Abstract

WAHED, ABDUS SHAKOOR. Efficient Estimation of The Survival Distribution and Related Quantities of Treatment Policies in Two-Stage Randomization Designs in Clinical Trials. (Under the supervision of Professor Anastasios A. Tsiatis, Ph.D.)

Two-stage designs are common in therapeutic clinical trials such as Cancer or AIDS treatments. In a two-stage design, patients are initially treated with one induction (primary) therapy and then depending upon their response and consent, are treated by a maintenance therapy, sometimes to intensify the effect of the first stage therapy. The goal is to compare different combinations of primary and maintenance (intensification) therapies to find the combination that is most beneficial. To achieve this goal, patients are initially randomized to one of several induction therapies and then if they are eligible for the second-stage randomization, are offered to be randomized to one of several maintenance therapies. In practice, the analysis is usually conducted in two separate stages which does not directly address the major objective of finding the best combination. Recently Lunceford et. al. (2002, *Biometrics*, **58**, 48-57) introduced ad hoc estimators for the survival distribution and mean restricted survival time under different treatment policies. These estimators are consistent but not efficient, and do not include information from auxiliary covariates. In this dissertation study we derive estimators that are easy to compute and are more efficient than previous estimators. We also show how to improve efficiency further by taking into account additional information from auxiliary variables. Large sample proper-

ties of these estimators are derived and comparisons with other estimators are made using simulation. We apply our estimators to a leukemia clinical trial data set that motivated this study.

EFFICIENT ESTIMATION OF THE SURVIVAL DISTRIBUTION AND RELATED QUANTITIES OF TREATMENT POLICIES IN TWO-STAGE RANDOMIZATION DESIGNS IN CLINICAL TRIALS

by

ABDUS SHAKOOR FAZLUL WAHED

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

2003

APPROVED BY:

Anastasios Tsiatis Chair of Advisory Committee	Marie Davidian
Dennis Boos	Daowen Zhang

This work is dedicated to

Sejuti, Sudipto, and Samia,

the three most wonderful gifts I received in my life,

and to

My Parents

Biography

Abdus Shakoor Fazlul Wahed (known to friends, colleagues and relatives by the name Wahed) was born to parents Mohammad Abdul Baset and Begum Ayesha Akhter in Jalial, a remote village in the southern district Noakhali of Bangladesh on July 1, 1969. He completed his high school education from the prestigious school Brother Andre High School and the college Chaumuhani Saleh Ahmed College. He obtained his B. Sc. (honors) in Statistics with minors in Mathematics and Economics and M. Sc. degree in Statistics from the nation's most renowned university, University of Dhaka, Bangladesh.

Before coming to the United States for higher education in the August of 1998, Wahed served on the faculty of the Department of Statistics, University of Dhaka as a lecturer for approximately three years. He obtained his second masters in Mathematical Statistics from the Department of Mathematical Sciences at Ball State University in July, 2000 and left Indiana to enter the doctoral program in Statistics at North Carolina State University. Upon completion of his doctoral degree, he will be joining as an Assistant Professor in the Department of Biostatistics at the University of Pittsburgh Graduate School of Public Health.

While serving at the University of Dhaka, Wahed married a beautiful girl, Samia Lopa on October 3, 1997 and have shared a wonderful life together since. Wahed and Samia have a boy, Sudipto, and a girl, Sejuti.

Acknowledgment

First of all, I would like to thank my advisor Dr. Anastasios A. Tsiatis for his constant support and guidance that made this work possible. I feel very fortunate to have him as my dissertation advisor and he was my non-stop source of inspiration. During the last two years, I received many valuable suggestions both in and outside the class from him. Most of the difficult, technical and conceptual materials became edible to me only because of his way of expressing it in a manner I guarantee that anybody would be able to understand.

Next I would like to thank my committee members Dr. Denis Boos, Dr. Marie Davidian and Dr. Dowen Zhang for their valuable time and service. I especially thank Dr. Davidian for her help with things such as computing, LaTex, and my job search, to name a few. She was there for me anytime I needed to ask a question no matter how busy her schedule was.

My wholehearted thanks goes to Dr. Sastry G. Pantula, who, as the director of the statistics graduate program, tried his best to make my stay at NC State a comfortable and productive one. He would reply to my mails right away even if I sent it at midnight. Thanks a lot, Sastry.

I take this opportunity to thank all the faculty in the Department of Statistics at NC State, who in some way, either by their instruction in class or by their work outside the class, have left an impact on my professional growth.

Thanks to Dr. Dale E. Umbach, Department of Mathematical Sciences, Ball State

University for his valuable suggestions regarding choosing a good graduate program. When I was about to toss a coin to choose between NC State and another school, he was the one who referred me to NC State. I now realize how good a decision that was.

I would like to extend thanks to my colleagues whose company made my stay at NC State warm, especially to Jimmy for helping me out in numerous occasions. I feel proud to have a friend like Jimmy.

Thanks to the nice staff of Statistics Department, Terry,"Stat-Mom" Janice, Linda, Miranda, Germane, Joanna, Karla to name a few who never hesitated to extend their hands in any situation. Terry, what would have we done without you?

My gratitude goes to my parents, siblings and in-laws back at home who took the pain of leaving me and my family alone for the last five years. I am grateful to my father Mohammad Abdul Baset for teaching me how to be a good citizen. All good deeds I do is because of him. I owe big time to my elder brother ABM Abdul Alim, whose constant encouragement helped me come across several barriers on my way to higher education. I am also indebted to my father-in-law Shamsul Huda Bhuiyan whose first question in every conversation in last five years was "Is your study going OK?".

Thanks to the friends in Bangladeshi community in the Triangle area for their support during the last several years, especially at times when we needed it most. Your company made us feel at home.

No word in any language will be enough to express my debt to my wife Samia

for her contribution towards this achievement. Besides being the principal source of inspiration, she made extraordinary sacrifices for this work to come into light. We both did our M.A. degree from Ball State University at the same time. She then sacrificed her desire for higher study and career to facilitate my further education. During the last several years, while I was at school, she stayed at home taking care of our children all by herself, sometimes in the weekends and nights as well. I wish there were some Degree that would recognize this kind of support. I just have one sentence for Samia "May Allah bless you.".

July 30, 2003

Raleigh

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In a two-stage design, patients are treated initially with an induction therapy followed by maintenance therapy at some later time. Depending on the clinical trial, the maintenance therapy may be offered to a subset of the patients, for example, those patients who showed some response to the induction treatment. Such designs are common in cancer and other clinical trials where treatments are usually combinations of some therapies. To compare different combinations of induction and maintenance therapies, two-stage randomized studies may be considered where patients are initially randomized to one of several induction therapies and then patients who are eligible for maintenance therapy are randomized to one of several maintenance therapies. The primary goal of such randomized studies is to determine the combination of induction and maintenance therapies that will result in the best prognosis such as the longest average survival. However, like the randomization scheme used in these trials, data analysis typically is separated into two parts: (i) comparing induction

therapies using all the data ignoring maintenance therapy and (ii) comparing only those individuals randomized to maintenance therapy. Neither of these two analyses directly addresses the question of finding the best combination of maintenance and induction treatments.

As an example, throughout this dissertation we will consider the Cancer and Leukemia Group B (CALGB) clinical trial which motivated the study. Protocol 8923 was a double-blind, placebo-controlled two-stage trial reported by Stone et. al. (1995) examining the effects of infusions of granulocyte-macrophage colony-stimulating factor (GM-CSF) after initial chemotherapy in 388 elderly patients with acute myelogenous lukemia (AML). Patients were randomized initially to GM-CSF or placebo following standard chemotherapy. Later, patients meeting the criteria for complete remission were offered a second randomization to one of two intensification treatments. Figure 1 explains the two stages of CALGB 8923 trial. The goal of the study is to examine the effect on survival of granulocyte-macrophage colony-stimulating-factor (GM-CSF) in addition to chemotherapy followed by subsequent intensification treatment.

More examples of two-stage clinical trials can be found in Thall et. al. (2002), Tummarello et al. (1994), and Joss et al. (1994).

For concreteness, we will consider the two-stage clinical trial where patients are initially randomized to one of the induction treatments, say A_1 or A_2, upon entry into the trial. Among those eligible for maintenance therapy, *i.e.,* meets the criteria for randomization to the second stage therapy, a second randomization is offered

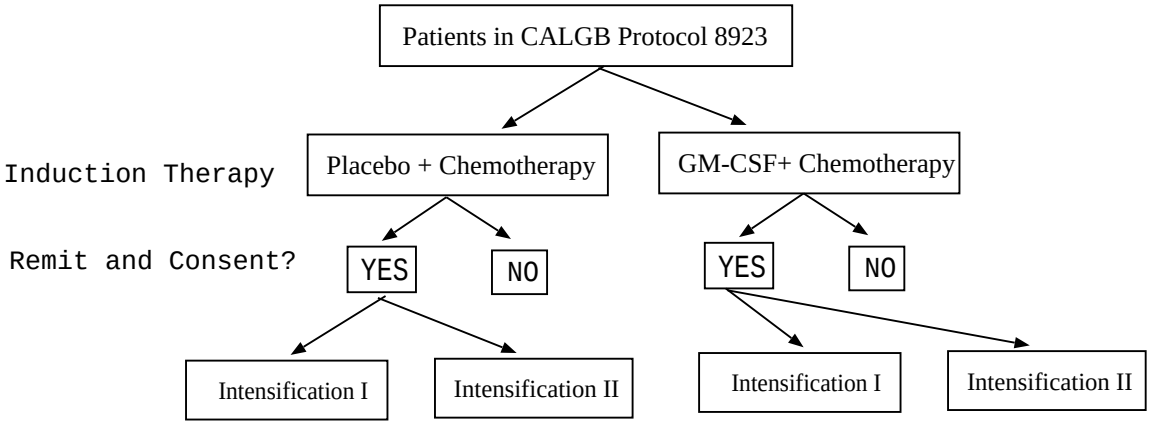

Figure 1.1: Two stages of Cancer and Lukemia Group B Protocol 8923 clinical trial.

to one of the maintenance therapies B_1 or B_2. Our objective is to compare the survival probabilities and related quantities associated with the treatment policies $A_jB_k, j, k = 1, 2$ where A_jB_k represents the policy "treat with A_j followed by B_k if the patient is eligible and consents to subsequent maintenance therapy". The way we defined the treatment policy suggests that we will be considering an intent-to-treat approach here. As is common in most survival analysis problem, the survival time is defined as the time from initial randomization until death.

Because of the randomizations involved, the problem easily fits into a missing data problem and taking this into account, Lunceford, Davidian and Tsiatis (2002) (subsequently referred to as LDT) derived the inverse-probability-of-missing-weighted (IPMW) estimator for the mean restricted survival time and the survival distribution for treatment policies in a two-stage clinical trial. They also proposed two other estimators which are more efficient than the IPMW estimator. These estimators

were defined in an ad hoc basis and did not include the most efficient estimator. By this we mean there exists estimators that are more efficient than LDT estimators. In this dissertation study we use the theory of Robins, Rotnitzky and Zhao (1994) to characterize the most efficient estimator for this problem and show how to derive estimators which are easily computable and are more efficient than the LDT estimator.

This has been done in steps. In Chapter 2 we derive the most efficient semiparametric estimator for the cases where there is no censoring and propose estimators that will be more efficient than LDT estimators. We applied the proposed estimators to the CALGB data set and also considered simulation studies to compare them to the LDT estimator. It turns out that the proposed estimators are always more efficient than the LDT estimators. In Chapter 3 we considered non-informative right-censored data and develop efficient procedures for estimating the survival distribution and related quantities for different treatment policies. Finally, in Chapter 4 we give a brief discussion.

Chapter 2

Efficient Estimation of The Survival Distribution and Related Quantities for Treatment Policies in Two-Stage Randomization Designs in Clinical Trials

The data that arises from a two stage randomization designs has enough complicated structure to make the derivation of the semiparametric efficient estimators difficult, even when there is no censoring. In this chapter we develop the methodologies for the cases where one observes the complete data. The preliminary notation developed in Chapter 1 will be used throughout this dissertation. The main goal of this chapter is to define efficient estimators for the survival distributions and related quantities such as mean survival time, survival probability at a time point t, etc. for policies

$A_jB_k, j, k = 1, 2$. Here is an outline of how this chapter is organized. In Section 2.1, using potential outcomes, we make explicit the model framework necessary to define the survival distribution for treatment policy A_jB_k. Also, in this section we elucidate all the assumptions made and give a brief review of available methodologies. In Section 2.2, we derive the class of all regular asymptotically linear estimators and find the most efficient estimator within this class. Section 2.3 describes the construction of feasible locally efficient estimators. Another strategy, where we derive efficient estimators within a restricted class of regular asymptotically linear estimators that are easy to compute, is described in Section 2.4. In Section 2.5 we apply the different estimators to estimate and test for differences in the mean survival time for the different combinations of induction/maintenance treatment regimes in the CALGB dataset. In Section 2.6 we report on results from several simulation studies comparing our estimators with the available estimators.

2.1 Model Framework and Notation

Since the data from the patients who receive induction treatment A_1 are independent of the data that are collected from patients with induction treatment A_2, it suffices to consider the two treatment policies that are associated with the induction treatment A_1; namely A_1B_1 and A_1B_2. (The methods follow analogously for policies A_2B_1 and A_2B_2). Thus, for the time being, we will only consider the case where each patient

in our sample received A_1 as the initial treatment and B_1 or B_2 as the subsequent treatment if they are eligible and consent. We will index individuals in our study by i, $i = 1, 2, \ldots, n$.

As in Lunceford et al. (2002), we conceptualize this problem through the use of a set of random variables some of which may not be observed for all individuals. Assume that each patient i has an associated set of random variables

$$\{R_i, (1 - R_i)T_{0i}, R_i T_i^R, R_i T_{1i}^*, R_i T_{2i}^*, R_i V_i\},$$

where,

R_i: the eligible/consent status if patient i were assigned to A_1; that is, $R_i = 1$ if patient i was eligible and would consent to subsequent maintenance treatment; $R_i = 0$, otherwise;

T_{0i}: the survival time of patient i if $R_i = 0$; that is, the survival time for a patient that was not eligible or refused subsequent maintenance treatment;

T_i^R: the time from initial randomization to the time he/she receives maintenance therapy and is defined only if $R_i = 1$;

T_{1i}^*: the survival time of patient i if the patient was eligible, willing to receive maintenance treatment and received treatment B_1;

T_{2i}^*: the survival time of patient i if the patient was eligible, willing to receive maintenance treatment and received treatment B_2; and

V_i: a vector of auxiliary covariates collected on individual i prior to their second randomization and is defined only if $R_i = 1$.

The auxiliary covariates V_i may include baseline covariates, as well. In the CALGB data, some examples of auxiliary covariates include elapsed time between response to the induction therapy and second randomization, age and white blood cell count. In actuality, if patient i were eligible and consented to the maintenance randomization ($R_i = 1$), we could not observe both T^*_{1i} and T^*_{2i} which is why these are referred to as counterfactuals (see Holland, 1986) or potential outcomes.

Continuing with this conceptualization, the survival time for patient i, if assigned to treatment policy A_1B_1, would be

$$T_{11i} = (1 - R_i)T_{0i} + R_iT^*_{1i}$$

and

$$T_{12i} = (1 - R_i)T_{0i} + R_iT^*_{2i}$$

if assigned to treatment policy A_1B_2. The variables (T_{11i}, T_{12i}) are also potential outcomes since they are not necessarily both observed for each individual, rather, they represent what might occur under policies, contrary to that to which the individual might actually be exposed.

Under this setup, the distribution of $T_{1k}, k = 1, 2$ represents the distribution of the potential survival time for the population, were all patients to be assigned to A_1B_k, realizing that some patients eligible for maintenance therapy B_k may refuse additional treatment, so that inference on features of these distributions addresses

directly the "intent-to-treat" question of interest. Our goal is to draw inference on the distribution of variables of interest T_{1k} from the observed data from a two-stage design described in Section 1.

In contrast to the potential outcomes defined above, the observed data can be characterized as

$$(R_i, R_i T_i^R, R_i V_i, R_i Z_i, T_i)$$

where R_i, T_i^R, V_i are defined exactly as above, but now Z_i denotes the B treatment assignment indicator, defined only if $R_i = 1$, where $Z_i = 1$, if assigned to treatment B_1, 0, if assigned to B_2 and T_i is the observed survival time. We make the reasonable assumption that the observed survival time for patient i is related to the potential outcomes by

$$T_i = (1 - R_i)T_{0i} + R_i \left\{ Z_i T_{1i}^* + (1 - Z_i) T_{2i}^* \right\}. \tag{2.1}$$

to reflect the belief that patient's i survival time would be T_{0i} if he/she didn't receive maintenance therapy, T_{1i}^* if he/she received B_1 as maintenance therapy and T_{2i}^* if he/she received B_2 as maintenance therapy.

In addition, we assume that

$$P(Z_i = 1 | R_i = 1, T_i^R, V_i, T_{1i}^*, T_{2i}^*) = P(Z_i = 1 | R_i = 1), \tag{2.2}$$

to reflect the fact that, by design, the second stage randomization is made independently of prognosis (T_{1i}^*, T_{2i}^*) or any pre-second stage randomization characteristics $\left(T_i^R, V_i\right)$ of the patient. We define $\pi_1 = P(Z_i = 1 | R_i = 1)$ and $\pi_2 = 1 - \pi_1 = P(Z_i =$

$0|R_i = 1)$ to denote the probability of being randomized to treatments B_1 or B_2 respectively.

Our primary goal is to estimate parameters involving the distribution of the treatment policy survival times T_{1k} for $k = 1, 2$. For example, we may want to estimate $\mu_{1k} = E\{h(T_{1k})\}$ for some function $h(\cdot)$ of T_{1k}. This allows us to consider the estimation of parameters such as the mean survival time or the survival distribution for treatment policy A_1B_k by taking $h(T_{1k}) = T_{1k}$, or $h(T_{1k}) = I(T_{1k} \geq t)$ respectively. One naive approach in estimating such quantities is to average the function $h(T_i)$ over those patients whose data are consistent with the treatment policies they are randomized to. Explicitly, one might use the estimator

$$\widehat{\mu_{1k}}^{NAIVE} = \left\{ \sum_{i=1}^{n} (1 - R_i + R_i X_{ki}) \right\}^{-1} \sum_{i=1}^{n} (1 - R_i + R_i X_{ki})\, h(T_i). \tag{2.3}$$

where X_k is the assignment indicator for treatment B_k, $X_1 = Z$ and $X_2 = (1 - Z)$. This estimator, as we will demonstrate, is biased.

In order to find unbiased and consistent estimators, we first need to show that the distribution of the potential outcome T_{1k} can be identified from the distribution of the observed random variables. It was shown in LDT that under assumptions (2.1) and (2.2) that

$$\mu_{1k} = E\{h(T_{1k})\} = E\left\{ \left(1 - R + \frac{RX_k}{\pi_k} \right) h(T) \right\}, \tag{2.4}$$

Relationship (2.4) leads to one of the estimators for μ_{1k} given by LDT, namely,

$$\widehat{\mu}_{1k} = \frac{1}{n} \sum_{i=1}^{n} \left(1 - R_i + \frac{R_i X_{ki}}{\pi_k} \right) h(T_i). \tag{2.5}$$

A useful way to think about this problem is as follows: If everyone in our sample were given treatment according to policy A_1B_k, then we would have complete data that could be used to estimate μ_{1k} in a straightforward fashion. Some individuals, however, were assigned treatment inconsistent with treatment policy A_1B_k, namely, those individuals randomized to receive the other maintenance therapy $B_{3-k},\ k = 1, 2$. The data from such individuals can be viewed as missing data for the purpose of estimating μ_{1k}. However, because of randomization, such individuals are similar prognostically to those randomized to treatment B_k. Consequently, by weighting the individuals randomized to treatment B_k by $\frac{1}{\pi_k}$, then, roughly speaking, the response of an individual randomized to treatment B_k counts for him/herself as well as the response of $(\frac{1}{\pi_k} - 1)$ similar individuals who have "missing data" with respect to treatment policy A_1B_k; i.e. those individuals randomized to the other treatment B_{3-k}. This also makes clear why the naive estimator given by (2.3), which does not weight, results in a biased estimator.

Other ad hoc estimators were also given by LDT. The problem we address in this paper is how to efficiently estimate parameters involving the distribution of the treatment policy survival times T_{1k} for $k = 1, 2$ using the observed data including the auxiliary covariates. Because this problem can be cast as a missing data problem, we can use the theory developed by Robins et al. (1994) to characterize the class of all estimators and find the most efficient estimator.

2.2 Efficient Estimator

Most estimators used in practice are regular asymptotically linear estimators (RAL). That is, the estimator minus the estimand can be approximated asymptotically by a sum of identically and independently distributed (iid) mean zero random variables. Specifically, an estimator $\hat{\eta}$ of the parameter η is asymptotically linear if

$$n^{1/2}(\hat{\eta} - \eta) = n^{-1/2} \sum_{i=1}^{n} \psi_i + o_p(1),$$

where $\psi_i, i = 1, \ldots, n$ are iid mean zero random variables and $o_p(1)$ denotes a term that converges in probability to zero. The random variable ψ_i is referred to as the i-th influence function of the estimator $\hat{\eta}$. It is clear from the representation above that the asymptotic variance of an asymptotically linear estimator is equal to the variance of the influence function. Consequently, the optimal estimator among a class of asymptotically linear estimators is the one whose influence function has the smallest variance. The restriction to regular estimators is a technical condition imposed to exclude estimators that have undesirable local properties. For details, the reader is referred to Newey (1990).

The estimator given by (2.5) is an example of an inverse-probability-of-missing-weighted (IPMW) estimator for μ_{1k}. The influence function for this estimator can be shown to be equal to

$$\left\{ (1 - R_i) + \frac{R_i X_{ki}}{\pi_k} \right\} h(T_i) - \mu_{1k}. \tag{2.6}$$

Using the semiparametric theory of Robins et al. (1994), all RAL estimators for μ_{1k}

have an influence function belonging to the class

$$\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}\{h(T_i)-\mu_{1k}\}+R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)f(T_i^R,V_i), \tag{2.7}$$

where $f(T_i^R, V_i)$ is an arbitrary function of T_i^R, the time to response to the induction therapy, and V_i, the vector of auxiliary covariates measured prior to the second randomization. Note that both T_i^R and V_i are defined only for patients i such that $R_i = 1$. The choice of the function $f(\cdot)$ will determine how efficient the corresponding estimator for μ_{1k} will be. The goal would be to appropriately define the function $f(\cdot)$ so that we can improve the efficiency of our estimators. The use of auxiliary information in gaining efficiency has previously been considered by several authors such as Robins and Rotnitzky (1992), Laan and Hubbard (1998, 1999), Xu and Zeger (2001), and Faucett et al. (2002).

The following proposition characterizes the most efficient influence function in the class of influence functions (2.7).

Proposition 1 *Among all influence functions in (2.7), the most efficient one is given by*

$$\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}h(T_i)-R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\theta_h(T_i^R,V_i)-\mu_{1k} \tag{2.8}$$

where $\theta_h(T_i^R, V_i) = E\left\{h(T_i)|T_i^R, V_i, R_i = 1, X_{ki} = 1\right\}$.

The proof of this proposition is given in Appendix A.1. If $\theta_h(T_i^R, V_i)$ were known (which is not the case in practice), then the estimating equation

$$\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}h(T_i)-R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\theta_h(T_i^R,V_i)-\mu_{1k}\right]=0 \tag{2.9}$$

could be used to find the efficient estimator

$$\widehat{\mu}_{1k}^{ME} = \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_i X_{ki}}{\pi_k}\right\}h(T_i) - R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\theta_h(T_i^R, V_i)\right]. \quad (2.10)$$

Theoretically, $\widehat{\mu}_{1k}^{ME}$ is referred to as optimal in the sense that it has the smallest variance among the class of all RAL estimators and its variance is said to achieve the semiparametric efficiency bound (Newey, 1990). Since the conditional expectation $\theta_h(T_i^R, V_i)$ is not known, it must be estimated from the data leading to locally efficient estimators.

2.3 Locally Efficient Estimators

If we want to use the estimator defined in (2.10), we need to estimate the conditional expectation $\theta_h(T_i^R, V_i)$ from the data. To do so, one can posit a regression model, linear or non-linear, of the form

$$E\left\{h(T_i)\ T|_i^R, V_i, R_i = 1, X_{ki} = 1\right\} = g(T_i^R, V_i, \boldsymbol{\gamma}) \quad (2.11)$$

in terms of a finite number of parameters $\boldsymbol{\gamma}$ which can be estimated using standard techniques such as least squares using the subset of the data for individuals $\{i : R_i = 1, X_{ki} = 1\}$. Then $\widehat{\theta}_h(T_i^R, V_i) = g(T_i^R, V_i, \widehat{\boldsymbol{\gamma}})$ can be substituted in (2.10) to give the locally efficient estimator of μ_{1k}

$$\widehat{\mu}_{1k}^{LE} = \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_i X_{ki}}{\pi_k}\right\}h(T_i) - R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)g(T_i^R, V_i, \widehat{\boldsymbol{\gamma}})\right]. \quad (2.12)$$

We give a brief argument in Appendix B to show that this estimator is consistent and asymptotically normal even if the posited regression relationship (2.11) is incorrectly

specified. In addition, if the posited regression model is correctly specified, then it will be the most efficient estimator for μ_{1k}. The way this estimator is constructed suggests that the efficiency gain for this estimator over IPMW or LDT estimators depends on how correlated the response time and the auxiliary variables are to the survival time among responders. As will be seen in simulation studies, the higher the correlation, the larger the gain. The variance of this estimator can be estimated by the sandwich variance given by

$$\widehat{\sigma^2_{LE}} = \frac{1}{n^2}\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}h(T_i) - R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)g(T_i^R,V_i,\widehat{\boldsymbol{\gamma}}) - \widehat{\mu}_{1k}^{LE}\right]^2. \tag{2.13}$$

For instance, if we assume that

$$g(T_i^R,V_i,\boldsymbol{\gamma}) = \gamma_0 + \gamma_1 T_i^R + \boldsymbol{\gamma}_2^T V_i$$

and estimate $\boldsymbol{\gamma} = (\gamma_0,\gamma_1,\boldsymbol{\gamma}_2^T)^T$ by the least squares estimates $\widehat{\boldsymbol{\gamma}}$ from the subset of data corresponding to the individuals $\{i : R_i = 1, X_{ki} = 1\}$, then we obtain the locally efficient estimator

$$\widehat{\mu}_{1k}^{LS} = \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}h(T_i) - R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)(\widehat{\gamma_0}+\widehat{\gamma_1}T_i^R+\widehat{\boldsymbol{\gamma}_2}^TV_i)\right]. \tag{2.14}$$

Alternatively, the conditional expectation $\theta_h(T_i^R,V_i)$ can be estimated by local non(semi)-parametric regression methods such as loess (Cleveland and Devlin, 1988) for cases with few independent variables or generalized additive models (GAM) (Hastie and Tibshirani, 1990) when considering many independent variables. Let us denote this estimator by $\tilde{\theta_h}(T_i^R,V_i)$. Substituting this estimated conditional expectation

in (2.10) we obtain another estimator for μ_{1k}. In Section 2.5, for the data analysis using several auxiliary variables, we have used the generalized additive model to estimate the conditional expectation and refered to the corresponding estimator of μ_{1k} as $\widehat{\mu}_{1k}^{GAM}$. In our simulation study, where only a single auxiliary variable is considered, we used the loess method and the corresponding estimator of μ_{1k} is refered to as $\widehat{\mu}_{1k}^{LOESS}$. These two estimators do not depend on the choice of a particular model but the slow convergence rates for local regression methods may have an effect on overall consistency and asymptotic normality which we investigate numerically in our simulation study. The variance of these estimators can be easily estimated by (2.13) by replacing $g(T_i^R, V_i, \widehat{\boldsymbol{\gamma}})$ with $\tilde{\theta}_h(T_i^R, V_i)$ and $\widehat{\mu}_{1k}^{LE}$ with $\widehat{\mu}_{1k}^{GAM}$ for GAM estimator and with $\widehat{\mu}_{1k}^{LOESS}$ for the LOESS estimator.

2.4 Improved Estimator

The construction of the first set of locally efficient estimators that were derived in Section 2.3 involves the selection of an appropriate model for the conditional expectation, which may or may not be correct. If the regression relationship (2.11) is incorrectly specified, then there is no guarantee that the estimator will gain efficiency over the IPMW estimator. Another approach that guarantees improved efficiency is to restrict the class of estimators to those whose influence functions are of the form

$$\left\{(1-R_i)+\frac{R_i X_{ki}}{\pi_k}\right\}\{h(T_i)-\mu_{1k}\}+R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\boldsymbol{\gamma}^T\boldsymbol{W}_i, \qquad (2.15)$$

where $\boldsymbol{W}_i$ is a fixed q-dimensional vector of prespecified functions of T_i^R and V_i, and $\boldsymbol{\gamma}$ is an arbitrary q-dimensional constant vector. The influence functions (2.15), for $\boldsymbol{\gamma} \in \mathcal{R}^q$, define a linear subspace of the space of influence functions (2.7) and the goal is to find the optimal estimator within this class; that is, to find the estimator whose influence function is the one within the class (2.15) with smallest variance. This entails finding the q-dimensional vector $\boldsymbol{\gamma}^{opt}$ which gives the smallest variance. Formalizing this as a multiple regression problem, $\boldsymbol{\gamma}^{opt}$ is given by

$$\begin{aligned} \boldsymbol{\gamma}^{opt} &= -\left[E\left\{R(Z_k-\pi_k)^2\boldsymbol{W}_i\boldsymbol{W}_i^T\right\}\right]^{-1} E\left[RZ_k(Z_k-\pi_k)\left\{h(T_i)-\mu_{1k}\right\}\boldsymbol{W}_i\right] \\ &= -\boldsymbol{\gamma}^{*}+\boldsymbol{\gamma}^{**}\mu_{1k} \end{aligned} \tag{2.16}$$

with

$$\boldsymbol{\gamma}^{*} = \left[E\left[R_i(X_{ki}-\pi_k)^2\boldsymbol{W}_i\boldsymbol{W}_i^T\right\}\right]^{-1} E\left\{R_iX_{ki}(X_{ki}-\pi_k)h(T_i)\boldsymbol{W}_i\right\}. \tag{2.17}$$

and

$$\boldsymbol{\gamma}^{**} = \left[E\left\{R_i(X_{ki}-\pi_k)^2\boldsymbol{W}_i\boldsymbol{W}_i^T\right\}\right]^{-1} E\left\{R_iX_{ki}(X_{ki}-\pi_k)\boldsymbol{W}_i\right\}. \tag{2.18}$$

If the coefficient vectors $\boldsymbol{\gamma}^{*}$ and $\boldsymbol{\gamma}^{**}$ were known, one could solve the estimating equation

$$\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}\left\{h(T_i)-\mu_{1k}\right\}-R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\left(\boldsymbol{\gamma}^{*}-\mu_{1k}\boldsymbol{\gamma}^{**}\right)^T\boldsymbol{W}_i\right]=0 \tag{2.19}$$

to obtain the optimal restricted estimator

$$\widehat{\mu_{1k}}^{MEL} = \frac{\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}h(T_i)-R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\boldsymbol{\gamma}^{*T}\boldsymbol{W}_i\right]}{\sum_{i=1}^{n}\left[\left\{(1-R_i)+\frac{R_iX_{ki}}{\pi_k}\right\}-R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\boldsymbol{\gamma}^{**T}\boldsymbol{W}_i\right]}. \tag{2.20}$$

It can be shown that substituting root-n consistent estimators for $\boldsymbol{\gamma}^*$ and $\boldsymbol{\gamma}**$ in (2.19) and (2.20) will yield estimators that are asymptotically equivalent to those where $\boldsymbol{\gamma}^*$ and $\boldsymbol{\gamma}**$ are known. Thus, replacing the expectations in (2.17) and (2.18) by their corresponding empirical averages we obtain estimates $\widehat{\boldsymbol{\gamma}^*}$ and $\widehat{\boldsymbol{\gamma}^{**}}$ which can then be substituted in (2.20) to obtain an improved estimator which we will denote by $\widehat{\mu_{1k}}^{IMP}$. The variance of this estimator can be estimated by the sandwich variance given by

$$\widehat{\sigma^2_{IMP}} = \frac{1}{n^2}\sum_{i=1}^{n}\left[\left\{ (1-R_i) \quad + \quad \frac{R_i X_{ki}}{\pi_k}\right\}\left\{h(T_i) - \widehat{\mu_{1k}}^{IMP}\right\} \right. \\ \left. - \quad R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\left(\widehat{\boldsymbol{\gamma}^*} - \widehat{\mu_{1k}}^{IMP}\widehat{\boldsymbol{\gamma}^{**}}\right)^T \boldsymbol{W}_i\right]^2 \quad (2.21)$$

This estimator is optimal within the restricted class of RAL estimators which have influence function in the class defined by (2.15). If we take $\boldsymbol{W}_i$ to be a scalar constant, then this reduces to the estimator $\widehat{\mu}''_{1k}$, one of the LDT estimators. From here on we will refer to the estimator $\widehat{\mu}''_{1k}$ as the LDT estimator.

It is also argued in Appendix B that the estimator $\widehat{\mu}^{IMP}_{1k}$ is consistent and asymptotically normal. It is easy to compute and because it is the most efficient estimator among a class of estimators that include the IPMW estimator and the LDT estimator, it is guaranteed to be at least as efficient as the best of these. Moreover, if the conditional expectation $\theta_h(T_i^R, V_i)$ given by (2.8) is equal to $\boldsymbol{\gamma}^T\boldsymbol{W}_i$ for some $\boldsymbol{\gamma}$, then $\widehat{\mu}^{IMP}_{1k}$ is the most efficient among all RAL estimators.

2.5 Analysis of CALGB 8923 data

We apply the methods developed for improving efficiency described in the previous sections to estimate the survival distribution and mean lifetime for the different combinations of induction/maintenance treatment policies using the data from CALGB 8923 described earlier in the Introduction. We also use Wald tests, constructed using these estimators, to test various treatment policy contrasts. There were 388 patients that participated in CALGB 8923. Of these, 79 out of 193 patients in the GM-CSF group and 90 out of 195 in the placebo group achieved remission (responded) and consented to further randomization to the intensification therapy; and, of these, 37 GM-CSF and 45 placebo patients were randomized to intensification therapy I and the rest to intensification therapy II. This study has matured and all 388 patients have been followed for at least 2521 days of whom 356 have died. Since we have complete data for 2521 days, we consider survival time restricted to 2521 days. Thus, in the analysis, we will estimate and test the mean restricted survival time for the different treatment policies.

In our analysis, there were three variables that we considered as auxiliary variables V_i; namely, time between the response and the second randomization, age and white blood cell count. For modeling the conditional expectation (2.11), first we used a least squares linear regression on these auxiliary variables plus the time of response T_i^R. Then we used proc GAM with default options in SAS to fit the generalized additive model in these four variables to estimate the conditional expectation (2.11). For

details of how this is implemented, the reader is refered to Chapter 5 of "SAS/STAT Software: Changes and Enhancements, Release 8.2". For modeling the conditional survival probability, logistic distribution was used in the generalized additive model. Similarly, for the improved estimator, we defined the prespecified vector function $\boldsymbol{W}_i$ as the column vector whose elements are the random variables T_i^R, time between the response and the second randomization, age, white blood cell count and a constant function identically equal to 1.

Table 2.1 shows the estimates of mean restricted survival time for each of the four treatment policies using the estimators NAIVE, IPMW, LDT, LS, IMP and GAM. Most estimators gave similar estimates except for the NAIVE estimator which underestimates the mean restricted lifetime reflecting the bias of this method. The IPMW estimator gave the largest estimated standard error as would be expected by the theory. In most cases the GAM estimator has the smallest estimated standard error, but, as noted earlier, we cannot be confident of the small sample accuracy of this estimated standard error because of the slow convergence rate of such smoothing methods. The other three estimators LDT, LS, and IMP gave very similar results both in terms of the estimates as well as the standard errors. Among responders, the time to response and the auxiliary covariates were weakly related to survival time (as seen in Figure 2.7)which explains why there was not any appreciable gain in efficiency of the LS and IMP estimators as compared to the LDT estimator. Similar conclusions follow for the survival probability estimates as is evident from the results in Table 2.2.

Estimated survival curves were computed using these different estimation tech-

niques for the four treatment policies. The treatment-policy specific curves using the IMP method are depicted in Figure 2.7. We also give the estimated survival curves using the different estimators for the policy GM-CSF/Maintenance I (see Figure 2.7). As expected by the theory, except for the NAIVE estimator, all other survival curves were similar. The same conclusion was reached for the other three treatment policies as well.

A Wald chi-square test of equality of treatment means ($H_0 : \mu_{11} = \mu_{21} = \mu_{12} = \mu_{22}$) did not show any significant differences. Similar conclusions, showing no significant difference in mean survival times, follow for comparing main effects of GM-CSF which is tested by the null hypothesis $H_1 : (\mu_{11} + \mu_{12})/2 = (\mu_{21} + \mu_{22})/2$ and a comparison of the two intensification therapies which is tested by the null hypothesis $H_2 : (\mu_{11} + \mu_{21})/2 = (\mu_{12} + \mu_{22})/2$. The conclusion is the same no matter which estimator we use except for the NAIVE estimators. Equivalent hypotheses were tested for survival probabilities and resulted in similar conclusions for survival probability at 548 days except that the test using the GAM estimator for the main effects of induction therapies turned out to be statistically significant. The results of all these tests are summarized in Table 2.3.

Test results for hypotheses related to the survival probability at other time points also showed no significant differences among the four policies.

2.6 Simulation Study

To assess the accuracy of the large sample properties of our estimators with moderate sample sizes and to compare the relative performance of the different estimators, we conducted several simulation experiments. For simplicity, in our simulation studies, we only allowed the survival time to depend on the response time T_i^R and did not consider any additional auxiliary variables. Since data from patients assigned to treatment A_1 are independent of data from patients who receive treatment A_2, we only simulate data for "A_1-patients".

We took R_i, the eligible/consent indicator, to be Bernoulli with $P(R_i = 1) = \pi_R$ and considered two different values of π_R, 0.5 and 0.7. When $R_i = 0$, a survival time T_{0i} is generated from an exponential distribution with mean λ truncated at b_2. When $R_i = 1$, treatment B assignment indicator Z_i is generated from Bernoulli(.5) distribution. Also when $R_i = 1$, a response time T_i^R is generated from an exponential distribution with mean α truncated at b_1. To examine the effect that correlation among responders, between the survival time and the auxiliary variables has on the relative efficiency of the various estimators, we considered a linear relationship between the survival time of responders and the auxiliary variable (response time in this case) generated by

$$T_{1i}^* = T_i^R + (\beta_1 + \beta_2 T_i^R)U_{1i} \tag{2.22}$$

$$T_{2i}^* = T_i^R + (\beta_1 + \beta_2 T_i^R)U_{2i} \tag{2.23}$$

where $U_{ji}, j = 1, 2$ is generated from a uniform$(0, \theta_j)$ distribution. The strength of

the correlation is determined by the choice of β_2. Finally we defined

$$T_i = (1 - R_i)T_{0i} + R_i \{Z_i T_{1i}^* + (1 - Z_i)T_{2i}^*\}$$

to generate the observed survival time for the i^{th}individual.

In the first simulation scenario, we considered $\lambda = 365, b_2 = 1095, \alpha = 365, b_1 = 730, \beta_1 = 1.0, \beta_2 = 1.0, \theta_1 = 1.5, \theta_2 = 1$ so that when $\pi_R = 0.5, \mu_{11} = 510.4$ days, $\mu_{12} = 433.4$ days and $\mu_{11} = 591.5$ days, $\mu_{12} = 483.6$ days when $\pi_R = 0.7$. We considered 5000 Monte-Carlo samples of sizes 200 and 500. Under this scenario, the correlation among responders between T_i^R and T_{1i}^* is .53, and .70 between T_i^R and T_{2i}^* representing moderate to high correlations.

For each of the 5000 simulated data sets, $\mu_{1k} = E(T_{1k})$ and $S_{1k}(t) = P(T_{1k} > t)$ were estimated for $k = 1, 2$ and for $t = 183$ days and $t = 548$ days representing an earlier and later time point of the study. All estimators discussed in earlier sections were considered. For improved and locally efficient estimators, linear regression of survival time on the response time was used. For the LOESS regression, the smoothing parameter was set to .4. Other values of the smoothing parameters such as .2, .3, .5, .6 were also considered. These gave similar results and are not presented here.

Tables 2.4 and 2.5 present the coverage probabilities for 95% Wald intervals and relative efficiencies for all the estimators under consideration with respect to the LDT estimator. The relative efficiencies are defined in terms of the ratio of the Monte-Carlo mean squared errors compared to the LDT estimator. The LS, IMP and LOESS estimators were always more efficient than the LDT estimator. The

IPMW estimator was by far the least efficient of these five estimators and the LS estimator was the most efficient. The NAIVE estimator, as expected, performs very poorly with coverage probability always below 60% (80%, for the survival probability estimates) and sometimes as low as 8% (13%). The LOESS estimator tended to have some bias with smaller sample sizes as evidenced by the poor coverage probabilities. The coverage probabilities for the LOESS estimator improved with increasing sample sizes. We believe this is due to the slow convergence rate for such local regression methods. We expect this difficulty would be exacerbated if one considered many auxiliary variables where we would run into the curse of dimensionality. In such cases, the use of generalized additive or spline models may be a useful alternative for estimating high dimensional conditional expectations with sufficient data.

From the above analysis it is clear that when the auxiliary variables are at least moderately correlated to the survival times among the responders, the locally efficient estimators gain sufficiently over the LDT or IPMW estimators. A second simulation scenario was also considered where the correlation between the time to response and time to death among the responders were weakly correlated (similar to those observed in the CALGB study). The setup for this simulation is the same as the first one except that the two equations (2.22) and (2.23) have been replaced by

$$T_{1i}^{*} = E(T_i^R) + \left\{\beta_1 + (2/3)E(T_i^R) + (\beta_2/3)T_i^R\right\} U_{1i} \tag{2.24}$$

$$T_{2i}^{*} = E(T_i^R) + \left\{\beta_1 + (2/3)E(T_i^R) + (\beta_2/3)T_i^R\right\} U_{2i}. \tag{2.25}$$

This allowed us to preserve the values for μ_{11} and μ_{12} from the first scenario, but with

a low correlation among responders (.17 and .18 respectively) between T_i^R and T_{1i}^*, and T_i^R and T_{2i}^*. The results for estimating the mean survival time for this scenario are given in Table 2.6. In this case, there is virtually no gain in using the locally efficient or the improved estimators. In terms of relative mean squared errors, the LS estimators always performed well while the LOESS and IMP estimators showed some loss of efficiency with smaller sample sizes. Also for small samples, the LOESS estimator had the worst coverage. For larger sample sizes, all the four estimators LDT, LS, IMP, and LOESS gave similar results.

2.7 Figures and Tables

Table 2.1: Estimates of mean restricted survival time in days (and corresponding standard error in the parenthesis) for different treatment policies for CALGB 8923 data.

Policy	LS	IMP	NAIVE	IPMW	LDT	GAM
GM-CSF/Intensification I	472(49)	469(49)	360(− − −)	441(62)	461(50)	423(46)
GM-CSF/Intensification II	487(61)	484(61)	396(− − −)	518(75)	492(62)	482(59)
placebo/Intensification I	562(60)	566(60)	486(− − −)	579(75)	579(61)	553(58)
placebo/Intensification II	587(64)	581(64)	481(− − −)	572(78)	572(65)	553(63)

Table 2.2: Estimated survival probability (corresponding standard error in the parenthesis) for different treatment policies in CALGB 8923 data.

Time	Policy	LS	IMP	NAIVE	IPMW	LDT	GAM
	GM-CSF/Intensification I	0.60(.036)	0.60(.036)	0.49(—)	0.57(.057)	0.59(.036)	0.55(.035)
183	GM-CSF/Intensification II	0.58(.038)	0.57(.038)	0.49(—)	0.60(.059)	0.58(.038)	0.55(.035)
days	placebo/Intensification I	0.64(.035)	0.63(.035)	0.53(—)	0.64(.059)	0.64(.035)	0.59(.035)
	placebo/Intensification II	0.58(.040)	0.57(.039)	0.49(—)	0.57(.057)	0.57(.040)	0.55(.036)
	GM-CSF/Intensification I	0.26(.038)	0.26(.038)	0.18(—)	0.25(.046)	0.26(.038)	0.24(.034)
548	GM-CSF/Intensification II	0.23(.039)	0.22(.039)	0.17(—)	0.24(.045)	0.23(.039)	0.22(.036)
days	Placebo/Intensification I	0.31(.041)	0.32(.041)	0.27(—)	0.33(.050)	0.33(.042)	0.32(.038)
	Placebo/Intensification II	0.32(.041)	0.32(.041)	0.25(—)	0.31(.048)	0.31(.041)	0.31(.038)

Table 2.3: P-values for different hypothesis for CALGB 8923 data.

	Testing mean restricted survival times						Testing survival prob. at $t = 548$ days					
Hyp.	LS	IMP	NAIVE	IPMW	LDT	GAM	LS	IMP	NAIVE	IPMW	LDT	GAM
H_0	0.70	0.69	0.19	0.51	0.63	0.40	0.52	0.44	0.22	0.53	0.43	0.33
H_1	0.17	0.16	0.03	0.12	0.15	0.13	0.11	0.08	0.02	0.08	0.08	0.04
H_2	0.68	0.75	0.59	0.67	0.81	0.51	0.70	0.57	0.66	0.75	0.39	0.62

Table 2.4: **Moderate and high correlation effect.** *Monte Carlo coverage probability and relative efficiency for estimators of mean survival time based on 5000 data sets: entries in parentheses are relative efficiencies (e.g., for the LS row,* $MSE(\widehat{\mu}_{1k}^{LDT})/MSE(\widehat{\mu}_{1k}^{LS})$*).*

	$\widehat{\mu}_{11}$				$\widehat{\mu}_{12}$			
	$n = 200$		$n = 500$		$n = 200$		$n = 500$	
Estimator	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$
NAIVE	40.5(0.25)	54.3(0.29)	08.2(0.11)	20.4(0.13)	60.9(0.35)	69.5(0.40)	27.1(0.16)	40.2(0.19)
IPMW	93.8(0.51)	93.8(0.48)	93.9(0.51)	94.4(0.48)	94.5(0.53)	94.7(0.48)	95.1(0.52)	95.1(0.47)
LDT	92.6(1.00)	93.6(1.00)	94.1(1.00)	94.4(1.00)	94.3(1.00)	94.3(1.00)	94.7(1.00)	95.1(1.00)
LOESS	89.4(1.09)	91.1(1.17)	92.4(1.18)	93.2(1.26)	90.7(1.15)	92.0(1.23)	93.1(1.22)	93.8(1.30)
IMP	91.5(1.15)	92.7(1.22)	93.4(1.22)	93.6(1.29)	92.2(1.18)	93.0(1.25)	94.1(1.24)	94.2(1.33)
LS	93.7(1.29)	94.4(1.36)	94.0(1.28)	94.4(1.34)	95.9(1.35)	96.1(1.42)	96.2(1.32)	96.5(1.40)

Table 2.5: **Moderate and high correlation effect.** *Monte Carlo coverage probability and relative efficiency for estimators of survival probability at $t = 548$ days based on 5000 data sets: entries in parentheses are relative efficiencies (e.g., for the LS row, $MSE\{\widehat{S}_{1k}^{LDT}(548)\}/MSE\{\widehat{S}_{1k}^{LS}(548)\}$). The true values are $S_{11}(548) = 0.38420, S_{12}(548) = 0.30922$ for 50% response and $S_{11}(548) = 0.46128, S_{12}(548) = 0.35702$ for 70% response.*

	$\widehat{S}_{11}(548)$				$S_{12}(548)$			
	$n = 200$		$n = 500$		$n = 200$		$n = 500$	
Estimator	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$
NAIVE	53.8(0.29)	65.4(0.34)	18.6(0.14)	12.6(0.17)	74.8(0.53)	80.9(0.61)	50.7(0.27)	37.9(0.34)
IPMW	94.4(0.68)	94.5(0.64)	93.9(0.70)	94.7(0.66)	94.8(0.79)	94.3(0.75)	94.9(0.79)	95.1(0.76)
LDT	94.3(1.00)	94.4(1.00)	94.1(1.00)	94.9(1.00)	94.6(1.00)	94.3(1.00)	94.6(1.00)	95.0(1.00)
LOESS	92.7(1.06)	93.3(1.11)	93.0(1.12)	94.9(1.17)	92.9(1.17)	93.2(1.25)	93.9(1.19)	95.0(1.27)
IMP	93.6(1.07)	93.9(1.12)	93.4(1.10)	94.5(1.16)	93.6(1.17)	93.8(1.24)	94.2(1.18)	94.9(1.25)
LS	94.2(1.15)	94.5(1.17)	94.1(1.15)	94.6(1.18)	94.8(1.23)	94.9(1.28)	94.9(1.22)	95.6(1.26)

Table 2.6: **Low correlation effect.** *Monte Carlo coverage probability and relative efficiency for estimators of mean survival time based on 5000 data sets: entries in parentheses are relative efficiencies (e.g., for the LS row,* $MSE(\widehat{\mu}_{1k}^{LDT})/MSE(\widehat{\mu}_{1k}^{LS})$*).*

	$\widehat{\mu}_{11}$				$\widehat{\mu}_{12}$			
	$n = 200$		$n = 500$		$n = 200$		$n = 500$	
Estimator	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$	$\pi_R = 0.5$	$\pi_R = 0.7$
NAIVE	28.7(0.15)	38.5(0.16)	02.5(0.06)	05.8(0.07)	49.8(0.19)	69.5(0.19)	13.3(0.09)	16.6(0.09)
IPMW	94.5(0.37)	94.1(0.31)	94.1(0.37)	94.5(0.31)	94.9(0.36)	94.7(0.28)	94.8(0.36)	94.7(0.27)
LDT	94.1(1.00)	94.2(1.00)	94.5(1.00)	94.5(1.00)	94.3(1.00)	94.3(1.00)	94.7(1.00)	95.1(1.00)
LOESS	91.9(0.91)	92.1(0.91)	93.1(0.96)	93.8(0.97)	92.1(0.90)	92.0(0.91)	94.0(0.96)	94.2(0.96)
IMP	93.0(0.92)	93.2(0.93)	94.0(0.97)	94.2(0.98)	92.9(0.91)	93.0(0.91)	94.2(0.95)	94.5(0.96)
LS	94.4(1.05)	94.6(1.06)	94.5(1.03)	94.8(1.03)	96.4(1.06)	96.1(1.06)	96.8(1.02)	96.9(1.02)

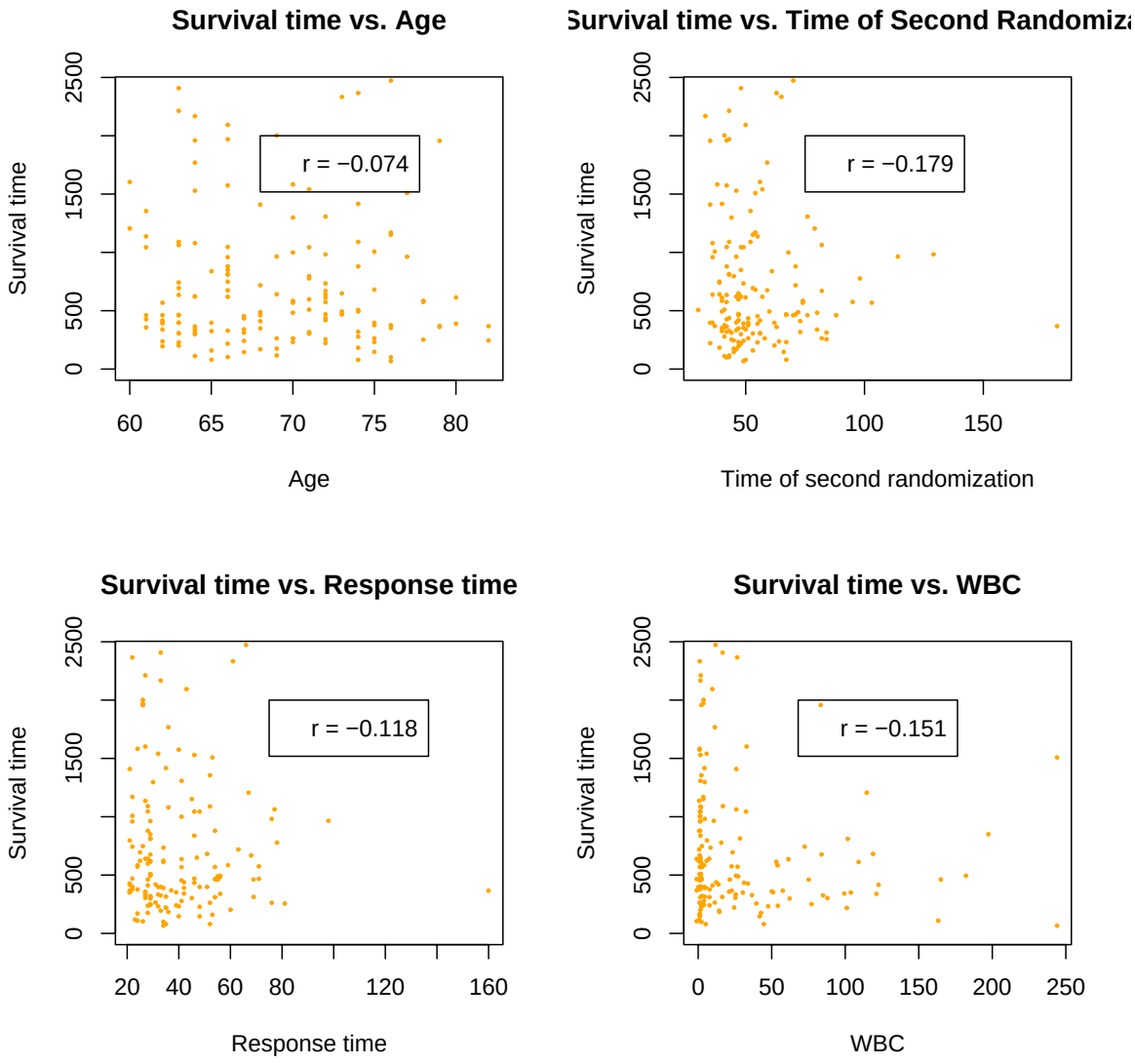

Figure 2.1: Scatter plot of survival time vs. various auxiliary variables from CALGB 8923 data.

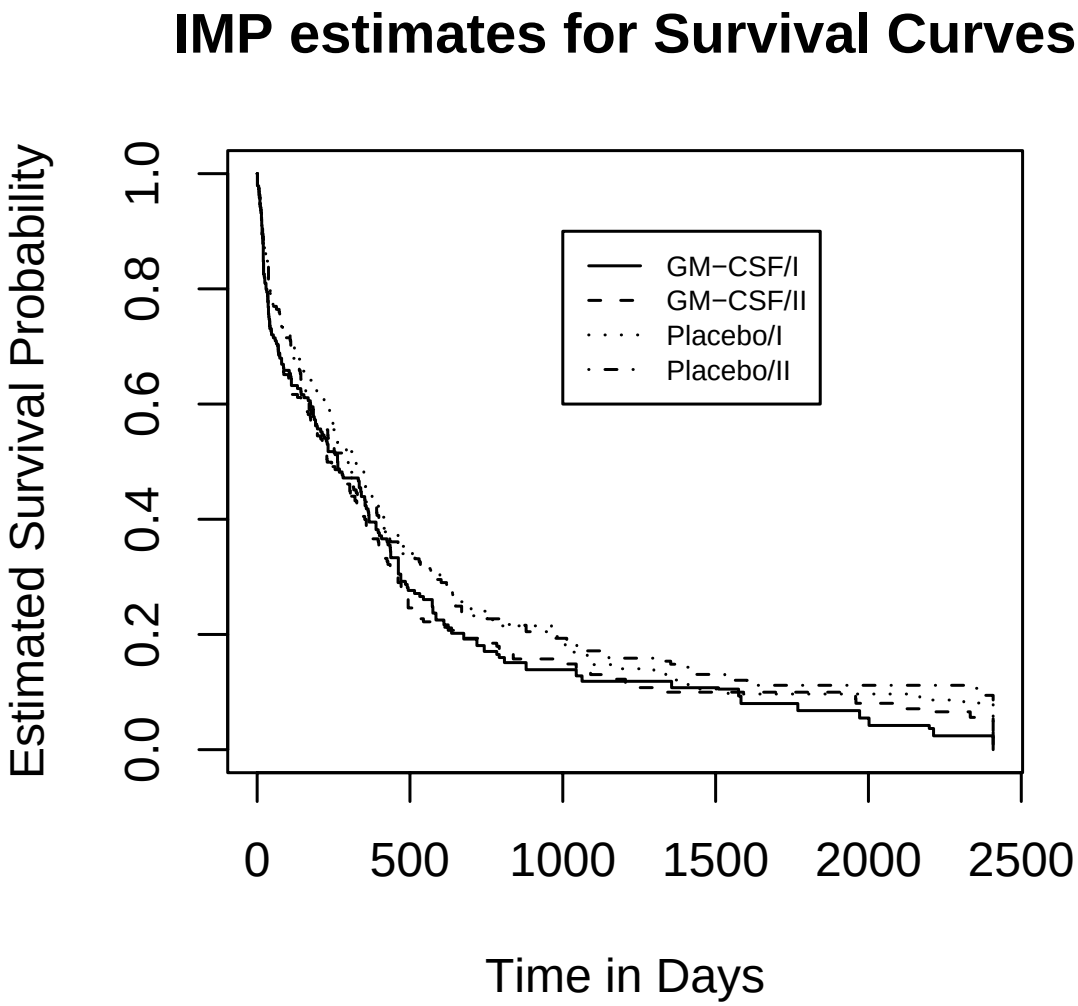

Figure 2.2: Estimated survival curves under IMP method.

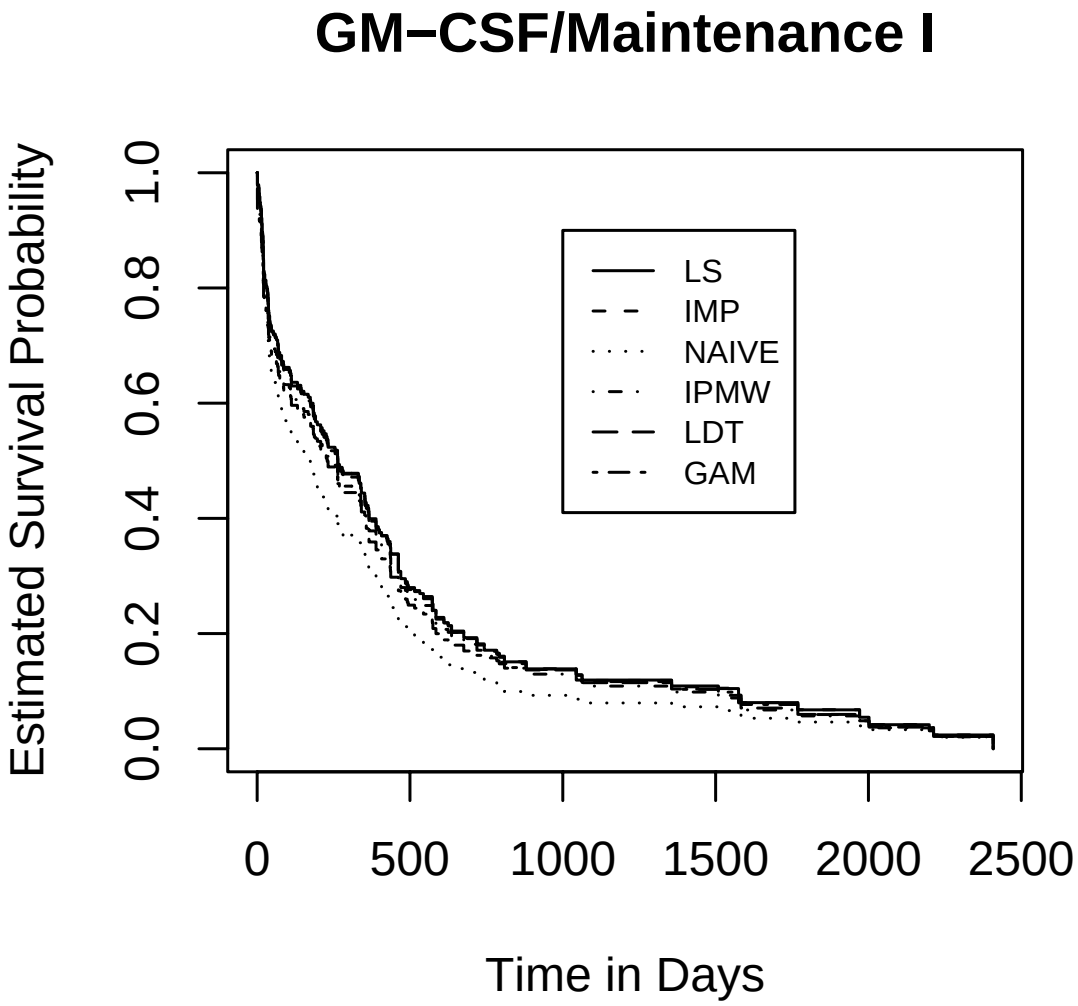

Figure 2.3: Estimated survival curves under GM-CSF/Maintenance I.

Chapter 3

Efficient Estimation of Survival Distribution and Related Quantities for Treatment Policies in Two-Stage Randomization Designs in Clinical Trials with Censored Data

So far we have considered issues related to the efficient estimation of survival distributions in two-Stage randomization designs in clinical trials when the data are observed completely. In many time-to-event studies, the data are subject to censoring due to patient drop-out, termination of study, etc. In this Chapter we will investigate the issue of efficiently estimating the survival distribution and related quantities in two-stage randomization designs with right-censored data. The problem here is more

complicated because, not only the survival time, but also the response to the first stage treatment and the time to response may be censored. We will deal with these issues in details in this chapter.

The chapter is organized as follows. In section 3.1, we modify our model derived in Chapter 2 to incorporate the right censoring. Also, additional assumptions are made related to censoring. In Section 3.2, we derive the class of all regular asymptotically linear estimators and find the most efficient estimator within this class for censored data. Efficient estimators within a restricted class of regular asymptotically linear estimators that are easy to compute, are derived in Section 3.3. In Section 3.4 we apply the different estimators to estimate and test for differences in the mean survival time for the different combinations of induction/maintenance treatment regimes in the CALGB dataset. In Section 3.5 we report on results from several simulation studies comparing our estimators with the available estimators. In the last section of this chapter, we present the Figures and Tables that were used in data analysis and in the simulation.

3.1 Model For Censored Data

Again, for simplicity, we will only consider the two treatment policies that are associated with the induction treatment A_1; namely A_1B_1 and A_1B_2. (The methods follow analogously for policies A_2B_1 and A_2B_2). Thus, for the time being, we will

only consider the case where each patient in our sample received A_1 as the initial treatment and B_1 or B_2 as the subsequent treatment if they are eligible and consent. We will index individuals in our study by i, $i = 1, 2, \ldots, n$.

A slight modification of the counterfactuals will be done to facilitate our purpose. Assume that each patient i has an associated set of random variables

$$\{R_i, (1-R_i)T_{0i}, (1-R_i)G^H(T_{0i}), R_iT_i^R, R_iT_{1i}^*, R_iT_{2i}^*, R_iG^H(T_{1i}^*), R_iG^H(T_{2i}^*)\},$$

where, as in Chapter 2

- R_i = the eligible/consent status if patient i were assigned to A_1; that is, $R_i = 1$ if patient i was eligible and would consent to subsequent maintenance treatment; $R_i = 0$, otherwise;
- T_{0i} = the survival time of patient i if $R_i = 0$; that is, the survival time for a patient that was not eligible or refused subsequent maintenance treatment;
- T_i^R = the time from initial randomization to the time he/she receives maintenance therapy and is defined only if $R_i = 1$;
- T_{1i}^* = the survival time of patient i if the patient was eligible, willing to receive maintenance treatment and received treatment B_1;
- T_{2i}^* = the survival time of patient i if the patient was eligible, willing to receive maintenance treatment and received treatment B_2;

and $G^H(u)$ denotes auxiliary information collected on individual i prior to time u. Specifically, $G^H(u)$ may contain covariates related to the patient survival (described

by V_i in Chapter 2). In the CALGB data, some examples of such auxiliary information comes through covariates including elapsed time between response to the induction therapy and second randomization, age and white blood cell count.

The main variables of interest would be the following survival times under policies $A_1B_k, k = 1, 2$:

$$T_{1ki} = (1 - R_i)T_{0i} + R_iT^*_{ki}, k = 1, 2.$$

As is evident from the expressions, the two variables (T_{11i}, T_{12i}) are not necessarily both observed for each individual, rather, they represent what might occur under policies, contrary to that to which the individual might actually be exposed. Therefore, inference about the distribution of $T_{1k}, k = 1, 2$ would apply to the population where all patients are assigned to the policy $A_1B_k, k = 1, 2$. This formulation recognizes the fact that in practice some patients eligible for maintenance therapy B_k may refuse additional treatment and hence allows us to consider the problem from an "intent-to-treat" point of view.

Since in most clinical trials, total follow-up time is limited, in this Chapter we will consider restricted survival time up to time L. This would mean that the variable T_{1k} will bear the meaning of $min(T_{1k}, L)$, unless stated otherwise.

The objective of the study is to draw inference on the distribution of variables of interest $T_{1k}, k = 1, 2$ or quantities related to the distribution of $T_{1k}, k = 1, 2$. In particular, one might be interested in answering the question: If everyone in a population were assigned to the policy A_1B_k, then how would there survival distribution

look like? Again, as in Chapter 2, we consider the specific problem of estimating $\mu_{1k} = h(T_{1k})$, where $h(.)$ can be any real-valued function of T_{1k}. This allows us to consider the estimation of parameters such as the mean survival time or the survival distribution for treatment policy A_1B_k by taking $h(T_{1k}) = T_{1k}$, or $h(T_{1k}) = I(T_{1k} \geq t)$ respectively.

If there were no censoring, then the observed data would be

$$(R_i, R_i T_i^R, G^H(T_i), R_i Z_i, T_i)$$

as was indicated in Chapter 2. We will assume, once again, that the observed survival time is related to the counterfactuals by Equation (2.1). To accommodate right censoring, suppose each patient i has an associated censoring time C_i. We will make the assumption that the censoring is non-informative. More explicitly, we assume that the distribution of C_i does not depend on any other variables in the sample or counterfactuals. That is,

$$P(C_i > t |\, R_i, R_i T_i^R, G^H(T_i), R_i Z_i, T_{1i}^*, T_{2i}^*) = P(C_i > t). \tag{3.1}$$

The censoring distribution may differ by the induction treatments A_1 and A_2, but since we are only considering data from A_1-patients, this difference is inconsequential. Let $K(t) = P(C_i > t)$ denote the survival distribution for the censoring time C_i. Since maximum follow-up time is greater than L, throughout the paper we will assume that there is always a positive probability of being censored at or beyond time L. i.e. $K(L) > 0$.

However, with the introduction of right censoring, the observed data can be written as

$$(U_i, \Delta_i, G^H(U_i)), i = 1, 2, \ldots, n.$$

where

- $U_i = min(T_i, C_i)$,
- $\Delta_i = I(T_i \leq C_i)$

and $G^H(u)$ as defined previously but with the censored data, $G^H(u)$ will also include information on whether the patient responded prior to time u or not; if they had responded, their time of response T_i^R, compliance status and the second treatment they are randomized to, if they responded and complied plus other auxiliary variables of interest.

Before we proceed further to construct estimators for $\mu_{1k} = E\{h(T_{1k}\})$ we note that by design, the probability of randomization to treatment B_k does not depend on counterfactuals, neither does it depend on the history of information collected prior to the randomization time. That is,

$$P(\, Z_i = 2 - k |\, R_i = 1, T_i^R, G^H(T_i^R), T_{1i}^*, T_{2i}^*) = P(Z_i = 2 - k | R_i = 1), k = 1, 2. \quad (3.2)$$

It may well depend on the induction treatment assignment. Since we are only considering patients from A_1 treatment, we will denote this randomization probability by π_k which is assumed to be known in our case. Define $X_{1i} = Z_i$ and $X_{2i} = 1 - Z_i$.

One naive approach in estimating μ_{1k} one could take is as follows: consider data from patients consistent with the treatment policy A_1B_k, that is individuals in the set $\{i : 1 - R_i + R_iX_{ki} = 1\}$ and use Kaplan-Meier estimator to approximate the survival distribution of T_{1k} and then use this distribution to estimate $\mu_{1k} = E\{h(T_{1k})\}$. Let us denote this estimator by $\widehat{\mu}_{1k}$. This estimator is expected to underestimate μ_{1k} because it does not account for the missingness due to the randomization to treatment B_{3-k}. We will investigate this in our simulation study.

In an attempt to derive unbiased estimators for μ_{1k}, in LDT it has been shown that, under assumptions (2.1) and (3.2),

$$E\left\{\frac{\Delta_i Q_{ki}}{K(U_i)}h(U_i)\right\} = \mu_{1k}, \tag{3.3}$$

where $Q_{ki} = 1 - R_i + R_iX_{ki}/\pi_k$. Equation (3.3) shows that the distribution of T_{1k} is identifiable from the distribution of observed data. This prompts to define the estimator

$$\widehat{\mu}_{1k} = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i Q_{ki}}{\widehat{K}(U_i)}h(U_i), \tag{3.4}$$

where

- $\widehat{K}(.)$ is the product limit estimator of the censoring survival distribution, namely, $\widehat{K}(t) = \prod_{u\leq t}(1 - dN^c(u)/Y(u))$
- $N^c(u) = \sum_{i=1}^{n} I(U_i \leq u, \Delta_i = 0)$, and
- $Y(u) = \sum_{i=1}^{n} I(U_i \geq u)$.

The rationale behind this estimator is as follows. For estimating quantities related to the distribution of T_{1k} the observed data can be treated as coarsened where the coarsening occurs from two different sources: first, because of the randomization to B treatment, some patients will be assigned to B_{3-k} treatment and hence data from such individuals will be missing for the purpose of estimating μ_{1k}; the second form of missingness occurs due to the fact that some individuals might be right censored. That is why there are two forms of weighting involved in (3.4). Consider the first weighting factor $Q_{ki} = 1 - R_i + R_i X_{ki}/\pi_k$. Those who are randomized to receive the maintenance therapy B_{3-k} are similar prognostically to those randomized to treatment B_k. Consequently, by weighting the individuals randomized to treatment B_k by $\frac{1}{\pi_k}$, then, roughly speaking, the response of an individual randomized to treatment B_k counts for him/herself as well as the response of $(\frac{1}{\pi_k} - 1)$ similar individuals who have "missing data" with respect to treatment policy $A_1 B_k$; i.e. those individuals randomized to the other treatment B_{3-k}. The second weighting factor $\frac{1}{\widehat{K}(U_i)}$is the usual form of weighting complete data by the inverse probability of censoring. We will refer to the estimator in (3.4) as the incomplete-probability-of-missing-weighted (IPMW) estimator.

Some other ad hoc estimators were proposed in LDT as well. These estimators, although consistent and asymptotic normal, do not include the most efficient estimator, in general. In the next section we consider the efficient estimation of parameters of interest related to the distribution of treatment policy survival times T_{1k} using the observed data. We make use of the methods developed by Robins et.

al. (1994) for semiparametric theory to characterize the most efficient estimator for $\mu_{1k} = E\{h(T_{1k})\}$.

3.2 Efficient Estimator

Again, as in Chapter 2, we will restrict ourselves to the class of estimators which are regular and asymptotically linear. Estimators will be derived by considering the class of influence functions for RAL estimators. Specifically, if we identify the influence function $\Psi_i(\theta)$ for estimating θ, one can just set the estimating equation $\sum_{i=1}^{n} \Psi_i(\theta) = 0$ and solve for θ to get the corresponding estimator. Asymptotic properties of such estimators are relatively easy to derive with asymptotic variance, at the least, estimated using the sandwich estimator.

Had we observed "full" data, i.e. where all patients were treated according to the policy A_1B_k, then the influence function for all RAL estimators would just be $T_i - \mu_{1k}$ leading to the natural estimator $\widehat{\mu}_{1k} = \sum_{i=1}^{n} h(T_i)/n$. For two-stage randomization studies, following the procedures shown in Chapter 2, all RAL estimators for μ_{1k} have an influence function belonging to the class

$$\Psi_i(f) = Q_{ki}\{h(T_i) - \mu_{1k}\} + (Q_{ki} - 1)f(T_i^R, G^H(T_i^R)), \tag{3.5}$$

where $f(T_i^R, G^H(T_i^R))$ is an arbitrary function of T_i^R, the time to response to the induction therapy, and other information collected prior to T_i^R. Steps analogous to the ones in Appendix A.1 can be followed to show that among all influence functions

in (3.5), the most efficient one is obtained when

$$f(T_i^R, G^H(T_i^R)) = -E\left\{ h(T_i) - \mu_{1k} | \, T_i^R, G^H(T_i^R), R_i = 1, X_{ki} = 1 \right\}.$$

Now for the case of right censored data, let us consider the data without censoring as the "full" data and let censoring be the source of missing data. In that case the class of all full-data influence functions is given by (3.5). Again, use of the theory for the missing data developed by Robins et. al. (1994) gives the following general form for the observed data influence functions for all RAL estimators for μ_{1k}:

$$\Psi_{obs}(f, g) = \frac{\Delta_i \Psi_i(f)}{K(U_i)} + \int \frac{g(u, G_i^H(u)) dM_i^c(u)}{K(u)} \tag{3.6}$$

where

- $K(.)$ = the survival function for the censoring time,
- $dM_i^c(u) = dI(U_i \leq u, \Delta_i = 0) - \lambda^c(u) Y_i(u) du$,
- $\lambda^c(u)$ = the hazard rate for the censoring distribution,
- $Y_i(u) = I(U_i \geq u)$, and
- $g(u, G^H(u))$ is an arbitrary real-valued functional of defined through the mapping

  $$h(u, .) : \mathcal{G}^H(u) \rightarrow \Re^1$$

  from $\mathcal{G}^H(u)$, the collection of data-histories up to time u, to the set of real numbers.

The influence function (3.6) is indexed by two arbitrary functions f and g. Choice of these two arbitrary functions will determine how efficient the corresponding RAL estimator will be. Therefore, the problem of finding the RAL estimator with minimum variance is equivalent to finding the optimal $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ for which the variance of $\Psi_{obs}(f,g)$ in (3.6) is minimum. We do this in steps. First we fix f and find the optimal g, say, g^{opt}. Then, within the class of optimal influence functions $\Psi_{obs}(f,g^{opt})$ with arbitrary f, search for the one with minimum variance. From the theory of monotone coarsening [see Robins et al. (1994) and Laan and Hubbard (1998)], we know that for fixed f, the optimal observed data influence function in the class of influence functions (3.6) is given by

$$\frac{\Delta_i \Psi_i(f)}{K(U_i)} + \int \frac{dM_i^c(u)}{K(u)} E\left[\Psi_i(f)|G^H(u)\right]. \tag{3.7}$$

Using the identity

$$\frac{\Delta_i}{K(U_i)} = 1 - \int_0^\infty \frac{dM_i^c(u)}{K(u)}$$

from Robins and Rotnitzky (1992) we can also write (3.7) in the form

$$\Psi_i(f) - \int \frac{dM_i^c(u)}{K(u)} \left[\Psi_i(f) - E\left\{\Psi_i(f)|G^H(u)\right\}\right]. \tag{3.8}$$

Next we find the optimal f for which the variance of (3.8) is minimum. In Appendix B.1, we have used the theory of Hilbert space and Martingales to show that the optimal influence function in the class of influence functions (3.8) is given by

$$\begin{aligned}\Psi_i &= \frac{\Delta_i}{K(U_i)} Q_{ki} h(U_i) + \int \frac{dM_i^c(u)}{K(u)} e_h\left\{G^H(u)\right\} \\ &- \frac{\Delta_i^*(Q_{ki}-1)}{K(U_i^*)} E\left\{h(T_i)|\, R_i = 1, X_{ki} = 1, T_i^R, G^H(T_i^R)\right\} - \mu_{1k},\end{aligned} \tag{3.9}$$

where

- $U_i^* = min(C_i, T_i^R)$, and
- $\Delta_i^* = I(C_i < T_i^R)$

are defined only if $R_i = 1$; and $e_h\left\{G^H(u)\right\}$ is given by

$$
\begin{aligned}
& e_h\left\{G^H(u)\right\} && (3.10)\\
&= R_i I(T_i^R < u)\frac{X_{ki}}{\pi_k} E\left(h(T_i)|R_i = 1, X_{ki} = 1, G^H(u), T_i \geq u\right)\\
&+ \left\{1 - R_i I(T_i^R < u)\right\} E[Q_{ki}h(T_i)|R_i I(T_i^R < u) = 0, G^H(u), T_i \geq u]. && (3.11)
\end{aligned}
$$

The fact that (3.9) has expectation zero can be shown as follows:

$$
\begin{aligned}
E\left[\Psi_i\right] &= E\left\{\frac{\Delta_i}{K(U_i)}Q_{ki}h(U_i)\right\} + E\left[\int \frac{dM_i^c(u)}{K(u)} e_h\left\{G^H(u)\right\}\right]\\
&- E\left[\frac{\Delta_i^*(Q_{ki}-1)}{K(U_i^*)} E\left\{h(T_i)|\, R_i = 1, X_{ki} = 1, T_i^R, G^H(T_i^R)\right\}\right] - \mu_{1k}, && (3.12)
\end{aligned}
$$

The first term on the right-hand-side of Equation (3.12) has expectation μ_{1k} by Equation (3.3). Consider the second term

$$
\begin{aligned}
E\left[\int \frac{dM_i^c(u)}{K(u)} e_h\left\{G^H(u)\right\}\right] &= \int E\left[\frac{dM_i^c(u)}{K(u)} e_h\left\{G^H(u)\right\}\right]\\
&= \int \frac{E\left[E\left\{dM_i^c(u) e_h\left\{G^H(u)\right\}\middle|\, R_i, X_{ki}, T_i^R, G^H(T_i^R)\right\}\right]}{K(u)}\\
&= \int \frac{E\left[e_h\left\{G^H(u)\right\} E\left\{dM_i^c(u)|\, R_i, X_{ki}, T_i^R, G^H(T_i^R)\right\}\right]}{K(u)}\\
&= 0.
\end{aligned}
$$

where the intermediate step follows because of the independence of C_i and other quantities and that $dM_i^c(u)$, being a martingale difference, has expectation zero. Let

us denote the third term in the RHS of (3.9) by T_3. Then,

$$\begin{aligned} E[T_3] &= E\left[\frac{\Delta_i^*(Q_{ki}-1)}{K(U_i^*)} E\left\{h(T_i)|\, R_i=1, X_{ki}=1, T_i^R, G^H(T_i^R)\right\}\right] \\ &= E\left[\frac{\Delta_i^*}{K(U_i^*)} E\left\{h(T_i)|\, R_i=1, X_{ki}=1, T_i^R, G^H(T_i^R)\right\}\right. \\ &\qquad \left. \times E\left\{Q_{ki}-1|\, R_i, T_i^R, G^H(T_i^R)\right\}\right] \\ &= 0. \end{aligned}$$

where the last line follows because

$$\begin{aligned} E\left\{Q_{ki}|\, R_i, T_i^R, G^H(T_i^R)\right\} &= E\left\{1-R_i+\frac{R_i X_{ki}}{\pi_k}\middle|\, R_i, T_i^R, G^H(T_i^R)\right\} \\ &= 1-R_i+\frac{R_i}{\pi_k} E\left\{X_{ki}|\, R_i, T_i^R, G^H(T_i^R)\right\} \\ &= 1-R_i+R_i\pi_k^{-1}\pi_k \\ &= 1. \end{aligned}$$

Putting all these together in (3.12), we have, $E[\Psi_i]=0$.

Theoretically, Ψ_i has the minimum variance among all influence functions of the RAL estimators of μ_{1k}. But observe that $E\left\{h(T_i)|\, R_i=1, X_{ki}=1\ , T_i^R, G^H(T_i^R)\right\}$ and $e_h\left\{G^H(u)\right\}$ are functions of the population quantities and hence are not known in practice. Therefore, to construct estimators for μ_{1k} using this influence function, one needs to estimate them from the observed data. Estimation of the former can be done by regressing $h(T_i)$ on the auxiliary variables observed prior to time T_i^R, in which case the model selection as well will influence the efficiency of the estimator which has been discussed in details in Chapter 2 for complete data. On the other hand, estimating the function $e_h\left\{G^H(u)\right\}$ can be a daunting task. Because of this

consideration, in the following section, we consider a simple problem where we restrict our search for the optimal estimator in a sub-class of the RAL estimators that contains the IPMW and the LDT estimators. Although this estimator may not be optimal, it is relatively easy to compute and is guaranteed to have smaller variance as compared to the IPMW and the LDT estimators.

3.3 Improved Estimator

Having found the optimal influence function (3.9), and realizing that construction of estimators based on this influence function is infeasible, we restrict our attention to estimators whose influence functions belong to the class

$$\begin{aligned} & \frac{\Delta_i}{K(U_i)} Q_{ki} \left\{ h(U_i) - \mu_{1k} \right\} + \frac{\Delta_i^*(Q_{ki}-1)}{K(U_i^*)} \boldsymbol{\gamma}^T \boldsymbol{W}_i \left\{ G^H(T_i^R) \right\} \\ + & \int \frac{dM_i^c(u)}{K(u)} \left[R_i I(T_i^R < u) \frac{X_{ki}}{\pi_k} \varphi_1(u) + (1 - R_i I(T_i^R < u)) \varphi_2(u) \right], \quad (3.13) \end{aligned}$$

where $\boldsymbol{W}_i \left\{ G^H(T_i^R) \right\}$ is a fixed q-dimensional vector of pre-specified functions of T_i^R and the auxiliary information $G^H(T_i^R)$ (for simplicity, from now on we will drop the arguments of $\boldsymbol{W}_i \left\{ G^H(T_i^R) \right\}$ and simply write $\boldsymbol{W}_i$); $\boldsymbol{\gamma}$ is an arbitrary q-dimensional real vector; and $\varphi_1(u)$ and $\varphi_2(u)$ are arbitrary functions of u. It is not hard to verify that (3.13) is a subspace of the class of influence functions (3.6) and resembles the optimal one given by (3.9). Also the influence function for IPMW estimator belongs to this class with $\boldsymbol{\gamma} = \varphi_1(u) = \varphi_2(u) = 0$, and if we take $\varphi_1(u) = \varphi_2(u) = 0$ and $\boldsymbol{\gamma}$ to be a real constant, we obtain influence function for the LDT estimators. The influence

functions (3.13), for $\boldsymbol{\gamma} \in \mathcal{R}^q$, $\varphi_1 : \mathcal{R}^+ \to \mathcal{R}$, and $\varphi_2 : \mathcal{R}^+ \to \mathcal{R}$ define a linear subspace of the space of influence functions (3.6). The goal is to find the optimal estimator within this class, i.e., to find the estimator whose influence function is the one within the class (3.13) with smallest variance. This restricted optimal influence function, will be at least as efficient as the influence functions of IPMW and LDT estimators.

Finding the optimal influence function in the class (3.13) is equivalent to determining the combination of $\boldsymbol{\gamma}$, $\varphi_1(\cdot)$, and $\varphi_2(\cdot)$ for which the variance of(3.13) is minimum. Using the projection theorem for vector spaces, we find that the optimal combination is given by $\boldsymbol{\gamma}^{opt} = -\boldsymbol{\gamma}_h + \boldsymbol{\gamma}_\mu \mu_{1k}$, $\varphi_1^{opt}(u) = \varphi_{1h}(u) - \mu_{1k}$, $\varphi_2^{opt}(u) = \varphi_{2h}(u) - \varphi_{2\mu}(u)\mu_{1k}$ where

$$\boldsymbol{\gamma}_h = \left[E\left\{ \frac{(Q_{ki}-1)^2 \boldsymbol{W}_i \boldsymbol{W}_i^T}{K(T_i^R)} \right\} \right]^{-1} E\left\{ \frac{Q_{ki}(Q_{ki}-1)h(T_i)\boldsymbol{W}_i}{K(T_i^R)} \right\}, \tag{3.14}$$

$$\boldsymbol{\gamma}_\mu = \left[E\left\{ \frac{(Q_{ki}-1)^2 \boldsymbol{W}_i \boldsymbol{W}_i^T}{K(T_i^R)} \right\} \right]^{-1} E\left\{ \frac{Q_{ki}(Q_{ki}-1)\boldsymbol{W}_i}{K(T_i^R)} \right\}, \tag{3.15}$$

$$\varphi_{1h}(u) = \frac{E\left\{ I(T_i^R < u \leq T_i) R_i X_{ki} h(T_i) \right\}}{E\left\{ I(T_i^R < u \leq T_i) R_i X_{ki} \right\}}, \tag{3.16}$$

$$\varphi_{2h}(u) = \frac{E\left[\left\{ (1-R_i) I(T_i^R \geq u) + \frac{R_i X_{ki}}{\pi} \right\} h(T_i) \right]}{E\left[I(T_i \geq u) \left\{ 1 - R_i I(T_i^R < u) \right\} \right]}, \tag{3.17}$$

and

$$\varphi_{2\mu}(u) = \frac{E\left\{ (1-R_i) I(T_i^R \geq u) + \frac{R_i X_{ki}}{\pi} \right\}}{E\left[I(T_i \geq u) \left\{ 1 - R_i I(T_i^R < u) \right\} \right]}. \tag{3.18}$$

Thus the optimal restricted influence function is given by (3.13) with $\boldsymbol{\gamma}$, $\varphi_1(\cdot)$, and $\varphi_2(\cdot)$ respectively replaced by $\boldsymbol{\gamma}^{opt}$, $\varphi_1^{opt}(\cdot)$, and $\varphi_2^{opt}(\cdot)$. If the expectations in expressions (3.15)-(3.18) were known, then the restricted optimal estimator would be

obtained by solving the estimating equation

$$\sum_{i=1}^{n}\left[\frac{\Delta_i}{\widehat{K}(U_i)}Q_{ki}\left\{h(U_i)-\mu_{1k}\right\}+\frac{\Delta_i^*(Q_{ki}-1)}{\widehat{K}(U_i^*)}\boldsymbol{\gamma}^{opt^T}\boldsymbol{W}_i\left\{G^H(U_i^*)\right\}\right.$$
$$\left.+\int\frac{dN_i^c(u)}{\widehat{K}(u)}\left\{\varphi_1^{opt}(u)L_{\varphi_1 i}(u)+\varphi_2^{opt}(u)L_{\varphi_2 i}(u)\right\}\right]=0, \tag{3.19}$$

where

$$L_{\varphi_1 i}(u)=R_iI(U_i^*<u)\frac{X_{ki}}{\pi_k}-\frac{\sum_{i=1}^n R_iI(U_i^*<u)\frac{X_{ki}}{\pi_k}I(U_i\geq u)}{\sum_{i=1}^n I(U_i\geq u)},$$

and

$$L_{\varphi_2 i}(u)=1-R_iI(U_i^*<u)-\frac{\sum_{i=1}^n\left\{1-R_iI(U_i^*<u)\right\}I(U_i\geq u)}{\sum_{i=1}^n I(U_i\geq u)}.$$

As these quantities are not known, we propose estimating them by IPMW estimators. For example, the estimators for $\boldsymbol{\gamma}_h$ and $\boldsymbol{\gamma}_\mu$ are defined by

$$\widehat{\boldsymbol{\gamma}}_h=\left[\sum_{i=1}^n\left\{\frac{\Delta_i(Q_{ki}-1)^2\boldsymbol{W}_i\boldsymbol{W}_i^T}{\widehat{K}(U_i)\widehat{K}(U_i^*)}\right\}\right]^{-1}\sum_{i=1}^n\left\{\frac{\Delta_iQ_{ki}(Q_{ki}-1)h(U_i)\boldsymbol{W}_i}{\widehat{K}(U_i)\widehat{K}(U_i^*)}\right\}, \tag{3.20}$$

$$\widehat{\boldsymbol{\gamma}}_\mu=\left[\sum_{i=1}^n\left\{\frac{\Delta_i(Q_{ki}-1)^2\boldsymbol{W}_i\boldsymbol{W}_i^T}{\widehat{K}(U_i)\widehat{K}(U_i^*)}\right\}\right]^{-1}\sum_{i=1}^n\left\{\frac{\Delta_iQ_{ki}(Q_{ki}-1)\boldsymbol{W}_i}{\widehat{K}(U_i)\widehat{K}(U_i^*)}\right\}, \tag{3.21}$$

and similarly for other parameters.

Substituting these estimators in (3.19) and solving for μ_{1k}, we obtain the restricted improved estimator, say, $\widehat{\mu}_{1k}^{IMP}$; namely,

$$\widehat{\mu}_{1k}^{IMP}=\frac{\sum_{i=1}^n\left[\frac{\Delta_iQ_{ki}}{\widehat{K}(U_i)}h(U_i)-\frac{\Delta_i^*(Q_{ki}-1)}{\widehat{K}(U_i^*)}\widehat{\boldsymbol{\gamma}}_h^T\boldsymbol{W}_i+\int\frac{dN_i^c(u)}{\widehat{K}(u)}\left\{\widehat{\varphi}_{1h}(u)L_{\varphi_1 i}(u)+\widehat{\varphi}_{2h}(u)L_{\varphi_2 i}(u)\right\}\right]}{\sum_{i=1}^n\left[\frac{\Delta_iQ_{ki}}{\widehat{K}(U_i)}-\frac{\Delta_i^*(Q_{ki}-1)}{\widehat{K}(U_i^*)}\widehat{\boldsymbol{\gamma}}_\mu^T\boldsymbol{W}_i+\int\frac{dN_i^c(u)}{\widehat{K}(u)}\left\{L_{\varphi_1 i}(u)+\widehat{\varphi}_{2\mu}(u)L_{\varphi_2 i}(u)\right\}\right]} \tag{3.22}$$

The asymptotic variance of $\widehat{\mu}_{1k}^{IMP}$ is $Var(\widehat{\mu}_{1k}^{IMP}) = \sigma^2/n$, where σ^2 is given by Equation B.35. This variance can then be estimated by

$$\begin{aligned} v(\widehat{\mu}_{1k}^{IMP}) &= \frac{1}{n}\left[n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\widehat{K}(U_i)}\left\{ Q_{ki}(h(U_i)-\widehat{\mu}_{1k}^{IMP}) + (Q_{ki}-1)\left(\widehat{\boldsymbol{\gamma}^{opt}}\right)^T \boldsymbol{W}_i \right\}^2 \right. \\ &\quad \left. + \int_0^L \frac{dN^c(u)}{\widehat{K}(u)Y(u)}\widehat{E}\left\{R_{1ki}(u)\right\}^2 \right], \end{aligned} \tag{3.23}$$

where

$$\widehat{E}\left\{R_{1ki}(u)\right\}^2 = n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\widehat{K}(U_i)}\widehat{R}_{1ki}^2(u), \tag{3.24}$$

$$\begin{aligned} \widehat{R}_{1ki}(u) &= (Q_{ki}-1)\left(\widehat{\boldsymbol{\gamma}^{opt}}\right)^T \boldsymbol{W}_i I(U_i^* \geq u) - \left\{ \widehat{\varphi_1^{opt}}(u)L_{\varphi_1 i}(u) + \widehat{\varphi_2^{opt}}(u)L_{\varphi_2 i}(u) \right. \\ &\quad \left. - Q_{ki}(h(U_i)-\widehat{\mu}_{1k}^{IMP}) + \widehat{G}_{1k}(u) \right\} I(U_i \geq u), \end{aligned} \tag{3.25}$$

and

$$\widehat{G}_{1k}(u) = \frac{\sum_{i=1}^{n}\Delta_i Q_{ki}\left(h(U_i)-\widehat{\mu}_{1k}^{IMP}\right) I(U_i \geq u)/\widehat{K}(U_i)}{n\widehat{S}(u)} \tag{3.26}$$

In Appendix B, we show that the estimator $\widehat{\mu}_{1k}^{IMP}$ is consistent and asymptotically normal. The construction of this estimator guarantees that it is asymptotically more efficient than the IPMW and the LDT estimators.

In order to test different hypotheses related to the means, we need to estimate the covariances between $\widehat{\mu}_{11}^{IMP}$ and $\widehat{\mu}_{12}^{IMP}$ as well. The covariance of the estimators $\widehat{\mu}_{11}^{IMP}$ and $\widehat{\mu}_{12}^{IMP}$, say, $cov(\widehat{\mu}_{11}^{IMP}, \widehat{\mu}_{12}^{IMP})$ is given by Equation B.37 which can be estimated

by

$$\widehat{cov}(\widehat{\mu}_{11}^{IMP}, \widehat{\mu}_{12}^{IMP}) = \frac{1}{n}\left[n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(U_i)} \left\{ Q_{1i}(h(U_i) - \widehat{\mu}_{11}^{IMP}) + (Q_{1i} - 1) \left(\widehat{\boldsymbol{\gamma}_1^{opt}} \right)^T \boldsymbol{W}_i \right\} \right.$$
$$\times \left\{ Q_{2i}(h(U_i) - \widehat{\mu}_{12}^{IMP}) + (Q_{2i} - 1) \left(\widehat{\boldsymbol{\gamma}_2^{opt}} \right)^T \boldsymbol{W}_i \right\}$$
$$\left. + \int_0^L \frac{dN^c(u)}{\widehat{K}(u)Y(u)} \widehat{E}\left\{ R_{11i}(u) R_{12i}(u) \right\}^2 \right], \tag{3.27}$$

where

$$\widehat{E}\left\{ R_{11i}(u) R_{12i}(u) \right\}^2 = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(U_i)} \widehat{R}_{11i}(u) \widehat{R}_{12i}(u), \tag{3.28}$$

where $\widehat{R}_{1ki}(u)$ is given by the Equation 3.25 for $k = 1, 2$, with the understanding that any function that is dependent on k will be changed accordingly.

3.4 Analysis of CALGB 8923 data

An application of our proposed methods has been considered by analyzing the CALGB 8923 data which motivated this study. There were 388 patients that participated in CALGB 8923. Of these, 79 out of 193 patients in the GM-CSF group and 90 out of 195 in the placebo group achieved remission (responded) and consented to further randomization to the intensification therapy; and, of these, 37 GM-CSF and 45 placebo patients were randomized to intensification therapy I and the rest to intensification therapy II. Since at the time of analysis there were only a few censored observations, we artificially terminated the study at time point T_0 to have a reasonable amount of censoring. We considered two different values of T_0: 2.5 (this ending

point was also considered by Lunceford et al. (2002)) and 1.7 years, respectively yielding approximately 30% and 50% of the patients censored. The goal is to estimate the mean restricted survival time and the survival probabilities at certain time points t for the four treatment policies for this clinical trial: GM-CSF/Maintenance I, GM-CSF/Maintenance II, Placebo/Maintenance I, Placebo/Maintenance II. For estimating the mean survival time, we restricted ourselves to the time point of 548 days (1.5 years).

In our analysis, there were four variables, namely, time between the response and the second randomization, age, white blood cell count that were used as auxiliary variables. For the improved estimator, we defined the prespecified vector function $\boldsymbol{W}_i$ as the column vector whose elements are the random variables T_i^R, time between the response and the second randomization, age, white blood cell count, and a constant function identically equal to 1.

Table 3.1 shows the estimates of mean restricted survival time for each of the four treatment policies using the estimators KM, IPMW, LDT, and IMP. The Kaplan-Meier estimator almost always gave an estimate that is smaller than the estimates using any other three methods which emphasizes our intuition that this estimator is biased. We will investigate this issue in details in our simulation study. Although in terms of estimates and its standard errors, the IMP, IPMW and LDT estimators provide similar results, the standard error for the IMP estimator seems to be the smallest. The IPMW estimator gave the largest estimated standard error as would be expected by the theory. Also we observe by moving from one column to the next

in Table 3.1 that, for this particular data set, as censoring gets heavier, the gain in efficiency in terms of standard error for IMP estimator over LDT or IMP estimator becomes larger. Similar conclusions follow for the survival probability estimates presented in Table 3.2.

Survival curves for different policies under different methods have been constructed. Figure 3.6 gives the images of survival curves for different policies under the improved method of estimation. The survival curves do not differ much across treatment policies.

Although the survival curves do not seem to differ across policies (as is evident in Table 3.2 and Figure 3.6), it seems, from the results in Table 3.1, that the mean restricted lifetimes may differ across treatment policies. To see this we have considered testing hypotheses regarding the treatment means for policies as well as for main effects and interactions. For this, we have used large sample Wald chi-square tests. For example, to test the equality of treatment means, we formulated the hypothesis $H_0 : \mu_{11} = \mu_{21} = \mu_{12} = \mu_{22}$ which is equivalent to testing the contrast $H_0 : C^T\mu = 0$ where $\mu = (\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22})^T$, and C is the matrix of constants

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}. \tag{3.29}$$

Then using the fact that $\widehat{\mu} = (\widehat{\mu}_{11}, \widehat{\mu}_{21}, \widehat{\mu}_{12}, \widehat{\mu}_{22})^T$ is asymptotically normally distributed with mean $\mu = (\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22})^T$ and with estimated variance, say, $v(\widehat{\mu})$, the large-sample Wald chi-square test statistic is given by $\widehat{\mu}^T C \left\{C^T v(\widehat{\mu}) C\right\}^{-1} C^T \widehat{\mu}$

which has an approximate χ^2-distribution with 3 d.f. Similar tests have been considered for the hypothesis $H_1 : (\mu_{11} + \mu_{12})/2 = (\mu_{21} + \mu_{22})/2$ to test the main effects of induction therapy, for the null hypothesis $H_2 : (\mu_{11} + \mu_{21})/2 = (\mu_{12} + \mu_{22})/2$ to test whether there is any difference in mean survival times under two intensification therapies. We also tested the equality of maintenance treatment effect within GM-CSF by the hypothesis $H_3 : \mu_{11} = \mu_{12}$ and within chemotherapy by $H_4 : \mu_{21} = \mu_{22}$. The tests have been done using all methods except the Kaplan-Meier estimator which will be shown to be biased in the next section. A test of interaction between Induction and Maintenance therapies is considered through the hypothesis $H_5 : (\mu_{11} - \mu_{12}) = (\mu_{21} - \mu_{22})$.

The results of all these tests are summarized in Table 3.3.

3.5 Simulation Study

Several simulation experiments have been carried out to assess the accuracy of the large-sample properties of our proposed improved estimator and also to compare its relative performance to the IPMW and the LDT estimators. For simplicity, in our simulation studies, we did not consider any auxiliary variables but allowed the survival time to depend on the response time T_i^R. Since data from patients assigned to treatment A_1 are independent of data from patients who receive treatment A_2, we conduct simulation experiment only with "A_1-patients". The experiment will be

analogous for "A_2-patients".

We took R_i, the eligible/consent indicator, to be Bernoulli with $P(R_i = 1) = \pi_R$ and considered two different values of π_R, 0.5 and 0.7. When $R_i = 0$, a survival time T_{0i} is generated from an exponential distribution with mean λ truncated at b_2. When $R_i = 1$, treatment B assignment indicator Z_i is generated from Bernoulli(.5) distribution. Also when $R_i = 1$, a response time T_i^R is generated from an exponential distribution with mean α truncated at b_1 and we take

$$T_{1i}^* = T_i^R + (\beta_1 + \beta_2 T_i^R)U_{1i}$$

$$T_{2i}^* = T_i^R + (\beta_1 + \beta_2 T_i^R)U_{2i}$$

where $U_{ji}, j = 1, 2$ is generated from a uniform$(0, \theta_j)$ distribution. Finally we defined

$$T_i = (1 - R)T_{0i} + R_i[Z_i T_{1i}^* + (1 - Z_i)T_{2i}^*]$$

to generate the observed survival time for the i^{th}individual.

In our simulation scenario, we restricted the lifetime to $L = 548$ days and considered $\lambda = 365, b_2 = 1095, \alpha = 365, b_1 = 548, \beta_1 = 1.0, \beta_2 = 1.0, \theta_1 = 1.5, \theta_2 = 1$ so that when $\pi_R = 0.5, \mu_{11} = 360$ days, $\mu_{12} = 337$ days and $\mu_{1k} = 395$ days, $\mu_{12} = 363$ days when $\pi_R = 0.7$. We considered 5000 Monte-Carlo samples of sizes 250 and 500.

The censoring variable C_i was generated, independently of all other random variables, from a uniform$(0, \theta_c)$. We considered two values of θ_c, $\theta_c = 730$ and $\theta_c = 913$ which yielded respectively 38% and 48% censoring for 50% response rate and 42% and 52% respective censoring for 70% response. This will give us the opportunity to compare different estimators in the presence of different levels of censoring.

For each of the 5000 simulated data sets, $\mu_{1k} = E(T_{1k}), k = 1, 2$ were estimated for $k = 1, 2$. In addition to estimating the mean restricted survival times, we have also considered estimating survival probabilities at several points of time. We estimated survival probabilities using different methods at time points 183 days (1/2 year), 365 days (1 year). The true values were $S_{11}(183) = 0.754(0.822)$ for $\pi_R = 0.5(0.7), S_{12}(183) = 0.735(0.795)$ for $\pi_R = 0.5(0.7), S_{11}(365) = 0.542(0.626)$ for $\pi_R = 0.5(0.7), S_{12}(365) = 0.477(0.535)$ for $\pi_R = 0.5(0.7)$. The two values of t have been considered to compare the performance of estimating the survival function at both earlier and later time point of the study. Tables 3.4, 3.5 and 3.6 present the coverage probabilities for 95% Wald intervals and relative efficiency for each of the four different estimators under consideration. The relative efficiency has been calculated by using the ratio of Monte-Carlo mean-squared errors. For instance, the entry 93.6(0.84) in the first LDT row of Table 3.4 refers to the case where in samples of size 250 from a population where 50% of the patients respond to the initial treatment, and on average 38% patients get censored, (i) the 95% Wald confidence interval for μ_{11} shows a coverage probability of 93.6% and (ii) the ratio of the Monte-Carlo mean-squared error of the improved estimator to the LDT estimator is .84, *i.e.,* $MSE(\widehat{\mu}_{11}^{IMP})/MSE(\widehat{\mu}_{11}^{LDT}) = .84$. The latter implies that for this special case, the IMP estimator gains 16% efficiency over the LDT estimator.

Consider the case of estimating the mean restricted life time. If we look at the coverage probabilities across the entries in Table 3.4, we find that they are comparable for the three estimators IPMW, LDT and IMP and does not deviate much from the

nominal level of 95%. The coverage probability for the Kaplan-Meier estimator is always smaller than the nominal level, sometimes as low as 42.5%. This is justified by the large biases we observed for this estimator during the simulation procedure. The superiority of the improved estimator over the KM, IPMW and LDT estimators is seen through the columns of relative efficiencies where the improved estimator achieves remarkable gain in efficiency over these three estimators. For the simulation scenarios considered, the gain in efficiency of the improved estimator over LDT estimator for the mean restricted lifetimes lies between 15%-30%. There are three sources from which the IMP estimator gains efficiency: (i) by including the auxiliary variables and (ii) by incorporating information from the censored patients and (iii) by extracting more information from the patients eligible but randomized to a treatment inconsistent with the policy in the second stage . (i) suggests that the stronger the correlation between the auxiliary variables and the survival time the more the gain should be in efficiency of improved estimator over LDT estimator. The survival time T_{1i} and T_{2i} has been generated in such a way that $T_{1i}(T_{2i})$ has a correlation of .53(.70) with the response time T_i^R representing from moderate to high correlations. This is why, we believe, the improved estimator for mean restricted lifetime for the policy A_1B_2 gains more efficiency than that of the policy A_1B_1 over the LDT estimator (Table 3.4). Also the results show that our improved estimator is more efficient than other three estimators, no matter how heavier the level of censoring is as is seen from the relative efficiencies for different estimators for different censoring percentage. (ii) suggests that the heavier the censoring, the better our estimator should gain over IPMW

estimator, which is the case in the simulation scenarios considered. It is also easy to check (iii) by considering different randomization probability for the second stage.

It is of importance to mention here that the inverse-probability-of-missing-weighted estimator performs very poorly in terms of relative efficiency. Almost always it is less efficient than all other estimators, even the naive Kaplan-Meier estimator. This is due to the fact that this estimator does only incorporate the complete observations weighted by the censoring survival distribution. One thing we should be cautious about though, although it seems that the Kaplan-Meier estimator is more efficient than the IPMW estimator, we would hesitate to recommend it because of its large bias and very poor coverage probability.

Increasing sample size from 250 to 500 helped improve the coverage probabilities for most of the estimators, but the relative efficiencies remained unaffected for IPMW, LDT and the improved estimators. The coverage probability and the efficiency for the Kaplan-Meier estimator drops down with the increase of sample size.

Now consider the results for the survival probabilities from Tables 3.5 and Tables 3.6. The pattern is similar to that in the case of mean survival time. Again, with a comparable coverage probability for a 95% Wald confidence interval, the improved estimator gains 7%-20% efficiency over the LDT estimator. The Kaplan-Meier estimator has the worst coverage (as expected by the theory) and the IPMW estimator has the smallest relative efficiency.

3.6 Figures and Tables

Table 3.1: Estimates of mean restricted survival time in days (and corresponding standard error in the parenthesis) for different treatment policies for CALGB 8923 data

	30% Censoring				50% Censoring			
Policy	KM	IPMW	LDT	IMP	KM	IPMW	LDT	IMP
GM-CSF/Int. I	219(−−)	282(45)	284(21)	266(20)	241(−−)	251(75)	272(27)	309(23)
GM-CSF/Int. II	227(−−)	280(44)	277(21)	286(20)	255(−−)	297(78)	274(27)	290(24)
placebo/Int. I	270(−−)	275(42)	291(21)	303(20)	275(−−)	293(72)	285(30)	281(27)
placebo/Int. II	272(−−)	319(45)	300(22)	314(21)	297(−−)	308(78)	317(31)	335(28)

Table 3.2: Estimated survival probability (corresponding standard error in the parenthesis) for different treatment policies in CALGB 8923 data

		30% Censoring				50% Censoring			
Time	Policy	KM	IPMW	LDT	IMP	KM	IPMW	LDT	IMP
183 days (1/2 year)	GM-CSF/Int. I	0.46(−−)	0.59(0.091)	0.60(0.043)	0.56(0.042)	0.52(−−)	0.53(0.143)	0.57(0.054)	0.61(0.048)
	GM-CSF/Int. II	0.47(−−)	0.59(0.090)	0.58(0.045)	0.59(0.044)	0.52(−−)	0.59(0.145)	0.55(0.059)	0.52(0.053)
	placebo/Int. I	0.55(−−)	0.58(0.088)	0.61(0.044)	0.62(0.041)	0.59(−−)	0.67(0.152)	0.65(0.058)	0.54(0.050)
	placebo/Int. II	0.53(−−)	0.62(0.090)	0.59(0.047)	0.61(0.046)	0.58(−−)	0.61(0.151)	0.62(0.062)	0.63(0.058)
365 days (1 year)	GM-CSF/Int. I	0.32(−−)	0.43(0.087)	0.43(0.052)	0.39(0.051)	0.36(−−)	0.38(0.142)	0.42(0.067)	0.57(0.048)
	GM-CSF/Int. II	0.32(−−)	0.40(0.086)	0.39(0.055)	0.41(0.053)	0.39(−−)	0.47(0.147)	0.43(0.062)	0.40(0.054)
	placebo/Int. I	0.36(−−)	0.39(0.083)	0.42(0.055)	0.42(0.052)	0.37(−−)	0.40(0.145)	0.39(0.093)	0.31(0.085)
	placebo/Int. II	0.37(−−)	0.48(0.089)	0.44(0.055)	0.46(0.054)	0.43(−−)	0.48(0.152)	0.49(0.076)	0.57(0.069)
511 days (1.4 year)	GM-CSF/Int. I	0.21(−−)	0.31(0.084)	0.32(0.056)	0.30(0.054)	0.22(−−)	0.28(0.143)	0.32(0.073)	0.39(0.050)
	GM-CSF/Int. II	0.21(−−)	0.28(0.082)	0.28(0.058)	0.32(0.055)	0.28(−−)	0.38(0.152)	0.34(0.070)	0.31(0.047)
	placebo/Int. I	0.28(−−)	0.29(0.075)	0.31(0.061)	0.28(0.056)	0.29(−−)	0.26(0.130)	0.25(0.101)	0.14(0.081)
	placebo/Int. II	0.32(−−)	0.40(0.085)	0.37(0.062)	0.37(0.059)	0.39(−−)	0.39(0.147)	0.41(0.098)	0.45(0.091)

Table 3.3: P-values for different hypothesis test for CALGB 8923 data when 30% observations are censored. $\mu_{jk} = E[T_{jk}]$ for mean survival and $\mu_{jk} = P[T_{jk} > t]$ for survival probabilities at time t, j = 1, 2.

Hypothesis	Testing mean restricted survival times			Testing survival prob. at $t = 365$ days		
	IPMW	LDT	IMP	IPMW	LDT	IMP
$H_0 : \mu_{11} = \mu_{21} = \mu_{12} = \mu_{22}$	0.89	0.87	0.33	0.92	0.86	0.84
$H_1 : (\mu_{11} + \mu_{12})/2 = (\mu_{21} + \mu_{22})/2$	0.56	0.59	0.26	0.75	0.77	0.49
$H_2 : (\mu_{11} + \mu_{21})/2 = (\mu_{12} + \mu_{22})/2$	0.71	0.91	0.17	0.79	0.86	0.59
$H_3 : \mu_{11} = \mu_{12}$	0.98	0.64	0.17	0.83	0.47	0.85
$H_4 : \mu_{21} = \mu_{22}$	0.57	0.63	0.51	0.55	0.68	0.57
$H_5 : (\mu_{11} - \mu_{12}) = (\mu_{21} - \mu_{22})$	0.68	0.51	0.73	0.57	0.43	0.78

Table 3.4: *Monte Carlo coverage probability for 95% Wald confidence intervals and relative efficiency for estimators of restricted mean survival time based on 5000 data sets: entries in parentheses are relative efficiencies with respect to the improved estimator.*

		50% response				70% response			
		38% Censoring		48% Censoring		42% Censoring		52% Censoring	
n	Estimator	$\widehat{\mu}_{11}$	$\widehat{\mu}_{12}$	$\widehat{\mu}_{11}$	$\widehat{\mu}_{12}$	$\widehat{\mu}_{11}$	$\widehat{\mu}_{12}$	$\widehat{\mu}_{11}$	$\widehat{\mu}_{12}$
250	IMP	93.5(1.00)	93.2(1.00)	93.2(1.00)	92.7(1.00)	93.9(1.00)	93.6(1.00)	93.4(1.00)	93.4(1.00)
	LDT	93.6(0.84)	93.9(0.78)	93.5(0.85)	93.7(0.84)	94.0(0.78)	94.0(0.70)	93.6(0.79)	94.0(0.72)
	IPMW	94.5(0.20)	94.1(0.20)	93.8(0.16)	94.0(0.17)	93.9(0.15)	93.8(0.15)	93.6(0.11)	93.6(0.12)
	KM	67.0(0.30)	81.8(0.30)	73.8(0.39)	85.7(0.62)	74.9(0.35)	88.1(0.54)	81.2(0.46)	90.7(0.72)
500	IMP	94.2(1.00)	94.7(1.00)	94.1(1.00)	94.5(1.00)	94.4(1.00)	94.7(1.00)	94.1(1.00)	94.6(1.00)
	LDT	94.6(0.84)	95.1(0.80)	94.7(0.85)	95.0(0.80)	94.3(0.79)	95.5(0.72)	94.3(0.79)	95.3(0.72)
	IPMW	95.1(0.20)	94.5(0.20)	94.6(0.15)	94.5(0.14)	95.1(0.14)	94.4(0.14)	95.0(0.11)	94.3(0.11)
	KM	42.5(0.17)	68.8(0.29)	53.0(0.22)	76.6(0.39)	53.8(0.20)	79.9(0.37)	65.7(0.27)	86.7(0.53)

Table 3.5: *Monte Carlo coverage probability for 95% Wald interval and relative efficiency for estimators of survival probabilities at $t = 183$ days, based on 5000 data sets: entries in parentheses are relative efficiencies with respect to the improved estimator.*

		50% response				70% response			
		38% Censoring		48% Censoring		42% Censoring		52% Censoring	
n	Estimator	$\widehat{S}_{11}(183)$	$\widehat{S}_{12}(183)$	$\widehat{S}_{11}(183)$	$\widehat{S}_{12}(183)$	$\widehat{S}_{11}(183)$	$\widehat{S}_{12}(183)$	$\widehat{S}_{11}(183)$	$\widehat{S}_{12}(183)$
250	IMP	94.0(1.00)	94.0(1.00)	93.6(1.00)	93.5(1.00)	93.9(1.00)	93.8(1.00)	93.6(1.00)	93.6(1.00)
	LDT	94.0(0.91)	94.0(0.89)	93.7(0.92)	93.7(0.93)	94.1(0.86)	94.0(0.84)	94.2(0.87)	94.1(0.86)
	IPMW	94.2(0.22)	94.1(0.25)	94.0(0.17)	94.1(0.19)	94.0(0.15)	94.1(0.18)	93.7(0.11)	94.0(0.14)
	KM	78.0(0.36)	83.7(0.47)	81.2(0.43)	86.6(0.57)	81.2(0.37)	87.1(0.52)	85.0(0.46)	90.3(0.65)
500	IMP	94.0(1.00)	93.9(1.00)	94.0(1.00)	93.7(1.00)	94.4(1.00)	94.1(1.00)	94.4(1.00)	93.7(1.00)
	LDT	94.4(0.91)	94.2(0.91)	94.3(0.92)	94.4(0.92)	94.4(0.86)	94.4(0.87)	94.7(0.87)	94.2(0.88)
	IPMW	95.2(0.22)	94.4(0.24)	94.8(0.17)	94.1(0.18)	94.8(0.15)	94.2(0.18)	94.6(0.11)	94.3(0.13)
	KM	59.3(0.22)	69.6(0.31)	66.0(0.27)	75.7(0.38)	64.8(0.23)	76.2(0.36)	71.7(0.29)	82.2(0.46)

Table 3.6: *Monte Carlo coverage probability for 95% Wald interval and relative efficiency for estimators of survival probabilities at $t = 365$ days, based on 5000 data sets: entries in parentheses are relative efficiencies with respect to the improved estimator.*

		50% response				70% response			
		38% Censoring		48% Censoring		42% Censoring		52% Censoring	
n	Estimator	$\widehat{S}_{11}(365)$	$\widehat{S}_{12}(365)$	$\widehat{S}_{11}(365)$	$\widehat{S}_{12}(365)$	$\widehat{S}_{11}(365)$	$\widehat{S}_{12}(365)$	$\widehat{S}_{11}(365)$	$\widehat{S}_{12}(365)$
250	IMP	93.8(1.00)	93.9(1.00)	93.3(1.00)	93.0(1.00)	93.3(1.00)	94.0(1.00)	93.0(1.00)	93.4(1.00)
	LDT	94.1(0.91)	94.4(0.85)	93.8(0.93)	94.5(0.89)	94.2(0.88)	94.6(0.81)	93.9(0.89)	94.3(0.82)
	IPMW	94.6(0.40)	94.4(0.42)	93.9(0.33)	94.2(0.36)	94.3(0.33)	94.0(0.35)	93.8(0.26)	93.9(0.29)
	KM	74.9(0.40)	88.7(0.70)	80.2(0.51)	91.0(0.87)	82.1(0.47)	92.4(0.80)	87.2(0.61)	93.5(0.96)
500	IMP	94.2(1.00)	94.9(1.00)	94.5(1.00)	94.7(1.00)	94.1(1.00)	94.9(1.00)	94.0(1.00)	94.7(1.00)
	LDT	94.5(0.90)	95.2(0.85)	94.7(0.91)	95.4(0.85)	94.4(0.88)	95.6(0.81)	94.6(0.87)	95.5(0.80)
	IPMW	95.0(0.40)	94.8(0.41)	95.1(0.31)	94.6(0.32)	95.1(0.33)	94.8(0.34)	94.9(0.25)	94.9(0.27)
	KM	55.8(0.24)	82.0(0.50)	65.2(0.31)	87.8(0.64)	67.3(0.30)	89.9(0.64)	76.6(0.40)	93.4(0.82)

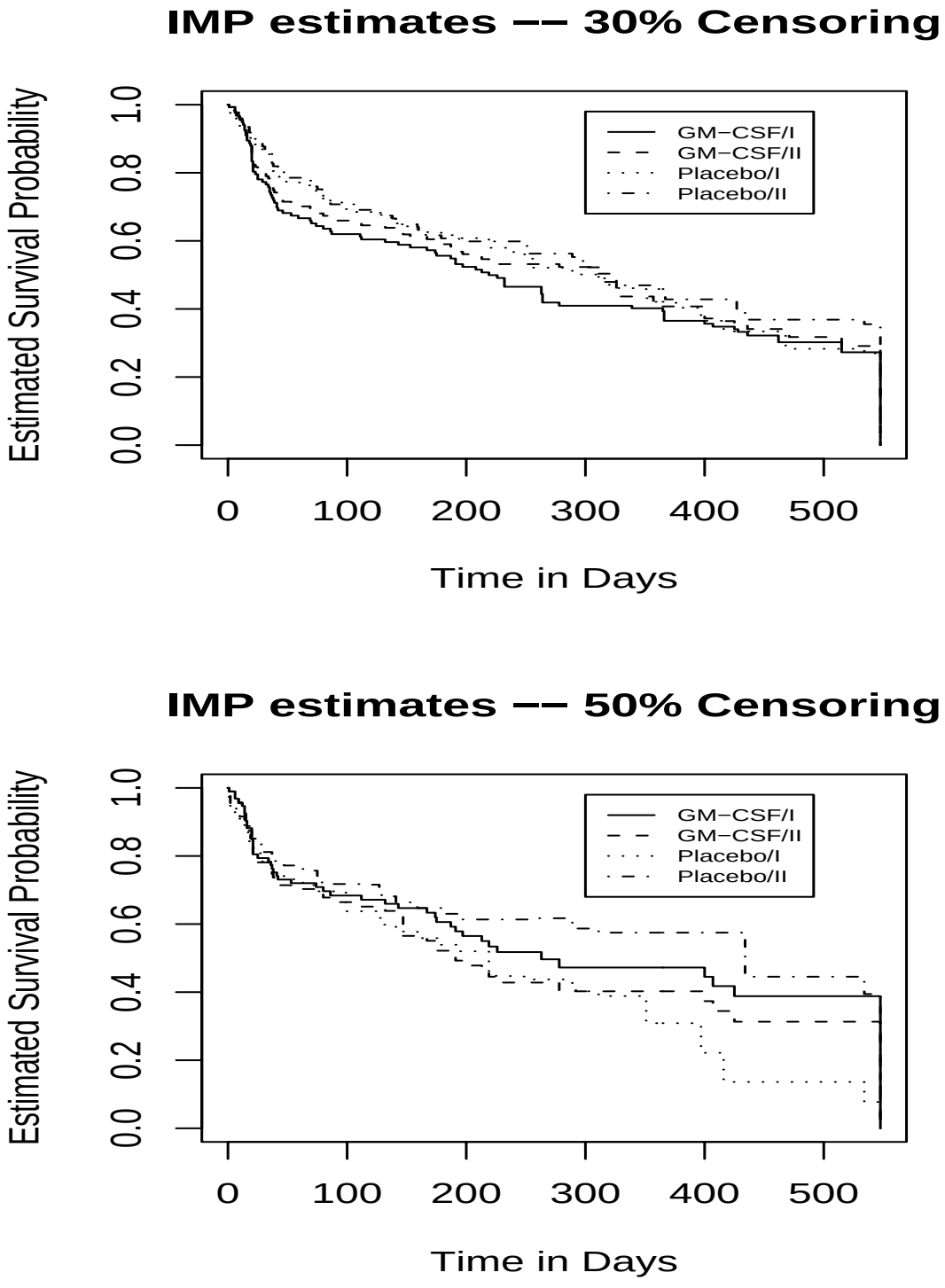

Figure 3.1: Estimated survival probabilities under IMP method for different treatment policies.

Chapter 4

Discussion

In this dissertation we have developed efficient estimation procedures for parameters related to the survival distribution of tratment policies in two stage randomization designs in clinical trials. In particular, we have derived the most efficient estimator for the quantities related to the survival distribution of patients who follows a given treatment policy. But the difficulty that arises, like in many other situations, is that the most efficient estimator (the estimator that attains the smallest variance) may not be feasible to calculate from the observed data. Considering this, for the case where one observes the complete data, we have proposed several estimators to approximate the most efficient estimator. These estimators, conditional on the fact that one can collect auxiliary information which are at least moderately correlated with the survival data among those eligible for the second randomization, performed very well as compared to the estimators defined by Lunceford et al. (2002). However, for a given data set they may not perform as well. Considering this, in a bid to improve

efficiency, we have proposed an estimator that is guaranteed to be more efficient than the LDT estimators and IPMW estimators. This improved estimator uses linear combinations of functionals of auxiliary information from the patients randomized to the second stage treatments. One versatility of this estimator is that varying the functionals to be used, one can try to find the best combination of functionals to be used for a given data set. Also this estimator is easy to compute and does not require any further model-based computation.

However, for the case with censored data, the most efficient estimator contains complicated functions of population quantities, and hence no attempt was made to approximate these complicated quantities. Rather we proposed a class of estimators which is easy to compute and is always more efficient than the LDT or IPMW estimators.

Bibliography

Cleveland, W. S. & Devlin, S. J. (1988). Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596–610.

Faucett, C. L., Schenker, N. & Taylor, J. M. G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to aids clinical trials data. *Biometrics* **58**, 37–47.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall. London.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–960.

Joss, R. A., Alberto, P., Bleher, E. A., Ludwig, C., Siegenthaler, P., Martinelli, G., Sauter, C., Schatzmann, E. & Senn, H. J. (1994). Combined-modality treatment of small-cell lung cancer: randomized comparison of three induction chemotherapies followed by maintenance chemotherapy with or without radiotherapy to

the chest. swiss group for clinical cancer research (sakk).. *Annals of Oncology* **5**, 921–928.

Laan, M. J. V. D. & Hubbard, A. (1998). Locally efficient estimation of the survival distribution with right censored data and covariates when collection of data is delayed. *Biometrika* **85**, 771–783.

Laan, M. J. V. D. & Hubbard, A. (1999). Locally efficient estimation of the quality adjusted lifetime distribution with right-censored data and covariates. *Biometrics* **55**, 530–536.

Luenberger, D. G. (1969). *Optimization by vector space methods*. John Wiley & Sons, Inc.. New York.

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

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**, 99–135.

Robins, J. M. & Rotnitsky, A. (1992). Recovery of information and adjustment of dependent censoring using surrogate markers. *in* N. Jewell, K. Dietz & V. Farewell, eds, 'AIDS Epidemiology-Methodological Issues'. Birkhauser. boston. pp. 297–331.

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

SAS Institute Inc. (2001). *SAS/STAT Software: Changes and Enhancements, Release 8.2*. SAS Institute Inc.. Cary, NC.

Stone, R. M., Berg, D. T., George, S. L., Dodge, R. K., Paciucci, P. A., Schulman, P., Lee, E. J., Moore, J. O., Powell, B. L. & Schiffer, C. A. (1995). Granulocyte-macrophage colony-stimulating factor after initial chemotherapy for elderly patients with primary acute myelogenous leukemia. *The New England Journal of Medicine* **332**, 1671–1677.

Thall, P. F., Sung, H.-G. & Estey, E. H. (2002). Selecting therapeutic strategies based on efficacy and death in multi-course clinical trials.. *Journal of the American Statistical Association* **97**, 29–39.

Tummarello, D., Mari, D., Graziano, F., Isidori, P., Cetto, G., Pasini, F., Santo, A. & Cellerino, R. (1994). A randomized, controlled phase iii study of cyclophosphamide, doxorubicin and vincristine with etposide (cav-e) or teniposide (cav-t), followed by recombinant a-interferon maintenance therapy orobservation, in small cell lung carcinoma patients with complete responses.. *Anticancer Research* **14**, 2221–2228.

Xu, J. & Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and time to events. *Applied Statistics* **50**, 375–387.

Appendix A

A.1 Proof of Proposition 1

As in Robins et al. (1994), we consider the Hilbert space $\mathcal{H}$ consisting of all mean zero random functions of the observed data with finite variance equipped with the covariance inner product. Within this space we define the closed linear subspace $\mathcal{U}$ consisting of random functions

$$R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) f(T_i^R, V_i),$$

where $f(T_i^R, V_i)$ is an arbitrary function with finite variance. Our aim is to find the function $f(\cdot, \cdot)$ which minimizes the variance in (3.5), or equivalently, to find the element in $\mathcal{U}$ which minimizes the distance from $\left\{ (1 - R_i) + \frac{R_i X_{ki}}{\pi_k} \right\} \{h(T_i) - \mu_{1k}\}$ to some element in $\mathcal{U}$. By the projection theorem for Hilbert spaces (Luenberger, 1969), the optimal $f(\cdot, \cdot)$ is given by the unique $f^*(\cdot, \cdot)$, where f^* satisfies

$$\begin{aligned} E \Bigg[& \left\{ (1 - R_i) + \frac{R_i X_{ki}}{\pi_k} \right\} \{h(T_i) - \mu_{1k}\} \\ & + \quad R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) f^*(T_i^R, V_i) \Bigg] \left\{ R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) f(T_i^R, V_i) \right\} = 0 \end{aligned} \qquad \text{(A.1)}$$

for all $f(\cdot,\cdot)$. Since $(1-R_i)R_i = 0$, Equation (A.1) can be simplified to

$$E\left[\frac{R_i X_{ki}(X_{ki}-\pi_k)}{\pi_k^2}\left\{h(T_i)-\mu_{1k}\right\} f(T_i^R, V_i) + R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)^2 f^*(T_i^R, V_i) f(T_i^R, V_i)\right] = 0 \tag{A.2}$$

We will compute the expectation on the LHS of Equation (A.2) by iterated conditional expectation. By conditioning the second term on T_i^R, V_i and R_i, we get,

$$E\left\{ R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)^2 f^*(T_i^R, V_i) f(T_i^R, V_i)\middle| R_i, T_i^R, V_i\right\} = R_i f^*(T_i^R, V_i) f(T_i^R, V_i)\left(\frac{1-\pi_k}{\pi_k}\right), \tag{A.3}$$

where we have used the fact that conditional on (R_i, T_i^R, V_i), X_{ki} is Bernoulli with probability π_k. Similarly,

$$\begin{aligned} & E\left[\frac{R_i X_{ki}(X_{ki}-\pi_k)}{\pi_k^2}\left\{h(T_i)-\mu_{1k}\right\} f(T_i^R, V_i)\middle| R_i, T_i^R, V_i\right] \\ = \;& \frac{R_i}{\pi_k^2} f(T_i^R, V_i) E\left[X_{ki}(X_{ki}-\pi_k)\left\{h(T_i)-\mu_{1k}\right\}\middle| R_i = 1, T_i^R, V_i\right] \\ = \;& \frac{R_i(1-\pi_k)}{\pi_k} f(T_i^R, V_i) E\left\{ h(T_i)-\mu_{1k}\middle| R_i = 1, X_{ki} = 1, T_i^R, V_i\right\}. \end{aligned} \tag{A.4}$$

Using (A.3) and (A.4) we rewrite (A.2) as

$$\begin{aligned} & E\left[R_i\left(\frac{1-\pi_k}{\pi_k}\right) E\left\{ h(T_i)-\mu_{1k}\middle| R_i = 1, X_{ki} = 1, T_i^R, V_i\right\}\right. \\ + \;& \left. R_i\left(\frac{1-\pi_k}{\pi_k}\right) f^*(T_i^R, V_i) f(T_i^R, V_i)\right] = 0 \end{aligned} \tag{A.5}$$

for all $f(T_i^R, V_i)$. Or, equivalently

$$E\left[R_i\left\{E\left[h(T_i)-\mu_{1k}\middle| R_i = 1, X_{ki} = 1, T_i^R, V_i\right] + f^*(T_i^R, V_i)\right\} f(T_i^R, V_i)\right] = 0 \tag{A.6}$$

for all $f(T_i^R, V_i)$. In order for (A.6) to hold for all $f(T_i^R, V_i)$, we must have

$$E\left\{ h(T_i)-\mu_{1k}\middle| R_i = 1, X_{ki} = 1, T_i^R, V_i\right\} + f^*(T_i^R, V_i) = 0, \tag{A.7}$$

Or, equivalently

$$\begin{aligned} f^*(T_i^R, V_i) &= -E\left\{ h(T_i) - \mu_{1k} |\, R_i = 1, X_{ki} = 1, T_i^R, V_i \right\} \\ &= \theta_h(T_i^R, V_i) - \mu_{1k}. \end{aligned} \tag{A.8}$$

Substituting $f^*(\cdot, \cdot)$ into (3.5) and simplifying further we get the most efficient influence function (2.8). This completes the proof.

A.2 Consistency and asymptotic normality

Consistency

Under mild regularity conditions, all the estimators $\widehat{\mu}_{1k}^{ME}$, $\widehat{\mu}_{1k}^{LE}$, $\widehat{\mu}_{1k}^{LS}$ and $\widehat{\mu}_{1k}^{IMP}$ are consistent. The consistency of $\widehat{\mu}_{1k}^{ME}$ follows from the fact that it is a sample average of iid quantities and hence WLLN applies. Consider $\widehat{\mu}_{1k}^{LE}$. We will assume that $\widehat{\boldsymbol{\gamma}}$ converges in probability to $\boldsymbol{\gamma}_0$. We write

$$\widehat{\mu}_{1k}^{LE} = \frac{1}{n} \sum_{i=1}^{n} \left[\left\{ (1 - R_i) + \frac{R_i X_{ki}}{\pi_k} \right\} h(T_i) - R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) g(T_i^R, V_i, \boldsymbol{\gamma_0}) \right] \tag{A.9}$$

$$- \frac{1}{n} \sum_{i=1}^{n} R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) \left\{ g(T_i^R, V_i, \widehat{\boldsymbol{\gamma}}) - g(T_i^R, V_i, \boldsymbol{\gamma}_0) \right\} \tag{A.10}$$

Now the term (A.9) is the sample average of iid quantities having mean μ_{1k} and constant variance, and hence converges to its mean μ_{1k} in probability. A Taylor's series expansion of $g(T_i^R, V_i, \widehat{\boldsymbol{\gamma}})$ in (A.10) shows that we can write the second term as

$$(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^T n^{-1} \sum_{i=1}^{n} R_i \left(\frac{X_{ki} - \pi_k}{\pi_k} \right) \left\{ \frac{\delta}{\delta \boldsymbol{\gamma}} g(T_i^R, V_i, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}} \right\} \tag{A.11}$$

where $\tilde{\boldsymbol{\gamma}}$ lies between $\widehat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}_0$. Now in addition if we assume that

$$Sup_{\boldsymbol{\gamma} \in \Gamma(\widehat{\boldsymbol{\gamma}},\tilde{\boldsymbol{\gamma}})} \frac{\delta}{\delta \boldsymbol{\gamma}} g(T_i^R, V_i, \boldsymbol{\gamma}) \leq H(T_i^R, V_i; \boldsymbol{\gamma}_0)$$

with $E\left\{H(T_i^R, V_i; \boldsymbol{\gamma}_0)\right\} < \infty$ where $\Gamma(\widehat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}})$ is the set of all $\boldsymbol{\gamma}$ that lies between $\widehat{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}$, then it follows that (A.11), hence (A.10) is $o_p(1)$. Combining these results, we see that $\widehat{\mu}_{1k}^{LE} \xrightarrow{p} \mu_{1k}$.

Since least squares estimators are consistent, $\widehat{\mu}_{1k}^{LS}$ being a special case of $\widehat{\mu}_{1k}^{LE}$ is also consistent. Similar arguments can be applied to show the consistency of $\widehat{\mu}_{1k}^{IMP}$.

Asymptotic normality

Under mild regularity conditions, all the estimators $\widehat{\mu}_{1k}^{ME}$, $\widehat{\mu}_{1k}^{LE}$, $\widehat{\mu}_{1k}^{LS}$ and $\widehat{\mu}_{1k}^{IMP}$ are asymptotically normal. Since $\widehat{\mu}_{1k}^{ME}$ is a sample average of iid quantities with finite variance, the asymptotic normality follows directly from the central limit theorem. We will sketch the proof for $\widehat{\mu}_{1k}^{LE}$. The rest follows immediately.

If $\widehat{\boldsymbol{\gamma}}$ is $n^{1/2}$-consistent, then similar arguments as the ones in the case of consistency can be used to write

$$n^{\frac{1}{2}}(\widehat{\mu}_{1k}^{LE} - \mu_{1k}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (Y_{1ki} - \mu_{1k}) + o_p(1) \tag{A.12}$$

where, $Y_{1ki} = \left\{(1 - R_i) + \frac{R_i X_{ki}}{\pi_k}\right\} h(T_i) - R_i \left(\frac{X_{ki} - \pi_k}{\pi_k}\right) g(T_i^R, V_i, \boldsymbol{\gamma_0})$ are iid with mean zero and variance $\sigma^2 = E(Y_{1ki}^2)$. Therefore, by Slutsky's theorem, $n^{\frac{1}{2}}(\widehat{\mu}_{1k}^{LE} - \mu_{1k}) \xrightarrow{d} N(0, \sigma^2)$. The asymptotic variance of $\widehat{\mu}_{1k}^{LE}$ can be estimated by (2.13).

Appendix B

B.1 Derivation of Most Efficient Influence Function for Censored Data

Let us define the filtration $\mathcal{F}_n(t)$ as the increasing sequence of sub-σ-algebras $\sigma\{I(C_i \leq u), u \leq t; G_i^H(t); R_i, X_{ki}, T_{1ki}, k = 1, 2; i = 1, \ldots, n\}$. Unless otherwise stated, this is the filtration with respect to which all the martingales are defined in this paper. For example, with respect to this filtration, the stochastic quantity $M_i^c(t)$ defined in Section 2.2 is a martingale process with the intensity function $\lambda^c(t)Y_i(t)$. In the following derivation, we exploit the powerful counting process and martingale theory described in Fleming and Harrington(1991).

For simplicity, let us define $h^*(T_{1k}) = h(T_{1k}) - \mu_{1k}$. Substituting (3.5) in (3.8), we

get

$$
\begin{aligned}
(3.8) \;&=\; Q_{ki}h^*(T_i) - \int \frac{dM_i^c(u)}{K(u)} \left[Q_{ki}h^*(T_i) - E\left\{ Q_{ki}h^*(T_i) | G^H(u) \right\} \right] && \text{(B.1)} \\
&+\; (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - \int \frac{dM_i^c(u)}{K(u)} \left[(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \right. && \text{(B.2)} \\
&-\; \left. E\left\{ (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) | G^H(u) \right\} \right] && \text{(B.3)}
\end{aligned}
$$

For a fixed u, the collection of information $G^H(u)$ for an individual i, will be different if he/she has responded and consented to second-stage randomization by time u from that if he/she has not responded. Also for a patient to contribute to the integral in (B.2-B.3), he/she needs to be at risk at time u. This prompts us to consider the following two cases.

Case I: $R_i = 1, T_i^R < u, T_i \geq u$.

In this case, given $G^H(u)$, we know X_{ki}, T_i^R and $G^H(T_i^R)$. Hence,

$$
E\left\{ (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) | G^H(u) \right\} = (Q_{ki}-1)f(T_i^R, G^H(T_i^R)). \qquad \text{(B.4)}
$$

Case II: $(R_i = 0, T_i \geq u) \cup (R_i = 1, T_i^R < u, T_i \geq u)$.

In this case, conditioning on $T_i^R, G^H(T_i^R)$, we find that

$$
E\left\{ (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) | G^H(u) \right\} = 0. \qquad \text{(B.5)}
$$

Cases I and II together gives

$$
E\left\{ (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) | G^H(u) \right\} = I(T_i^R < u)(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \qquad \text{(B.6)}
$$

, for $T_i \geq u$. Consequently the stochastic integral in (B.2-B.3) turns out to be

$$\begin{aligned}
&\int \frac{dM_i^c(u)}{K(u)} \left[(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - E\left\{(Q_{ki}-1)f(T_i^R, G^H(T_i^R))|G^H(u)\right\}\right] \\
= &\int \tfrac{dM_i^{*c}(u)I(T_i \geq u)}{K(u)} \left[(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - I(T_i^R < u)(Q_{ki}-1)f(T_i^R, G^H(T_i^R))\right] \\
= &\int \tfrac{dM_i^{*c}(u)I(T_i \geq u)I(T_i^R \geq u)}{K(u)}(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \qquad \text{(B.7)}
\end{aligned}$$

where $M_i^{*c}(u)$ is the martingale process defined through the martingale increments $dM_i^{*c}(u) = d\left(I(C_i \leq u)\right) - \lambda^c(u)I(C_i \geq u)du$. Since $R_i I(T_i \geq u)I(T_i^R \geq u) = R_i I(T_i^R \geq u)$, this implies that the stochastic integral in (B.2-B.3) equals to

$$\int \frac{dM_i^{*c}(u)}{K(u)} I(T_i^R \geq u)(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \qquad \text{(B.8)}$$

This enables us to write the influence function (3.8) in the form

$$\begin{aligned}
&Q_{ki}h^*(T_i) - \int \frac{dM_i^c(u)}{K(u)} \left[Q_{ki}h^*(T_i) - E\left\{Q_{ki}h^*(T_i)|G^H(u)\right\}\right] \\
+ \;& (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \\
- \;& \int \frac{dM_i^{*c}(u)}{K(u)}(Q_{ki}-1)I(T_i^R \geq u)f(T_i^R, G^H(T_i^R)). \qquad \text{(B.9)}
\end{aligned}$$

By defining another martingale difference

$$dM_i^{R_c}(u) = dI(U_i^* \leq u, \Delta_i^* = 0) - \lambda^c(u)I(U_i^* \geq u)du,$$

we can re-write (B.9) as

$$\begin{aligned}
&Q_{ki}h^*(T_i) - \int \frac{dM_i^c(u)}{K(u)} \left[Q_{ki}h^*(T_i) - E\left\{Q_{ki}h^*(T_i)|G^H(u)\right\}\right] \\
+ \;& (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - \int \frac{dM_i^{R_c}(u)}{K(u)}(Q_{ki}-1)f(T_i^R, G^H(T_i^R)). \qquad \text{(B.10)}
\end{aligned}$$

Notice that we have used the relationship $R_i dM_i^{R_c} c(u) = R_i I(T_i^R \geq u) dM_i^{*c}(u)$ to arrive at the above result. To find the optimal influence function, as in Robins et al. (1994) , we consider the Hilbert space $\mathcal{H}$ consisting of all mean zero random functions of the observed data with finite variance equipped with the covariance inner product. Within this space we define the closed linear subspace $\mathcal{U}$ consisting of random functions

$$\begin{aligned}
&(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - \int \frac{dM_i^{R_c}(u)}{K(u)}(Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \\
&= \quad (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) - \int \frac{dM_i^{*c}(u)}{K(u)}(Q_{ki}-1)I(T_i^R \geq u)f(T_i^R, G^H(T_i^R)),
\end{aligned}$$

where $f(T_i^R, G^H(T_i^R))$ is an arbitrary function, with finite variance. Our aim is to find the function $f(\cdot,\cdot)$ which minimizes the variance in (B.10), or equivalently, to find the element in $\mathcal{U}$ which minimizes the distance (square root of variance) from

$$Q_{ki}h^*(T_i) - \int \frac{dM_i^c(u)}{K(u)} \left[Q_{ki}h^*(T_i) - E\left\{ Q_{ki}h^*(T_i) | G^H(u) \right\} \right]$$

to some element in $\mathcal{U}$. By the projection theorem for Hilbert spaces (Luenberger, 1969), the optimal $f(\cdot,\cdot)$ is given by the unique $f^*(\cdot,\cdot)$, where f^* satisfies

$$\begin{aligned}
&E\left\{ Q_{ki}h^*(T_i) - \int \frac{dM_i^c(u)}{K(u)} \left[Q_{ki}h^*(T_i) - E\left\{ Q_{ki}h^*(T_i) | G^H(u) \right\} \right] \right. \\
&+ \quad (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \\
&- \quad \left. \int \frac{dM_i^{*c}(u)}{K(u)}(Q_{ki}-1)I(T_i^R \geq u)f(T_i^R, G^H(T_i^R)) \right\} \left\{ (Q_{ki}-1)f(T_i^R, G^H(T_i^R)) \right. \\
&- \quad \left. \int \frac{dM_i^{*c}(u)}{K(u)}(Q_{ki}-1)I(T_i^R \geq u)f^*(T_i^R, G^H(T_i^R)) \right\} = 0 \qquad \text{(B.11)}
\end{aligned}$$

for all $f(\cdot,\cdot)$.

To simplify the LHS of Equation (B.11), let us characterize the general form of functionals of $G^H(u)$. A typical functional of $G^H(u)$ can be written as

$$R_i I(T_i^R < u) h_1(G^H(u)) + \left[(1 - R_i) + R_i I(T_i^R \geq u)\right] h_2(G^H(u)) \tag{B.12}$$

for some functions $h_1(\cdot, \cdot, \cdot)$ and $h_2(\cdot)$. Consequently, any such functional, when multiplied by

$$I(T_i^R \geq u)(Q_{ki} - 1) f(T_i^R, G^H(T_i^R))$$

leaves us with

$$I(T_i^R \geq u)(Q_{ki} - 1) f(T_i^R, G^H(T_i^R)) h_2(G^H(u)).$$

In other words,

$$\begin{aligned} & I(T_i^R \geq u)(Q_{ki} - 1) f(T_i^R, G^H(T_i^R)) \times \text{functional of } G^H(u) \\ = \;& I(T_i^R \geq u)(Q_{ki} - 1) f(T_i^R, G^H(T_i^R)) h_2(G^H(u)). \end{aligned} \tag{B.13}$$

Note that (B.13) is defined only for individuals whose response time T_i^R goes beyond T_i^R and hence $G^H(u)$ will be contained in $G^H(T_i^R)$. Thus, for a given u, conditioning on $G^H(T_i^R)$, we can show that the RHS of (B.13) has expectation zero. This fact , along with the identities $R_i(1 - R_i) = 0$, $dM_i^c(u) = I(T_i \geq u) dM_i^{*c}(u)$ and other

standard martingale properties enables us to write (B.11) as

$$
\begin{aligned}
& E\left[Q_{ki}(Q_{ki}-1)h^*(T_i)f(T_i^R, G^H(T_i^R))\right] \\
& + \quad E\left[Q_{ki}(Q_{ki}-1)h^*(T_i)f(T_i^R, G^H(T_i^R))\int \frac{\lambda^c(u)du}{K(u)} I(T_i \geq u) I(T_i^R \geq u)\right] \\
& + \quad E\left[Q_{ki}^2 f^*(T_i^R, G^H(T_i^R))f(T_i^R, G^H(T_i^R))\right] \\
& + \quad E\left[Q_{ki}^2 f^*(T_i^R, G^H(T_i^R))f^*(T_i^R, G^H(T_i^R)f(T_i^R, G^H(T_i^R))\int \frac{\lambda^c(u)du}{K(u)} I(T_i^R \geq u)\right] = 0
\end{aligned}
\tag{B.14}
$$

for all $f(\cdot,\cdot)$.

Using sequential conditional expectations and other model assumptions such as (3.1) and (2.2), we can show that the above is equivalent to

$$
\begin{aligned}
& E\left[R_i\left(\frac{1-\pi_k}{\pi_k}\right)E(h^*(T_i)|R_i=1, X_{ki}=1, G^H(T_i^R))f(T_i^R, G^H(T_i^R))\right] \\
& + \quad E\left[R_i\left(\frac{1-\pi_k}{\pi_k}\right)\int \frac{\lambda^c(u)du}{K(u)} I(T_i^R \geq u)\right. \\
& \times \quad \left. E(h^*(T_i)|R_i=1, X_{ki}=1, G^H(T_i^R))f(T_i^R, G^H(T_i^R))\right] \\
& + \quad E\left[R_i\left(\frac{1-\pi_k}{\pi_k}\right)f^*(T_i^R, G^H(T_i^R))f(T_i^R, G^H(T_i^R))\right] \\
& + \quad E\left[R_i\left(\frac{1-\pi_k}{\pi_k}\right)\int \frac{\lambda^c(u)du}{K(u)} I(T_i^R \geq u) f^*(T_i^R, G^H(T_i^R))f(T_i^R, G^H(T_i^R))\right] = 0.
\end{aligned}
\tag{B.15}
$$

Or equivalently,

$$
\begin{aligned}
& E\left[R_i\left\{f^*(T_i^R, G^H(T_i^R)) + E\left(h^*(T_i)|R_i=1, X_{ki}=1, G^H(T_i^R)\right)\right\}\right. \\
& \times \quad \left.\left(1+\int \frac{\lambda^c(u)}{K(u)} I(T_i^R \geq u)\right)\right] = 0
\end{aligned}
\tag{B.16}
$$

for all $f(\cdot,\cdot)$s. The last equation leads us to the optimal solution

$$f^*(T_i^R, G^H(T_i^R)) = -E\left(h^*(T_i)|R_i = 1, X_{ki} = 1, G^H(T_i^R)\right) \tag{B.17}$$

Therefore, substituting the optimal f in Equation (B.10) and rearranging, the optimal influence function for estimating μ_{1k} is given by

$$\frac{\Delta_i Q_{ki} h^*(T_i)}{K(U_i)} + \int \frac{dM_i^c(u)}{K(u)} e_{h^*}\left\{G^H(u)\right\} - \frac{\Delta_i^*(Q_{ki}-1)}{K(U_i^*)} E\left(h^*(T_i)|R_i = 1, X_{ki} = 1, G^H(T_i^R)\right), \tag{B.18}$$

where $e_{h^*}\left\{G^H(u)\right\} = E\left\{Q_{ki}h^*(T_i)|G^H(u)\right\}$. The term $e_{h^*}\left\{G^H(u)\right\}$in (3.9) needs to be explained a little bit more to make any sense in reality. Again, we will consider the two cases separately.

Case I: $R_i = 1, T_i^R < u, T_i \geq u$. In this case, $E\left[(1-R_i)h^*(T_i)|G^H(u)\right] = 0$ and $E\left[\frac{R_i X_{ki}}{\pi_k} h^*(T_i)|G^H(u)\right] = \frac{X_{ki}}{\pi_k} E\left(h^*(T_i)|R_i = 1, X_{ki} = 1, G^H(u), T_i \geq u\right)$. And hence, $e_{h^*}\left\{G^H(u)\right\} = R_i I(T_i^R < u)\frac{X_{ki}}{\pi_k} E\left(h^*(T_i)|R_i = 1, X_{ki} = 1, G^H(u), T_i \geq u\right)$.

Case II: $(R_i = 0, T_i \geq u) \cup (R_i = 1, T_i^R \geq u, T_i \geq u)$. In this case,

$$\begin{aligned} & E\left[Q_{ki}h^*(T_i)|G^H(u)\right] \\ = \; & \left\{1 - R_i I(T_i^R < u)\right\} E[Q_{ki}h^*(T_i)|R_i I(T_i^R < u) = 0, G^H(u), T_i \geq u]. \end{aligned} \tag{B.19}$$

Cases I and II together gives

$$e_{h^*}\left\{G^H(u)\right\} \tag{B.20}$$

$$\begin{aligned} = \; & R_i I(T_i^R < u)\frac{X_{ki}}{\pi_k} E\left(h^*(T_i)|R_i = 1, X_{ki} = 1, G^H(u), T_i \geq u\right) \\ + \; & \left\{1 - R_i I(T_i^R < u)\right\} E[Q_{ki}h^*(T_i)|R_i I(T_i^R < u) = 0, G^H(u), T_i \geq u]. \end{aligned} \tag{B.21}$$

Notice that one can write $e_{h^*}\left\{G^H(u)\right\}$ as $e_h\left\{G^H(u)\right\} - e_1\left\{G^H(u)\right\}\mu_{1k}$, where

$$\begin{aligned} e_1\left\{G^H(u)\right\} &= \left[R_i I(T_i^R < u)\frac{X_{ki}}{\pi_k} + 1 - R_i I(T_i^R < u)\right] \\ &= \left[1 + R_i\left(\frac{X_{ki}-\pi_k}{\pi_k}\right) - R_i I(T_i^R \geq u)\left(\frac{X_{ki}-\pi_k}{\pi_k}\right)\right] \end{aligned} \quad \text{(B.22)}$$

and $e_h\left\{G^H(u)\right\}$ is given by (**??**). Therefore, by substituting $e_{h^*}\left\{G^H(u)\right\}$ in (B.18) and simplifying further by separating out the coefficients of μ_{1k}, we obtain the optimal influence function (3.9).

B.2 Consistency and Asymptotic Normality for the Estimators in Censored Data

B.2.1 Consistency

Under mild regularity conditions, the estimator $\widehat{\mu}_{1k}^{IMP}$ is consistent. To see that, we write $\widehat{\mu}_{1k}^{IMP}$ as $A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h})/A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu})$ where $A_n(\widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h})$ is defined as

$$\begin{aligned} & A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i Q_{ki}}{\widehat{K}(U_i)}h(U_i) - n^{-1}\sum_{i=1}^{n}\frac{\Delta_i^*(Q_{ki}-1)}{\widehat{K}(U_i^*)}\widehat{\boldsymbol{\gamma}}_h^T \boldsymbol{W}_i \\ &+ n^{-1}\sum_{i=1}^{n}\int\frac{dN_i^c(u)}{\widehat{K}(u)}\left\{\widehat{\varphi}_{1h}(u)L_{\varphi 1 i}(u) + \widehat{\varphi}_{2h}(u)L_{\varphi 2 i}(u)\right\} \end{aligned} \quad \text{(B.23)}$$

The first term on the RHS is the same as the estimator $\widehat{\mu_{1k}}^{IPMW}$, which converges in probability to μ_{1k} as was argued in the Appendix A of Lunceford et. al.(2002).

Consider the second term on the RHS of (B.23), which can be written as

$$\widehat{\boldsymbol{\gamma}}_h^T \left[n^{-1} \sum_{i=1}^{n} \frac{\Delta_i^*(Q_{ki}-1)\boldsymbol{W}_i}{K(T_i^R)} + n^{-1} \sum_{i=1}^{n} \frac{\widehat{K}(T_i^R) - K(T_i^R)}{\widehat{K}(T_i^R)K(T_i^R)} \Delta_i^*(Q_{ki}-1)\boldsymbol{W}_i \right]. \quad \text{(B.24)}$$

The first term in the square bracket is the sample average of iid random variables having expectation zero and hence is $o_p(1)$. The second term can be shown to be bounded by

$$\frac{Sup_{u \le L} \left| \widehat{K}(u) - K(u) \right|}{\widehat{K}(L)K(L)} \left| n^{-1} \sum_{i=1}^{n} \Delta_i^*(Q_{ki}-1)\boldsymbol{W}_i \right| \quad \text{(B.25)}$$

By the arguments made in Chapter 6 of Fleming and Harrington (1991), $\widehat{K}(u)$ converges uniformly to the survival distribution $K(u)$ for $u \le L$ and hence $\frac{Sup_{u \le L} \left| \widehat{K}(u) - K(u) \right|}{\widehat{K}(L)K(L)}$ is $o_p(1)$. It is easy to show that the other absolute term in (B.25) is also $o_p(1)$. Applying these results in combination with the fact that $\widehat{\boldsymbol{\gamma}}_h$ is a consistent estimator, we see that (B.24) is $o_p(1)$. Using the martingale process $M_i^c(t)$ defined previously, the third term in the RHS of (B.23) can be written as

$$n^{-1} \sum_{i=1}^{n} \int \frac{dM_i^c(u)}{\widehat{K}(u)} \left\{ \widehat{\varphi}_{1h}(u) L_{\varphi_1 i}(u) + \widehat{\varphi}_{2h}(u) L_{\varphi_2 i}(u) \right\}. \quad \text{(B.26)}$$

Using the cosistency property of IPMW estimators $\widehat{\varphi}_{1h}(u)$ and $\widehat{\varphi}_{1h}(u)$ and the properties of martingales, we can show that the term (B.26) converges in probability to zero. Thus $A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h}) \xrightarrow{p} \mu_{1k}$. Similar arguments can be applied to show that $A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu}) \xrightarrow{p} 1$ implying that $\widehat{\mu}_{1k}^{IMP}$ is a consistent estimator for μ_{1k}.

B.2.2 Asymptotic Normality

To derive the asymptotic distribution of $\widehat{\mu}_{1k}^{IMP}$, we write

$$n^{1/2}(\widehat{\mu}_{1k}^{IMP} - \mu_{1k}) = \frac{n^{1/2}\left\{A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h}) - A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu})\mu_{1k}\right\}}{A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu})} \quad \text{(B.27)}$$

Since $A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu}) \xrightarrow{p} 1$, by Slutsky's theorem, the asymptotic distribution of $n^{1/2}(\widehat{\mu}_{1k}^{IMP} - \mu_{1k})$ will be the same as the asymptotic distribution of

$$n^{1/2}\left\{A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h}) - A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu})\mu_{1k}\right\}.$$

i.e.,

$$\begin{aligned}
& n^{1/2}(\widehat{\mu}_{1k}^{IMP} - \mu_{1k}) \\
\overset{ad}{\equiv}\ & n^{1/2}\left\{A_n(h(.), \widehat{\boldsymbol{\gamma}}_h, \widehat{\varphi}_{1h}, \widehat{\varphi}_{2h}) - A_n(1, \widehat{\boldsymbol{\gamma}}_\mu, 1, \widehat{\varphi}_{2\mu})\mu_{1k}\right\} \\
=\ & n^{1/2}A_n(h^*(.), \widehat{\boldsymbol{\gamma}}^{opt}, \widehat{\varphi}_1^{opt}, \widehat{\varphi}_2^{opt}) \\
=\ & n^{-1/2}\sum_{i=1}^{n}\frac{\Delta_i Q_{ki}}{\widehat{K}(U_i)}h^*(U_i) + n^{-1/2}\sum_{i=1}^{n}\frac{\Delta_i^*(Q_{ki}-1)}{\widehat{K}(U_i^*)}\left(\widehat{\boldsymbol{\gamma}}^{opt}\right)^T \boldsymbol{W}_i \\
+\ & n^{-1/2}\sum_{i=1}^{n}\int\frac{dN_i^c(u)}{\widehat{K}(u)}\left\{\widehat{\varphi}_1^{opt}(u)L_{\varphi_1 i}(u) + \widehat{\varphi}_2^{opt}(u)L_{\varphi_2 i}(u)\right\} \qquad \text{(B.28)}
\end{aligned}$$

where $h^*(u) = h(u) - \mu_{1k}$, $\widehat{\boldsymbol{\gamma}}^{opt} = -\widehat{\boldsymbol{\gamma}}_h + \widehat{\boldsymbol{\gamma}}_\mu\mu_{1k}$, $\widehat{\varphi}_1^{opt}(u) = \widehat{\varphi}_{1h}(u) - \mu_{1k}$, $\widehat{\varphi}_2^{opt}(u) = \widehat{\varphi}_{2h}(u) - \widehat{\varphi}_{2\mu}(u)\mu_{1k}$, and $\overset{ad}{\equiv}$ stands for equivalent in asymptotic distribution.

Using Equation (A.5) of Lunceford et al. (2002) the first term in (B.28) can be written as

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\frac{\Delta_i Q_{ki}}{\widehat{K}(U_i)}h^*(U_i) \\
=\ & n^{-1/2}\sum_{i=1}^{n}Q_{ki}h^*(T_i) - n^{-1/2}\sum_{i=1}^{n}\int_0^L\frac{Q_{ki}h^*(T_i) - G_{1k}(u)}{K(u)}dM_i^c(u) + o_p(1)
\end{aligned} \qquad \text{(B.29)}$$

where

$$G_{1k}(u) = \frac{E\left\{h^*(T_{1k})I(T_{1k} \geq u)\right\}}{P(T > u)}.$$

Now consider the second term of (B.28). In a similar fashion, using the martingale difference sequence $\left\{dM_i^{R_c}(u), i = 1 \cdots n\right\}$, we can have the following result:

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \frac{\Delta_i^*(Q_{ki} - 1)}{\widehat{K}(U_i^*)} \left(\widehat{\boldsymbol{\gamma}}_h^{opt}\right)^T \boldsymbol{W}_i \\ = \;& n^{-1/2} \sum_{i=1}^{n} (Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i - n^{-1/2} \sum_{i=1}^{n} \int_0^L \frac{(Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i}{K(u)} dM_i^{R_c}(u) + o_p(1) \end{aligned} \tag{B.30}$$

The third and final term in (B.28),

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \int \frac{dN_i^c(u)}{\widehat{K}(u)} \left\{\widehat{\varphi}_1^{opt}(u) L_{\varphi_1 i}(u) + \widehat{\varphi}_2^{opt}(u) L_{\varphi_2 i}(u)\right\} \\ = \;& n^{-1/2} \sum_{i=1}^{n} \int \frac{dM_i^c(u)}{\widehat{K}(u)} \left\{\widehat{\varphi}_1^{opt}(u) L_{\varphi_1 i}(u) + \widehat{\varphi}_2^{opt}(u) L_{\varphi_2 i}(u)\right\} \\ = \;& n^{-1/2} \sum_{i=1}^{n} \int \frac{dM_i^c(u)}{K(u)} \left\{\varphi_1^{opt}(u) L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u) L_{\varphi_2 i}^*(u)\right\} + o_p(1) \end{aligned} \tag{B.31}$$

where

$$L_{\varphi_1 i}^*(u) = R_i I(U_i^* < u) \frac{X_{ki}}{\pi_k} - E\left\{R_i I(T_i^R < u) \frac{X_{ki}}{\pi_k}\right\},$$

and

$$L_{\varphi_2 i}^*(u) = 1 - R_i I(U_i^* < u) - E\left\{1 - R_i I(T_i^R < u)\right\}.$$

Substituting (B.29)-(B.31) in (B.28), we have ,

$$
\begin{aligned}
& n^{1/2}(\widehat{\mu}_{1k}^{IMP} - \mu_{1k}) \\
& \overset{ad}{\equiv} \; n^{-1/2} \sum_{i=1}^{n} \Big[\Big\{ Q_{ki} h^*(T_i) + (Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i \Big\} \\
& \qquad - \int_0^L \frac{dM_i^{R_c}(u)}{K(u)} (Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i \\
& \qquad + \int_0^L \frac{dM_i^{c}(u)}{K(u)} \left\{ \varphi_1^{opt}(u) L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u) L_{\varphi_2 i}^*(u) - Q_{ki} h^*(T_i) + G_{1k}(u) \right\} \Big] + o_p(1) \\
& = \; n^{-1/2} \sum_{i=1}^{n} \Big\{ Q_{ki} h^*(T_i) + (Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i \Big\} \\
& \qquad - n^{-1/2} \sum_{i=1}^{n} \int_0^L \frac{dM_i^{*c}(u)}{K(u)} \Big[(Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i I(T_i^R \geq u) \\
& \qquad - \left\{ \varphi_1^{opt}(u) L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u) L_{\varphi_2 i}^*(u) - Q_{ki} h^*(T_i) + G_{1k}(u) \right\} I(T_i \geq u) \Big] + o_p(1) \\
& = \; n^{-1/2} \sum_{i=1}^{n} \Psi_{ki}^{IMP} + o_p(1) \qquad \text{(B.32)}
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi_{ki}^{IMP} \; = \; & \Big\{ Q_{ki} h^*(T_i) + (Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i \Big\} \\
& - \int_0^L \frac{dM_i^{*c}(u)}{K(u)} \Big[(Q_{ki} - 1) \left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i I(T_i^R \geq u) \\
& - \left\{ \varphi_1^{opt}(u) L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u) L_{\varphi_2 i}^*(u) - Q_{ki} h^*(T_i) + G_{1k}(u) \right\} I(T_i \geq u) \Big]
\end{aligned}
\qquad \text{(B.33)}
$$

is the influence function for the estimator $\widehat{\mu}_{1k}^{IMP}$ and has expectation zero. Thus we have been able to show that $n^{1/2}$-times the estimator minus the parameter is equal to $n^{-1/2}$-times sum of mean-zero i.i.d random variables plus a term of $o_p(1)$. Therefore $\widehat{\mu}_{1k}^{IMP}$ is asymptotically linear estimator. Accordingly, $n^{1/2}(\widehat{\mu}_{1k}^{IMP} - \mu_{1k}) \xrightarrow{d} N(0, \sigma^2)$, where σ^2 is given by the variance of the influence function Ψ_{ki}^{IMP}. Namely, by using

the independence of the first and second term of Ψ_{ki}^{IMP},

$$\begin{aligned}
\sigma^2 &= Var(\Psi_{ki}^{IMP}) = E\left(\Psi_{ki}^{IMP}\right)^2 \\
&= E\left\{Q_{ki}h^*(T_i) + (Q_{ki}-1)\left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i\right\}^2 \\
&\quad + E\left\{\int_0^L \frac{dM_i^{*c}(u)}{K(u)}\left[(Q_{ki}-1)\left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i I(T_i^R \geq u)\right.\right. \\
&\quad \left.\left. - \left\{\varphi_1^{opt}(u)L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u)L_{\varphi_2 i}^*(u) - Q_{ki}h^*(T_i) + G_{1k}(u)\right\} I(T_i \geq u)\right]\right\}^2.
\end{aligned} \tag{B.34}$$

Using martingale properties, (B.34) can be simplified to

$$\sigma^2 = E\left\{Q_{ki}h^*(T_i) + (Q_{ki}-1)\left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i\right\}^2 + \int_0^L \frac{\lambda^c(u)du}{K(u)} E\left\{R_{1ki}(u)\right\}^2, \tag{B.35}$$

where

$$\begin{aligned}
R_{1ki}(u) &= (Q_{ki}-1)\left(\boldsymbol{\gamma}^{opt}\right)^T \boldsymbol{W}_i I(T_i^R \geq u) - \left\{\varphi_1^{opt}(u)L_{\varphi_1 i}^*(u) + \varphi_2^{opt}(u)L_{\varphi_2 i}^*(u)\right. \\
&\quad \left. - Q_{ki}h^*(T_i) + G_{1k}(u)\right\} I(T_i \geq u).
\end{aligned} \tag{B.36}$$

The covariance of the two estimators $\widehat{\mu}_{11}^{IMP}$ and $\widehat{\mu}_{12}^{IMP}$ might be of importance for the testing purposes. Entirely similar argument can be applied to show that

$$\begin{aligned}
cov(\widehat{\mu}_{11}^{IMP}, \widehat{\mu}_{12}^{IMP}) &= n^{-1}cov(\Psi_{1i}^{IMP}, \Psi_{1i}^{IMP}) = n^{-1}E\left(\Psi_{1i}^{IMP}\Psi_{2i}^{IMP}\right) \\
&= E\left[\left\{Q_{1i}\left(h(T_i) - \mu_{11}\right) + (Q_{1i}-1)\left(\boldsymbol{\gamma}_1^{opt}\right)^T \boldsymbol{W}_i\right\}\right. \\
&\quad \left.\times \left\{Q_{2i}\left(h(T_i) - \mu_{12}\right) + (Q_{2i}-1)\left(\boldsymbol{\gamma}_2^{opt}\right)^T \boldsymbol{W}_i\right\}\right] \\
&\quad \int_0^L \frac{\lambda^c(u)du}{K(u)} E\left\{R_{11i}(u)R_{12i}(u)\right\}.
\end{aligned} \tag{B.37}$$

where $\boldsymbol{\gamma}_1^{opt}$ and $\boldsymbol{\gamma}_2^{opt}$ are versions of $\boldsymbol{\gamma}^{opt}$ for $k = 1, 2$, and R_{11i} and R_{12i} are obtained

from Equation B.36 respectively replacing k by 1 and 2 understanding that quantities dependent on μ_{1k} will be changed accordingly.