ABSTRACT

HAGER, REBECCA SARAH. Optimal Dynamic Treatment Regimes from a Classification Perspective for Two Stage Studies with Survival Data. (Under the direction of Dr. Marie Davidian and Dr. Anastasios Tsiatis.)

Clinicians often make multiple treatment decisions for patients over the course of their disease. A dynamic treatment regime formalizes the treatment decision making process through rules at each decision time which map a patient's observed history to the set of available treatment options. The value of a regime is the mean outcome if a population were treated in accordance with that regime.

We consider the situation where the outcome of interest is a function of a possibly censored survival time. We present doubly-robust augmented inverse probability weighted estimators of the value of a given treatment regime in both a one stage and two stage study, either with or without censoring of the survival time. The augmentation terms capture back information from patients who either were censored or not consistent with the regime of interest, thus improving efficiency of the estimator.

The goal is to find the optimal treatment regime, which is the set of rules that would result in the most favorable outcome on average. By recasting the optimization of the value as a classification problem, we are able to estimate an optimal regime using well-studied classification techniques, such as support vector machines. A backward iterative method estimates an optimal sequence of treatment rules using value estimates of the available treatment options within the classification framework. A simulation study and an application of these methods to a cancer data set are presented.

Optimal Dynamic Treatment Regimes from a Classification Perspective for Two Stage Studies with Survival Data

by
Rebecca Sarah Hager

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

2016

APPROVED BY:

Dr. Eric Laber	Dr. Daowen Zhang
Dr. Marie Davidian Co-chair of Advisory Committee	Dr. Anastasios Tsiatis Co-chair of Advisory Committee

DEDICATION

To my family.

BIOGRAPHY

Rebecca Sarah Hager was raised in Glenmont, New York. She graduated from Bethlehem Central High School in 2007 and then attended the State University of New York at Buffalo. There, she was able to participate in multiple research projects and discovered statistics through her studies in mathematics. She graduated in 2011 with a B.S. in Mathematics and a minor in Biostistics. Rebecca went on to join the Department of Statistics at North Carolina State University and received a Master of Statistics degree in 2013. She will earn her Ph.D. in Statistics in 2016 under the direction of Drs. Davidian and Tsiatis.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors Marie Davidian and Butch Tsiatis. I am honored to have the privilege to learn from you both. In addition to abundant technical support, your patience and kindness allowed me to learn and grow throughout this process. There aren't enough words to express my gratitude. I hope that I will represent you well in my future endeavors.

I would like to further thank Marie for being a key figure on my journey to becoming a Statistician. She organizes the SIBS program at NC State which I attended after my junior year, and my experiences there cemented my decision to pursue statistics in graduate school. Once at NC State, Marie chose me to be on her training grant from which I received invaluable practical experience working at Duke Clinical Research Institute. Advising me with Butch on this dissertation is the capstone experience. Thank you for providing me with these opportunities that have helped me succeed.

I would also like to thank Eric Laber, Daowen Zhang, and Osman Ozaltin for taking the time to serve on my committee, and Shannon Holloway who provided me with useful code. I appreciate all of your comments and help which added to this dissertation.

I would like to thank Karen Pieper, Phil Schulte, and Daniel Wojdyla for all of your support during and after my time at Duke Clinical Research Institute. I appreciate the time you've spent helping me develop my skills and ensuring my success in the future.

There are too many students to name who have helped me over the last five years at NC State. I would like to thank everyone, and especially thank Caleb Browning and Rachel Marceau West for being there during the both the difficult and good times. I couldn't have done it without the two of you. I would also like to thank Alison McCoy for always being a happy presence at school and available to help with anything.

From my time at University at Buffalo, I would like to thank John Ringland, Surajit Sen, and Chris Andrews for introducing me to research and encouraging me to pursue a Ph.D. You provided me with the fundamentals that made the transition to graduate school easy and set me up for success in my current research.

Lastly, I would like to thank my family for always supporting me. My parents, Tom and Susan, have always encouraged me to follow my dreams and they do everything they can to help me realize those dreams. There isn't enough space to detail all the ways you two have helped, but know that I appreciate it all. I have always looked up to my big brother, Ben. Thank you for always being there, and especially for all the technical support over the years. I would also like to thank Ben along with my sister-in-law, Amy, for sending me pictures of my adorable nephew, Nathan. Those pictures always bring me a smile and especially helped on the hard days.

Thank you again to everyone for all of your support. I'm grateful to have had all of your help on this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Personalized Medicine and Dynamic Treatment Regimes

Personalized medicine considers patient heterogeneity in the population when choosing a best treatment for an individual patient. Physicians take into account many patient attributes, such as age, sex, historical medical information, etc. when deciding how to best treat a patient, because different patients may do better on different treatments. For example, the best cancer treatment for a 70-year-old male may not also be the best for a 20-year-old female. The goal of personalized medicine is to tailor treatment to patients based on their individual characteristics. It is of interest to formalize this treatment decision making process while accounting for patient differences.

For patients with chronic or long term illnesses such as AIDS, cancer, and heart disease,

clinicians often need to make multiple treatment decisions over the course of the disease. A dynamic treatment regime (DTR) is a sequence of decision rules that dictates which treatment to give an individual at each decision point using patient information. Before the first treatment, baseline characteristics such as age, sex, weight, history of disease, etc. are available to clinicians. After each treatment, additional information can be collected on patients such as their responses, toxicities, and other adverse events. At each treatment decision time, a DTR is able to use the information accrued up to that time to select the best treatment for each patient.

The main objective is to find the optimal DTR. This is the regime that, if used to select treatment for the entire patient population, would lead to an outcome on average that would be better than if any other regime had been followed. Currently, there is little work on estimating optimal DTRs for censored survival data. When patients are censored, the response of interest is not observed, but there is valuable information available that can still be of use. The challenge is how to incorporate these data along with those of other patients. We propose a method for estimation of an optimal DTR for a two stage study with survival outcomes.

1.2 Data for Estimating Regimes

It is of interest to evaluate how well DTRs perform and be able to find the optimal DTR. To do this, longitudinal data with information on patients at each treatment decision time are needed. Observational data may be used, but this can present many complications because the data were not collected with the intention of evaluating treatment regimes. Relevant data may not be collected or may not be collected in the same way for all patients

in an observational study. The proposed methods can be applied to observational data that satisfy the assumptions in Sections 2.2 and 3.2, but the focus of this dissertation will be on analyzing data from a clinical study.

A Sequential Multiple Assignment Randomized Trial (SMART) (Murphy, 2005; Lavori et al., 2000; Lavori and Dawson, 2004) is a clinical trial that results in data that can be used to estimate an optimal dynamic treatment regime. In a K stage SMART, patients are randomized to the available treatment options at each of the K decision points in the disease process. There could be any number of available treatments at each decision point, but we focus on only two possible treatment options. The multiple randomizations in the design allow for the evaluation of a complete treatment regime as opposed to evaluating the treatment options separately for each stage.

Powell et al. (2010) evaluate the North American Leukemia Intergroup Study C9710, which has a two stage SMART design as depicted schematically in Figure 1.1. The trial studied 518 patients with acute promyelocytic leukemia (APL) where the outcome of interest was time to event, denoted by T. An event was defined as failure to achieve complete remission (CR), relapse after achieving CR, or death, whichever comes first. Powell et al. (2010) specifically study event-free survival at 3 years.

In Study C9710, chemotherapy treatment for APL is split into three phases: induction, consolidation, and maintenance. Induction therapy aims to achieve remission, consolidation therapy intensifies the treatment to eliminate any remaining cancer cells in the body, and maintenance therapy ensures that the patient stays in remission (American Cancer Society, 2016).

The trial was comprised of $K = 2$ decisions points: selection of consolidation therapy

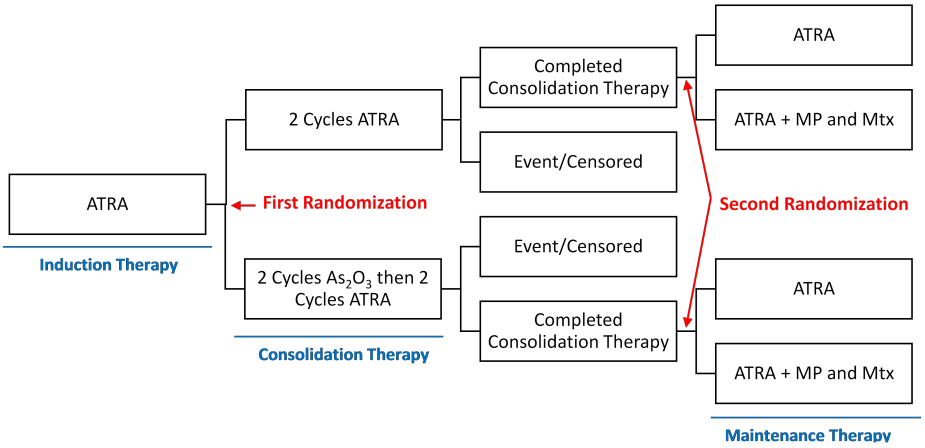

Figure 1.1 North American Leukemia Intergroup Study C9710 design.

and subsequent selection of maintenance therapy. Thus, all patients in the trial received the same induction therapy of all-trans-retinoic acid (ATRA), but were randomized to consolidation and maintenance therapies. The information available on a patient before the first treatment assignment is denoted by H_1. This includes baseline covariates such as age, sex, race, and risk group based on different blood measurements. After induction therapy, patients were first randomized to one of two different consolidation therapies: ATRA, coded as 0, or a combination of ATRA and arsenic trioxide (As_2O_3), coded as 1. We denote the treatment options available at this decision point by $\mathscr{A}_1 = \{0,1\}$, and A_1 is the option in $\mathscr{A}_1$ to which a patient was randomized.

A patient's history available before the second treatment is denoted by H_2 which is comprised of H_1, A_1, and any other additional covariate information collected on the individual before maintenance therapy. These additional covariates include different toxicities and adverse events such as hemorrhages and infections. Patients who completed consolida-

tion therapy were then randomized a second time to two different maintenance therapies: ATRA alone, coded as 0, or ATRA in combination with oral methotrexate (Mtx) and mercaptopurine (MP), coded as 1. We denote the available maintenance treatment options by $\mathscr{A}_2 = \{0,1\}$, and A_2 is the treatment to which a patient was randomized.

Although Study C9710 has a SMART design, the original intention was not to find the best regime, but rather to compare treatments at each stage separately. The main goal of the trial was to test whether adding arsenic trioxide to the consolidation therapy would improve event-free survival at three years. A secondary aim was to compare the efficacy and toxicity of the different maintenance therapies.

Powell et al. (2010) conclude that ATRA with arsenic trioxide is the better consolidation therapy, but were unable to detect a statistically significant difference between the maintenance therapies. As we demonstrate in Section 5.2, the proposed methods support the conclusions from this study and provide additional insights about the best treatment regime.

1.3 Outcomes

There are many different measures based on a time to event outcome, T, that could be of interest when comparing two treatment options at a single stage. A common analysis is to compare hazard functions under the two treatments using a hazard ratio. This analysis makes the proportional hazards assumption that the ratio is approximately constant over time. However, if the proportional hazards assumption is violated, then this ratio is not meaningful (Uno et al., 2014). Because this assumption is not always valid in practice, other methods of comparison are needed.

Another approach is to evaluate differences or ratios of the expectation of a monotone function $g(T)$ of the survival time under the treatments. In Study C9710, investigators were interested in event-free survival time up to 3 years due to finite patient follow up time. This is an example of a restricted survival time. Restricted survival time for an individual is a function of the event time, $g(T) = \min(T, L)$, for some fixed restriction time L. In Study C9710, 3 years is the cut off time L. The expectation of restricted survival time over all patients leads to an estimate of the mean restricted survival time for a given L. Alternatively, we could consider survival status of a patient at time L, $g(T) = I(T \geq L)$. The expectation of survival status over all patients leads to an estimate of the survival probability at time L. This can be extended further by estimating the whole survival function.

Uno et al. (2014) explain why using differences or ratios of the means of these functions of event times under the two treatments are good measures to examine. These measures are often clinically meaningful, easy to interpret, and can be estimated without making assumptions about the proportionality of the hazards functions under the two treatments. The mean restricted survival time, $\mu^L = E\{\min(T, L)\}$, can be estimated nonparametrically by taking the area under the Kaplan-Meier estimate of the survival curve up to time L, $\mu^L = \int_0^L S(t)dt$, or using an inverse probability weighted (IPW) estimator, both of which are simple to calculate (Uno et al., 2014). IPW estimators (Horvitz and Thompson, 1952; Robins et al., 1994) are used to account for incompletely observed data by weighting the sample to more accurately reflect the population.

Tools are available to analyze these measures when comparing two treatments in a single stage study, but this becomes more complicated in a multistage study. The sequential randomizations of a SMART study are a core feature of the design that provides data

to determine the best *sequence* of treatment decisions. Instead of examining the stages separately, there is a need for a way to evaluate the full sequence of treatments.

1.4 Embedded Regimes

A *treatment regime*, $d = (d_1, d_2)$, is a set of rules that determine which patients should receive which treatment based on their history. Each rule is a function which maps a patient's history to the set of possible treatments, $d_1 : H_1 \rightarrow \mathscr{A}_1$ and $d_2 : H_2 \rightarrow \mathscr{A}_2$. The rule d_1 uses baseline covariates to decide which treatment to assign at the first stage. The rule d_2 uses all information accrued to determine which treatment to assign at the second stage. The set containing all of the possible treatment regimes is denoted by $\mathscr{D}$.

In Study C9710, there are four *embedded* treatment regimes. At each stage, patients are randomized to either receive treatment 0 or treatment 1, and no historical information is considered. The four embedded regimes in the study are (i) receive treatment 0 at the first stage, followed by treatment 0 at the second stage if consolidation therapy was completed, denoted by $d^1 = (0,0)$; (ii) receive treatment 0 at the first stage, followed by treatment 1 at the second stage if consolidation therapy was completed, denoted by $d^2 = (0,1)$; (iii) receive treatment 1 at the first stage, followed by treatment 0 at the second stage if consolidation therapy was completed, denoted by $d^3 = (1,0)$; and (iv) receive treatment 1 at the first stage, followed by treatment 1 at the second stage if consolidation therapy was completed, denoted by $d^4 = (1,1)$.

Outcomes consistent with following regime d^1 are observable only for patients who actually receive treatment 0 at both stages, and similarly for regimes d^2, d^3, and d^4. If a patient receives treatment 0 at the first stage and then has an event before the second stage

treatment, he or she is considered to be consistent with both regimes d^1 and d^2. Care must be taken when incorporating these patients in the analysis of embedded regimes.

Lunceford et al. (2002) provide three different ways to estimate either the cumulative distribution, using $g(T) = I(T \leq L)$, or the mean restricted survival time, using $g(T) = \min(T, L)$, of each embedded regime in a two stage study. The first estimator is similar to the classic inverse probability weighted estimator. Outcomes of patients who are censored or not consistent with the regime of interest are weighted by 0. Outcomes of patients who are consistent with the regime for first stage treatment, but do not receive a second stage treatment are weighted by 1 to represent only themselves. Outcomes of patients who are consistent with the full regime are inversely weighted by the probability of not being censored and the probability of receiving that sequence of treatments. These patients are meant to represent themselves and also those patients with similar history who were censored or received other treatment. The paper gives two ways to handle these weighted values for all of the patients: either take the usual average, or the average using a probabilistically adjusted sample size. The third estimator in Lunceford et al. (2002) combines these two estimators in an ad hoc way. Variance estimates are given for all estimators so Wald tests can be used for comparisons of regimes.

Wahed and Tsiatis (2004) build on this work by providing the locally efficient augmented inverse probability weighted (AIPW) estimator when there is no censoring of the outcome. The augmentation term is included to take advantage of information from patients who were not consistent with the embedded regime of interest. The paper allows estimation of any function of survival time, $g(T)$, and shows how the variance can be estimated by the sandwich variance formula. They also take a different perspective by finding the optimal

estimator within a restricted class of estimators.

Wahed and Tsiatis (2006) continue using this new perspective, providing the best regular asymptotic linear estimator within a restricted class of estimators for functions of the survival time when there is censoring of the outcome. This estimator adds another augmentation term to capture back more information from patients who were censored. Covariances necessary for Wald tests are also provided.

These methods all provide ways to estimate and compare the embedded regimes. However, the embedded regimes do not take into account patient-specific information. We would like to be able to find the optimal treatment regime and expand the search to regimes that are more personalized.

1.5 Optimal Regime

The optimal dynamic treatment regime is the set of rules that, if used to select treatment for the entire patient population, leads to the most favorable result on average. There are a few different types of methods researchers have used to find the best dynamic treatment regime. Q- and A-learning methods (Schulte et al., 2014) posit models for the expected outcome, referred to as the reward, or the contrast function at each stage. The contrast function is the difference in expected outcome under each treatment. Backwards recursion is used to find the best treatment regime. This involves first solving for the best rule at the last stage, and then working backwards to iteratively find the best rule at the previous stages. The best rule at any given stage is the one that maximizes the current and future rewards given the past history and assuming that the optimal regime will be followed in future stages. Details of these methods can be found in Schulte et al. (2014).

Goldberg and Kosorok (2012) extend Q-learning to the case with censoring and a flexible number of stages. One issue with methods such as Q- and A- learning is that the models must be correctly specified in order to obtain consistent estimators of the outcome or contrast function. If the estimators are not consistent, then the decision rules derived from these methods may not be optimal. When there are multiple stages, correctly specifying the models is nearly impossible to do. Computation also becomes a problem as the number of stages and complexity of the optimization problem increases. Because of this, semiparametric methods are an attractive alternative.

Zhang et al. (2012a) present what we will refer to as the value search method where efficient, doubly robust, augmented IPW estimators are used to estimate the value for any given regime for a single treatment decision. The value of a regime d is defined as the mean response if a population were treated in accordance with regime d. The double robustness property will be discussed in detail later. The value estimator is directly maximized over a restricted class of rules using either a grid search or genetic algorithm (Zhang et al., 2012a) to solve for the best treatment rule.

Zhang et al. (2013) extend this estimator to the K stage case. They present the perspective of viewing the data as having monotone missingness and construct estimators using semiparametric theory from Tsiatis (2006). The optimal treatment regime is estimated by finding the regime that maximizes the estimated outcome via grid search over all stages' treatment rules. The sandwich variance estimator is used in both papers. It is clear that directly solving the maximization problem using grid search can become computationally difficult very quickly.

Zhang et al. (2012b) formally show for the one decision problem that finding the treat-

ment rule that optimizes the expected outcome is equivalent to solving a weighted classification problem. This breakthrough allows for easier computation because existing, well-studied classification methods can be used, such as support vector machines (SVM) and classification and regression trees (CART). Using SVM or CART implicitly defines a restricted class of regimes in which to search. Only an estimate of the contrast function is needed for this technique, which is discussed in detail in Section 4.2. The paper uses an augmented IPW estimator to obtain an efficient estimator of the value of a regime when there is no censoring of the outcome. These value estimators are used to estimate the contrast function which is then utilized within the classification framework.

Zhao et al. (2012) present a method called Outcome Weighted Learning (OWL), which also views the issue of finding the one stage optimal rule as a weighted classification problem. SVM is used to estimate the optimal treatment rule, but this formulation of the classification problem is different from that in Zhang et al. (2012b). The weight used in OWL is the reward inversely weighted by the probability of receiving the observed treatment. Zhao et al. (2015b) extend this method to multiple stages with a technique called Backward Outcome Weighted Learning (BOWL). This method solves for the best dynamic treatment regime using backwards recursion as used in Q- and A-learning. At each stage, SVM is used to find the optimal treatment rule to simplify computation. This method assumes all patients are observed for the full sequence of treatment.

None of these semiparametric methods are appropriate when the outcome is censored. Zhao et al. (2015a) extend OWL in one stage by adding an augmentation term to their IPW weight to account for censoring, and again solve for the optimal rule using SVM.

Bai et al. (2013, 2016) build on Zhang et al. (2012b) by providing another augmentation

term to obtain the locally efficient estimator of the outcome in a one stage study with censoring. This estimator uses one more augmentation term than the Zhao et al. (2015a) paper, which improves efficiency. The classification perspective is taken, and SVM is used to estimate the optimal treatment regime.

Jiang et al. (2016) take a different approach to the computation issue. They aim to maximize the t-year survival probability by estimating the survival function of a given regime using an augmented inverse probability weighted Kaplan-Meier estimator. The search is restricted to only linear regimes, taking the same approach as many of the previous papers. Because the estimated function is not smooth, kernel smoothers are used to more easily obtain the maximizer.

The methods developed in this dissertation extend the work of Zhang et al. (2012b) and Bai et al. (2013, 2016) to the two stage case both with and without censoring using a classification perspective. When there is no censoring, the proposed estimator of the outcome is similar to the AIPW estimator used in Zhang et al. (2013), which is more efficient than BOWL from Zhao et al. (2015b). However, unlike in Zhang et al. (2013), we solve for an optimal regime using a backwards iterative strategy and take the classification perspective at each stage so SVM can be used for computation. This combines the best parts of all of the existing methodology presented.

Unlike Jiang et al. (2016), the proposed method allows for maximization of the value, or the expected outcome of any monotone function of survival time. The proposed method allows for the length of the first stage to differ across patients and uses information from patients who did not receive a second stage treatment due to the event or censoring occurring first. The proposed method accommodates more realistic situations that occur in

practice because of its ability to account for these factors and censoring.

1.6 Outline

In Chapter 2 we describe how to estimate the value of any given regime in a one stage study when there is not censoring and also when there is censoring of the survival time. In Chapter 3 we extend these estimators to a two stage study. In Chapter 4, we describe the framework and method schema for finding an optimal dynamic treatment regime. Finally, in Chapter 5 we present a simulation study and data analysis of Study C9710, as well as a brief conclusion.

CHAPTER 2

ESTIMATING THE VALUE OF A RULE IN A ONE STAGE STUDY

In this chapter, we discuss how to estimate the value of a treatment rule d in a study where only one treatment decision is made. The value is defined as the expected outcome if all patients had received treatment consistent with rule d.

2.1 Notation

Consider a one stage randomized study with n patients. Let X be a vector of patient baseline characteristics that are available prior to treatment. This can also be referred to as the history, denoted by $H = \{X\}$, and let $\tilde{H} = (1, H^T)^T$. The history includes all relevant information

available to a clinician when making a treatment decision. Assume that the set $\mathscr{A} = \{0, 1\}$ represents the two available treatment options coded by 0 and 1, and denote the observed treatment received by A. Time to an event is denoted by T.

A treatment rule, denoted by d, is a way to assign treatment based on a patient's history. Formally, rule d is a function that maps values of H to $\mathscr{A}$. A patient with history $H = h$ who is consistent with rule d receives treatment 1 if $d(h) = 1$ and treatment 0 if $d(h) = 0$. The set of all possible treatment rules is denoted by $\mathscr{D}$. Define the variable $C_{d,i} = \mathrm{I}\{A_i = d(H_i)\}$ as the indicator that patient i receives treatment that is consistent with the treatment that regime d prescribes. When $C_d = 1$, the patient is consistent with receiving treatment according to rule d, and when $C_d = 0$, the patient does not receive treatment consistent with rule d. The probability of receiving treatment consistent with rule d given a patient's history is defined as $\pi_d(H) = P\{C_d = 1 \mid H\}$.

For any treatment $a \in \mathscr{A}$, the *potential outcome* $T^*(a)$ is the event time that would be achieved if a patient were to receive treatment a. For any treatment rule, $d \in \mathscr{D}$, the potential outcome $T^*(d)$ is defined as $T^*(d) = T^*(1)\mathrm{I}\{d(H) = 1\} + T^*(0)\mathrm{I}\{d(H) = 0\}$. This is the event time for an arbitrary individual in the population if that individual were given treatment, possibly contrary to fact, according to rule d. Potential outcomes are hypothetical variables that are not always realized for all patients in the sample.

The primary outcome of interest is $g\{T^*(d)\}$, where $g(\cdot)$ is a known monotone function. We specifically focus on the function $g\{T^*(d)\} = \min\{T^*(d), L\}$, or the truncated event time through time L. In most studies, patients are followed for a finite amount of time, so we are limited to considering the truncated or restricted lifetime.

The goal is to use the observed data, $O_i = \{X_i, A_i, T_i\}$, $i = 1, ..., n$, to estimate the value of

rule d, defined as the expected outcome were all patients to receive treatment consistent with rule d, $V(d) = \mathrm{E}\left[g\left\{T^*(d)\right\}\right]$. We also wish to find the optimal rule d^{opt}, which is the rule that maximizes this value, $V(d^{\text{opt}}) \geq V(d)$ for all $d \in \mathscr{D}$. This chapter focuses on estimating the value of any given fixed rule. Chapter 4 discusses how to use these value estimates to find the optimal rule.

2.2 Assumptions

The goal is to estimate the value of rule d, which is defined in terms of potential outcomes. Several assumptions are necessary to use the observed sample data to estimate this quantity.

The *consistency* assumption (Robins, 2004) states that the outcome observed is equivalent to the patient's potential outcome under the treatment that he or she actually received, $T = T^*(1)A + T^*(0)(1-A)$.

The *stable unit treatment value* assumption (Rubin, 1978) states that a patient's outcomes and covariates are not affected by the process of how treatments are assigned to patients. An example of when this assumption is violated is when the treatments are vaccines, where a patient's outcome could be affected by whether or not those around him are vaccinated.

The *no unmeasured confounders* or *sequential randomization* assumption (Robins, 2004) requires that potential outcomes are independent of treatment given the patient's past observed covariates. This is written as $\{T^*(0), T^*(1)\} \perp\!\!\!\perp A \mid H$, where $\perp\!\!\!\perp$ denotes "independent of." This assumption is automatically satisfied in a randomized study due to the design.

Lastly, the *positivity* assumption (Schulte et al., 2014) ensures that the probability of

receiving any one treatment is not 0 or 1. Formally, this states $P(A = a \mid H = h) > 0$ for all $a \in \mathscr{A}, h$. If this assumption is violated, it would make estimation of treatment rules for certain subsets of patients impossible. The positivity assumption should be satisfied with a well-designed randomized study. We assume that all of these assumptions hold in what follows.

2.3 No Censoring Case

We first estimate the value of rule d for a one stage study when the outcome has no censoring. As a result of the consistency assumption, $C_d = 1$ indicates that the observed outcome is equivalent to the potential outcome under rule d, $T = T^*(d)$. When $C_d = 0$, $T^*(d)$ is not observed. In this way, the variable C_d can be viewed as representing different levels of missingness in the data.

Note that this type of missingness is either missing completely at random, independent of all observed and unobserved variables, or missing at random, dependent only on observed variables, depending on the rule of interest and design of the study. According to semiparametric theory for monotone missing data in Tsiatis (2006), all consistent regular asymptotically linear estimators for the value can be written as a mean,

$$\hat{V}(d) = n^{-1} \sum_{i=1}^{n} \Psi_i(d), \tag{2.1}$$

$$\text{where } \Psi_i(d) = \frac{C_{d,i}}{\pi_d(H_i)} g(T_i) - \frac{C_{d,i} - \pi_d(H_i)}{\pi_d(H_i)} m(H_i), \tag{2.2}$$

for any function $m(\cdot)$ of H. Tsiatis (2006) also shows that the most efficient version of this estimator is obtained when $m(H_i) = \mathrm{E}\left[g\left\{ T_i^*(d) \right\} \mid H_i \right]$.

There are two levels of missingness in the one stage problem without censoring, $C_d = 1$ or 0, and the Ψ function in (2.2) has two terms. We call these the kernel term and augmentation term, denoted by $\Psi_i(d) = \text{Kernel}_i + \text{Aug}_i$.

2.3.1 Kernel Term

The first term of (2.2) is the kernel term,

$$\text{Kernel}_i = \frac{C_{d,i}}{\pi_d(H_i)} g(T_i). \tag{2.3}$$

The consistency probability can be estimated by $\hat{\pi}_d(H) = d(H)\hat{\pi}(H) + \{1 - d(H)\}\{1 - \hat{\pi}(H)\}$ where $\hat{\pi}(H)$ is an estimate of the propensity score, $\pi(H) = P(A = 1 \mid H)$. In a randomized study, the propensity score is known from the treatment assignment scheme. However, some efficiency is gained by estimating this using patient data. To estimate the propensity score, one could posit a model, for example, the logistic regression model $\log[\pi(H)/\{1 - \pi(H)\}] = \alpha^T \tilde{H}$. The parameters, α, can be estimated by the maximum likelihood estimator $\hat{\alpha}$, and the estimated propensity score is then $\hat{\pi}(H; \hat{\alpha}) = \exp(\hat{\alpha}^T \tilde{H}) / \left\{1 + \exp(\hat{\alpha}^T \tilde{H})\right\}$. One example of a nonparametric estimator of the propensity score is the sample mean of the observed A_i, $\hat{\pi}(H) = \hat{\pi} = n^{-1} \sum_{i=1}^{n} A_i$.

The kernel term is equivalent to the version of (2.2) where $m(H) = 0$, so $\hat{V}^{\text{IPW}}(d) = n^{-1} \sum_{i=1}^{n} \widehat{\text{Kernel}}_i$ is a consistent estimator of the value of rule d as long as the propensity score model is correctly specified. We refer to this estimator as the inverse probability weighted (IPW) estimator. Our interest is in the mean outcome if all patients had received treatment consistent with rule d. Because this outcome is not observed for patients with

$C_d = 0$, those patients' outcomes are weighted by 0. Outcomes of patients who do receive treatment consistent with d are inversely weighted by the probability of receiving that treatment. These patients represent themselves as well as patients with similar history who do not receive treatment according to rule d.

The IPW estimator is a consistent estimator of the value, but this only includes information from patients who are consistent with rule d. This does not take advantage information from many patients in the sample. Some efficiency is gained in using this patient data for the estimation of the propensity score, but further gains are possible.

2.3.2 Augmentation Term

The augmentation term is added to capture back some information from patients who are not consistent with rule d. The second term in (2.2) with $m(H_i) = \mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_i\right]$ is

$$\mathrm{Aug}_i = -\frac{C_{d,i} - \pi_d(H_i)}{\pi_d(H_i)} \mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_i\right]. \tag{2.4}$$

By the sequential randomization assumption, $\mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_i\right] = \mathrm{E}\left\{g(T_i) \mid H_i, A_i = d(H_i)\right\}$. To estimate the expectation, first posit a model for the event time hazard function which is defined as

$$\lambda(t \mid H, A) = \lim_{dt \to 0} (dt)^{-1} P(t \leq T < t + dt \mid T \geq t, H, A). \tag{2.5}$$

We can posit a Cox proportional hazards model,

$$\lambda(t \mid H, A, \beta) = \lambda_0(t) \exp\left\{\beta_1^T H + A\left(\beta_2^T \tilde{H}\right)\right\},$$

where $\lambda_0(t)$ is the baseline hazard function. The parameters, $\beta = (\beta_1^T, \beta_2^T)^T$, are estimated by maximizing the partial likelihood (Cox, 1972; Cox, 1975) using the observed data from all patients. The baseline hazard function can be estimated using the Breslow estimator (Breslow, 1972). The estimated model $\hat{\lambda}(t \mid H, A) = \lambda(t \mid H, A, \hat{\beta})$ is used to estimate the expected outcome if all patients had received treatment according to rule d,

$$\begin{aligned} \widehat{\mathrm{E}}\left\{g(T) \mid H, A = d(H)\right\} &= \int_0^\infty g(t)\left[-d\hat{S}\left\{t \mid H, A = d(H)\right\}\right] \\ \text{where } \hat{S}\left\{t \mid H, A = d(H)\right\} &= \exp\left[-\int_0^t \hat{\lambda}\left\{w \mid H, A = d(H)\right\} d\,w\right]. \end{aligned} \tag{2.6}$$

When analyzing restricted lifetime, it is shown in Appendix A.1 that the expectation can be estimated by $\widehat{\mathrm{E}}\left\{g(T) \mid H, A = d(H)\right\} = \int_0^L t\left[-d\hat{S}\left\{t \mid H, A = d(H)\right\}\right] + L\hat{S}\left\{L \mid H, A = d(H)\right\}$.

Including the augmentation term with $m(H) = \mathrm{E}\left\{g(T) \mid H, A = d(H)\right\}$ increases the efficiency compared to the kernel term alone. The estimator for the value using the augmentation term is called an augmented inverse probability weighted (AIPW) estimator. Noting again that $\Psi_i(d) = \mathrm{Kernel}_i + \mathrm{Aug}_i$, the value of a given rule d can be estimated by $\hat{V}^{\mathrm{AIPW}}(d) = n^{-1}\sum_{i=1}^n \widehat{\Psi}_i(d)$ where

$$\widehat{\Psi}_i(d) = \frac{C_{d,i}}{\hat{\pi}_d(H_i)} g(T_i) - \frac{C_{d,i} - \hat{\pi}_d(H_i)}{\hat{\pi}_d(H_i)} \widehat{\mathrm{E}}\left\{g(T_i) \mid H_i, A_i = d(H_i)\right\}. \tag{2.7}$$

The AIPW estimator is doubly robust in that it is a consistent estimator of the value of rule d if either the propensity score or event time model are correctly specified. In a randomized study, the treatment assignment probabilities are known, so consistency is guaranteed for both the IPW and AIPW estimators.

2.4 Censoring Case

We have shown how to estimate the value of fixed rule d in a one stage study without censoring. We now adapt this estimator to allow for censoring of the event time.

2.4.1 Additional Notation and Assumptions

Additional notation is needed when allowing for censoring of the event time. Censoring time is denoted by C. Assume that censoring is noninformative, or independent of event time given all observed covariates, $C \perp\!\!\!\perp T \mid H$. Define $U = \min(T, C)$ as the observed time to an event or censoring, whichever occurs first. When analyzing restricted lifetime, the observed time is the minimum of event time, censoring time, and cutoff time, $g(U) = \min(T, C, L)$.

Define $\Delta = I(T \leq C)$ as the indicator that an event occurred before censoring. When using restricted lifetime, this is defined to also include the possibility of the cutoff time, L, occurring before censoring, $\Delta = I\{\min(T, L) \leq C\}$. If the cutoff time occurs before an event or censoring, then the restricted survival time is known to be $g(T) = L$, even if the patient was ultimately censored in the study.

The goal is to estimate the value of the rule d using the observed data $O_i = \{X_i, A_i, U_i, \Delta_i\}$, $i = 1, ..., n$. However, $g\{T^*(d)\}$ is only observed when both $C_d = 1$ and $\Delta = 1$. Now there are two types of possible missingness: inconsistency with rule d and censoring. We still want to use information from these types of patients in the estimation of the value of a rule.

We adapt (2.2) to have three terms, $\Psi_i(d) = \text{Kernel}_i + \text{Aug}_i + \text{CAug}_i$: One kernel term for patients without a censored event time who are consistent with rule d, a term to augment for patients who are not consistent with d, and a second augmentation term to capture

back information from patients with a censored event time who are consistent with d.

2.4.2 Censoring Model

A model for the censoring distribution is necessary. The censoring hazard function is

$$\alpha(u \mid H, A) = \lim_{du \to 0} (du)^{-1} P(u \leq U < u + du, \Delta = 0 \mid U \geq u, H, A).$$

We choose to model the censoring hazard using a Cox proportional hazards model,

$$\alpha(u \mid H, A, \gamma) = \alpha_0(u) \exp\left\{\gamma_1^T H + A\left(\gamma_2^T \tilde{H}\right)\right\}.$$

Estimates for the parameters $\gamma = (\gamma_1^T, \gamma_2^T)^T$ and the baseline hazard function $\alpha_0(u)$ can be obtained using standard survival methods by reversing the roles of failure time and censoring; i.e., using the data $\{X_i, A_i, U_i, 1-\Delta_i\}$, $i = 1, ..., n$ to maximize the partial likelihood (Cox, 1972; Cox, 1975) and find the Breslow estimator (Breslow, 1972). The censoring survival function can be estimated by $\hat{K}(u \mid H, A) = \exp\left\{-\int_0^u \hat{\alpha}(w \mid H, A) dw\right\}$ where $\hat{\alpha}(u \mid H, A) = \alpha(u \mid H, A, \hat{\gamma})$. The censoring survival function $K(u \mid H, A)$ represents the probability that a patient has not been censored up to time u.

2.4.3 Kernel Term

The kernel term is similar to the kernel for the no censoring case,

$$\text{Kernel}_i = \frac{\Delta_i}{K(U_i \mid H_i, A_i)} \frac{C_{d,i}}{\pi_d(H_i)} g(U_i).$$

We adjust (2.3) for censored data by including an additional weight. This inversely weights the outcome of patients who are not censored by the probability of not being censored up to their event time, $K(U \mid H, A)$. Outcomes of patients who are censored are now weighted by 0. The probability $\pi_d(H)$ can be estimated the same way as in the no censoring case. The kernel is nonzero only for patients who are consistent with rule d and not censored, or those for whom $g\{T^*(d)\}$ is observed.

The estimator $\hat{V}^{\text{IPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\text{Kernel}}_i$ is an IPW estimator that is a consistent estimator of the value of rule d if both the propensity score and censoring models are correctly specified. Efficiency can be improved by including information from additional patients through augmentation terms.

2.4.4 First Augmentation Term

The first augmentation term is written the same as (2.4) for the no censoring case,

$$\text{Aug}_i = -\frac{C_{d,i} - \pi_d(H_i)}{\pi_d(H_i)}\text{E}\left[g\left\{T_i^*(d)\right\} \mid H_i\right].$$

This term augments to recover information from patients who are not consistent with rule d. Again, under sequential randomization, $\text{E}\left[g\left\{T_i^*(d)\right\} \mid H_i\right] = \text{E}\left\{g(T_i) \mid H_i, A_i = d(H_i)\right\}$. To estimate the expectation, it is necessary to adjust for censoring in the event time hazard model in (2.5), $\lambda(u \mid H, A) = \lim_{du \to 0}(du)^{-1}P(u \le U < u + du, \Delta = 1 \mid U \ge u, H, A)$. Because there is noninformative censoring, the parameters in the Cox proportional hazards model can be estimated by maximizing the partial likelihood (Cox, 1972; Cox, 1975) and the baseline hazard can be estimated using the Breslow estimator (Breslow, 1972). The estimation

of the expectation proceeds as in (2.6).

The estimator $\hat{V}^{\text{AIPW}}(d) = n^{-1}\sum_{i=1}^{n}\left(\widehat{\text{Kernel}}_i + \widehat{\text{Aug}}_i\right)$ is an AIPW estimator of the value of rule d. According to theory about the monotone missingness problem, this estimator is consistent if either the censoring model and the propensity score are correctly specified, or if the event model is correctly specified. However, efficiency can be improved further by including one more augmentation term.

2.4.5 Second Augmentation Term

The second augmentation term is used to capture back some information from patients who are consistent with rule d, but whose event time is censored. Tsiatis (2006) and Bai et al. (2013) show that this augmentation term is

$$\text{CAug}_i = \frac{C_{d,i}}{\pi_d(H_i)}\int_0^\infty \frac{dM_i(w \mid H_i, A_i)}{K(w \mid H_i, A_i)}\text{E}\left[g\left\{T_i^*(d)\right\} \mid T_i^*(d) > w, H_i\right]$$

where $dM_i(u \mid H_i, A_i) = dN_i(u) - \alpha(u \mid H_i, A_i)Y_i(u)$ is the martingale increment with counting process $N_i(u) = I(U_i \le u, \Delta_i = 0)$, and at-risk indicator $Y_i(u) = I(U_i \ge u)$. By the sequential randomization assumption, the expectation in this augmentation term is equivalent to $\text{E}\left\{g(T_i) \mid U_i > w, H_i, A_i = d(H_i)\right\}$. Rewriting this as

$$\text{E}\left\{g(T_i) \mid U_i > w, H_i, A_i = d(H_i)\right\} = \frac{\int_w^\infty g(u)[-dS\{u \mid H_i, A_i = d(H_i)\}]}{S\{w \mid H_i, A_i = d(H_i)\}},$$

it is clear how to estimate this term from the models previously discussed.

Noting again that $\Psi_i(d) = \text{Kernel}_i + \text{Aug}_i + \text{CAug}_i$, the value of a given rule d is estimated

by $\hat{V}^{\text{CAIPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d)$ where

$$\begin{aligned}\widehat{\Psi}_i(d) = \frac{\Delta_i}{\hat{K}(U_i \mid H_i, A_i)}\frac{C_{d,i}}{\hat{\pi}_d(H_i)}g(U_i) - \frac{C_{d,i}-\hat{\pi}_d(H_i)}{\hat{\pi}_d(H_i)}\widehat{\mathrm{E}}\left\{g(T_i) \mid H_i, A_i = d(H_i)\right\} \\ + \frac{C_{d,i}}{\hat{\pi}_d(H_i)}\int_0^\infty \frac{d\hat{M}_i(w \mid H_i, A_i)}{\hat{K}(w \mid H_i, A_i)}\widehat{\mathrm{E}}\left\{g(T_i) \mid U_i > w, H_i, A_i = d(H_i)\right\}.\end{aligned} \tag{2.8}$$

This is the locally efficient estimator for the value of rule d. The estimator is doubly robust as it is a consistent estimator of the value if either both the propensity score model, $\pi(H)$, and the censoring hazard model, $\alpha(u \mid H, A)$, are correctly specified or if the event hazard model, $\lambda(u \mid H, A)$, is correctly specified (Bai et al., 2013). In a clinical study, the propensity score model is known by design, and a well-executed study should only have administrative censoring due to time limitations. When this is the case, consistency of $\hat{V}^{\text{CAIPW}}(d)$ is guaranteed.

CHAPTER 3

ESTIMATING THE VALUE OF A REGIME IN A TWO STAGE STUDY

In this chapter, we discuss how to estimate the value of a treatment regime $d = (d_1, d_2)$ for a two stage study both with and without censoring of the event time.

3.1 Notation

Consider a two stage SMART study with n patients. Let X_1 be a vector of baseline characteristics available prior to the first treatment. This can also be referred to as a patient's history prior the first stage treatment, denoted by $H_1 = \{X_1\}$, and let $\tilde{H}_1 = (1, H_1^T)^T$. The set H_1 includes all of the relevant information available to a clinician when making the first

treatment decision. Let A_1 represent the first stage treatment a patient receives, and assume that the set $\mathscr{A}_1 = \{0,1\}$ represents the two available treatments at the first stage, coded by 0 and 1.

Let X_2 be a vector of additional covariates measured after the first stage treatment and prior to the second stage treatment. Define the random variable τ as the length of the first stage, or the time from entry into the study to the time until event or the second stage treatment decision, whichever occurs first. The value of τ can differ across patients, and therefore, the times at which patients are treated are not required to be fixed as assumed in Jiang et al. (2016).

Let R be a variable that classifies the second stage status of each patient. Let $R = 0$ define patients who are not observed to receive treatment at the second stage. In Study C9710, some patients have an event occur or are censored before the second stage, which means they do not receive a second stage treatment. For these patients, τ represents the time to event. Patients that do not have an event occur prior to the second stage are classified according to the set of treatments they are eligible to receive at the second stage. For example, if patients are responding to treatment from the first stage, they may be eligible to receive maintenance therapies at the second stage. Patients who are not responding to treatment should only be eligible to receive salvage treatments at the second stage. It would not be ethical to treat nonresponders with maintenance therapy, and may not make sense to treat responders with salvage therapy.

Classify these patients using the notation $R = 1,...,B$ where B is the total number of treatment identifying groups. The variable $R > 0$ defines which set of second stage treatments each patient is eligible to be assigned, $\mathscr{A}_{2R} = \{0_R, 1_R\}$. The set $\mathscr{A}_2 = \bigcup_{r=1}^{B} \mathscr{A}_{2r}$

represents all possible treatments that could be received at the second stage. As an example, $\mathscr{A}_{21} = \{0_1, 1_1\}$ could be the set of possible maintenance therapies, and $\mathscr{A}_{22} = \{0_2, 1_2\}$ could be the set of possible salvage therapies. The set $\mathscr{A}_2 = \{0_1, 1_1, 0_2, 1_2\}$ would contain both the maintenance and salvage therapies. In Study C9710, every patient is eligible to receive the same second stage treatments, so R only takes values of 0 or 1.

The treatments a patient is eligible to receive can be determined based on the patient's history just prior to the second stage, $H_2 = \{X_1, A_1, \tau, R, X_2\}$. The set $H_{2r} = \{X_1, A_1, \tau, r, X_{2r}\}$ is a subset of H_2 that contains the part of X_2 observed when $R = r$, denoted by X_{2r}. Also define $\tilde{H}_{2r} = (1, H_{2r}^T)^T$ when $R = r$. For patients with $R > 0$, τ represents the time to second stage treatment assignment. The variable A_{2R} represents the treatment a patient receives at the second stage, and define $A_2 = \sum_{r=1}^{B} I(R = r)A_{2r}$.

Event time is denoted by T. For patients with $R > 0$, define additional life as the time to event after second stage treatment, $\text{AL} = T - \tau$. The outcome $g(T)$ for any monotone function $g(\cdot)$ is of specific interest.

Often restricted lifetime of patients is of interest. This is defined as $g(T) = \min(T, L)$ where L is the cutoff time. Define $g(\tau) = \min(\tau, L)$ as the observed time to second stage treatment, event, or cutoff time, whichever occurs first. When restricted lifetime is of interest, it is necessary to redefine $R_i = R_i I(\tau_i < L)$ as the new second stage status of patient i. If the cutoff time, L, occurs before τ_i, then patient i should be reclassified by $R_i = 0$ because the event of interest occurs before the second stage treatment decision. If the cutoff occurs after τ_i, then the second stage status remains the same.

A *treatment regime*, $d = (d_1, d_2)$, is a set of rules that determine which patients should receive which treatment based on their history. Each rule is a function that maps a patient's

history to the set of possible treatments, $d_1 : H_1 \rightarrow \mathscr{A}_1$ and $d_2 : H_2 \rightarrow \mathscr{A}_2$. The rule d_1 uses patient baseline covariates to decide which treatment to assign at the first stage. The rule d_2 uses all of the information accrued up to the second stage to determine which treatment to assign at the second stage. Because each patient could be eligible to receive different treatments at the second stage, the rule d_2 is comprised of B different rules for each value of R, $d_2(H_2) = \sum_{r=1}^{B} I(R = r) d_{2r}(H_{2r})$ where d_{2r} is a rule that maps values of H_{2r} to $\mathscr{A}_{2r}$ when $R = r$. The set containing all of the possible treatment regimes (d_1, d_2) is denoted by $\mathscr{D}$.

Define the variable $C_{d,1,i} = \mathrm{I}\{A_{1i} = d_1(H_{1i})\}$ as the indicator that patient i receives first stage treatment that is consistent with rule d_1. Define $C_{d,2,i} = \mathrm{I}\{A_{1i} = d_1(H_{1i}), A_{2i} = d_2(H_{2i})\}$ as the indicator that patient i receives treatment that is consistent with regime d at both stages. The probability of receiving treatment consistent with rule d_1 at the first stage is defined as $\pi_{d_1}(H_1) = P\{A_1 = d_1(H_1) \mid H_1\}$ which is equivalent to $\pi_{C_{d,1}}(H_1) = P(C_{d,1} = 1 \mid H_1)$. The probability of receiving treatment consistent with rule d_2 at the second stage is defined as $\pi_{d_2}(H_2) = \sum_{r=1}^{B} I(R = r) P\{A_{2r} = d_{2r}(H_{2r}) \mid H_{2r}\}$. The probability of receiving treatment consistent with the full regime d is defined as $\pi_{C_{d,2}}(H_2) = P(C_{d,2} = 1 \mid H_2) = \pi_{d_1}(H_1)\pi_{d_2}(H_2)$.

The set of all potential outcomes is defined as $W^* = \left[\tau^*(a_1), R^*(a_1), I\{R^*(a_1) > 0\} X_2^*(a_1), T^*(a_1, a_2) \text{ for all } (a_1, a_2) \in \mathscr{A}_1 x \mathscr{A}_2\right]$. These are the outcomes that would have been observed if a patient received treatment a_1 at the first stage and a_2 at the second stage for any of the possible treatment combinations. Also denote $S^*(d_1) = \left[\tau^*(d_1), R^*(d_1), I\{R^*(d_1) > 0\} X_2^*(d_1)\right]$ as the set of potential outcomes of the variables observed after the first stage treatment but before second stage treatment if a patient had received treatment according to rule d_1 at the first stage.

The potential outcome $T^*(d) = T^*(d_1, d_2)$ is the event time that would be achieved if a pa-

tient had received treatment consistent with regime d, and $g\{T^*(d)\}$ is the outcome of interest. This quantity is observed only for patients in the sample who actually received treatment consistent with regime d. For all other patients, this quantity can be considered to be missing. The goal is to use the observed data $O_i = \{X_{1i}, A_{1i}, \tau_i, R_i, I(R_i > 0)X_{2i}, I(R_i > 0)A_{2i}, T_i\}$, $i = 1, ..., n$, to estimate the value of regime d, which is defined as $V(d) = \mathrm{E}\big[g\{T^*(d)\}\big] = \mathrm{E}\big[g\{T^*(d_1, d_2)\}\big]$.

3.2 Assumptions

The goal is to estimate the value of regime d, which is defined in terms of potential outcomes. Several assumptions are necessary to use the observed sample data to estimate this quantity. These assumptions are generalizations of the assumptions in Section 2.2 for the one stage problem.

The *consistency* assumption (Robins, 2004) states that the outcomes and covariates observed are equivalent to a patient's potential outcome under the treatments actually received, $R = R^*(A_1)$, $\tau = \tau^*(A_1)$, $X_2 = X_2^*(A_1)$, and $T = T^*(A_1, A_2)$.

The *stable unit treatment value* (Rubin, 1978) assumption states that a patient's outcomes and covariates are not affected by the process of how treatments were assigned to patients. As in the one stage case, one example of when this assumption is violated is when the treatments are vaccines, where a patient's outcome could be affected by whether or not those around him were vaccinated.

The *no unmeasured confounders* or *sequential randomization* assumption (Robins, 2004) requires that potential outcomes are independent of treatment given the patient's past covariates. Formally, this is written as $A_1 \perp\!\!\!\perp W^* \mid H_1$ and $A_2 \perp\!\!\!\perp W^* \mid H_2$, where $\perp\!\!\!\perp$ denotes

"independent of." This assumption is automatically satisfied in a SMART study due to the design.

Lastly, the *positivity* assumption (Schulte et al., 2014) ensures that the probability of receiving any one eligible treatment is not 0 or 1. Formally, this states $P(A_1 = a_1 \mid H_1) > 0$ if $a_1 \in \mathscr{A}_1$ and $P(A_{2r} = a_{2r} \mid H_{2r}) > 0$ if $a_{2r} \in \mathscr{A}_{2r}$ for each $R = r$. If this assumption is violated, it would make estimation of regimes for certain subsets of patients impossible. The positivity assumption should be satisfied with a well-designed and executed SMART study. We will assume that all of these assumptions hold in what follows.

3.3 No Censoring Case

In the two stage problem with no censoring, there are three levels of missingness. Patients with $R > 0$ and $C_{d,2} = 1$ are consistent with the full regime of interest $d = (d_1, d_2)$, so all covariates and outcomes under this regime are observed. Patients with $R = 0$ and $C_{d,1} = 1$ are also considered to be consistent with regime d. Patients with $R > 0$, $C_{d,1} = 1$, and $C_{d,2} = 0$ are only consistent with the first treatment rule, d_1, so the final outcome under regime d is not observed, though baseline and intermediate covariates are observed. Patients with $C_{d,1} = 0$ are not consistent with the first rule in the regime, so only their baseline covariates under regime d are observed.

As in the one stage problem, semiparametric theory for monotone missing data in Tsiatis (2006) shows that all consistent regular asymptotically linear estimators for the value,

$V(d) = \mathrm{E}\left[g\left\{T^*(d)\right\}\right]$, can be written as

$$\hat{V}(d) = n^{-1}\sum_{i=1}^{n}\Psi_i(d), \tag{3.1}$$

where $\Psi_i(d) = \text{Kernel}_i + \text{Aug1}_i + \text{Aug2}_i$. Because there are three levels of consistency, the Ψ function has three parts: a kernel term and two augmentation terms. The kernel term uses information from patients for whom the outcome under regime d is observed, and the augmentation terms are added to recover information from patients for whom the outcome under regime d is not observed. We discuss each of these terms and how to estimate them in detail.

3.3.1 Kernel

The kernel uses information from patients who are consistent with the regime $d = (d_1, d_2)$. There are two different types of patients that satisfy this requirement: patients who receive treatment consistent with regime d at both stages, $\left\{i : R_i > 0, C_{d,2,i} = 1\right\}$, and those who receive treatment consistent with rule d_1 at the first stage, but then have an event occur before receiving second stage treatment, $\left\{i : R_i = 0, C_{d,1,i} = 1\right\}$. The kernel has two terms to represent each type of patient,

$$\text{Kernel}_i = I(R_i = 0)\frac{C_{d,1,i}}{\pi_{C_{d,1}}(H_{1i})}g(T_i) + \sum_{r=1}^{B} I(R_i = r)\frac{C_{d,2,i}}{\pi_{C_{d,2}}(H_{2i})}g(T_i). \tag{3.2}$$

Recall that $\pi_{C_{d,1}}(H_1) = \pi_{d_1}(H_1) = P\left\{A_1 = d_1(H_1) \mid H_1\right\}$. This can be estimated by $d_1(H_1)\hat{\pi}_1(H_1) + \left\{1 - d_1(H_1)\right\}\left\{1 - \hat{\pi}_1(H_1)\right\}$ where $\hat{\pi}_1(H_1)$ is an estimator of the probability of receiving treatment 1 at the first stage, $\pi_1(H_1) = P(A_1 = 1 \mid H_1)$. Because $\pi_{C_{d,2}}(H_2) = \pi_{d_1}(H_1)\pi_{d_2}(H_2)$ and

$\pi_{d_2}(H_2) = \sum_{r=1}^{B} I(R=r)P\{A_{2r} = d_{2r}(H_{2r}) \mid H_{2r}\}$, the probabilities $P\{A_{2r} = d_{2r}(H_{2r}) \mid H_{2r}\}$ for each $R = r$ are the only terms left to be estimated. These probabilities can be estimated by $d_{2r}(H_{2r})\hat{\pi}_{2r}(H_{2r}) + \{1 - d_{2r}(H_{2r})\}\{1 - \hat{\pi}_{2r}(H_{2r})\}$ where $\hat{\pi}_{2r}(H_{2r})$ is an estimate of the probability of receiving treatment 1_r at the second stage for group $R = r$, $\pi_{2r}(H_{2r}) = P(A_{2r} = 1_r \mid H_{2r})$. The propensity scores, $\pi_1(H_1)$ and $\pi_{2r}(H_{2r})$ for $r = 1, ..., B$, are known in a SMART study because the design is known. However, some efficiency is gained by using patient data to estimate these quantities with a model or nonparametric estimator as discussed in Section 2.3.1.

The kernel inversely weights outcomes of patients consistent with regime d by the probability of receiving those treatments to represent both themselves and patients with a similar history who do not receive treatment consistent with regime d. Outcomes of patients that receive treatment inconsistent with regime d are weighted by zero.

The IPW estimator of the value of regime d, $\hat{V}^{\text{IPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\text{Kernel}}_i$, is a consistent estimator as long as the propensity score models are correctly specified. However, efficiency can be gained by adding augmentation terms to take advantage of more patient data.

3.3.2 First Augmentation Term

The first augmentation term recovers information from patients who are consistent with receiving treatment according to rule d_1 at the first stage, but are not consistent with d_2 at the second stage, $\left\{i : R_i > 0, C_{d,1,i} = 1, C_{d,2,i} = 0\right\}$. This term is written as

$$\text{Aug1}_i = -\sum_{r=1}^{B} I(R_i = r)\frac{C_{d,1,i}}{\pi_{C_{d,1}}(H_{1i})}\frac{I\{A_{2i} = d_2(H_{2i})\} - \pi_{d_2}(H_{2i})}{\pi_{d_2}(H_{2i})}\text{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}, S_i^*(d_1)\right]. \quad (3.3)$$

In order to estimate this term, an estimator for $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, S_i^*(d_1)\right]$ is needed. The augmentation term is only nonzero when $A_1 = d_1(H_1)$, so by the sequential randomization assumption, the expectation for these patients is equivalent to $\mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\}$ when $R = r$. To estimate this expectation, posit a model for additional life. Only patients who are able to receive second stage treatment are used to estimate this model, and the value of τ is known and considered fixed. The hazard function for additional life for each $r = 1, ..., B$ is

$$\lambda_{2r}(t \mid H_{2r}, A_{2r}) = \lim_{dt \to 0} (dt)^{-1} P(t \leq \mathrm{AL} < t + dt \mid \mathrm{AL} \geq t, H_{2r}, A_{2r}, R = r). \tag{3.4}$$

One possible model for the additional life hazard function is the Cox proportional hazards model,

$$\lambda_{2r}(t \mid H_{2r}, A_{2r}, \beta_{2r}) = \lambda_{20r}(t) \exp\left\{\beta_{21r}^T H_{2r} + A_{2r}\left(\beta_{22r}^T \tilde{H}_{2r}\right)\right\}. \tag{3.5}$$

The parameters $\beta_{2r} = (\beta_{21r}^T, \beta_{22r}^T)^T$ are estimated by maximizing the partial likelihood (Cox, 1972; Cox, 1975) using the observed data from patients who are categorized by $R = r$. The baseline hazard function can be estimated using the Breslow estimator (Breslow, 1972). The estimated hazard function, $\lambda_{2r}(t \mid H_{2r}, A_{2r}, \hat{\beta}_{2r}) = \hat{\lambda}_{2r}(t \mid H_{2r}, A_{2r})$, can be fit to the patient data, but substituting $A_{2r} = d_{2r}(H_{2r})$ instead of the observed second stage treatment. This estimator is used to calculate the expected outcome,

$$\begin{aligned} \widehat{\mathrm{E}}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\} &= \int_0^\infty g(\tau + t)\left[-d\hat{S}_{2r}\left\{t \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\}\right] \\ \text{where } \hat{S}_{2r}\left\{t \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\} &= \exp\left[-\int_0^t \hat{\lambda}_{2r}\left\{w \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\} dw\right]. \end{aligned} \tag{3.6}$$

Note that this augmentation term is only nonzero for patients with $A_1 = d_1(H_1)$, so this expectation estimates the outcome had the patient been consistent with the full regime d. When dealing with the special case of optimizing mean restricted lifetime, it is shown in Appendix A.2 that this expectation can be written as

$$\widehat{\mathrm{E}}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\} = \tau + \int_0^{L-\tau} \hat{S}_{2r}\left\{t \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\} dt.$$

The expectation in (3.6) can be substituted as an estimator for $\mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}, S_i^*(d_1)\right]$ in (3.3) to complete the estimation of the augmentation term.

3.3.3 Second Augmentation Term

The last augmentation term captures back information from the last group of patients, $\left\{i : C_{d,1,i} = 0\right\}$, who are not consistent with even the first treatment prescribed by regime d,

$$\mathrm{Aug2}_i = -\frac{I\left\{A_{1i} = d_1(H_{1i})\right\} - \pi_{d_1}(H_{1i})}{\pi_{d_1}(H_{1i})} \mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}\right]. \tag{3.7}$$

The only component left to estimate is $\mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}\right]$, or the expected outcome had patient i received treatment consistent with regime d given only baseline covariates. Estimating this expectations takes several steps.

Because the expectation is conditional only on baseline covariates, all possible second stage statuses, $R = 0, ..., B$, that could occur need to be accounted for. To do this, first develop a model for the cause-specific hazard function for τ. Recall that τ can represent either time to second stage treatment, $R > 0$, or time to event, $R = 0$, whichever occurs first. For each

$r = 0, 1, ..., B$, the cause-specific hazard function is

$$\lambda_{1r}(t \mid H_1, A_1) = \lim_{dt \to 0} (dt)^{-1} P(t \leq \tau < t + dt, R = r \mid \tau \geq t, H_1, A_1).$$

We choose to model these hazard functions for each $R = r$ with a Cox proportional hazards model such as

$$\lambda_{1r}(t \mid H_1, A_1, \beta_{1r}) = \lambda_{10r}(t) \exp\left\{\beta_{11r}^T H_1 + A_1 \left(\beta_{12r}^T \tilde{H}_1\right)\right\}.$$

All patients' data are used in estimating each model. Then $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1\right]$ can be estimated by integrating over values of τ,

$$\begin{aligned} &\int_0^\infty g(t)\lambda_{10}\left\{t \mid H_1, A_1 = d_1(H_1)\right\} S_{1\cdot}\left\{t \mid H_1, A_1 = d_1(H_1)\right\} dt + \\ &\sum_{r=1}^B \int_0^\infty \left(\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, A_1 = d_1(H_1), R = r, \tau = v\right] \lambda_{1r}\left\{v \mid H_1, A_1 = d_1(H_1)\right\}\right. \\ &\left. \times S_{1\cdot}\left\{v \mid H_1, A_1 = d_1(H_1)\right\} dv\right), \end{aligned} \tag{3.8}$$

where $S_{1\cdot}(t \mid H_1, A_1) = \exp\left\{-\sum_{r=0}^B \int_0^t \lambda_{1r}(w \mid H_1, A_1) dw\right\}$ is the all-cause survival function. Intuitively, (3.8) represents the sum of the estimated expected outcomes for each possibility of R multiplied by the probability of being categorized into group R. The estimators $\hat{\lambda}_{1r}(t \mid H_1, A_1) = \lambda_{1r}(t \mid H_1, A_1, \hat{\beta}_{1r})$ and $\hat{S}_{1\cdot}(t \mid H_1, A_1)$ can be substituted into this equation. Further details of the calculation of (3.8) in the special case where restricted lifetime is of interest are presented in Appendix A.3.

The only component left to estimate is $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, A_1 = d_1(H_1), R = r, \tau\right]$ for any given τ when $R = r > 0$. Estimating this conditional expectation is not straightforward.

Note that

$$\begin{aligned}&\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, A_1 = d_1(H_1), R = r, \tau\right] = \\ &\qquad \mathrm{E}\left\{\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, A_1 = d_1(H_1), R = r, \tau, X_{2r}\right] \mid H_1, A_1 = d_1(H_1), R = r, \tau\right\}.\end{aligned} \tag{3.9}$$

Due to the sequential randomization assumption, the inner expectation on the right-hand side of (3.9) is equal to $\mathrm{E}\left\{g(T) \mid H_1, A_1 = d_1(H_1), R = r, \tau, X_{2r}, A_2 = d_2(H_2)\right\}$. In (3.6), we show how to compute and estimate $\mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\}$ when $R = r$. Therefore, we define $V_{2r}(H_{2r}) = \mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\}$ and $Q_{1r}(H_1, A_1, \tau) = \mathrm{E}\left\{V_{2r}(H_{2r}) \mid H_1, A_1, \tau, R = r\right\}$. Then the left-hand side of (3.9) is equal to $Q_{1r}\left\{H_1, A_1 = d_1(H_1), \tau, R = r\right\}$ by the sequential randomization assumption.

Therefore, one solution for estimation is to posit a model $Q_{1r}(H_1, A_1, \tau, R = r; \xi_r)$ for each $r = 1, ..., B$ in terms of parameters ξ_r. For each patient i such that $R_i = r > 0$, we estimate the predicted value $V_{2r}(H_{2ri})$ and then regress these predicted values on H_{1i}, A_{1i}, and τ_i to obtain estimators for ξ_r using, say, least squares. Then $Q_{1r}(H_1, A_1, \tau, R = r; \hat{\xi}_r)$ can be substituted into (3.8) for $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1, A_1 = d_1(H_1), R = r, \tau\right]$ to obtain the estimator for $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1\right]$.

The value of regime d can be estimated by $\hat{V}^{\mathrm{AIPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d)$ where

$$\begin{aligned}\widehat{\Psi}_i(d) = I(R_i = 0)\frac{C_{d,1,i}}{\hat{\pi}_{C_{d,1}}(H_{1i})}g(T_i) + \sum_{r=1}^{B} I(R_i = r)\Bigg(&\frac{C_{d,2,i}}{\hat{\pi}_{C_{d,2}}(H_{2i})}g(T_i) \\ -\frac{C_{d,1,i}}{\hat{\pi}_{C_{d,1}}(H_{1i})}&\left[\frac{I\left\{A_{2i} = d_2(H_{2i})\right\} - \hat{\pi}_{d_2}(H_{2i})}{\hat{\pi}_{d_2}(H_{2i})}\right]\hat{\mathrm{E}}\left\{g(T_i) \mid H_{2ri}, A_{2ri} = d_{2r}(H_{2ri})\right\}\Bigg) \\ -\frac{I\left\{A_{1i} = d_1(H_{1i})\right\} - \hat{\pi}_{d_1}(H_{1i})}{\hat{\pi}_{d_1}(H_{1i})}&\widehat{\mathrm{E}}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}\right].\end{aligned} \tag{3.10}$$

Including both augmentation terms increases the efficiency compared to the IPW estimator.

This is referred to as the AIPW estimator, which is doubly robust in that it is a consistent estimator of the value of regime d if either the propensity score models or event hazard models are correctly specified.

3.4 Censoring Case

We now adapt the previous estimator to allow for censoring of the event time. Censoring can occur in either the first or second stages, which adds further complexity to the estimation.

3.4.1 Additional Notation

We introduce additional notation when allowing for censoring of the event time. Define $U = \min(T, C)$ as the observed outcome, or the minimum of event time and censoring time, C. Denote $\Delta = I(T \leq C)$ as the indicator that U is the observed event time instead of the censoring time. Because censoring can also occur during the first stage, define $U^1 = \min(\tau, C)$ as the possibly censored first stage observed time and $\Gamma = I(\tau \leq C)$ as the indicator that censoring does not occur during the first stage before τ. When analyzing restricted lifetime, redefine these as $\Delta = I\left\{g(T) \leq C\right\}$ and $\Gamma = I\left\{g(\tau) \leq C\right\}$. Define $O_i = \left\{X_{1i}, A_{1i}, U_i^1, \Gamma_i, \Gamma_i R_i, \Gamma_i I(R_i > 0) X_{2i}, \Gamma_i I(R_i > 0) A_{2i}, U_i, \Delta_i\right\}$, $i = 1, ..., n$ as the observed data and $O_i(u)$ as the data that are observed for patient i up through time u.

The variable $C_{d,1,i} = \mathrm{I}\{A_{1i} = d_1(H_{1i})\}$ is defined the same as in the no censoring case. However, now define $\tilde{C}_{d,2,i} = \mathrm{I}\{A_{1i} = d_1(H_{1i}), A_{2i} = d_2(H_{2i}), C_i > \tau_i\}$ as the indicator that patient i is not censored before the second stage and receives treatment that is consistent with regime d at both stages. The probability of this is $\pi_{\tilde{C}_{d,2}}(H_2) = P\left\{\tilde{C}_{d,2} = 1 \mid H_2\right\} =$

$\pi_{d_1}(H_1)\pi_{d_2}(H_2)P(C > \tau \mid H_1)$. Estimating the terms $\pi_{d_1}(H_1)$ and $\pi_{d_2}(H_2)$ can be done as described in Section 3.3.1, and an estimate of the censoring distribution is needed to estimate $P(C > \tau \mid H_1)$.

Assume that censoring is noninformative, so the hazard of a patient being censored at time u given he is at risk at time u will only depend on the data observed through time u. The censoring hazard function can be written as

$$\alpha\{u \mid O(u)\} = \lim_{du \to 0} (du)^{-1} P\{u \leq U < u + du, \Delta = 0 \mid U \geq u, O(u)\}.$$

Define the censoring survival function as $K\{u \mid O(u)\} = \exp\left[-\int_0^u \alpha\{w \mid O(w)\}\, dw\right]$. Now the probability that $\tilde{C}_{d,2} = 1$ can be written as $\pi_{\tilde{C}_{d,2}}(H_2) = \pi_{d_1}(H_1)\pi_{d_2}(H_2)K\{\tau \mid O(\tau)\}$. The censoring distribution can be estimated by positing a model, such as the Cox proportional hazards model, and estimating that model using patient data.

In the two stage problem that allows for censoring of the event time, $g\{T^*(d)\}$ is only observed for patients who are both consistent with regime d and who do not have their event time censored. There are four ways in which patients could violate these conditions: (i) a patient could be consistent with receiving treatment according to the first rule, d_1, but not consistent with receiving treatment according to rule d_2; (ii) A patient could not be consistent with receiving treatment according to the first rule, d_1; (iii) a patient could be consistent with regime d, but censored after second stage treatment; (iv) a patient could be consistent with rule d_1, but censored before second stage treatment. For all of these patients, $g\{T^*(d)\}$ is not observed. There will be an augmentation term for each type of patient in order to capture back some information from them. Therefore, $\Psi_i(d) = \text{Kernel}_i + \text{Aug1}_i + \text{Aug2}_i + \text{CAug1}_i + \text{CAug2}_i$.

3.4.2 Kernel

The kernel has two terms similar to (3.2). Only patients who are not censored have a nonzero kernel. There are two types of patients represented in this term; those who do not receive second stage treatment, but are consistent with rule d_1, $\left\{i : R_i = 0, C_{d,1,i} = 1, \Gamma_i = 1\right\}$, and those who are consistent with regime d at both stages, $\left\{i : R_i > 0, \tilde{C}_{d,2,i} = 1, \Delta_i = 1\right\}$. In order to adjust for censoring, we inversely weight the terms in (3.2) by the probability of not being censored before the event time,

$$\begin{aligned}\text{Kernel}_i = I(R_i = 0)\frac{\Gamma_i}{K\left\{U_i \mid O_i(U_i)\right\}}\frac{C_{d,1,i}}{\pi_{C_{d,1}}(H_{1i})}g(U_i) + \\ \sum_{r=1}^{B} I(R_i = r)\frac{\Delta_i}{K\left\{U_i \mid O_i(U_i)\right\}/K\left\{\tau_i \mid O_i(\tau_i)\right\}}\frac{\tilde{C}_{d,2,i}}{\pi_{\tilde{C}_{d,2}}(H_{2i})}g(U_i).\end{aligned}$$

Again, the second term of the kernel represents patients who are not censored with $\tilde{C}_{d,2,i} = 1$. This implies that patients are not censored in the first stage, the probability of which is accounted for in $\pi_{\tilde{C}_{d,2}}(H_2)$. Therefore, we only need to inversely weight by the probability of not being censored up to event time U given that it is known censoring didn't occur before time τ, hence the weight $K\left\{U_i \mid O_i(U_i)\right\}/K\left\{\tau_i \mid O_i(\tau_i)\right\}$.

As in the no censoring case, $\hat{V}^{\text{IPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\text{Kernel}}_i$ is a consistent IPW estimator of the value of regime d as long as the propensity score and censoring models are correctly specified, though efficiency can be improved by including augmentation terms. A SMART study will likely satisfy the assumption of both models being correctly specified. The propensity score models are known by design, and a well-executed study should only have administrative censoring due to time limitations. In this case, the censoring distribution can be estimated by Kaplan-Meier or a Cox proportional hazards model.

3.4.3 First Augmentation Term

The first augmentation term is similar to (3.3) and captures back information from patients who are consistent with receiving treatment according to rule d_1 at the first stage, but are not consistent with rule d_2 at the second stage, $\left\{i : C_{d,1,i} = 1, \Gamma_i = 1, \tilde{C}_{d,2,i} = 0\right\}$. This assumes that the patient is not censored during the first stage, so we include an additional weight. This inversely weights the outcome of patients not censored in the first stage, $\Gamma = 1$, by the probability of not being censored up to time τ, making the augmentation term

$$\text{Aug1}_i = -\sum_{r=1}^{B}\left(I(R_i = r)\frac{\Gamma_i}{K\left\{\tau_i \mid O_i(\tau_i)\right\}}\frac{C_{d,1,i}}{\pi_{C_{d,1}}(H_{1i})}\frac{I\left\{A_{2i} = d_2(H_{2i})\right\} - \pi_{d_2}(H_{2i})}{\pi_{d_2}(H_{2i})} \times \mathrm{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}, S_i^*(d_1)\right]\right).$$

As in the no censoring case, the expectation is equivalent to $\mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r} = d_{2r}(H_{2r})\right\}$ when $R = r$ by the sequential randomization assumption and the fact that this expectation is only calculated when $A_1 = d_1(H_1)$. The additional life hazard function in (3.4) adjusted for censoring is

$$\lambda_{2r}(u \mid H_{2r}, A_{2r}) = \lim_{du \to 0}(du)^{-1} P(u \leq U - U^1 < u + du, \Delta = 1 \mid U - U^1 \geq u, H_{2r}, A_{2r}, R = r).$$

The expectation can be estimated by positing a model for this hazard function as in (3.5). If using the Cox proportional hazards model, the parameters can be estimated by maximizing the partial likelihood (Cox, 1972; Cox, 1975) and baseline hazard can be estimated with the Breslow estimator (Breslow, 1972). Then use the estimated model to estimate the expectation as described in (3.6).

3.4.4 Second Augmentation Term

The second augmentation term recovers information from patients who are not consistent with even the first rule in regime d, $\{i : C_{d,1,i} = 0\}$. This is written just as in (3.7) for the no censoring case,

$$\text{Aug2}_i = -\frac{I\{A_{1i} = d_1(H_{1i})\} - \pi_{d_1}(H_{1i})}{\pi_{d_1}(H_{1i})} \text{E}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}\right].$$

The estimation differs from that of (3.7) in that censoring needs to be accounted for in the cause-specific hazard functions. This is written for $r = 1, ..., B$ as,

$$\lambda_{1r}(u \mid H_1, A_1) = \lim_{du \to 0} (du)^{-1} P(u \leq U^1 < u + du, R = r, \Gamma = 1 \mid U^1 \geq u, H_1, A_1).$$

From here, the estimation of the expectation follows the same as described in Section 3.3.3.

The estimator $\hat{V}^{\text{AIPW}}(d) = n^{-1} \sum_{i=1}^{n} \left(\widehat{\text{Kernel}}_i + \widehat{\text{Aug1}}_i + \widehat{\text{Aug2}}_i\right)$ is an AIPW estimator. According to theory for the monotone missingness problem, this estimator is consistent if either the censoring models and the propensity score models are correctly specified, or if the event models are correctly specified. However, efficiency can be further improved by including additional augmentation terms to account for censoring.

3.4.5 Third and Fourth Augmentation Terms

The third and fourth augmentation terms capture back some information from patients who are censored in each stage. Define the martingale increment as $dM_i\{u \mid O_i(u)\} = dN_i(u) - \alpha\{u \mid O_i(u)\} Y_i(u) du$ with counting process $N_i(u) = I(U_i \leq u, \Delta_i = 0)$ and at-

risk indicator $Y_i(u) = I(U_i \geq u)$. The first censoring term augments for patients who are consistent with regime d, but censored in the second stage, $\left\{i : \tilde{C}_{d,2,i} = 1, \Delta_i = 0\right\}$. This term is

$$\text{CAug1}_i = \sum_{r=1}^{B} \left(I(R_i = r) \frac{\tilde{C}_{d,2,i}}{\pi_{\tilde{C}_{d,2}}(H_{2i})} \times \int_{\tau_i}^{U_i} \frac{dM_i\{w \mid O_i(w)\}}{K\{w \mid O_i(w)\}/K\{\tau_i \mid O_i(\tau_i)\}} \text{E}\left[g\left\{T_i^*(d)\right\} \mid T_i^*(d) > w, H_{1i}, S_i^*(d)\right] \right).$$

Due to the sequential randomization assumption and the fact that the expectation is only calculated when $\tilde{C}_{d,2,i} = 1$, the expectation can be written as $\text{E}\left\{g(T_i) \mid U_i > \omega, H_{2ri}, A_{2ri}\right\}$. See Appendix A.4 for further calculations of the estimator of this expectation when examining mean restricted lifetime. The second censoring term augments for patients who are consistent with the first rule in the regime, but are censored in the first stage, $\left\{i : C_{d,1,i} = 1, \Gamma_i = 0\right\}$. This term is

$$\text{CAug2}_i = \frac{C_{d,1,i}}{\pi_{C_{d,1}}(H_{1i})} \int_0^{\tau_i} \frac{dM_i\{v \mid O_i(v)\}}{K\{v \mid O_i(v)\}} \text{E}\left[g\left\{T_i^*(d)\right\} \mid T_i^*(d) > v, H_{1i}\right].$$

Each part of these augmentation terms can be estimated from the methods described previously.

Using the kernel and the four augmentation terms, we can estimate the value of regime

d when censoring of the event time occurs by $\hat{V}^{\text{CAIPW}}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d)$ where

$$
\begin{aligned}
\widehat{\Psi}_i(d) = I(R_i = 0)&\frac{\Gamma_i}{\hat{K}\{U_i \mid O_i(U_i)\}}\frac{C_{d,1,i}}{\hat{\pi}_{C_{d,1}}(H_{1i})}g(U_i) \\
&+\sum_{r=1}^{B} I(R_i = r)\Bigg[\frac{\Delta_i}{\hat{K}\{U_i \mid O_i(U_i)\}/\hat{K}\{\tau_i \mid O_i(\tau_i)\}}\frac{\tilde{C}_{d,2,i}}{\hat{\pi}_{\tilde{C}_{d,2}}(H_{2i})}g(U_i) \\
&-\frac{\Gamma_i}{\hat{K}\{\tau_i \mid O_i(\tau_i)\}}\frac{C_{d,1,i}}{\hat{\pi}_{C_{d,1}}(H_{1i})}\frac{I\{A_{2i} = d_2(H_{2i})\} - \hat{\pi}_{d_2}(H_{2i})}{\hat{\pi}_{d_2}(H_{2i})} \\
&\qquad\times\widehat{\mathrm{E}}\left\{g(T_i) \mid H_{2ri}, A_{2ri} = d_{2r}(H_{2ri})\right\} \\
&+\frac{\tilde{C}_{d,2,i}}{\hat{\pi}_{\tilde{C}_{d,2}}(H_{2i})}\int_{\tau_i}^{U_i}\frac{d\hat{M}_i\{w \mid O_i(w)\}}{\hat{K}\{w \mid O_i(w)\}/\hat{K}\{\tau_i \mid O_i(\tau_i)\}}\widehat{\mathrm{E}}\left\{g(T_i) \mid U_i > \omega, H_{2ri}, A_{2ri}\right\}\Bigg] \\
&-\frac{I\{A_{1i} = d_1(H_{1i})\} - \hat{\pi}_{d_1}(H_{1i})}{\hat{\pi}_{d_1}(H_{1i})}\widehat{\mathrm{E}}\left[g\left\{T_i^*(d)\right\} \mid H_{1i}\right] \\
&+\frac{C_{d,1,i}}{\hat{\pi}_{C_{d,1}}(H_{1i})}\int_{0}^{\tau_i}\frac{d\hat{M}_i\{v \mid O_i(v)\}}{\hat{K}\{v \mid O_i(v)\}}\widehat{\mathrm{E}}\left[g\left\{T_i^*(d)\right\} \mid U_i^1 > v, H_{1i}\right].
\end{aligned}
\tag{3.11}
$$

This is an augmented IPW estimator which we will refer to as the CAIPW estimator to differentiate it from the AIPW estimator that does not have augmentation terms to capture back information from patients who are consistent with regime d, but are censored. According to theory for the monotone missingness problem, this estimator is consistent if either the censoring models and the propensity score models are correctly specified, or if the event models are correctly specified. In a clinical study, the propensity score models are known by design, and a well-executed study should only have administrative censoring due to time limitations. When this is the case, consistency of $\hat{V}^{\text{CAIPW}}(d)$ is guaranteed. Including all of the augmentation terms increases the efficiency of the estimator compared to the IPW and AIPW estimators.

3.5 Inference

The proposed method can estimate the value of any regime, but making inference about these values and comparing regimes is also of interest. To do this, we need a way to quantify error. We present a simple first attempt to estimate the variance under certain regularity conditions, with the understanding that these conditions may not hold in practice and improvements should be made in the future.

3.5.1 Variance Estimates

Recall that the value of a given regime d can be estimated by $\hat{V}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d)$. By the argument in Appendix B, the variance of the estimated value can be estimated by the standard sandwich variance formula,

$$\hat{\sigma}^2\left\{\hat{V}(d)\right\} = n^{-2}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d) - \hat{V}(d)\right\}^2. \tag{3.12}$$

As stated in Bai (2016), this result holds under the non exceptional law. In the one stage problem, this means that the set $\{h : \mathrm{CF}(h) = 0\}$ has probability zero. In the two stage problem, both sets $\{h_1 : \mathrm{CF}_1(h_1) = 0\}$ and $\{h_2 : \mathrm{CF}_2(h_2) = 0\}$ should have probability zero. This assumption is violated when there is no treatment effect, which means that under the null hypothesis, the variance result may not hold.

Semiparametric methods were used to estimate $V(d)$, and the variance result is contingent on the models being correctly specified. Bai et al. (2016) describe the effects on the sandwich variance estimator if these models are incorrectly specified.

The assumptions necessary to use these variance calculations may not be valid in some situations. We approach these estimates as a first attempt, which seem to work well in the simulation study presented. More work in the future is needed to explore and improve these estimators under violations of these assumptions. Bootstrap techniques such as those in Chakraborty et al. (2013) could be explored to alleviate these issues.

3.5.2 Testing

It is often of interest whether one treatment regime is better than another. A simple paired t-test is used to formally compare the estimated values of two regimes. When comparing regime d^1 and d^2, use test statistic

$$t = \frac{n^{-1}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d^1) - \widehat{\Psi}_i(d^2)\right\}}{\mathrm{SE}\left\{\widehat{\Psi}(d^1) - \widehat{\Psi}(d^2)\right\}/\sqrt{n}},$$

where SE is the standard error of the differences of the estimated Ψ functions. This test statistic follows the Student's t-distribution with $n-1$ degrees of freedom. This test will be used to compare the embedded regimes in the data analysis.

CHAPTER 4

ESTIMATING AN OPTIMAL REGIME

In this chapter, we discuss in more detail the value search method and classification perspective mentioned in Chapter 1. We then describe how to use these tools to estimate an optimal treatment regime.

4.1 Value Search Method

In Chapters 2 and 3 we show how to estimate the value of any given treatment regime in either a one stage or two stage study. However, the main objective is to find an optimal regime, or the set of rules that, if used to select treatment for the entire patient population, leads to the most favorable result on average. An estimate of the optimal regime would be the regime that has the highest estimated value of all regimes, assuming without loss of

generality that larger values of $g(T)$ are desirable. However, this is a highly complex problem when searching across all possible regimes in $\mathscr{D}$.

A different approach is to restrict the search to a specified class of regimes indexed by parameter η, denoted by $\mathscr{D}_\eta$, and then find the optimal regime in that class. This rule is denoted by $d_{\eta^{\text{opt}}}$ where $\eta^{\text{opt}} = \arg\max_\eta V(d_\eta)$. The true optimal rule may not be in this restricted class, but the optimal rule in $\mathscr{D}_\eta$ may still be of clinical interest. The restricted class can be chosen to make rules more interpretable for clinicians or to reduce cost. Let us focus on the one stage problem. For example, we can restrict the search to rules involving linear combinations of the covariates, which takes form $\mathscr{D}_\eta = \left\{d_\eta; d_\eta(H) = I\left(\eta^T \tilde{H} > 0\right)\right\}$ where $\eta = (\eta_0, \eta_1, ..., \eta_p)^T$ and p is the number of covariates in H. This rule is simple, understandable, and would be easy to put into practice.

Taking this approach is not a disadvantage compared to other methods. Define the mean outcome by $\mathrm{E}\left\{g(T) \mid H, A\right\} = Q(H, A)$ and the contrast function by $\mathrm{E}\left\{g(T) \mid H, A = 1\right\} - \mathrm{E}\left\{g(T) \mid H, A = 0\right\} = \mathrm{CF}(H)$. As stated in Chapter 1, Q-learning for the one stage problem involves positing a model for the mean outcome, $Q(H, A; \theta)$. The specification of this Q model implicitly restricts the search to the class of regimes with form $I\left\{Q(H, 1; \hat{\theta}) > Q(H, 0; \hat{\theta})\right\}$, which can be indexed by η through parameter θ. A-learning similarly posits a model for the contrast function, $\mathrm{CF}(H; \zeta)$, which again implicitly restricts the search to the class of regimes with form $I\left\{\mathrm{CF}(H; \hat{\zeta}) > 0\right\}$.

As an example, one could posit the model $Q(H, A; \theta) = \theta_0 + \theta_1 X_1 + A_1(\theta_2 + \theta_3 X_1)$ for the expected outcome in a one stage problem. Similarly, one could posit the corresponding model $\mathrm{CF}(H; \zeta) = \zeta_0 + \zeta_1 X_1$ for the contrast function. This implies that the estimated optimal regime will have form $I(\eta_0 + \eta_1 X_1 > 0)$, which defines the class $\mathscr{D}_\eta$. The parameter η_0 could

be estimated by $\hat{\theta}_2$ or $\hat{\zeta}_0$ and η_1 could be estimated by $\hat{\theta}_3$ or $\hat{\zeta}_1$ depending on which method is used.

If the Q or contrast function model is correctly specified, then the true optimal rule is in the restricted class of regimes indexed by η, $d^{\text{opt}} \in \mathscr{D}_\eta$. However, if the model is misspecified, then d^{opt} may or may not be in $\mathscr{D}_\eta$. The estimated optimal rule, $\hat{d}^{\text{opt}}$, may not even be the optimal rule in class $\mathscr{D}_\eta$ if using Q-learning. As the number of treatment decision points increases, specifying these models correctly gets increasingly difficult, which is why robust semiparametric methods are proposed in Chapters 2 and 3.

Again, the new approach is to search for the best rule within a restricted class of rules; that is, the rule that maximizes an estimator for the value $V(d_\eta) = \mathrm{E}\left[g\left\{T^*(d_\eta)\right\}\right]$ among all regimes in the restricted class, $d_\eta \in \mathscr{D}_\eta$. Thus, we refer to this as a value search method. This approach is more direct in stating the restrictions imposed on the form of the estimated optimal rule.

4.2 Classification Perspective

We describe semiparametric estimators for $V(d)$ for any given rule d in Chapters 2 and 3. Using the value search method described in the previous section, we can estimate the value of a rule parameterized by η by $\hat{V}(d_\eta)$ and we want to estimate $\hat{\eta}^{\text{opt}} = \arg\max_\eta \hat{V}(d_\eta)$. However, $\hat{V}(d_\eta)$ is a non-smooth, non-convex function, hence it is difficult to maximize. For a low dimensional parameter space, a grid search or genetic algorithm may be used (Zhang et al., 2012a; Zhang et al., 2013), however, this becomes computationally prohibitive as the number of parameters increases. This leads to the development of the classification perspective.

Estimators of the value can be written as

$$\begin{aligned}\hat{V}(d_\eta) &= n^{-1}\sum_{i=1}^{n}\Psi_i(d_\eta)\\ &= n^{-1}\sum_{i=1}^{n}\left[d_\eta(H_i)\Psi_i(1)+\left\{1-d_\eta(H_i)\right\}\Psi_i(0)\right]\\ &= n^{-1}\sum_{i=1}^{n}d_\eta(H_i)\{\Psi_i(1)-\Psi_i(0)\}+n^{-1}\sum_{i=1}^{n}\Psi_i(0).\end{aligned}$$

Thus, maximizing the estimated value in terms of η is equivalent to maximizing $n^{-1}\sum_{i=1}^{n} d_\eta(H_i)\widehat{\mathrm{CF}}_i$ where $\widehat{\mathrm{CF}}_i = \Psi_i(1)-\Psi_i(0)$ can be viewed as a predictor for the contrast function $\mathrm{CF}(H_i) = \mathrm{E}\left\{g(T_i)\mid H_i, A_i = 1\right\} - \mathrm{E}\left\{g(T_i)\mid H_i, A_i = 0\right\}$. Because $d_\eta(H)$ is a binary function, Zhang et al. (2012b) are able to prove the identity

$$\sum_{i=1}^{n} d_\eta(H_i)\widehat{\mathrm{CF}}_i = -\sum_{i=1}^{n}|\widehat{\mathrm{CF}}_i|\left\{I\left(\widehat{\mathrm{CF}}_i>0\right)-d_\eta(H_i)\right\}^2 + \text{ terms not including } \eta. \tag{4.1}$$

We can view $Z_i = I(\widehat{\mathrm{CF}}_i > 0)$ as a binary classifier label, and $W_i = |\widehat{\mathrm{CF}}_i|$ as the weight. Now, it is clear that $\mathrm{E}\left[W\left\{Z-d_\eta(H)\right\}^2\right]$ is the expected weighted misclassification error under rule d_η. Therefore, Zhang et al. (2012b) transform the optimization of the value into the classification problem

$$\hat{\eta}^{\mathrm{opt}} = \arg\min_{\eta}\sum_{i=1}^{n}|\widehat{\mathrm{CF}}_i|\left\{I\left(\widehat{\mathrm{CF}}_i>0\right)-d_\eta(H_i)\right\}^2. \tag{4.2}$$

We can heuristically motivate this formulation. If we knew the true contrast function, $I\{\mathrm{CF}(H_i)>0\}$ would be the best treatment that patient i should receive. Also, $|\mathrm{CF}(H_i)|$ would represent the loss if patient i is misclassified and receives the incorrect treatment.

Therefore, patient outcomes that differ more based on which treatment they receive would be weighted more in choosing an optimal rule.

In practice, solving (4.2) for the optimal rule only requires an estimate of the contrast function, $\widehat{\text{CF}}$. Recasting the optimization as a classification problem allows us to use well-studied classification techniques such as support vector machines (SVM) (Cortes and Vapnik, 1995) and classification and regression trees (CART) (Breiman et al., 1984). This simplifies computation and allows us to use efficient optimization algorithms and software packages that have already been developed. The non-smooth, non-standard optimization issues do not go away, but much work has already gone into efficient ways to handle these issues in the classification framework.

The choice of classification technique determines the restricted class of regimes that is searched when estimating an optimal regime. For example, linear SVM can estimate a linear rule of the form $d_\eta(H) = I\left\{\eta^T \tilde{H} > 0\right\}$ if only the variables in H are entered into the model. Therefore, this restricts the search for the optimal regime to the class $\mathscr{D}_\eta = \left\{d_\eta; d_\eta(H) = I\left(\eta^T \tilde{H} > 0\right)\right\}$. If a rule of form $d_\eta = I(\eta_1^T X_1 > 0) I(\eta_2^T X_2 > 0)$ was desired instead, CART can be used. The proposed method takes advantage of using this classification perspective. We choose to use SVM for illustration, though other classification methods can also be used.

4.3 Support Vector Machine (SVM)

In what follows, we consider the restricted class $\mathscr{D}_\eta = \left\{d_\eta; d_\eta(H) = I\left(\eta^T \tilde{H} > 0\right)\right\}$ and use SVM to estimate an optimal regime. Thus, we first give more technical details on this technique. The minimization of weighted classification error can be solved using weighted

SVM with hinge loss function, $\phi(x) = (1-x)^+ = \max(1-x, 0)$, and ℓ_1 penalization (Cortes and Vapnik, 1995). The linear support vector machine uses a hinge loss function instead of the 0-1 loss in (4.2) to make the problem convex and more computationally feasible (Zhao et al., 2012). Using the hinge loss function will result in a classification rule that is a consistent estimator of the true optimal rule $I\{\text{CF}(H) > 0\}$, proved in Appendix C. This property is called Fisher consistency. The weighted classification problem in (4.2) can be rewritten as

$$\hat{\eta}^{\text{opt}} = \arg\min_{\eta} \sum_{i=1}^{n} \left|\widehat{\text{CF}}_i\right| \phi\left(Y_i \eta^T \tilde{H}_i\right) + k \sum_{j=1}^{p} |\eta_j| \tag{4.3}$$

where $Y_i = 2\text{I}(\widehat{\text{CF}}_i \geq 0) - 1$ transforms the contrast function to take values $\{-1, 1\}$, used often in the area of classification, instead of $\{0, 1\}$. The vector $\eta = \left(\eta_0, \eta_1, ..., \eta_p\right)^T$ represents the intercept and coefficients of p covariates in H. This formulation assumes that the treatment rule has linear form, $d_\eta(H) = I(\eta^T \tilde{H} \geq 0)$. The ℓ_1 penalization can be used for variable selection and to prevent overfitting the rule to the data. The penalty term, k, can be chosen by cross validation.

For computational purposes, (4.3) can be converted into a linear convex optimization problem, seen in Appendix D (Bai et al., 2016). In this form, the system is easily solved using linear programming software. Only an estimate of the contrast function is needed for the SVM machinery to estimate an optimal rule.

4.4 Method Schema

Thus far, we have introduced how to estimate the value of a given regime d for both the one and two stage problems, and we have described the value search method and classification

perspective. We combine these ideas to formulate a method that estimates an optimal treatment regime for a two stage study.

The proposed method estimates an optimal dynamic treatment regime in a backward iterative fashion similar to that of Backward Outcome Weighted Learning (BOWL) in Zhao et al. (2015b). We present the steps of the proposed method here, assuming that data are available from a completed two stage SMART study.

1. **Estimate the second stage contrast function**

   Start at the second stage with the goal to estimate an optimal rule given that all past history H_2 is known. Only use patients who receive treatment at the second stage, $R > 0$. The second stage contrast function is

$$\mathrm{CF}_2(H_2) = \mathrm{E}\left[g\left\{T^*(A_1,1)\right\} \mid H_2\right] - \mathrm{E}\left[g\left\{T^*(A_1,0)\right\} \mid H_2\right].$$

   This represents the difference in expected outcome had the patient taken treatment 1 at the second stage instead of treatment 0 given all past history. Estimating each expectation is equivalent to estimating the value of a rule in the one stage problem. The patient-level contrast is estimated by $\widehat{\mathrm{CF}}_{2i} = \widehat{\Psi}_{2i}(1) - \widehat{\Psi}_{2i}(0)$. Estimate $\Psi_{2i}(1)$ and $\Psi_{2i}(0)$ by substituting rules $d_2(H_2) = 1$, "give treatment 1 at the second stage regardless of patient history", and $d_2(H_2) = 0$, "give treatment 0 at the second stage regardless of patient history", for $d(H)$ respectively into either (2.7) or (2.8) depending on whether the event time is censored or not. In these equations, use $g(T) - g(\tau)$ instead of $g(T)$ when $g(\cdot)$ is a function such as $g(T) = T$ or $g(T) = \min(T, L)$. Only additional life needs to be optimized because τ is fixed and known at the second stage. Using $g(T)$

only adds more noise with the addition of τ. All history H_2 can be used in place of H, and A_2 in place of A. Be sure to account for the different groups classified by R in the calculations if there are values of R greater than 1.

With censored data, substitute $\hat{K}(u \mid H_i, A_i)/\hat{K}(\tau_i \mid H_i, A_i)$ for each $\hat{K}(u \mid H_i, A_i)$ in (2.8). Only second stage patients are included this analysis, so it is known that patients were not censored before time τ. Similarly, only integrate the martingale term from τ_i to U_i instead of from 0 to infinity. It is also appropriate to integrate to infinity, but the martingale increment for patient i is 0 for all times greater than U_i.

2. **Use SVM to estimate an optimal second stage rule**

   At the second stage, a different rule is estimated for each $R = r > 0$. Once the second stage contrast function is estimated, input this into the SVM framework presented in Section 4.3. For each $r = 1, ..., B$, solve

$$\hat{\eta}_{2r} = \arg\min_{\eta_{2r}} \sum_{\{i:R_i=r\}} \left|\widehat{\mathrm{CF}}_{2i}\right| \phi\left(Y_{2i}\eta_{2r}^T \tilde{H}_{2i}\right) + k_{2r} \sum_{j=1}^{p_{2r}} |\eta_{2rj}|, \tag{4.4}$$

   where $Y_{2i} = 2\mathrm{I}\left(\widehat{\mathrm{CF}}_{2i} \geq 0\right) - 1$ and p_{2r} is the number covariates in H_{2r} when $R = r$. This results in an estimated rule at the second stage, denoted by $\hat{d}_{2r}(H_{2r}) = I(\hat{\eta}_{2r}^T \tilde{H}_{2r} \geq 0)$. Define the general second stage rule as $\hat{d}_2(H_2) = \sum_{r=1}^{B} I(R = r)\hat{d}_{2r}(H_{2r})$. Any patient's information can be input into this rule to estimate which treatment would have been optimal at the second stage given his or her history.

3. **Estimate the first stage contrast function**

   After estimating an optimal rule at the second stage, now move back to estimate an

optimal rule at the first stage. The first stage contrast function is

$$\mathrm{CF}_1(H_1) = \mathrm{E}\left[g\left\{T^*(1,\hat{d}_2)\right\} \mid H_1\right] - \mathrm{E}\left[g\left\{T^*(0,\hat{d}_2)\right\} \mid H_1\right].$$

This contrast function represents the difference in expected outcome had a patient received treatment 1 rather than treatment 0 at the first stage, assuming for both cases that treatment at the second stage will be assigned according to rule $\hat{d}_2$.

Estimating each expectation is equivalent to estimating the value of a regime for the two stage problem. The patient-level contrast is estimated $\widehat{\mathrm{CF}}_{1i} = \widehat{\Psi}_{1i}(1,\hat{d}_2) - \widehat{\Psi}_{1i}(0,\hat{d}_2)$. Estimate the Ψ functions by substituting regimes $d = (1,\hat{d}_2)$ and $d = (0,\hat{d}_2)$ respectively into either (3.10) or (3.11) depending on whether the event time is censored or not.

4. **Use SVM to estimate an optimal first stage rule**

   Once the first stage contrast functions are estimated, use SVM to obtain an estimator of the first stage rule, $\hat{d}_1(H_1) = I(\hat{\eta}_1^T \tilde{H}_1 > 0)$ by solving

$$\hat{\eta}_1 = \arg\min_{\eta_1} \sum_{i=1}^{n} \left|\widehat{\mathrm{CF}}_{1i}\right| \phi\left(Y_{1i}\eta_1^T \tilde{H}_{1i}\right) + k_1 \sum_{j=1}^{p_1} |\eta_{1j}|, \tag{4.5}$$

   where $Y_{1i} = 2\mathrm{I}\left(\widehat{\mathrm{CF}}_{1i} \geq 0\right) - 1$ and p_1 are the number of covariates in H_1. Again, any patient's information can be input into the rule to estimate which treatment would have been optimal at the first stage based on baseline covariates. At this step, an optimal treatment regime has been estimated, denoted by $\hat{d} = \left(\hat{d}_1, \hat{d}_2\right)$.

5. **Estimate the value of the estimated regime**

   Using the techniques in Chapter 3, the value of the estimated optimal regime $\hat{d}$ can

be estimated by substituting $\hat{d}$ into either equation (3.10) or (3.11) depending on whether the event time is censored or not, and calculating $\hat{V}(\hat{d}) = \sum_{i=1}^{n} \widehat{\Psi}_i(\hat{d})$.

4.5 Inference

In Section 3.5.1, we discuss how to obtain variance estimates for the estimated value of a fixed regime. More care needs to be taken when estimating the variance of the estimated value of the estimated optimal regime. Under certain regularity conditions, Zhang et al. (2012a) showed that

$$n^{\frac{1}{2}} \left\{ \hat{V}(\hat{d}^{\text{opt}}) - V(d^{\text{opt}}) \right\} = n^{\frac{1}{2}} \left\{ \hat{V}(d^{\text{opt}}) - V(d^{\text{opt}}) \right\} + o_p(1).$$

Because of this, the asymptotic variance of the value estimate, $\hat{V}(\hat{d}^{\text{opt}})$ can be estimated with the asymptotic variance of $\hat{V}(d^{\text{opt}})$. Recalling that the value of regime $\hat{d}^{\text{opt}}$ can be estimated by $\hat{V}(\hat{d}^{\text{opt}}) = n^{-1} \sum_{i=1}^{n} \widehat{\Psi}_i(\hat{d}^{\text{opt}})$, this can be estimated by the sandwich variance formula,

$$\hat{\sigma}^2 \left\{ \hat{V}(\hat{d}^{\text{opt}}) \right\} = n^{-2} \sum_{i=1}^{n} \left\{ \widehat{\Psi}_i(\hat{d}^{\text{opt}}) - \hat{V}(\hat{d}^{\text{opt}}) \right\}^2. \tag{4.6}$$

The issues discussed in Section 3.5.1 about the non exceptional law and correct model specification still hold. Again, the assumptions needed to use these variance calculations may not be valid in some situations. We approach these estimates as a first attempt, which seem to work well in the simulation study presented. More work in the future is needed to explore and improve these estimators under violations of these assumptions.

CHAPTER 5

SIMULATIONS AND DATA ANALYSIS

5.1 Simulation Study

To evaluate the proposed methods, we simulate data to represent a two stage SMART study where patients are randomized to one of two treatments with equal probability at each stage. Every patient who survives up to the second stage is eligible to receive the same set of treatments, so R only takes values 0 and 1.

5.1.1 Data Generation

We simulate a two stage SMART study with $n = 500$ patients. There are two baseline covariates, one standard normal and one uniform. There are two possible treatment options at the first stage, coded by 0 and 1. These two treatments are randomly assigned with equal

probability, and A_1 represents the first stage treatment that a patient receives. For each patient, we simulate the hazard of death, $\lambda_{10}(H_1, A_1)$, and hazard of surviving until the second stage treatment, $\lambda_{11}(H_1, A_1)$. These hazard functions are dependent on the baseline covariates and first stage treatment. Each time to event is a random exponential variable with corresponding hazard.

$$\begin{aligned} X_{11} \sim \mathrm{N}(\mu = 0, \sigma = 1), X_{12} \sim \mathrm{Uniform}(0,1) \\ A_1 \sim \mathrm{Bernoulli}(p = 0.5) \\ \lambda_{10}(H_1, A_1) = \exp\{-5.5 + 0.15X_{11} + 0.2X_{12} - A_1(1.3 + 1.55X_{11} - 2.6X_{12})\} \\ \lambda_{11}(H_1, A_1) = \exp\{-4.2 + 0.15X_{11} + 0.2X_{12} - A_1(-1.2 + 1.1X_{11} + 2.2X_{12})\} \\ \tau_{\mathrm{D}} \sim \mathrm{Exponential}\{\lambda_{10}(H_1, A_1)\}, \tau_{\mathrm{SS}} \sim \mathrm{Exponential}\{\lambda_{11}(H_1, A_1)\}, \tau = \min(\tau_{\mathrm{D}}, \tau_{\mathrm{SS}}) \end{aligned} \tag{5.1}$$

If time to death, τ_{D}, is greater than the time to second stage, τ_{SS}, then $\tau = \tau_{\mathrm{SS}}$ and the patient moves on to receive second stage treatment, $R = 1$. If time to death is less than or equal to the time to second stage, then $T = \tau = \tau_{\mathrm{D}}$ and the patient does not move on to the second stage, $R = 0$. In this simulation scenario, approximately 70% of patients survive long enough to receive a second stage treatment.

For patients who survive to the second stage, we generate a Bernoulli covariate, X_2, that occurs after the first treatment, but before the second treatment. This is dependent on the baseline covariates and first stage treatment. There are two possible treatment options at the second stage, coded by 0 and 1. These two treatments are randomly assigned with equal probability, and A_2 represents the second stage treatment that a patient receives. Additional life is the length of time that the patient survives past time τ. The additional life

hazard is dependent on all previous history.

$$\begin{aligned} X_2 &\sim \text{Bernoulli}(p = p_{X_2}) \text{ where } p_{X_2} = \text{expit}\{0.2 + 0.5X_{11} + 0.4X_{12} + 0.6A_1\} \\ A_2 &\sim \text{Bernoulli}(p = 0.5) \\ \lambda_2(H_2, A_2) &= \exp\{-3.5 + 0.12X_{11} - 0.42X_{12} - 0.52X_2 \\ &\quad - A_2(0.55 + 0.8X_{11} - 0.6X_{12} - 0.2A_1 - 0.65X_2)\} \\ \text{AL} &\sim \text{Exponential}\{\lambda_2(H_2, A_2)\} \end{aligned} \tag{5.2}$$

The event time for patients with $R = 1$ is $T = \tau + \text{AL}$. We are interested in mean restricted lifetime with restriction time $L = 150$, so we calculate $g(\tau) = \min(\tau, L)$ and $g(T) = \min(T, L)$.

Censoring is assumed to be noninformative. We generate censoring time, C, to be independent of all observed and unobserved data. Let $U = \min(T, C)$ and $U^1 = \min(\tau, C)$. The variables $\Delta = I\left\{g(T) \leq C\right\}$ and $\Gamma = I\left\{g(\tau) \leq C\right\}$ are indicators that the event times are observed prior to censoring. We examine how the method performs under various censoring distributions presented in Table 5.1.

Table 5.1 Censoring distributions used in simulations.

Cens. Scheme	Distribution
1	$\lambda_C = \exp(0.003)$, $C \sim \text{Exponential}(\lambda_C)$
2	$\lambda_C = \exp(0.008)$, $C \sim \text{Exponential}(\lambda_C) + 50$
3	$C \sim \text{Uniform}(0, 400)$
4	$C \sim \text{Uniform}(50, 250)$

Censoring schemes 2 and 4 were created to simulate the situation where there is a lag in censoring, or censoring does not occur for the first 50 units in time. Table 5.2 presents

the approximate percentage of censored patients for each of these distributions.

Table 5.2 Approximate percentage of censored patients for each distribution in Table 5.1. Overall presents the average percentage of censored patients across both stages. First stage presents the average percentage of censored patients in the first stage. Second stage presents the average percentage of censored patients in the second stage among patients who survive up to the second stage, or patients with $R = 1$.

Cens. Scheme	Overall (%)	First Stage (%)	Second Stage (%)
1	19	12	12
2	21	10	17
3	19	11	12
4	18	8	15

5.1.2 True Optimal Regime

The proposed method aims to estimate the optimal treatment regime. The performance of this estimated optimal regime should be compared to the true optimal regime. At the second stage, it is clear that the optimal treatment rule is $I(0.55+0.8X_{11}-0.6X_{12}-0.2A_1-0.65X_2>0)$ because a smaller hazard is better. We normalize the parameter vector of the treatment rule so that there is identifiability and we can compare treatment rules. For parameter vector $\eta=(\eta_0,...,\eta_p)$, the normalized vector is $\eta/|\eta|$ where $|\eta|=(\eta_0^2+...+\eta_p^2)^{1/2}$ is the vector norm. After normalization, the optimal second stage rule is $d_2^{\text{opt}}(H_2)=I(0.41+0.60X_{11}-0.45X_{12}-0.15A_1-0.49X_2>0)$.

At the first stage, time to death and time to the second stage treatment are competing risks, and it is not obvious what the optimal rule is. To find the optimal rule, d_1^{opt}, we conduct a grid search over all possible normalized treatment rules of the form $I(\eta_{10}+$

$\eta_{11}X_{11}+\eta_{12}X_{12}>0$). Because the normalized rule requires that $\eta_{10}^2+\eta_{11}^2+\eta_{12}^2=1$, we restrict the search to the surface of the unit sphere.

First, we create a grid to search in terms of two parameters as seen in Figure 5.1. For illustration purposes, the sample depicted in Figure 5.1 is five times more sparse than the sample actually used. The absolute value of third parameter is determined by the normalization restriction. Both the positive and negative values of the third parameter are checked, so each point in Figure 5.1 represents two different rules.

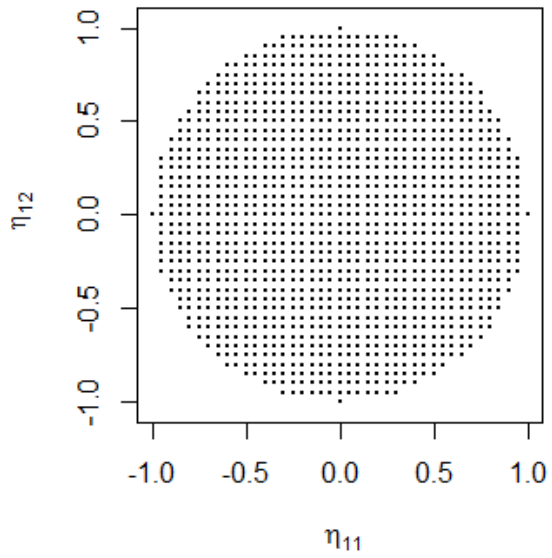

Figure 5.1 Grid search in two of three normalized parameters in the first stage rule.

We use $m=1,000$ Monte Carlo simulations to obtain the value of each rule covered

by the grid. We generate m new data sets as described in (5.1) and (5.2), except use a rule from the grid to determine which treatment each patient receives at the first stage, and use the optimal rule to determine which treatment each patient receives at the second stage. We then calculate the value, or the mean restricted lifetime for each data set. This is called g-computation (Robins, 1986). The optimal rule is the one with the largest Monte Carlo average value.

Figure 5.2 is a heat map of the results from this simulation. The red areas on the sphere indicate rules that have higher values while white represents rules with lower values. This search technique found that the optimal rule at the first stage is $d_1^{\text{opt}}(H_1) = I(-0.197 + 0.98X_{11} + 0.03X_{12} > 0)$.

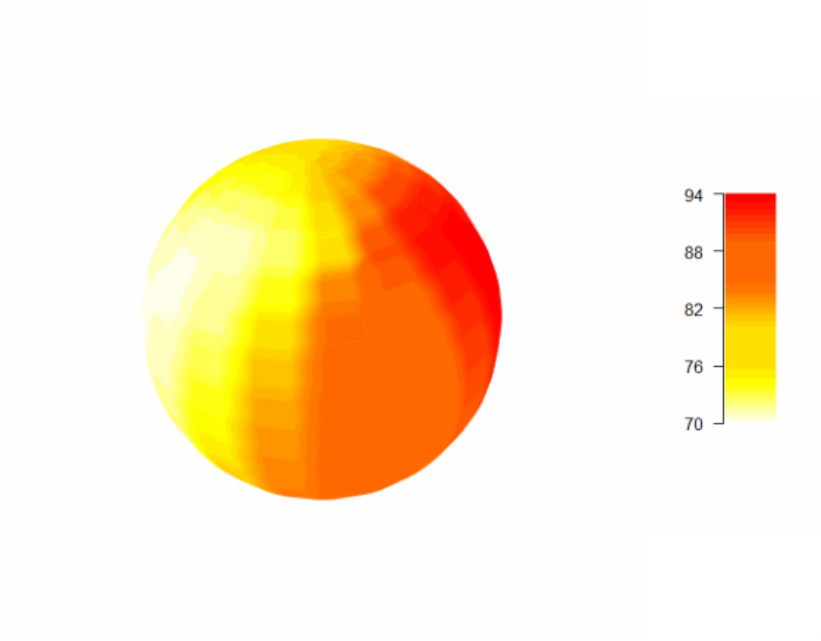

Figure 5.2 Monte Carlo average value across unit sphere of possible first stage rules. Red indicates rules that have higher values while white represents rules with lower values.

Using 100,000 more Monte Carlo simulations, the mean restricted lifetime, or the value of the optimal treatment regime is 93.9. If the optimal regime were to be used to assign treatment, about 40% of patients in the simulation should receive treatment 1 and about 60% of patients should receive treatment 0 at each stage. A traditional clinical trial testing whether one treatment is better than the other might not pick up any treatment effect, or would recommend giving the same treatment to every patient. In the latter case, the treatment would be correct for only 60% of patients if the better treatment is chosen.

5.1.3 Modeling Choices

We have shown how to generate the data for the analysis. To implement the proposed method, we need to posit and estimate event hazard models, propensity scores, and censoring hazard models. The following list presents the model choices for implementation of the method used in the simulations:

- Event hazard functions at the second stage for the no censoring and censoring cases:

$$\lambda_2(u \mid H_2, A_2) = \lim_{du \to 0} (du)^{-1} P(u \le \mathrm{AL} < u + du \mid \mathrm{AL} \ge u, H_2, A_2, R = 1)$$
$$\lambda_2(u \mid H_2, A_2) = \lim_{du \to 0} (du)^{-1} P(u \le U - U^1 < u + du, \Delta = 1 \mid U - U^1 \ge u, H_2, A_2, R = 1).$$

  Posit a Cox proportional hazards model:

$$\lambda_2(u \mid H_2, A_2, \beta_2) = \lambda_{20}(u) \exp\left\{\beta_{21} X_{11} + \beta_{22} X_{12} + \beta_{23} A_1 + \beta_{24} X_2 + A_2(\beta_{25} + \beta_{26} X_{11} + \beta_{27} X_{12} + \beta_{28} A_1 + \beta_{29} X_2)\right\},$$

  where $\lambda_{20}(u)$ is an unspecified baseline hazard function.

- Cause-specific event hazard functions at the first stage for each R=r when r=0,1 for the no censoring and censoring cases:

$$\lambda_{1r}(u \mid H_1, A_1) = \lim_{du \to 0} (du)^{-1} P(u \leq \tau < u + du, R = r \mid \tau \geq u, H_1, A_1)$$
$$\lambda_{1r}(u \mid H_1, A_1) = \lim_{du \to 0} (du)^{-1} P\left(u \leq U^1 < u + du, R = r, \Gamma = 1 \mid U^1 \geq u, H_1, A_1\right).$$

  Posit a cause-specific Cox proportional hazards model:

$$\lambda_{1r}(u \mid H_1, A_1, \beta_{1r}) = \lambda_{10r}(u) \exp\left\{\beta_{11r} X_{11} + \beta_{12r} X_{12} + A_1\left(\beta_{13r} + \beta_{14r} X_{11} + \beta_{15r} X_{12}\right)\right\},$$

  where $\lambda_{10r}(u)$ is an unspecified baseline hazard function.

- Propensity score at the second stage: Sample mean, $\hat{\pi}_2(H_2) = (\sum_{i=1}^n R_i)^{-1} \sum_{\{i:R_i=1\}} A_{2i}$.

- Propensity score at the first stage: Sample mean, $\hat{\pi}_1(H_1) = n^{-1} \sum_{i=1}^n A_{1i}$.

- Censoring hazard function:

$$\alpha_2\{u \mid O(u)\} = \lim_{du \to 0} (du)^{-1} P\{u \leq U < u + du, \Delta = 0 \mid U \geq u, O(u)\}.$$

  Posit a Cox proportional hazards model:

$$\alpha_2\left(u \mid H_1, \gamma_2\right) = \alpha_{20}(u) \exp\left\{\gamma_{21} X_{11} + \gamma_{22} X_{12} + A_1(\gamma_{23} + \gamma_{24} X_{11} + \gamma_{25} X_{12})\right\}$$
$$\hat{K}_2(u \mid H_1) = \exp\left\{-\int_0^u \alpha_2(w \mid H_1, \hat{\gamma}_2) dw\right\},$$

  where $\alpha_{20}(u)$ is an unspecified baseline hazard function.

- First stage censoring hazard function:

$$\alpha_1\{u \mid O(u)\} = \lim_{du\to 0}(du)^{-1} P\left\{u \le U^1 < u+du, \Gamma = 0 \mid U^1 \ge u, O(u)\right\}.$$

  Posit a Cox proportional hazards model:

$$\alpha_1\left(u \mid H_1, \gamma_1\right) = \alpha_{10}(u)\exp\left\{\gamma_{11}X_{11} + \gamma_{12}X_{12} + A_1(\gamma_{13} + \gamma_{14}X_{11} + \gamma_{15}X_{12})\right\}$$
$$\hat{K}_1(u \mid H_1) = \exp\left\{-\int_0^u \alpha_1(w \mid H_1, \hat{\gamma}_1)dw\right\},$$

  where $\alpha_{10}(u)$ is an unspecified baseline hazard function.

The true censoring distribution is not dependent on any patient covariates. This is nested within the chosen models. The censoring survival distribution, $K\{u \mid O(u)\}$, is estimated by $\hat{K}_1(\tau \mid H_1)\hat{K}_2(u \mid H_1)/\hat{K}_2(\tau \mid H_1)$ when $u > \tau$ for better accuracy. If $u \le \tau$, then we estimate $K\{u \mid O(u)\}$ with $\hat{K}_1(u \mid H_1)$.

We also note that we do not use any penalization term when using SVM. Several tuning parameters were tried, but the penalization does not make a difference, as the number of covariates is low.

5.1.4 Results - No Censoring

We first present simulation results for the case where there is no censoring of the event time. The outcome of interest is mean restricted lifetime with $L = 150$. We compare two different strategies: the AIPW and IPW estimator for the contrast function in the SVM framework. The BOWL method in Zhao et al. (2015b) does not account for the possibility that an event can occur before the second stage treatment. However, the proposed IPW method is similar

in spirit to BOWL, but does adjust for this possibility. See Appendix E for details.

We compare how the estimated regimes perform for each method by the following:

1. Simulate $s = 1,000$ data sets and estimate a treatment regime for each data set using each method. Calculate the estimated value of each estimated regime, $\hat{V}(\hat{d}_l)$, $l = 1,...,s$, as described in step 5 of Section 4.4 and the estimated standard error as described in Section 4.5. Then $\hat{V}(\hat{d}) = s^{-1}\sum_{l=1}^{s}\hat{V}(\hat{d}_l)$.

2. For each estimated rule, $l = 1,...,s$, use the g-computation algorithm on $m = 10,000$ simulated data sets to obtain the value of the estimated regime for each data set, $V_j(\hat{d}_l)$, $j = 1,...,m$. See Appendix F for details on g-computation.

3. The Monte Carlo average value of each regime $l = 1,...,s$ is $V(\hat{d}_l) = m^{-1}\sum_{j=1}^{m}V_j(\hat{d}_l)$. Then $V(\hat{d}) = s^{-1}\sum_{l=1}^{s}V(\hat{d}_l)$

The value of the estimated regime is a measure that can only be obtained by simulation, however both the estimated value and estimated standard error can be obtained in practice. These results can be seen in Table 5.3. For comparison, the value of the optimal rule is 93.9.

Table 5.3 Value simulation results using 1000 Monte Carlo data sets for the AIPW and IPW methods in the no censoring case. $V(\hat{d})$ presents the Monte Carlo average and standard deviation of the value of the estimated regime using 10000 Monte Carlo data sets for each of the 1000 estimated regimes. $\hat{V}(\hat{d})$ presents the Monte Carlo average and standard deviation of the estimated value of the estimated regime. SE presents the Monte Carlo average and standard deviation of the estimated standard error of $\hat{V}(\hat{d})$. Coverage is based on 95% Wald-type confidence intervals for $V(d^{\text{opt}}) = 93.9$.

Method	$V(\hat{d})$	$\hat{V}(\hat{d})$	SE	Coverage
AIPW	92.5 (1.7)	94.7 (3.6)	3.7 (0.2)	0.96
IPW	89.6 (3.5)	96.6 (7.2)	7.8 (0.4)	0.96

The AIPW method performs better in terms of the value of the estimated regime, and also has a much smaller standard error. The value and estimated value of the regime are the best measures of performance to evaluate the methods because the goal is to maximize the value. However, we can also examine how often the estimated rules agrees with the optimal rules, seen in Table 5.4. This is computed for the $m = 10,000$ g-computation simulations in step 2 for each of the $s = 1,000$ estimated rules.

In addition to presenting each stage separately, we present the average proportion of patients for whom the treatments selected by both the estimated first and second stage rules agree with those selected by the optimal regime for a patient. This is only calculated for patients who receive a second stage treatment.

Table 5.4 Agreement with the optimal rule using 10000 Monte Carlo data sets for each of the 1000 estimated regimes using the AIPW and IPW methods in the no censoring case. Agree Stage 1 and Agree Stage 2 present the Monte Carlo average proportion of patients for whom the treatment selected by estimated rule agrees with that selected by the optimal rule in each stage. Both Stages presents the Monte Carlo average proportion of second stage patients, $R = 1$, for whom both treatments selected by the estimated regime agrees with those selected by the optimal regime. Agree Overall presents the Monte Carlo average proportion of patients for whom the either treatments selected by the estimated regime agrees with those selected by the optimal regime when $R = 1$ or the treatment selected by estimated first stage rule agrees with that selected by the optimal first stage rule when $R = 0$.

Method	Agree Stage 1	Agree Stage 2	Both Stages	Agree Overall
AIPW	0.90	0.87	0.78	0.80
IPW	0.79	0.80	0.63	0.66

We also present the average proportion of patients for whom the treatments selected by the estimated regime agrees with those selected by the optimal regime for as long as the patient is alive. For patients included only in the first stage analysis, this is when the

treatment selected by the estimated first stage rule agrees with the treatment selected by the optimal first stage rule. For patients in the second stage analysis, this is when treatments selected by both estimated rules agree with those selected by the optimal regime. This is a measure of how often we expect to treat a patient in total agreement with the optimal rule. Clearly, the AIPW method performs much better than the IPW method across all measures.

One last measure we can examine is the how well each method estimates the coefficients in the regime. The results for the second and first stage parameters are presented in Tables 5.5 and 5.6 respectively.

Table 5.5 Estimated second stage rule using 1000 Monte Carlo data sets for the AIPW and IPW methods in the no censoring case. $\hat{\eta}_{20}, \hat{\eta}_{21}, \hat{\eta}_{22}, \hat{\eta}_{23}$, and $\hat{\eta}_{24}$ present the Monte Carlo average parameter estimates with Monte Carlo standard deviations in parentheses for each method. The optimal parameters are $(\eta_{20}, \eta_{21}, \eta_{22}, \eta_{23}, \eta_{24}) = (0.41, 0.60, -0.45, -0.15, -0.49)$.

Method	$\hat{\eta}_{20}$	$\hat{\eta}_{21}$	$\hat{\eta}_{22}$	$\hat{\eta}_{23}$	$\hat{\eta}_{24}$
AIPW	0.32 (0.21)	0.55 (0.17)	-0.37 (0.29)	-0.14 (0.20)	-0.45 (0.20)
IPW	0.23 (0.33)	0.45 (0.23)	-0.29 (0.39)	-0.13 (0.30)	-0.39 (0.30)

Table 5.6 Estimated first stage rule using 1000 Monte Carlo data sets for the AIPW and IPW methods in the no censoring case. $\hat{\eta}_{10}, \hat{\eta}_{11}$, and $\hat{\eta}_{12}$ present the Monte Carlo average parameter estimates with Monte Carlo standard deviations in parentheses for each method. The optimal first stage parameters are $(\eta_{10}, \eta_{11}, \eta_{12}) = (-0.197, 0.98, 0.03)$.

Method	$\hat{\eta}_{10}$	$\hat{\eta}_{11}$	$\hat{\eta}_{12}$
AIPW	-0.18 (0.33)	0.77 (0.22)	-0.17 (0.43)
IPW	-0.27 (0.49)	0.48 (0.33)	-0.17 (0.56)

The AIPW method yields more accurate estimates of the parameters with lower standard deviations in both stages. These results show that the AIPW method is an improvement over the IPW method, and thus BOWL, when there is no censoring of the event time.

5.1.5 Results - Censoring

We also examine how the proposed method performs when there is censoring of the event time. In Chapters 2 and 3, we present three different consistent estimators: IPW, AIPW, and CAIPW. We present the simulation results including the value, estimated value, estimated standard error, and coverage in Table 5.7 just as in the case without censoring. For comparison, we include the AIPW estimator for the no censoring case in the results as well. Again, the value obtained by the optimal regime is 93.9.

Table 5.7 Value simulation results using 1000 Monte Carlo data sets for the CAIPW, AIPW, and IPW methods in the censoring case. Cens. Scheme is the distribution from which the censoring time is generated according to Table 5.1. All other quantities are analogous to those in Table 5.3.

Cens. Scheme	Method	$V(\hat{d})$	$\hat{V}(\hat{d})$	SE	Coverage
None	AIPW	92.5 (1.7)	94.7 (3.6)	3.7 (0.2)	0.96
1	CAIPW	92.1 (2.1)	94.3 (4.0)	4.0 (0.2)	0.96
	AIPW	90.5 (3.0)	95.0 (5.4)	6.8 (0.6)	0.99
	IPW	88.4 (4.0)	97.6 (8.4)	10.2 (0.8)	0.98
2	CAIPW	92.2 (2.0)	93.9 (4.0)	4.1 (0.2)	0.96
	AIPW	89.2 (3.6)	95.8 (6.6)	9.4 (1.2)	0.997
	IPW	87.8 (4.1)	98.4 (9.3)	12.2 (1.4)	0.99
3	CAIPW	92.2 (2.0)	94.4 (3.8)	4.0 (0.2)	0.96
	AIPW	90.4 (3.1)	95.0 (5.4)	7.0 (0.6)	0.99
	IPW	88.6 (3.8)	98.0 (8.5)	10.2 (0.8)	0.97
4	CAIPW	92.3 (1.9)	94.0 (3.9)	4.0 (0.2)	0.95
	AIPW	89.6 (3.5)	95.8 (6.3)	8.6 (1.0)	0.99
	IPW	88.3 (4.0)	99.0 (9.0)	11.5 (1.1)	0.98

These results show that the CAIPW estimator performs the best and achieves a value very close to the value that the AIPW estimator achieves in the no censoring case. The CAIPW estimator is the least biased and has lower standard errors compared to the AIPW and IPW estimators. The coverage is close to the desired 0.95 for the CAIPW estimator, but the AIPW and IPW estimators have higher coverage than desired for all of the censoring distributions. This is because the estimated standard errors are larger than the Monte Carlo standard deviation, so even though the value estimators are more biased, the large estimated standard error makes the coverage high.

As in the simulations without censoring, we also examine how often the estimated regime agrees with the true optimal regime in Table 5.8.

Table 5.8 Agreement with the optimal rule using 10000 Monte Carlo data sets for each of the 1000 estimated regimes using the CAIPW, AIPW, and IPW methods in the censoring case. Cens. Scheme is the distribution from which the censoring time is generated according to Table 5.1. All other quantities are analogous to those in Table 5.4.

Cens. Scheme	Method	Agree Stage 1	Agree Stage 2	Both Stages	Agree Overall
None	AIPW	0.90	0.87	0.78	0.80
1	CAIPW	0.88	0.85	0.75	0.78
	AIPW	0.82	0.80	0.67	0.69
	IPW	0.76	0.76	0.58	0.61
2	CAIPW	0.88	0.85	0.75	0.78
	AIPW	0.78	0.76	0.59	0.63
	IPW	0.74	0.74	0.56	0.59
3	CAIPW	0.88	0.85	0.75	0.78
	AIPW	0.81	0.80	0.65	0.69
	IPW	0.76	0.77	0.59	0.62
4	CAIPW	0.88	0.86	0.76	0.78
	AIPW	0.79	0.78	0.62	0.65
	IPW	0.75	0.76	0.58	0.61

Table 5.8 shows that as more augmentation terms are added, the estimated rules agree more with the optimal rules. Results were similar for the CAIPW estimator for all of the censoring distributions, but the AIPW estimator performs best when the censoring is not lagged. The CAIPW estimator performs almost as well as the AIPW estimator in the no censoring case. This shows that the estimator does well in capturing back the information lost from censoring.

The last evaluation we present is the parameter estimates. Tables 5.9 and 5.10 show the parameter estimates and standard deviations for the second and first stage rules respectively.

Table 5.9 Estimated second stage rule using 1000 Monte Carlo data sets for the CAIPW, AIPW, and IPW methods in the censoring case. Cens. Scheme is the distribution from which the censoring time is generated according to Table 5.1. All other quantities are analogous to those in Table 5.5. The optimal second stage parameters are $(\eta_{20}, \eta_{21}, \eta_{22}, \eta_{23}, \eta_{24}) = (0.41, 0.60, -0.45, -0.15, -0.49)$.

Cens.	Method	$\hat{\eta}_{20}$	$\hat{\eta}_{21}$	$\hat{\eta}_{22}$	$\hat{\eta}_{23}$	$\hat{\eta}_{24}$
None	AIPW	0.32 (0.21)	0.55 (0.17)	-0.37 (0.29)	-0.14 (0.20)	-0.45 (0.20)
1	CAIPW	0.31 (0.25)	0.54 (0.17)	-0.33 (0.32)	-0.15 (0.22)	-0.44 (0.21)
	AIPW	0.24 (0.35)	0.47 (0.23)	-0.26 (0.38)	-0.14 (0.28)	-0.40 (0.28)
	IPW	0.17 (0.41)	0.40 (0.26)	-0.20 (0.42)	-0.13 (0.33)	-0.36 (0.33)
2	CAIPW	0.31 (0.25)	0.54 (0.17)	-0.33 (0.32)	-0.15 (0.22)	-0.44 (0.20)
	AIPW	0.17 (0.42)	0.40 (0.25)	-0.22 (0.43)	-0.13 (0.32)	-0.35 (0.32)
	IPW	0.15 (0.44)	0.37 (0.25)	-0.19 (0.44)	-0.11 (0.34)	-0.33 (0.35)
3	CAIPW	0.30 (0.26)	0.53 (0.18)	-0.36 (0.32)	-0.14 (0.22)	-0.43 (0.21)
	AIPW	0.20 (0.39)	0.46 (0.24)	-0.28 (0.38)	-0.13 (0.28)	-0.38 (0.29)
	IPW	0.16 (0.40)	0.42 (0.25)	-0.23 (0.43)	-0.13 (0.33)	-0.33 (0.32)
4	CAIPW	0.31 (0.24)	0.54 (0.17)	-0.36 (0.31)	-0.14 (0.22)	-0.44 (0.20)
	AIPW	0.18 (0.42)	0.43 (0.24)	-0.26 (0.40)	-0.13 (0.29)	-0.36 (0.31)
	IPW	0.14 (0.43)	0.40 (0.25)	-0.22 (0.44)	-0.12 (0.32)	-0.31 (0.34)

Table 5.10 Estimated first stage rule using 1000 Monte Carlo data sets for the CAIPW, AIPW, and IPW methods in the censoring case. Cens. Scheme is the distribution from which the censoring time is generated according to Table 5.1. All other quantities are analogous to those in Table 5.6. The optimal first stage parameters are $(\eta_{10}, \eta_{11}, \eta_{12}) = (-0.197, 0.98, 0.03)$.

Cens. Scheme	Method	$\hat{\eta}_{10}$	$\hat{\eta}_{11}$	$\hat{\eta}_{12}$
None	AIPW	-0.18 (0.33)	0.77 (0.22)	-0.17 (0.43)
1	CAIPW	-0.17 (0.37)	0.72 (0.26)	-0.19 (0.46)
	AIPW	-0.25 (0.45)	0.59 (0.32)	-0.19 (0.50)
	IPW	-0.25 (0.53)	0.43 (0.34)	-0.18 (0.57)
2	CAIPW	-0.16 (0.36)	0.73 (0.25)	-0.19 (0.46)
	AIPW	-0.27 (0.50)	0.50 (0.34)	-0.16 (0.54)
	IPW	-0.24 (0.55)	0.40 (0.34)	-0.16 (0.59)
3	CAIPW	-0.19 (0.36)	0.73 (0.25)	-0.16 (0.46)
	AIPW	-0.28 (0.45)	0.58 (0.32)	-0.16 (0.51)
	IPW	-0.29 (0.53)	0.43 (0.34)	-0.15 (0.57)
4	CAIPW	-0.19 (0.36)	0.74 (0.24)	-0.16 (0.45)
	AIPW	-0.28 (0.49)	0.50 (0.33)	-0.17 (0.53)
	IPW	-0.26 (0.53)	0.42 (0.34)	-0.15 (0.57)

Again, the CAIPW estimator performs best in terms of bias and has the lowest standard deviations. These estimates also get close to the AIPW estimates in the no censoring case.

The value, estimated value, standard errors, agreement with the optimal regime, and parameter estimates all show that the CAIPW estimator performs the best compared to the AIPW and IPW estimators in the case where the event time can be censored. The CAIPW performance also comes close to the AIPW estimator in the no censoring case, indicating that the extra augmentation terms do well in capturing back information lost due to censoring.

5.2 Data Analysis

We now apply the proposed methods to the North American Leukemia Intergroup Study C9710 which studies patients with acute promyelocytic leukemia as described in Chapter 1. The primary endpoint in this study is event-free survival at 3 years, where an event is defined as failure to achieve complete remission (CR), relapse after achieving CR, or death, whichever comes first.

Patients all received the same induction therapy. Then, there were two different times at which patients were randomized to different treatment options. Patients were first randomized to a consolidation therapy of either all-trans-retinoic acid (ATRA) alone, coded as 0, or ATRA in combination with arsenic trioxide, coded as 1. Patients who did not experience an event or censoring before the second stage were randomized again to a maintenance therapy of either ATRA alone, coded as 0, or ATRA in combination with oral methotrexate (Mtx) and mercaptopurine (MP), coded as 1 (Powell et al., 2010).

Baseline covariates available on patients include age, gender, race, ethnicity (Hispanic vs non-Hispanic), ECOG performance status, risk group, white blood cell count (WBC), platelet count, serum creatinine, and hemoglobin. Second stage covariates include any adverse events and toxicities that occurred. In the following analyses, only adverse events that are experienced by at least 5% of patients are considered. There are only 23 such adverse events, and a table of these with their 8-digit Medical Dictionary for Regulatory Activities (medDRA) code is in Appendix G.2.

The adverse event variables are coded to include the stage in which the toxicity occurs followed by the 8-digit medDRA code of the toxicity or adverse event. “I” means the adverse

event occurred during the induction phase. "C" means the adverse event occurred during the consolidation phase. "IC" means the adverse event occurred either during the induction or consolidation phases. For example, IC10028813 represents the adverse event of nausea occurring during either induction or consolidation treatment.

Powell et al. (2010) evaluate this trial by testing the treatments at each stage separately. Kaplan-Meier estimates of the survival distributions for each consolidation treatment are compared using log-rank tests. The paper determines that the consolidation therapy of ATRA and arsenic is better than ATRA alone for 3 year event-free survival. There are not enough events to power the analysis to determine a difference in maintenance therapies, though there is some suggestion that ATRA in combination with MP and Mtx is better.

We seek to check if the proposed methods can confirm the results about the consolidation therapy in Powell et al. (2010), and also determine if any further conclusions can be made. The data are modified as detailed in Appendix G.1. After this data cleaning, there are 468 patients in the first stage and 312 patients in the second stage available for analysis.

5.2.1 No Censoring Case

We examine mean restricted survival time with a cutoff time of $L = 1100$ days, or about 3 years, to be comparable to the outcome evaluated in Powell et al. (2010). There is very little censoring of the event times of the patients before $L = 1100$: 29 patients (6%) overall, 14 patients (3%) in the first stage, and 15 patients (5%) in the second stage. We adapt the data in order to use the proposed no censoring method. The 8 patients whose censoring time is below 600 are dropped from the analysis, and patients whose censoring time is larger than 600 have their event time set to $L = 1100$. Thus, there are 460 patients analyzed in the first

stage and 309 patients analyzed in the second stage.

The event hazard functions at each stage and the propensity scores must be estimated. The propensity scores are estimated by the sample proportion of patients who received treatment 1 at each stage. We posit a Cox proportional hazards model for the additional life hazard function, $\lambda_2(t \mid H_2, A_2)$. All baseline covariates, adverse events, treatments, and all interactions with treatment at the second stage are considered. Stepwise variable selection is used to reduce the number of variables. After variable selection, we estimate the following model:

$$\lambda_2(t \mid H_2, A_2, \beta_2) = \lambda_{20}(t) \exp\big\{\beta_{21} A_1 + \beta_{22}\text{Risk.group} + \beta_{23}\text{ECOG.Performance.Status} + \beta_{24} A_2 + \beta_{25}\text{IC10028813} + \beta_{26}\text{I10012457} + \beta_{27}\text{C10035528} + \beta_{28}(A_2 \times \text{IC10028813})\big\}.$$

Similarly, the cause-specific first stage event hazard functions, $\lambda_{1r}(t \mid H_1, A_1)$, need to be estimated for $R = r$ when $r = 0, 1$. We posit a Cox proportional hazards model starting with all baseline covariates, first stage treatment, and interactions involving first stage treatment. Stepwise variable selection is used to reduce the number of variables, and we estimate the following models:

$$\lambda_{10}(t \mid H_1, A_1, \beta_{10}) = \lambda_{10,0}(t) \exp(\beta_{10,1}\text{WBC} + \beta_{10,2}\text{ECOG.Performance.Status})$$

$$\lambda_{11}(t \mid H_1, A_1, \beta_{11}) = \lambda_{11,0}(t) \exp\big\{\beta_{11,1} A_1 + \beta_{11,2}\text{Race} + \beta_{11,3}\text{Age} + \beta_{11,4}\text{Ethnicity} + \beta_{11,5}\text{Hemoglobin} + \beta_{11,6}\text{Platelet} + \beta_{11,7}(A_1 \times \text{Race})\big\}.$$

SVM is used to estimate optimal treatment rules. We use 10-fold cross validation repeated 10 times to choose the tuning parameters k_1 and k_2 in (4.5) and (4.4) respectively. 10-fold cross validation partitions the data into 10 different parts. Nine of those parts make

up the training set used to estimate the treatment rule parameters. The remaining partition, the test set, is used the calculate the weighted classification error when using the parameters estimated by the training set. This is done a total of 10 times so each partition is used as a test set once.

We repeat this 10-fold cross validation a total of 10 times for each tuning parameter, using different partitions of the data each time. The tuning parameter chosen is the one with the lowest weighted classification error, defined for the second stage as

$$k_2 = \arg\min_k \sum_{l=1}^{10}\sum_{j=1}^{10}\sum_{\{i\in \text{test}_{lj}\}} \left[W_{2i} \left\{ \text{sign}(\hat{\eta}_{2,k,l,j}^T \tilde{H}_{2i}) - Y_{2i} \right\}^2 \right],$$

where $W_{2i} = |\widehat{\text{CF}}_{2i}|$, $Y_{2i} = 2\widehat{\text{CF}}_{2i} - 1$ are calculated for the j^{th} test set in the l^{th} 10-fold cross validation repetition, and $\hat{\eta}_{2,k,l,j}$ are the parameters estimated using the j^{th} training set in the l^{th} 10-fold cross validation repetition with tuning parameter k. Similarly at the first stage,

$$k_1 = \arg\min_k \sum_{l=1}^{10}\sum_{j=1}^{10}\sum_{\{i\in \text{test}_{lj}\}} \left[W_{1i} \left\{ \text{sign}(\hat{\eta}_{1,k,l,j}^T \tilde{H}_{1i}) - Y_{1i} \right\}^2 \right],$$

where $W_{1i} = |\widehat{\text{CF}}_{1i}|$, $Y_{1i} = 2\widehat{\text{CF}}_{1i} - 1$ are calculated for the j^{th} test set in the l^{th} 10-fold cross validation repetition, and $\hat{\eta}_{1,k,l,j}$ are the parameters estimated using the j^{th} training set in the l^{th} 10-fold cross validation repetition with tuning parameter k.

The results of the no censoring method are shown in Table 5.11. This table presents the tuning parameter chosen by cross validation, the estimated rule, and how many patients should have received each treatment under the estimated rule for each stage.

At the first stage, it is estimated that ATRA in combination with arsenic trioxide is the optimal consolidation therapy, regardless of patient history. According to the estimated

Table 5.11 Data analysis results in the case without censoring. For each stage, k is the tuning parameter chosen by 10 repetitions of 10-fold cross validation. $\hat{d}$ presents the estimated regime. Trt 1 and Trt 0 show how many patients should have received treatment 1 and treatment 0 respectively according to the estimated regime, $\hat{d}$. $\hat{V}(\hat{d}) = 980$ with an estimated standard error of 21 days.

Stage	k	$\hat{d}$	Trt 1	Trt 0
1	17380	$\hat{d}_1(H_1) = 1$	460	0
2	3810	$\hat{d}_2(H_2) = I(0.45 - 0.89\text{IC}10043607 > 0)$	278	31

regime, there is a small subset of patients who experience the adverse event of thrombosis/thrombus/embolism and should receive ATRA alone for maintenance therapy. All other patients should be treated with ATRA in combination with MP and Mtx. If all patients had followed the estimated optimal treatment regime, the estimated mean restricted survival time is $\hat{V}(\hat{d}) = 980$ with an estimated standard error of 21 days.

We compare the estimated optimal regime to all four of the embedded regimes in the trial using the test in Section 3.5.2. Table 5.12 lists the embedded regimes, their estimated value and standard error, the t-statistic when compared to the estimated regime, and the one-tailed p-value. These tests show that the estimated regime results in a significantly higher estimated value than all of the embedded regimes except for regime $(1,1)$. This is not surprising as the estimated regime treats patients similarly to $(1,1)$.

Note that the estimated regime is increasingly correlated with the embedded regimes moving down Table 5.12. This is because the embedded regimes increasingly agree with the estimated regime on how to treat patients, which causes the standard errors of the difference to be smaller. This results in a smaller test statistic when comparing the estimated regime to $(0,1)$ as opposed to $(1,0)$ even though the estimated difference is larger.

Table 5.12 Estimated regime tested against the embedded regimes in the case without censoring. $\hat{V}(d)$ presents the estimated value of the embedded regime with the estimated standard error in parentheses, t presents the test statistic when the embedded regime is compared to the estimated regime, and p-value presents the one-tailed p-value in the direction of the estimated value of the estimated regime being larger than the estimated value of the embedded regime. $\hat{V}(\hat{d}) = 980$ with an estimated standard error of 21 days.

d	$\hat{V}(d)$	t	p-value
(0,0)	834 (29)	4.24	<0.0001
(0,1)	919 (24)	1.97	0.02
(1,0)	958 (23)	2.13	0.02
(1,1)	973 (22)	1.24	0.11

Because we could not conclude that the estimated regime is better than all of the embedded regimes, we also test regime (1,1) against all of the other embedded regimes. This regime has the highest estimated value of all the embedded regimes and corresponds with the conclusions about the best consolidation therapy and the theory about the best maintenance therapy made in Powell et al. (2010). The results of these tests are presented in Table 5.13.

Table 5.13 Embedded regime (1,1) tested against the other embedded regimes in the no censoring case. $\hat{V}(d)$ presents the estimated value of the embedded regime with the estimated standard error in parentheses, t presents the test statistic when the embedded regime is compared to regime (1,1), and p-value presents the one-tailed p-value in the direction of the estimated value of regime (1,1) being larger than the estimated value of the other embedded regime. $\hat{V}(1,1) = 973$ with an estimated standard error of 22 days.

d	$\hat{V}(d)$	t	p-value
(0,0)	834 (29)	3.97	<0.0001
(0,1)	919 (24)	1.70	0.04
(1,0)	958 (23)	1.23	0.11

From these results, we conclude that the regime (1,1) result in a higher estimated estimated value than (0,0) and (0,1), but we can not conclude that (1,1) has a higher estimated value than (1,0).

5.2.2 Censoring Case

We carry out another analysis using the proposed method that can handle censoring of the event time. We do not drop any patients or change their event times for this analysis. As mentioned previously, there is a very low rate of censoring at $L = 1100$ days. We extended the cutoff to $L = 2500$ when there is a 19% rate of censoring overall, 6% in the first stage and 20% in the second stage. Again, 10 repetitions of 10-fold cross validation is used to choose each tuning parameter, k_1 and k_2.

Estimates of the event hazard functions at each stage, the censoring hazard functions, and the propensity scores are needed. The propensity scores are estimated by the sample proportion of patients who received treatment 1 at each stage. We posit a Cox proportional hazards model for the additional life hazard function, $\lambda_2(u \mid H_2, A_2)$. All baseline covariates, adverse events, treatments, and all interactions with treatment at the second stage are considered. Stepwise variable selection is used to reduce the number of variables. After variable selection, we estimate the following model:

$$\begin{aligned}\lambda_2(u \mid H_2, A_2, \beta_2) = \lambda_{20}(u) \exp\big\{&\beta_{21}A_1 + \beta_{22}\text{Risk.group} + \beta_{23}\text{ECOG.Performance.Status} \\ &+ \beta_{24}\text{I10012457} + \beta_{25}A_2 + \beta_{26}\text{I10028813} + \beta_{27}\text{C10035528} + \beta_{28}\text{IC10035528} \\ &+ \beta_{29}\text{Platelet} + \beta_{2,10}(A_2 \times \text{I10028813})\big\}.\end{aligned}$$

Similarly, the cause-specific first stage event hazard functions, $\lambda_{1r}(u \mid H_1, A_1)$, must be

estimated for $R = r$ when $r = 0, 1$. We posit a Cox proportional hazards model starting with all baseline covariates, first stage treatment, and interactions involving first stage treatment. Stepwise variable selection is used to reduce the number of variables, and we estimate the following models:

$$\begin{aligned}\lambda_{10}(u \mid H_1, A_1, \beta_{10}) = \lambda_{10,0}(u) \exp(&\beta_{10,1}\text{WBC} + \beta_{10,2}\text{ECOG.Performance.Status} + \beta_{10,3}A_1 \\ &+ \beta_{10,4}\text{Creatinine}) \\ \lambda_{11}(u \mid H_1, A_1, \beta_{11}) = \lambda_{11,0}(u) \exp\big\{&\beta_{11,1}A_1 + \beta_{11,2}\text{Race} + \beta_{11,3}\text{Age} + \beta_{11,4}\text{Ethnicity} \\ &+ \beta_{11,5}\text{Hemoglobin} + \beta_{11,6}\text{Platelet} + \beta_{11,7}(A_1 \times \text{Race})\big\}.\end{aligned}$$

We posit a Cox proportional hazards model for both censoring models as well. The function $\alpha_2(u \mid H_1, A_1)$ is the censoring hazard function for the event time U and $\alpha_1(u \mid H_1, A_1)$ is the censoring hazard function for the first stage event time U^1. Both models first consider all baseline covariates, first stage treatment, and interactions involving first stage treatment. Stepwise variable selection is used to reduce the number of variables, and we estimate the following models:

$$\begin{aligned}\alpha_1(u \mid H_1, A_1, \gamma_1) &= \alpha_{10}(u) \exp(\gamma_{11}A_1 + \gamma_{12}\text{WBC}) \\ \alpha_2(u \mid H_1, A_1, \gamma_2) &= \alpha_{20}(u) \exp(\gamma_{21}\text{Age} + \gamma_{22}A_1 + \gamma_{23}\text{Ethnicity} + \gamma_{24}\text{ECOG.Performance.Status} \\ &\quad + \gamma_{25}\text{Gender}).\end{aligned}$$

Table 5.14 presents the results at each stage in the same format as in Table 5.11. The estimated rules are the same as those in the no censoring case.

Again, it is estimated that ATRA in combination with arsenic trioxide therapy is the optimal consolidation therapy, regardless of patient history. Just as in the no censoring

Table 5.14 Data analysis results in the censoring case. All quantities are analogous to those in Table 5.11. $\hat{V}(\hat{d}) = 2176$ with an estimated standard error of 55 days.

Stage	k	$\hat{d}$	Trt 1	Trt 0
1	21650	$\hat{d}_1(H_1) = 1$	468	0
2	16660	$\hat{d}_2(H_2) = I(0.45 - 0.89\text{IC}10043607 > 0)$	281	31

case, according to the estimated regime, there is a small subset of patients who experience the adverse event of thrombosis/thrombus/embolism and should receive ATRA alone for maintenance therapy. All other patients should be treated with ATRA in combination with MP and Mtx. If all patients had followed the estimated optimal treatment regime, the estimated mean restricted survival time is $\hat{V}(\hat{d}) = 2176$ with an estimated standard error of 55 days.

We compare the estimated optimal regime to all four of the embedded regimes in the trial. From the results in Table 5.15, the estimated regime results in a significantly higher estimated value than all of the embedded regimes except for regime $(1,1)$. This is the same conclusion as in the no censoring case.

Table 5.15 Estimated regime tested against the embedded regimes in the censoring case. All quantities are analogous to those in Table 5.12. $\hat{V}(\hat{d}) = 2176$ with an estimated standard error of 55 days.

d	$\hat{V}(d)$	t	p-value
(0,0)	1610 (83)	5.87	<0.0001
(0,1)	1845 (71)	3.76	<0.0001
(1,0)	2078 (61)	2.45	0.007
(1,1)	2152 (58)	1.12	0.13

Because we could not conclude that the estimated regime is better than all of the embedded regimes, we also test regime (1,1) against all of the other embedded regimes, as we did in the no censoring case. The results of these tests are presented in Table 5.16.

Table 5.16 Embedded regime (1,1) tested against the other embedded regimes in the censoring case. All quantities are analogous to those in Table 5.13. $\hat{V}(1,1) = 2152$ with an estimated standard error of 58 days.

d	$\hat{V}(d)$	t	p-value
(0,0)	1610 (83)	5.49	<0.0001
(0,1)	1845 (71)	3.40	0.0004
(1,0)	2078 (61)	1.63	0.05

From these results, we conclude that regime (1,1) results in a significantly higher estimated value than regimes (0,0) and (0,1). The p-value is on the edge of significance when comparing (1,1) to (1,0).

This data analysis shows that the proposed methods are able to improve upon the methods used in the original paper. The proposed methods are able to compare sequences of treatments, while the original analysis was only able to choose a best treatment at one stage. We have also identified a small subset of patients in the second stage for clinicians to investigate further to see if they should receive different maintenance therapy from the rest of the population.

5.3 Conclusions

In Chapters 2 and 3, we present semiparametric methods to estimate the value of a treatment regime in the one and two stage problems both when there is and is not censoring of the event time. These semiparametric estimators are doubly robust and guaranteed to be consistent under conditions that should automatically be satisfied in a well-run SMART study.

In Chapter 4, we discuss how to estimate an optimal treatment regime. We present the value search method which restricts the search to a chosen class of regimes. This makes finding the optimal regime less daunting, and the choice of class of regimes could reduce the cost and make the estimated regime more easily interpretable. Competing methods implicitly restrict the class of regimes based on model choices, so the value search perspective is not a disadvantage in comparison.

We also show how Zhang et al. (2012b) transforms the problem of finding the rule that optimizes the value into a classification problem. This allows us to use well-studied and already developed techniques, such as SVM, to handle this non-smooth, non-convex problem.

We use all of these tools to create a backward iterative method that uses semiparametric estimators of the contrast function in conjunction with SVM to estimate an optimal treatment rule at each stage. This method can maximize the value for any monotone function of survival time, $g(T)$.

Simulations show that this method performs very well in that the estimated regime achieves a value close to that of the optimal regime. Also, the method used when there

is censoring does well at recovering the data lost due to censoring. The data analysis of Study C9710 is able to confirm the results in the original paper, as well as provide additional insights.

There is still more work to do. Simulation studies to evaluate how well the penalty term in SVM is able to select the correct variables should be explored. More work can also be done on the standard error estimators. As discussed, the estimators we present are only valid under specific regularity conditions that may not hold. Work should also be done to extend this method to more than two stages and allow for more than two treatment options at each stage.

BIBLIOGRAPHY

American Cancer Society (2016). *Typical treatment of acute lymphocytic leukemia*. URL: http://www.cancer.org/cancer/leukemia-acutelymphocyticallinadults/detailedguide/leukemia-acute-lymphocytic-treating-typical-treatment (visited on 07/11/2016).

Bai, X., Tsiatis, A. A., and O'Brien, S. M. (2013). Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* **69** (4), 830–839.

Bai, X., Tsiatis, A. A., Lu, W., and Song, R. (2016). Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime Data Analysis*, in press.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor Francis.

Breslow, N. E. (1972). Discussion on D.R. Cox (1972) paper. *Journal of the Royal Statistical Society, Series B* **34** (2), 216–217.

Chakraborty, B., Laber, E. B., and Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69** (3), 714–723.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20** (3), 273–297.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **34** (2), 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62** (2), 269–276.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *The Annals of Statistics* **40** (1), 529–560.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of The American Statistical Association* **47** (260), 663–685.

Jiang, R., Lu, W., Song, R., and Davidian, M. (2016). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B*, in press.

Lavori, P. W. and Dawson, R. (2004). Dynamic treatment regimes: practical design considerations. *Clinical Trials* **1**, 9–20.

Lavori, P. W., Dawson, R., and Rush, A. J. (2000). Flexible treatment strategies in chronic disease: clinical and research implications. *Biological Psychiatry* **48** (6), 605–614.

Lunceford, J. K., Davidian, M., and Tsiatis, A. A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics* **58** (1), 48–57.

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine* **24**, 1455–1481.

Powell, B., Moser, B., Stock, W., Gallagher, R. E., Willman, C. L., Stone, R. M., Rowe, J. M., Coutre, S., Feusner, J. H., Gregory, J., Couban, S., Appelbaum, F. R., Tallman, M. S., and Larson, R. A. (2010). Arsenic trioxide improves event-free and overall survival for adults with acute promyelocytic leukemia: North American Leukemia Intergroup Study C9710. *Blood* **116** (19), 3751–3757.

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

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. *Proceedings of the Second Seattle Symposium in Biostatistics*. Ed. by Lin, D. Y. and Heagerty, P. J. **179**. New York, NY: Springer New York, 189–326.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of The American Statistical Association* **89** (427), 846–866.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* **6** (1), 34–58.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science* **29** (4), 640–661.

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

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L., Skali, H., Solomon, S., Jacobus, S., Hughes, M., Packer, M., and Wei, L. J. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32** (22), 2380–2385.

Wahed, A. S. and Tsiatis, A. A. (2004). Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics* **60** (1), 124–133.

Wahed, A. S. and Tsiatis, A. A. (2006). Semiparametric efficient estimation of survival distributions in two-stage randomization designs in clinical trials with censored data. *Biometrika* **93** (1), 163–177.

Yang, X., Song, Q., and Wang, Y. (2007). A weighted support vector machine for data classification. *International Journal of Pattern Recognition and Artificial Intelligence* **21** (5), 961–976.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012a). A robust method for estimating optimal treatment regimes. *Biometrics* **68** (4), 1010–1018.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. B. (2012b). Estimating optimal treatment regimes from a classification perspective. *Stat* **1** (1), 103–114.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** (3), 681–694.

Zhao, Y. Q., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107** (449), 1106–1118.

Zhao, Y. Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015a). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* **102** (1), 151–168.

Zhao, Y. Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015b). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association* **110** (510), 583–598.

APPENDICES

APPENDIX A

RESTRICTED SURVIVAL TIME

A.1 Mean Calculations

The proposed methods can be applied to any monotone function of survival time, $g(T)$. One such function is restricted lifetime, defined as $g(T) = \min(T, L)$ for cutoff time L. Mean restricted lifetime is defined as $\mu^L = \int_0^L S(t)dt$ for survival function $S(t)$. Mean survival time $\mu = \int_0^\infty S(t)dt$ is shown to be equivalent to $\mu = \int_0^\infty t f(t)dt$ using integration by parts as follows:

$$u = t \rightarrow du = dt,$$

$$dv = f(t)dt \equiv -dS(t) \rightarrow v = -S(t).$$

$\int_0^\infty u\,dv = uv|_0^\infty - \int_0^\infty v\,du$ is equivalent to $\int_0^\infty t f(t)dt = -tS(t)|_0^\infty + \int_0^\infty S(t)dt = \int_0^\infty S(t)dt.$

However, when calculating mean restricted lifetime, it is necessary to integrate the survival function to L instead of infinity. Use integration by parts again to see the relation:

$$\int_0^L t f(t)dt = -tS(t)|_0^L + \int_0^L S(t)dt = -LS(L) + \int_0^L S(t)dt.$$

So mean restricted survival time is $\mu^L = \int_0^L S(t)dt = \int_0^L t f(t)dt + LS(L)$.

The same relation can be seen when finding the expected value directly. The restricted survival time can be written as $g(T) = \min(T, L) = TI(T < L) + LI(T \geq L)$. Then integrating to find the expected value of this function,

$$\begin{aligned}\mu^L = \mathrm{E}\{g(T)\} = \int_0^\infty \{tI(t < L) + LI(t \geq L)\} f(t)dt &= \int_0^L t f(t)dt + L\int_L^\infty f(t)dt \\ &= \int_0^L t f(t)dt + LS(L) \\ &= \int_0^L t\{-dS(t)\} + LS(L).\end{aligned}$$

A.2 Second Stage Expectation

When optimizing mean restricted lifetime, we can examine the expectation $\mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r}\right\}$ when $R = r$ more closely to see the intricacies of the calculation. Only patients with $R > 0$ are included in the calculation, so it is know that $\tau < L$. Rewriting the restricted survival time in terms of τ, additional life (AL), and L results in

$$\begin{aligned} g(T) = g(\mathrm{AL}+\tau) = \min(\mathrm{AL}+\tau, L) &= I\{\mathrm{AL}+\tau < L\}(\mathrm{AL}+\tau) + I\{\mathrm{AL}+\tau \geq L\} L \\ &= I\{\mathrm{AL} < L-\tau\}(\mathrm{AL}+\tau) + I\{\mathrm{AL} \geq L-\tau\} L. \end{aligned}$$

Because τ and L are considered fixed at the second stage, additional life is the only random variable. Rewrite the expectation in Equation 3.6 for each $R = r$, $r = 1, ..., B$ as

$$
\begin{aligned}
\mathrm{E}\left\{g(T) \mid H_{2r}, A_{2r}\right\} &= \int_0^{\infty} g(t+\tau)\{-dS_{2r}(t \mid H_{2r}, A_{2r})\} \\
&= \int_0^{L-\tau} (t+\tau)\{-dS_{2r}(t \mid H_{2r}, A_{2r})\} + \int_{L-\tau}^{\infty} L\{-dS_{2r}(t \mid H_{2r}, A_{2r})\} \\
&= \int_0^{L-\tau} t\{-dS_{2r}(t \mid H_{2r}, A_{2r})\} + \tau\{1 - S_{2r}(L-\tau \mid H_{2r}, A_{2r})\} \\
&\qquad + (L-\tau+\tau)S_{2r}(L-\tau \mid H_{2r}, A_{2r}) \\
&= \int_0^{L-\tau} t\{-dS_{2r}(t \mid H_{2r}, A_{2r})\} + (L-\tau)S_{2r}(L-\tau \mid H_{2r}, A_{2r}) \\
&\qquad + \tau\{1 - S_{2r}(L-\tau \mid H_{2r}, A_{2r})\} + \tau S_{2r}(L-\tau \mid H_{2r}, A_{2r}) \\
&= \tau + \int_0^{L-\tau} S_{2r}(t \mid H_{2r}, A_{2r})\, du.
\end{aligned}
$$

The last step uses properties of the expectation of restricted lifetime seen in Appendix A.1.

A.3 First Stage Expectation

In the special case of analyzing restricted lifetime, we can rewrite the estimator of $\mathrm{E}\left[g\left\{T^*(d)\right\} \mid H_1\right]$ in (3.8) as:

$$
\begin{aligned}
&\int_0^L t\,\lambda_{10}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t \\
&\quad + L \int_L^\infty \lambda_{10}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t \\
&\quad + \sum_{r=1}^{B} \left[\int_0^L Q_{1r}\left\{X_1, A_1 = d_1(H_1), \tau = t, R = r; \hat{\xi}_r\right\} \lambda_{1r}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} \right. \\
&\qquad \left. \times S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t + L \int_L^\infty \lambda_{1r}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t \right] \\
&= \int_0^L t\,\lambda_{10}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t + L S_{1.}\left\{L; X_1, A_1 = d_1(H_1)\right\} \\
&\quad + \sum_{r=1}^{B} \int_0^L Q_{1r}\left\{X_1, A_1 = d_1(H_1), \tau = t, R = r; \hat{\xi}_r\right\} \lambda_{1r}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} \\
&\qquad \times S_{1.}\left\{t \mid X_1, A_1 = d_1(H_1)\right\} d\,t.
\end{aligned}
$$

The last term was simplified using the fact that

$$\begin{aligned}
&\sum_{r=1}^{B} L \int_{L}^{\infty} \lambda_{1r}\{t \mid X_1, A_1 = d_1(H_1)\} S_{1.}\{t \mid X_1, A_1 = d_1(H_1)\}\, d\,t \\
&\quad = L \int_{L}^{\infty} \left[\sum_{r=1}^{B} \lambda_{1r}\{t \mid X_1, A_1 = d_1(H_1)\}\right] S_{1.}\{t \mid X_1, A_1 = d_1(H_1)\}\, d\,t \\
&\quad = L \int_{L}^{\infty} \frac{-d\,S_{1.}\{t \mid X_1, A_1 = d_1(H_1)\}}{S_{1.}\{t \mid X_1, A_1 = d_1(H_1)\}} S_{1.}\{t \mid X_1, A_1 = d_1(H_1)\} \\
&\quad = L S_{1.}\{L \mid X_1, A_1 = d_1(H_1)\}.
\end{aligned}$$

A.4 Conditional Expectation

The censoring augmentation terms include the expectation of restricted survival time conditional on the fact a patient already survived up to a specific time, ω. The time ω is in terms of total time, while the integration is over additional life after the second stage treatment. We can rewrite the expectation for each $R = r$ as:

$$
\begin{aligned}
\mathrm{E}\left\{g(T) \mid U > \omega, H_{2r}, A_{2r}\right\} &= \frac{\int_{\omega-\tau}^{\infty} g(u+\tau)\{-dS_{2r}(u \mid H_{2r}, A_{2r})\}}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&= \frac{\int_{\omega-\tau}^{L-\tau} (u+\tau)\{-dS_{2r}(u \mid H_{2r}, A_{2r})\}}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} + \frac{\int_{L-\tau}^{\infty} L\{-dS_{2r}(u \mid H_{2r}, A_{2r})\}}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&= \frac{\int_{\omega-\tau}^{L-\tau} u\{-dS_{2r}(u \mid H_{2r}, A_{2r})\}}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&\quad + \tau \frac{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r}) - S_{2r}(L-\tau \mid H_{2r}, A_{2r})}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&\quad + (L-\tau+\tau)\frac{S_{2r}(L-\tau \mid H_{2r}, A_{2r})}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&= \tau + \frac{\int_{\omega-\tau}^{L-\tau} u\{-dS_{2r}(u \mid H_{2r}, A_{2r})\} + (L-\tau)S_{2r}(L-\tau \mid H_{2r}, A_{2r})}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&= \tau + \frac{(\omega-\tau)S_{2r}(\omega-\tau \mid H_{2r}, A_{2r}) + \int_{\omega-\tau}^{L-\tau} S_{2r}(u \mid H_{2r}, A_{2r})du}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})} \\
&= \omega + \frac{\int_{\omega-\tau}^{L-\tau} S_{2r}(u \mid H_{2r}, A_{2r})du}{S_{2r}(\omega-\tau \mid H_{2r}, A_{2r})}.
\end{aligned}
$$

The second to last step uses the same idea as in Appendix A.1. Do integration by parts, except integrate from $\omega-\tau$ to $L-\tau$ instead:

$$
\int_{\omega-\tau}^{L-\tau} t f(t)dt = -tS(t)\Big|_{\omega-\tau}^{L-\tau} + \int_{\omega-\tau}^{L-\tau} S(t)dt = -(L-\tau)S(L-\tau) + (\omega-\tau)S(\omega-\tau) + \int_{\omega-\tau}^{L-\tau} S(t)dt.
$$

APPENDIX B

SANDWICH VARIANCE ESTIMATOR

For inference, the variance of the estimated value of regime d, $\hat{V}(d) = n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d)$, is necessary. Assume that $\hat{V}(d)$ is a consistent estimator of $V(d)$, so $n^{\frac{1}{2}}\left\{\hat{V}(d) - V(d)\right\} \to o_p(1)$. This can be rewritten as:

$$\begin{aligned}
n^{\frac{1}{2}}\left\{\hat{V}(d) - V(d)\right\} &= n^{\frac{1}{2}}\left\{n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d) - n^{-1}n\hat{V}(d)\right\} + n^{\frac{1}{2}}\left\{\hat{V}(d) - V(d)\right\} \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d) - \hat{V}(d)\right\} + n^{\frac{1}{2}}\left\{\hat{V}(d) - V(d)\right\} \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d) - \hat{V}(d)\right\} + o_p(1) \\
&= n^{\frac{1}{2}}\left\{n^{-1}\sum_{i=1}^{n}\widehat{\Psi}_i(d) - \hat{V}(d)\right\} + o_p(1).
\end{aligned}$$

This formulation shows that $\varphi(H_i) = \widehat{\Psi}_i(d) - \hat{V}(d)$ is the influence function which has

mean 0. By the central limit theorem,

$$n^{\frac{1}{2}}\left\{\hat{V}(d)-V(d)\right\}\rightarrow N\left[0,\mathrm{E}\left\{\varphi^2(H)\right\}\right],$$

so $\mathrm{Var}\left[n^{\frac{1}{2}}\left\{\hat{V}(d)-V(d)\right\}\right]$ can be estimated by $n^{-1}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d)-\hat{V}(d)\right\}^2$. This implies that $\mathrm{Var}\left\{\hat{V}(d)\right\}$ can be estimated by $n^{-2}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d)-\hat{V}(d)\right\}^2$. For large n, this is similar to the sample variance estimate of $\hat{V}(d)$ which is $\{n(n-1)\}^{-1}\sum_{i=1}^{n}\left\{\widehat{\Psi}_i(d)-\hat{V}(d)\right\}^2$.

APPENDIX C

HINGE LOSS FISHER CONSISTENCY

Using SVM to find an estimated optimal treatment rule, we rewrite (4.2) and solve

$$\hat{\eta} = \arg\min_{\eta} \sum_{i=1}^{n} \left|\widehat{\mathrm{CF}}_i\right| \phi\left\{Y_i f_\eta(H_i)\right\},$$

where $Y_i = 2\mathrm{I}\left(\widehat{\mathrm{CF}}_i \geq 0\right) - 1$. Note that when $Y_i = 1$, $\widehat{\mathrm{CF}}_i \geq 0$, and when $Y_i = -1$, $\widehat{\mathrm{CF}}_i < 0$. The estimated rule will have form $\hat{d}(H) = I\left\{f_{\hat{\eta}}(H) \geq 0\right\}$. Assume that the estimated contrast function is a consistent estimator,

$$\mathrm{E}\left(\widehat{\mathrm{CF}} \mid H\right) = \mathrm{CF}(H) = \mathrm{E}\left\{g(T) \mid H, A = 1\right\} - \mathrm{E}\left\{g(T) \mid H, A = 0\right\}.$$

Though $\mathrm{CF}(H)$ is not known, the true optimal treatment rule is $I\{\mathrm{CF}(H) \geq 0\}$. Therefore, if $\mathrm{CF}(H) \geq 0$, then $f_{\hat{\eta}}(H)$ should be greater than 0 in order to agree with the true optimal

rule. Likewise, if $\text{CF}(H) < 0$, then $f_{\hat{\eta}}(H)$ should be less than 0 in order to agree with the true optimal rule. We check to see if the rule estimated from using SVM with the hinge loss function is a consistent estimator of the true optimal rule. For easier notation, we drop the subscript η from $f_\eta(H)$.

Using the law of iterated expectations, the expectation of the risk function can be written as

$$\begin{aligned}\text{E}\left[\left|\widehat{\text{CF}}\right| \phi\left\{Y f(H)\right\}\right] &= \text{E}\left(\text{E}\left[\left|\widehat{\text{CF}}\right| \phi\left\{Y f(h)\right\} \mid H = h\right]\right) \\ &= \text{E}\Big(\text{E}\Big[I\left(\widehat{\text{CF}} \geq 0\right)\widehat{\text{CF}}\left\{1 - f(h)\right\}^+ - I\left(\widehat{\text{CF}} < 0\right)\widehat{\text{CF}}\left\{1 + f(h)\right\}^+ \mid H = h\Big]\Big).\end{aligned}$$

Recall that $\widehat{\text{CF}}$ is dependent on H, A, and T. Consequently, the value $f(H)$ which minimizes the expected risk function is given by the value $f(h)$ which minimizes the inner expectation,

$$\text{E}\left\{I\left(\widehat{\text{CF}} \geq 0\right)\widehat{\text{CF}} \mid H = h\right\}\left\{1 - f(h)\right\}^+ + \text{E}\left\{-I\left(\widehat{\text{CF}} < 0\right)\widehat{\text{CF}} \mid H = h\right\}\left\{1 + f(h)\right\}^+,$$

for each h. Note that each term is this equation is greater than or equal to 0. We will examine two separate cases in the following proof. For ease of notation, we rename the terms as follows:

$$\begin{aligned} e_1 &\equiv \text{E}\left\{I\left(\widehat{\text{CF}} \geq 0\right)\widehat{\text{CF}} \mid H = h\right\} \\ e_2 &\equiv \text{E}\left\{-I\left(\widehat{\text{CF}} < 0\right)\widehat{\text{CF}} \mid H = h\right\} \\ x &\equiv f(h). \end{aligned}$$

Case 1:

$$\begin{aligned}\text{Assume } \mathrm{E}\left\{I\left(\widehat{\mathrm{CF}}\geq 0\right)\widehat{\mathrm{CF}} \mid H=h\right\} &\geq \mathrm{E}\left\{-I\left(\widehat{\mathrm{CF}}<0\right)\widehat{\mathrm{CF}} \mid H=h\right\}\\ \implies \mathrm{E}\left\{I\left(\widehat{\mathrm{CF}}\geq 0\right)\widehat{\mathrm{CF}}+I\left(\widehat{\mathrm{CF}}<0\right)\widehat{\mathrm{CF}} \mid H=h\right\} &\geq 0\\ \implies \mathrm{E}\left(\widehat{\mathrm{CF}} \mid H=h\right)\geq 0 \implies \mathrm{CF}(h) &\geq 0.\end{aligned}$$

In this case, the true optimal rule when $H=h$, $I\{\mathrm{CF}(h)\geq 0\}$, dictates that treatment 1 is optimal. The estimated rule, $I\left\{f(h)\geq 0\right\}$, should agree with the true optimal rule, so $f(h)$ should be greater than or equal to 0.

Check:

In this case $e_1 \geq e_2$ by the original assumption. Finding the rule that minimizes the risk function is equivalent to

$$\arg\min_x \left\{e_1(1-x)^+ + e_2(1+x)^+\right\} = \arg\min_x \left[e_2\left\{(1+x)^+ + (1-x)^+\right\} + (e_1-e_2)(1-x)^+\right].$$

Note that each term is nonnegative, so if each part has the same minimizer, then that would minimize the entire function:

$$\begin{aligned}&\arg\min_x e_2\{(1+x)^+ + (1-x)^+\} \implies x\in[-1,1]\\ &\arg\min_x (e_1-e_2)(1-x)^+ \implies x\geq 1.\end{aligned} \tag{C.1}$$

These two statements together imply that the minimizer is $x\equiv f(h)=1$ when $e_1 > e_2$, which agrees with the sign of the contrast function. Note that when $e_1 = e_2$, there is no unique minimizer, however $\mathrm{CF}(h)=0$ in this case, so no one treatment is better than the other. Figure C.1 graphs the two functions in (C.1) to illustrate that both are minimized at $x=1$.

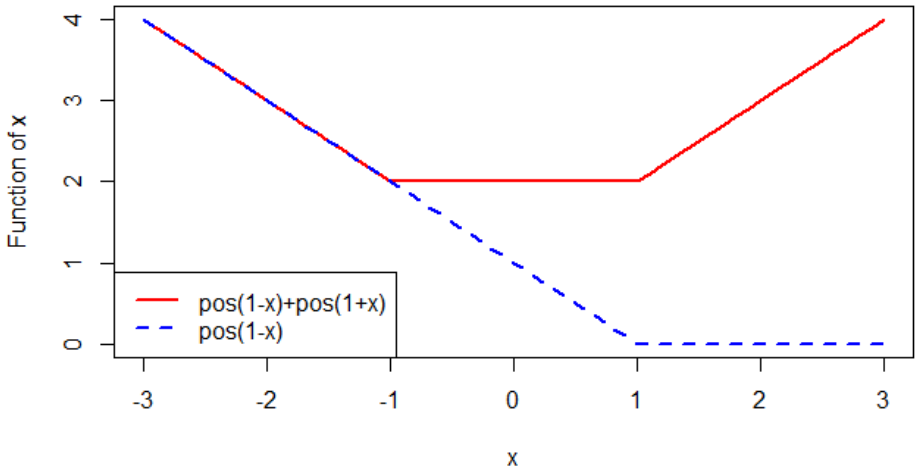

Figure C.1 Graph of functions minimized in (C.1) for case 1.

Case 2:

$$\begin{aligned}\text{Assume } \mathrm{E}\left\{I\left(\widehat{\mathrm{CF}}\geq 0\right)\widehat{\mathrm{CF}} \mid H=h\right\} &< \mathrm{E}\left\{-I\left(\widehat{\mathrm{CF}}<0\right)\widehat{\mathrm{CF}} \mid H=h\right\} \\ \implies \mathrm{E}\left\{I\left(\widehat{\mathrm{CF}}\geq 0\right)\widehat{\mathrm{CF}}+I\left(\widehat{\mathrm{CF}}<0\right)\widehat{\mathrm{CF}} \mid H=h\right\} &< 0 \\ \implies \mathrm{E}\left(\widehat{\mathrm{CF}} \mid H=h\right)<0 \implies \mathrm{CF}(h)&<0.\end{aligned}$$

In this case, the true optimal rule when $H=h$, $I\{\mathrm{CF}(h)>0\}$, dictates that treatment 0 is optimal. The estimated rule, $I\left\{f(h)>0\right\}$, should agree with the true optimal rule, so $f(h)$ should be less than 0.

Check:

In this case $e_1 < e_2$ by the original assumption. Finding the rule that minimizes the risk function is equivalent to

$$\arg\min_x \left\{e_1(1-x)^+ + e_2(1+x)^+\right\} = \arg\min_x \left[e_1\left\{(1-x)^+ + (1+x)^+\right\} + (e_2-e_1)(1+x)^+\right].$$

Note that each term is positive, so if each part has the same minimizer, then that would minimize the entire function:

$$\begin{aligned} &\arg\min_x e_1\{(1-x)^+ + (1+x)^+\} \implies x \in [-1,1] \\ &\arg\min_x (e_2 - e_1)(1+x)^+ \implies x \leq -1. \end{aligned} \tag{C.2}$$

These two statements together imply that the minimizer is $x \equiv f(h) = -1$ which agrees with the sign of the contrast function. Figure C.2 graphs the two functions in (C.2) to illustrate that both are minimized at $x = -1$.

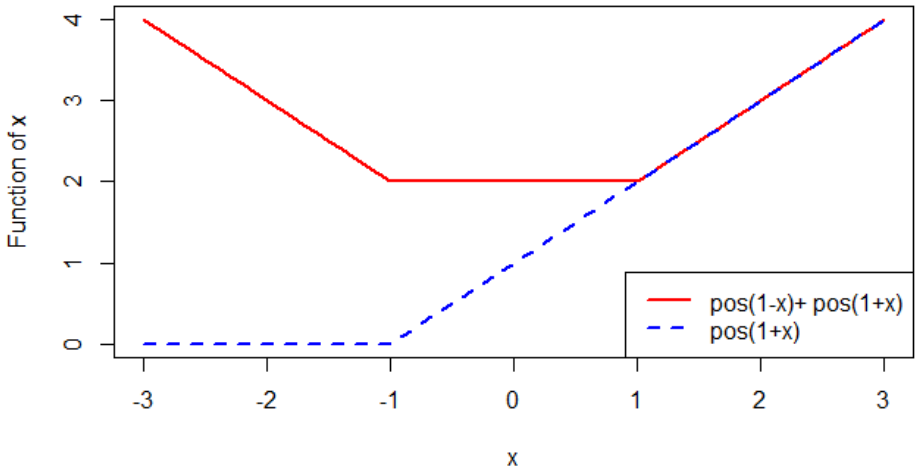

Figure C.2 Graph of functions minimized in (C.2) for case 2.

These cases prove that the rule estimated by weighted SVM is a consistent estimator for the true optimal rule.

APPENDIX D

SVM - LINEAR CONVEX OPTIMIZATION

The optimization in (4.3),

$$\min_{\eta} \sum_{i=1}^{n} \left|\widehat{\mathrm{CF}}_i\right| \max\left(1 - Y_i \eta^T \tilde{H}_i, 0\right) + k \sum_{j=1}^{p} |\eta_j|,$$

is equivalent to a linear convex optimization problem through the introduction of slack variables. Full details are shown in the supplementary material of Bai et al. (2016). The linear convex optimization problem is

$$\min_{\eta} \sum_{i=1}^{n} \left|\widehat{\mathrm{CF}}_i\right| \zeta_i + k \sum_{j=1}^{p} \delta_j \text{ subject to } \begin{cases} \zeta_i > 0, i = 1, ..., n \\ \zeta_i > 1 - Y_i \eta^T \tilde{H}_i, i = 1, ..., n \\ \delta_j \geq \eta_j, j = 1, ..., p \\ \delta_j \geq -\eta_j, j = 1, ..., p. \end{cases}$$

APPENDIX E

BOWL AND IPW ESTIMATOR

BOWL is another backward iterative method. Mathematically, the BOWL method is very similar to the proposed method using IPW estimators of the value. For ease of notation, assume that event time is of interest, so $g(T) = T$. At the second stage, BOWL minimizes the classification problem

$$\sum_{i=1}^{n} \frac{\mathrm{AL}_i}{P(A_{2i} = a_{2i} \mid H_{2i})} \phi \left\{ A'_{2i} f_2(H_{2i}) \right\} + k_2 ||\eta_2||^2,$$

where $\phi(x) = \max(1-x, 0)$ is the hinge loss function and $A'_{2i} = 2A_{2i} - 1$. The treatment denoted by A'_2 is coded by -1 and 1 in comparison to A_2 which is coded by 0 and 1. Note that the BOWL method uses the ℓ_2 penalization instead of the ℓ_1 penalization on η. The

classifier is A'_{2i} and the weight is

$$\text{Weight}_i = \frac{\text{AL}_i}{P(A_{2i} = a_{2i} \mid H_{2i})}.$$

The IPW method at the second stage minimizes

$$\sum_{\{i:R_i>0\}} \left|\widehat{\text{CF}}_{2i}\right| \phi\left[\left\{2I(\widehat{\text{CF}}_{2i} \geq 0) - 1\right\} f_2(H_{2i})\right] + k_2 \sum_{j=1}^{p} |\eta_{2j}|.$$

The classifier is $2I(\widehat{\text{CF}}_{2i} \geq 0) - 1$. The IPW estimator of the contrast function can be calculated as described in Section 4.4 by substituting AL_i for T_i in (2.3) to estimate the Ψ function under each treatment, so $\widehat{\text{CF}}_{2i} = \hat{\Psi}(1) - \hat{\Psi}(0)$. Equation (2.3) is zero under the treatment that the patient did not actually receive and positive under the treatment that the patient did receive. If $A_{2i} = 1$, then $\widehat{\text{CF}}_{2i} > 0$, and if $A_{2i} = 0$, then $\widehat{\text{CF}}_{2i} < 0$. Therefore, the IPW classifier will agree with the BOWL classifier at the second stage. The second stage weight for the IPW method is

$$\begin{aligned}
&\text{Weight}_i = \left| \frac{I(A_{2i} = 1)}{\pi_2(H_{2i})} \text{AL}_i - \frac{I(A_{2i} = 0)}{1 - \pi_2(H_{2i})} \text{AL}_i \right|. \\
&\text{When } A_{2i} = 1, \text{Weight}_i = \frac{\text{AL}_i}{\pi_2(H_{2i})}. \\
&\text{When } A_{2i} = 0, \text{Weight}_i = \frac{\text{AL}_i}{1 - \pi_2(H_{2i})}. \\
&\implies \text{Weight}_i = \frac{\text{AL}_i}{P(A_{2i} = a_{2i} \mid H_{2i})}.
\end{aligned}$$

This shows that the seconds stage weights are the same for both methods. Going back to the first stage, BOWL minimizes

$$\sum_{i=1}^{n} \frac{I\left\{A_{2i} = \hat{d}_2(H_{2i})\right\} T_i}{P(A_{1i} = a_{1i} \mid H_{1i}) P(A_{2i} = a_{2i} \mid H_{2i})} \phi\left\{A'_{1i} f_1(H_{1i})\right\} + k_1 ||\eta_1||^2,$$

where $A'_{1i} = 2A_{1i} - 1$. The classifier is A'_{1i}, and the weight is

$$\text{Weight}_i = \frac{I\left\{A_{2i} = \hat{d}_2(H_{2i})\right\} T_i}{P(A_{1i} = a_{1i} \mid H_{1i}) P(A_{2i} = a_{2i} \mid H_{2i})}.$$

This method does not allow for the possibility that a patient does not receive a second stage treatment. The IPW method adjusts for that by minimizing

$$\sum_{i=1}^{n} \left|\widehat{\text{CF}}_{1i}\right| \phi\left[\left\{2I(\widehat{\text{CF}}_{1i} \geq 0) - 1\right\} f_1(H_{1i})\right] + k_1 \sum_{j=1}^{p} |\eta_{1j}|,$$

using (3.2) to estimate the Ψ functions that make up the contrast function. By a similar argument as in the second stage, the first stage classifier is the same for both IPW and BOWL. The weight can be rewritten as

$$\begin{aligned}
\text{Weight}_i = \Bigg| & I(R_i = 0)\left\{\frac{I(A_{1i} = 1)}{\pi_1(H_{1i})} T_i - \frac{I(A_{1i} = 0)}{1 - \pi_1(H_{1i})} T_i\right\} \\
& + \sum_{r=1}^{B} I(R_i = r)\left[\frac{I\left\{A_{1i} = 1, A_{2i} = \hat{d}_2(H_{2i})\right\}}{\pi_1(H_{1i}) \pi_{\hat{d}_2}(H_{2i})} T_i - \frac{I\left\{A_{1i} = 0, A_{2i} = \hat{d}_2(H_{2i})\right\}}{\{1 - \pi_1(H_{1i})\} \pi_{\hat{d}_2}(H_{2i})} T_i\right]\Bigg|. \\
\text{When } A_{2i} = 1, \text{Weight}_i = & I(R_i = 0)\frac{I(A_{1i} = 1)}{\pi_1(H_{1i})} T_i + \sum_{r=1}^{B} I(R_i = r)\frac{I\left\{A_{1i} = 1, A_{2i} = \hat{d}_2(H_{2i})\right\}}{\pi_1(H_{1i}) \pi_{\hat{d}_2}(H_{2i})} T_i. \\
\text{When } A_{2i} = 0, \text{Weight}_i = & I(R_i = 0)\frac{I(A_{1i} = 0)}{1 - \pi_1(H_{1i})} T_i + \sum_{r=1}^{B} I(R_i = r)\frac{I\left\{A_{1i} = 0, A_{2i} = \hat{d}_2(H_{2i})\right\}}{\{1 - \pi_1(H_{1i})\} \pi_{\hat{d}_2}(H_{2i})} T_i. \\
\implies \text{Weight}_i = & I(R_i = 0)\frac{T_i}{P(A_{1i} = a_{1i} \mid H_{1i})} + \sum_{r=1}^{B} I(R_i = r)\frac{I\left\{A_{2i} = \hat{d}_2(H_{2i})\right\} T_i}{P(A_{1i} = a_{1i} \mid H_{1i}) P(A_{2i} = a_{2i} \mid H_{2i})}.
\end{aligned}$$

It is clear that the weights are very similar, except the IPW estimator is able to better represent real data.

APPENDIX F

G-COMPUTATION

G-computation (Robins, 1986) can be used because the true data generation scheme is known in the simulation studies. To find the Monte Carlo average value of regime $\hat{d} = (\hat{d}_1, \hat{d}_2)$, do the following:

1. Generate baseline covariates X_{11} and X_{12} as described in (5.1).
2. Treatment at the first stage is $A_1 = \hat{d}_1(H_1)$.
3. Check if $\hat{d}_1(H_1)$ agrees with $d_1^{\text{opt}}(H_1)$.
4. Generate τ based on τ_{D} and τ_{SS} as described in (5.1) and generate X_2 as described in (5.2).
5. Treatment at the second stage is $A_2 = \hat{d}_2(H_2)$.

6. Check if $\hat{d}_2(H_2)$ agrees with $d_2^{\text{opt}}(H_2)$.

7. Generate additional life as described in (5.2) and calculate event time, $T = \tau$ if $R = 0$ or $T = \tau + \text{AL}$ if $R = 1$.

8. Average $g(T)$ over all patients to get the value for this data set. This is considered to be a sample from the distribution of $\text{E}\left[g\left\{T^*(d)\right\}\right]$.

9. For each stage, calculate the proportion of patients for whom the treatment assigned by the estimated rule agrees with the treatment that would have been assigned by the optimal rule.

10. Repeat for m different data sets under regime $\hat{d}$.

11. Average the values and agreement proportions of the m Monte Carlo simulations.

APPENDIX G

DATA ANALYSIS

G.1 Data Cleaning

There is a total of n=518 patients in the North American Leukemia Intergroup Study C9710. For the data analysis, the following patients were removed:

- 57 children (age <15)
- 7 patients who have more than 2 covariates missing
- 6 patients who do not have an induction date

This leaves $n = 468$ patients remaining. The first 50 patients in the study had different maintenance treatment options. These patients are removed from the second stage analysis. There are 15 patients who are missing a maintenance start date and were also removed from

the second stage analysis. After removing these patients, there are 312 patients available for the second stage analysis.

Several covariates are adjusted as follows:

- The number of race groups are reduced to only include "White", "Hispanic American", "Black/African American", and "Other." Patients previously classified as "Asian", "Native Hawaiian or Pacific Islander", "American Indian or Alaska native", "Indian Subcontinent", or "Multiple Races Reported" are now classified as "Other."

- 1 patient has a platelet count of 5300 which is much higher than all other patients who had a range of 1-232. This observation is divided by 100 to be consistent with the range given in Powell et al. (2010).

- 4 patients have hemoglobin levels of 87, 92, 81, and 80 which are much higher than the rest of the patients who had a range of 4.3-14.6. It is believed these were entered as g/L instead of g/dL. We divide these hemoglobin observations by 10 to correct the units.

- 7 patients have creatinine levels of 76, 95, 87, 92, 90, 52, and 53 which are much higher than the rest of the patients who had a range of 0.1-10.4. It is believed these were entered as μmol/L instead of mg/dL. These creatinine observations are divided by 88.4 to correct the units.

The data after these adjustments are consistent with the information presented in Powell et al. (2010).

There are 446 of the 468 patients who have complete data. Missing baseline covariates are imputed as follows:

- Race is set to "Other" for the 8 patients for which it was missing.
- 5 patients are missing creatinine. A linear model to estimate creatinine using all other baseline covariates is estimated from the complete cases. The missing creatinine values are imputed using this model.
- 1 patient is missing white blood cell count (WBC). A linear model to estimate WBC using all baseline covariates except for WBC and Risk group is estimated from the complete cases. The missing WBC is imputed using this model.
- 4 patients are missing ECOG performance status. A linear model to estimate ECOG performance status using all other baseline covariates is estimated from the complete cases. The missing ECOG performance status values are imputed using this model.
- 4 patients are missing hemoglobin. A linear model to estimate hemoglobin using all other baseline covariates is estimated from the complete cases. The missing hemoglobin values are imputed using this model.

G.2 Variable Meanings

The North American Leukemia Intergroup Study C9710 data used in the data analysis contains baseline covariates collected at the time of registration into the trial and adverse event data collected during the induction and consolidation treatment phases. The adverse events are coded with an 8-digit Medical Dictionary for Regulatory Activities (medDRA) code. Table G.1 presents the baseline covariates and Table G.2 presents the medDRA code, adverse event category, and averse event which can be used to decode the variables in Chapter 5.

Table G.1 Baseline covariates available from the North American Leukemia Intergroup Study C9710.

Variable	Meaning
Age	Age at registration
Gender	1=Male, 2=Female
Race	1=White, 2=Hispanic American, 3=Black/African American, 9=Other
Ethnicity	1=Hispanic, 2=Non-Hispanic, 9=Unknown
WBC	White blood cell count at registration ($10^3/\mu$L)
Platelet	Platelet count at registration ($10^3/\mu$L)
ECOG.performance.status	ECOG performance status
Risk.group	1=Low (WBC <= 10 and Platelet > 40), 2=Intermediate (WBC <= 10 and Platelet <=40), 3=High (WBC > 10)
Creatinine	Serum creatinine (mg/dL)
Hemoglobin	Hemoglobin (g/dL)

Table G.2 medDRA codes and clinical meanings of the adverse events experienced by at least 5% of patients in the North American Leukemia Intergroup Study C9710.

medDRA	Adverse Event Category	Adverse Event
10002646	Gastrointestinal	Anorexia
10012457	Dermatology/skin	Rash/desquamation
10012745	Gastrointestinal	Diarrhea
10013442	Coagulation	Disseminated intravascular coagulation
10013972	Pulmonary	Dyspnea (shortness of breath)
10016288	Infection	Febrile neutropenia, fever of unknown origin without clinically or microbiologically documented infection (ANC <1.0 x 10e9/L, fever >=38.5 degrees C)
10018876	Blood/bone marrow	Hemoglobin
10019218	Pain	Headache
10020637	Metabolic/laboratory	Glucose serum-high (hyperglycemia)
10020947	Metabolic/laboratory	Calcium serum-low (hypocalcemia)
10021015	Metabolic/laboratory	Potassium serum-low (hypokalemia)
10021143	Pulmonary/upper respiratory	Hypoxia
10021842	Infection/febrile neutropenia	Infection without neutropenia
10024285	Blood/bone marrow	Leukocytes (total WBC)
10025327	Blood/bone marrow	Lymphopenia
10028813	Gastrointestinal	Nausea
10029363	Blood/bone marrow	Neutrophils/granulocytes (ANC/AGC)
10033359	Blood/bone marrow	Transfusion: packed red blood cells
10035528	Blood/bone marrow	Platelets
10035543	Blood/bone marrow	Transfusion: platelets
10043607	Vascular	Thrombosis/thrombus/embolism
90004060	Hemorrhage	Hemorrhage/bleeding with grade 3 or 4 thrombocytopenia
90004070	Infection/febrile neutropenia	Infection (documented clinically or microbiologically) with grade 3 or 4 neutropenia (ANC <1.0 x 10e9/L)