OR estimator will be severely biased.

Also as shown in Table D.1, the unadjusted treatment-specific Kaplan Meier estimator is biased because it failed to adjust for confounding.

Table D.1: Simulation 1, no stratified sampling, sample size N=2000.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}(u)$	Bias	Monte-Carlo SE	Bootstrap SE	Coverage Rate
IPW (Correct)	0.5	0.693	0.694	0.001	0.013	0.014	0.96
	1.0	0.523	0.525	0.002	0.015	0.016	0.95
	1.5	0.408	0.408	0.001	0.017	0.017	0.95
	2.0	0.324	0.326	0.001	0.017	0.017	0.95
	2.5	0.262	0.263	0.001	0.017	0.018	0.95
	3.0	0.214	0.216	0.002	0.018	0.019	0.96
IPW (Incorrect)	0.5	0.693	0.662	-0.031	0.014	0.014	0.38
	1.0	0.523	0.490	-0.034	0.015	0.016	0.39
	1.5	0.408	0.37477	-0.033	0.016	0.016	0.45
	2.0	0.324	0.295	-0.029	0.016	0.016	0.54
	2.5	0.262	0.235	-0.027	0.016	0.016	0.63
	3.0	0.214	0.192	-0.022	0.016	0.017	0.76
OR (Correct)	0.5	0.693	0.694	0.000	0.012	0.013	0.95
	1.0	0.523	0.524	0.001	0.015	0.015	0.94
	1.5	0.408	0.408	0.000	0.016	0.015	0.94
	2.0	0.324	0.325	0.001	0.016	0.016	0.95
	2.5	0.262	0.262	0.000	0.016	0.016	0.95
	3.0	0.214	0.215	0.001	0.016	0.017	0.95
OR (Incorrect)	0.5	0.694	0.662	-0.031	0.013	0.014	0.36
	1.0	0.523	0.490	-0.033	0.015	0.015	0.38
	1.5	0.408	0.375	-0.032	0.016	0.016	0.45
	2.0	0.324	0.295	-0.029	0.01547	0.016	0.52
	2.5	0.262	0.236	-0.026	0.016	0.016	0.62
	3.0	0.214	0.192	-0.022	0.016	0.016	0.73
Kaplan Meier	0.5	0.693	0.644	-0.050	0.014	0.014	0.06
	1.0	0.523	0.467	-0.057	0.015	0.015	0.04
	1.5	0.408	0.351	-0.057	0.016	0.015	0.05
	2.0	0.324	0.272	-0.053	0.015	0.015	0.09
	2.5	0.262	0.213	-0.049	0.015	0.015	0.11
	3.0	0.214	0.171	-0.043	0.015	0.015	0.21

Appendix E

Derivation of Equation (2.4) in Chapter 2

In this appendix, we derive the equation (2.4) in Chapter 2

$$
\begin{aligned}
& \mathrm{Var}[T\{\mathbf{O}, S^{(1)}(u)\}|N_1, \ldots, N_K] \\
= & \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \mathrm{Var}\left(\frac{\sum_{i:\epsilon_i=k} R_i \left[\phi\{F_i, S^{(1)}(u)\} - h(V_i)\right]}{n_k} + \frac{\sum_{i:\epsilon_i=k} h(V_i)}{N_k} \middle| N_k \right) \\
= & \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \mathrm{Var}\left(\sum_{i:\epsilon_i=k, R_i=1} \left[\frac{\phi\{F_i, S^{(1)}(u)\} - h(V_i)}{n_k} + \frac{h(V_i)}{N_k} \right] + \sum_{i:\epsilon_i=k, R_i=0} \frac{h(V_i)}{N_k} \middle| N_k \right) \\
= & \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \left\{ \mathrm{Var}\left(\sum_{i:\epsilon_i=k, R_i=1} \left[\frac{\phi\{F_i, S^{(1)}(u)\}}{n_k} - h(V_i)\left(\frac{1}{n_k} - \frac{1}{N_k}\right) \right] \middle| N_k \right) \right. \\
& \qquad \left. + \mathrm{Var}\left\{ \sum_{i:\epsilon_i=k, R_i=0} \frac{h(V_i)}{N_k} \right\} \right\} \\
= & \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \left[n_k \left\{ \mathrm{Var}_k(\phi) \left(\frac{1}{n_k}\right)^2 - 2\mathrm{Cov}_k(\phi, h) \left(\frac{1}{n_k}\right)\left(\frac{1}{n_k} - \frac{1}{N_k}\right) \right. \right. \\
& \qquad \left. \left. + \mathrm{Var}_k(h) \left(\frac{1}{n_k} - \frac{1}{N_k}\right)^2 \right\} + (N_k - n_k) \mathrm{Var}_k(h) \left(\frac{1}{N_k}\right)^2 \right] \\
= & \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \left[\mathrm{Var}_k(\phi) \left(\frac{1}{n_k}\right) - 2\mathrm{Cov}_k(\phi, h) \left(\frac{1}{n_k} - \frac{1}{N_k}\right) \right. \\
& \qquad \left. + \mathrm{Var}_k(h) \left\{ n_k \frac{(N_k - n_k)^2}{n_k^2 N_k^2} + \frac{N_k - n_k}{N_k^2} \right\} \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \left\{ \mathrm{Var}_k(\phi) \left(\frac{1}{n_k}\right) - 2\mathrm{Cov}_k(\phi, h) \left(\frac{1}{n_k} - \frac{1}{N_k}\right) + \mathrm{Var}_k(h) \left(\frac{1}{n_k} - \frac{1}{N_k}\right) \right\} \\
&= \sum_{k=1}^{K} \left(\frac{N_k}{n}\right)^2 \left[\{\mathrm{Var}_k(\phi) - 2\mathrm{Cov}_k(\phi, h) + \mathrm{Var}_k(h)\} \left(\frac{1}{n_k} - \frac{1}{N_k}\right) + \mathrm{Var}_k(\phi) \left(\frac{1}{N_k}\right) \right].
\end{aligned}
$$

Appendix F

Double Robustness of the AIPWCC Stratified Estimator

This appendix demonstrates the double robustness property of the AIPWCC stratified estimator through some simulation studies. The data generating scheme is the same as in section 4: namely, $B \sim Bernoulli(0.5), W \sim N(0,1)$ $X_2 \sim N(0,1)$, mutually independent, $X_1 = (B,W)^T$, and $X = (X_1^T, X_2)^T$. The treatment assignment propensity is generated by $P(Z=1|X) = \frac{\exp(0.1B+0.1W+0.5W^2+0.5X_2)}{1+\exp(0.1B+0.1W+0.5W^2+0.5X_2)}$, and the treatment-specific hazard functions are given by $\lambda_{Z=1}(t|X) = \exp(-1+0.1B+0.1W+0.5W^2+0.5X_2)$ and $\lambda_{Z=0}(t|X) = \exp(-0.5+0.1B+0.1W+0.5W^2+0.5X_2)$, respectively, and the independent censoring variable C is distributed as Uniform$(0,4)$. The covariate X_1 is observed on all patients, with sample size $5,000$, while X_2 is only observed on a subsample, with sample sizes $n_{11} = n_{12} = n_{01} = n_{02} = 300$ for each (Z,B) combination.

We considered three scenarios: "correct-correct" scenario, "wrong propensity, correct regression" scenario and "correct propensity, wrong regression" scenario, defined similarly to those in the first set of simulations in Section 1.5. To be specific, in the "correct-correct" scenario, X_1 is used to fit all models in $\phi_1^*(F)$, X is used to fit all models in $\phi_1^*(V)$. In the "wrong propensity, correct regression" scenario, (B,W) is used to fit $\pi(X), K_c^{(1)}(r,X)$ in $\phi_1^*(F)$, X_1 is used to fit $H^{(1)}(r,X)$ in $\phi_1^*(F)$, (B,W,X_2) is used to fit $\pi(X), K_c^{(1)}(r,X)$ in $\phi_1^*(V)$, X is used to fit $H^{(1)}(r,X)$ in $\phi_1^*(V)$. In the "correct propensity, wrong regression" scenario, X_1 is used to fit $\pi(X), K_c^{(1)}(r,X)$ in $\phi_1^*(F)$, (B,W) is used to fit $H^{(1)}(r,X)$ in $\phi_1^*(F)$, X is used to fit $\pi(X), K_c^{(1)}(r,X)$ in $\phi_1^*(V)$, (B,W,X_2) is used to fit $H^{(1)}(r,X)$ in $\phi_1^*(V)$.

Table F.1 shows the bias, estimated standard error and the coverage rate of the AIPWCC stratified estimator $\hat{S}_{strat}^{(1)}(u)$ under the three scenarios. All three estimators are consistent, demonstrating the double robustness property of the stratified AIPWCC estimator.

Table F.1: Simulation 2, estimators for $\hat{S}^{(1)}_{strat}(u)$ at selected time points when there is subsample to collect additional covariates. Sample size N=5000, $n_{11} = n_{12} = n_{01} = n_{02} = 300$ samples randomly selected from K=4 strata. Three estimators corresponding to (1) no misspecification (2) misspecify the propensity model and (3) misspecify the regression model are compared to illustrating double robustness property.

	u	$S^{(1)}(u)$	$\hat{S}^{(1)}_{strat}(u)$	Bias	Monte-Carlo SE	Estimated SE	Coverage Rate
	0.5	0.693	0.694	0.000	0.009	0.009	0.96
Correct	1.0	0.523	0.523	0.000	0.011	0.011	0.95
Propensity	1.5	0.408	0.408	0.000	0.012	0.012	0.95
Correct	2.0	0.324	0.324	0.000	0.012	0.012	0.95
Regression	2.5	0.262	0.262	0.000	0.013	0.013	0.95
	3.0	0.214	0.212	-0.001	0.013	0.013	0.94
	0.5	0.693	0.694	0.000	0.009	0.009	0.96
Wrong	1.0	0.523	0.523	0.000	0.010	0.010	0.95
Propensity	1.5	0.408	0.408	0.000	0.011	0.011	0.94
Correct	2.0	0.324	0.324	0.000	0.011	0.011	0.94
Regression	2.5	0.262	0.262	0.000	0.012	0.011	0.93
	3.0	0.214	0.213	-0.001	0.013	0.012	0.93
	0.5	0.693	0.694	0.000	0.009	0.009	0.95
Correct	1.0	0.523	0.523	0.000	0.011	0.011	0.93
Propensity	1.5	0.408	0.408	0.000	0.012	0.011	0.94
Wrong	2.0	0.324	0.324	0.000	0.012	0.012	0.95
Regression	2.5	0.262	0.262	0.000	0.013	0.012	0.94
	3.0	0.214	0.213	-0.001	0.014	0.013	0.94