

Semi-Parametric Methods for Competing Risks Data with Applications in Organ Transplantation

by

Ludi Fan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2013

Doctoral Committee:

Professor Douglas E. Schaebel, Chair
Professor Thomas M. Braun
Associate Research Professor Brenda W. Gillespie
Assistant Professor Christopher J. Sonnenday

© Ludi Fan 2013

All Rights Reserved

For my parents

ACKNOWLEDGEMENTS

First I would like to thank my dissertation committee: Drs. Doug Schaubel, Tom Braun, Brenda Gillespie and Chris Sonnenday. I owe my deepest gratitude to my advisor Dr. Schaubel, who has been tremendously and consistently supportive of me during my entire time in graduate school, starting with instructing the course that got me interested in survival analysis. Thank you for guiding me through the field of competing risks and the dissertation writing process. I enjoyed our meetings and discussions immensely and I have learned much from you about what it means to be a methodological statistician, a healthcare researcher, and a mentor. I consider myself lucky to have had the opportunity to learn from and work with you.

Drs. Braun and Gillespie have both been extremely crucial both as instructors and as committee members, offering insightful comments and suggestions. Thank you for carefully going through the drafts, helping me clarify my ideas, and asking the thought-provoking questions. I especially appreciate your efforts to challenge me to think beyond the immediate scope of the work and to make connections between the different sections of the dissertation and future work.

A special acknowledgement is due to Dr. Sonnenday, who has provided crucial clin-

ical knowledge and the medical motivation for the dissertation. Your first-hand experience with kidney transplantation helped me focus on the real world implications of the statistical methods. I am grateful for your perspective of providing optimal patient healthcare, which reminds me of why we ultimately do what we do.

I would also like to thank Dr. Jack Kalbfleisch, as well as Drs. Schaubel and Sonnenday for their financial support, making it possible for me to finish what I have started.

There were many friends and colleagues who have made graduate school a memorable experience, in particular, Jennifer Lin, Maria Larkina, Doug Fuller, Erin Shellman, Bryan Mayer, Rena Sun, Kevin He, John Li, Nabihah Tayob, Laura Fernandes, Elsie Grace, Julia Kotlyar, Amanda Wilke, Rick Ma, Gong Qi, Erin Leister, Matt Jones, and Vesela Gateva. Thank you for all the good times.

I would also like to thank Carl Miller, a constant companion through much of the writing of this dissertation, the only other person in the trenches with me. Thank you for the life lessons and your kindness, encouragement, and love. You have helped me more than you will ever know.

Lastly, I would like to thank the people who mean the world to me. My parents, Yinshui Fan and Fuying Lu, laid the foundation for my graduate studies. Thank you for your unwavering support and unconditional love. You were an integral part of this process and it is to you that this dissertation is dedicated.

TABLE OF CONTENTS

| | |
|--|------|
| DEDICATION | ii |
| ACKNOWLEDGEMENTS | iii |
| LIST OF FIGURES | vii |
| LIST OF TABLES | viii |
| LIST OF APPENDICES | x |
| CHAPTER | |
| I. Introduction | 1 |
| II. Comparing Cumulative Incidence Functions Between Non-Randomized Groups | 5 |
| 2.1 Introduction | 5 |
| 2.2 Proposed Methods | 10 |
| 2.3 Asymptotic Properties | 15 |
| 2.4 Simulation Studies | 18 |
| 2.5 Application | 23 |
| 2.6 Discussion | 26 |
| III. Comparing Cumulative Incidence Functions using Weighted Counting Processes | 29 |

| | | |
|------------|--|------------|
| 3.1 | Introduction | 29 |
| 3.2 | Proposed Methods | 33 |
| 3.3 | Asymptotic Properties | 38 |
| 3.4 | Simulation Studies | 41 |
| 3.5 | Application | 45 |
| 3.6 | Discussion | 53 |
| | | |
| IV. | Semi-Parametric Methods for Modeling the Subdistribution Hazard using Multiple Imputation | 56 |
| 4.1 | Introduction | 56 |
| 4.2 | Proposed Methods | 63 |
| | 4.2.1 Set-up and Models | 63 |
| | 4.2.2 Imputing Censoring Times | 66 |
| | 4.2.3 Inference Procedures | 70 |
| 4.3 | Simulation Studies | 71 |
| 4.4 | Application | 73 |
| 4.5 | Discussion | 75 |
| | | |
| V. | Conclusion | 78 |
| | | |
| | APPENDICES | 82 |
| | BIBLIOGRAPHY | 126 |

LIST OF FIGURES

Figure

| | | |
|-----|---|-----|
| 2.1 | Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ for 1, 3, and 5 years post wait-listing, for $j = 1, \dots, 6$ | 24 |
| 3.1 | Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ for various $t \in [0, 5]$ years post wait-listing, for $j = 1, \dots, 6$ | 46 |
| 3.2 | Analysis of SRTR data: $\hat{\delta}_{11}(t)$ with 95% confidence interval for various $t \in [0, 5]$ years post wait-listing. | 48 |
| 3.3 | Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ with 95% confidence interval of for various $t \in [0, 5]$ years post wait-listing, for $j = 2, \dots, 6$ | 49 |
| 3.4 | Analysis of SRTR data: Change in $\hat{\delta}_{11}(t)$ over various time intervals, for $t \in [0, 5]$ years post wait-listing. | 50 |
| 3.5 | Analysis of SRTR data: Change in $\hat{\delta}_{41}(t)$ for various $t \in [0, 5]$ years post wait-listing. | 51 |
| 3.6 | Analysis of SRTR data: Change in $\hat{\delta}_{61}(t)$ for various $t \in [0, 5]$ years post wait-listing. | 52 |
| B.1 | Analysis of SRTR data: Boxplot of $\hat{\delta}_{j1}(t)$ for $t = 1$, $t = 3$, and $t = 5$ years. Whiskers represent minimum and maximum. | 103 |
| B.2 | Analysis of SRTR data: Histogram of $\hat{\delta}_{j1}(t)$ at $t = 5$ years post wait-listing. | 104 |

LIST OF TABLES

Table

| | | |
|-----|---|----|
| 2.1 | Baseline cause-specific hazards for the simulation study | 19 |
| 2.2 | Performance of $\hat{\delta}_{j1}(t)$ based on 500 simulations of Configuration 1, with bias, empirical standard deviation (ESD), the bootstrap standard error (BSE), and the 95% confidence interval coverage probabilities (CP). | 21 |
| 2.3 | Performance of $\hat{\delta}_{j1}(t)$ based on 500 simulations of Configuration 2, with bias, empirical standard deviation (ESD), the bootstrap standard error (BSE), and the 95% confidence interval coverage probabilities (CP). | 22 |
| 2.4 | Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ with 95% confidence limits for 1, 3, and 5 years post wait-listing, for $j = 1, \dots, 6$ | 25 |
| 3.1 | Simulation results using Configuration 1: estimate for effect of center j on CIF of cause $k = 1$, with bias, empirical standard deviation (ESD), bootstrap standard error (BSE), and 95% confidence interval coverage probabilities (CP). | 43 |
| 3.2 | Simulation results using Configuration 2: estimate for effect of center j on CIF of cause $k = 1$, with bias, empirical standard deviation (ESD), bootstrap standard error (BSE), and 95% confidence interval coverage probabilities (CP). | 44 |

| | | |
|-----|---|----|
| 4.1 | Simulation results for proposed estimator for five configurations, with bias, asymptotic standard error (ASE), empirical standard deviation (ESD), and 95% confidence interval coverage probabilities (CP). | 72 |
| 4.2 | Analysis of SRTR data using proposed method with $M = 10$ | 74 |

LIST OF APPENDICES

Appendix

| | | |
|----|---|-----|
| A. | Proof of Theorem 2.1 | 83 |
| B. | Analysis of SRTR data utilizing $J = 58$ OPOs for Chapter III | 102 |
| C. | Proof of Theorem 3.1 | 105 |

CHAPTER I

Introduction

Competing risks data arise naturally in many biomedical studies. Often the subject is at risk for many types of events; with the occurrence of one event precluding the subsequent occurrence of all other types of events. Some key quantities in the competing risks setting are the cumulative incidence function, the cause-specific hazard and, more recently, the subdistribution hazard. In many practical settings, one of the possible causes is of chief interest, although competing causes must be taken into account.

There are two frameworks applied to the competing risks setting. In one approach, a latent failure time exists for each event, though it may not ever be observed. The minimum of the latent event times is the observed event time and, consequently, also determines the observed event type. The quantities associated with this framework are identifiable if the latent failure times are independent. This framework has been referred to as the ‘alarm clock’ model, where many alarms are set, but the one that goes off first is the only one that is heard and observed.

In the second framework, the competing risks data are made up of an observation time and an event type indicator for each individual. If a particular event is observed to occur, the event times for the other causes are not unobserved; they simply do not exist for the other causes and are undefined. The functions involved in this approach are the cause-specific hazard (CSH) and the cumulative incidence function (CIF), known as the crude functions. For the crude functions to be identifiable, no assumptions need to be made about independent causes. In this dissertation we will adopt the second framework, with the CIF serving as the key function to compare groups and characterize covariate effects.

Suppose that cause k is the event type of interest. The cause-specific hazard can be thought of as the rate at which subjects die due to cause k , among those still at risk (alive). The cumulative incidence is the cumulative probability of cause k , acknowledging that death from other causes prevents the occurrence of cause k . In this proposal we develop methods for analyzing observational time-to-event data in the presence of competing risks. We propose two methods that compare the cumulative incidence functions among subgroups. The third method uses multiple imputation to model the subdistribution hazard through Cox regression (*Cox*, 1972).

It is often of interest to compare outcomes between subgroups of subjects. In the presence of observational data, group assignment is typically not randomized, so that adjustment must be made for differences in covariate distributions across groups. The second chapter proposes a measure to contrast group-specific CIFs. The proposed method assumes proportional cause-specific hazards, which are estimated through

Cox models stratified by organ procurement organizations (OPO). The effect measure compares the average CIF of each OPO to the average CIF that would have resulted if that particular OPO had cumulative incidence equal to that of the national average.

Like the second chapter, the third chapter aims to compare the CIFs among subgroups of subjects from an observational study. However, this section proposes a measure which, based on direct standardization, contrasts the population average cumulative incidence under two scenarios: (i) subjects are distributed across groups as per the existing population and (ii) all subjects are members of a particular group. The proposed comparison of CIFs has a strong connection to measures used in the causal inference literature. The proposed methods are nonparametric in the sense that no models are assumed for the cause-specific hazards or the subdistribution function. Observed event counts are weighted using Inverse Probability of Treatment Weighting (IPTW) and Inverse Probability of Censoring Weighting (IPCW).

The fourth chapter describes a multiple imputation method for competing risks data that creates a ‘censoring-complete’ dataset such that all the potential censoring times are known. For individuals who experienced a competing risk, we impute what their censoring times would have been had death not prevented their occurrence. Regression methods for the subdistribution hazard model developed by *Fine and Gray* (1999) can then be applied to the censoring-complete data. The censoring hazard is estimated by a Cox proportional hazards model, with the baseline hazard computed by a linear interpolation of the Breslow estimator, instead of the usual step

function. We use the subdistribution hazard to estimate the effect of explanatory variables on the outcome of interest.

For each of the proposed methods, large sample properties are derived and the finite-sample properties are evaluated using simulations. We apply each method to national kidney transplantation data from the Scientific Registry of Transplant Recipients (SRTR). A patient in need of a kidney transplant is placed on a wait-list, which are maintained by the OPOs. OPOs are geographically based, usually covering patients in an area smaller than or the size of a state. Each OPO maintains its own wait-list and is responsible for allocating organs to those on the wait-list as kidneys become available. There are 58 OPOs in the United States, which are in turn grouped into 11 Regions that may consist of a couple states. In this dissertation, we focus on Region 10, which includes Michigan and Ohio.

Patients on the wait-list can experience events that prevent them from ever receiving a transplant. In this case, the competing risks are death while on the wait-list and removal from the wait-list due to deterioration of patient health. If either of the competing risks is observed, the OPO has failed to provide the patient with a transplant in time, and thus analyses on such data must incorporate this information. Since OPOs represent different parts of the country, the patient characteristics, such as race and age, may vary quite a bit. Thus valid comparisons between OPOs require covariate adjustment.

CHAPTER II

Comparing Cumulative Incidence Functions Between Non-Randomized Groups

2.1 Introduction

In many studies with time to event data, comparisons among groups are of key interest. These groups can be defined by country, state, region or treatment center, for example. A natural grouping that often occurs in medical data is defined by treatment centers or hospitals. It is often of interest to study differences in outcomes by location, since geographic and center-based disparities can be present in health outcomes, and would be of concern in settings where uniformity in quality of care is a priority. There are a number of existing methods for estimating group or center effects. *Harrington and Fleming* (1982) developed a class of non-parametric methods based on the rank test that accommodates comparisons of more than two groups. *Dabrowska and Doksum* (1988) developed estimation and inference methods

for comparison of the generalized odds-rate model in the two-group setting. Most of these methods assume, either explicitly or implicitly, that all events are of the same cause.

Often in survival data, the terminating event is due to one of several possible causes. Experiencing a particular type of event precludes an individual from experiencing any other types of events. These events are said to be competing risks. An example of the competing risks setting would be time until death, where death is caused by cardiovascular disease, cancer, and all other causes. This data structure occurs frequently, and has to be explicitly acknowledged in models fitted to said data. Analyses that ignore the competing risks aspect of the data by not differentiating between the types of events may lose information on the covariate effects and lead to inaccurate interpretation of results. Several methods have been proposed for analysis of competing risks data. *Benichou and Gail* (1990) developed a method for estimating the absolute risk of an event during a time interval for a particular covariate pattern. The authors investigated the results of using the exponential, piecewise exponential, and Cox models (*Cox*, 1972) for the hazard functions. *Cheng et al.* (1998) investigated the prediction of cumulative incidence functions (CIF) and simultaneous confidence bands. *Fine and Gray* (1999) developed regression methods for the subdistribution hazard function.

There are few methods for comparing groups or centers in the presence of competing risks. *Gray* (1988) developed a method to compare CIFs among multiple groups, non-parametrically and unadjusted for covariates. *Zhang and Fine* (2008)

then summarized differences for several transformations of the CIF for two groups. We propose methods that compare centers in the presence of competing risks. The methods are targeted at observational data and, therefore, account for differences in group-specific covariate distributions.

The data that motivated the methods proposed in this chapter arise from the end-stage renal disease (ESRD) setting. A patient who experiences kidney failure typically begins receiving dialysis and, if medically suitable, the patient may subsequently be placed on a wait-list for deceased-donor kidney transplantation. Patients often have to wait to receive an organ transplant because there are not nearly enough donor kidneys to accommodate all wait-listed patients. For the purpose of solid organ transplantation, the United States is divided into 11 Regions, each subdivided into donation service areas. For each donation service area, an organ procurement organization (OPO) maintains a wait-list (i.e., each OPO has its own wait-list) and is responsible for allocating organs to patients within its corresponding area, based on several factors. We use the general terms group and center to indicate any factor that subdivides the study population. In the context of organ transplantation, the unit of our interest is an OPO. This setting corresponds to a competing risks framework because a patient on a wait-list can experience a number of other events that preclude a transplant. The competing risks are death while on wait-list and removal from the wait-list, which occurs when a patient's general health has declined so much that he/she is not expected to make a recovery from the surgery itself. Thus both events indicate that the OPO has failed to allocate an organ to the patient fast enough for the patient to survive.

To be wait-listed by an OPO, a patient often must be able to travel to the location at which the transplant will occur within a relatively short period of time after being notified that a particular organ is available. Thus, since most individuals do not have unlimited resources to travel long distances within short notice, where a given patient can be put on a wait-list is largely determined by where he/she resides. A natural question to ask is how each OPO performs relative to the national average. A natural quantity reflecting an OPO's performance is the average probability that a patient on a wait-list will receive a transplant, acknowledging that the patient may instead die while on, or be removed from, the wait-list. In this chapter, our goal is to propose a metric that will compare the average experience of wait-listed patients at an OPO to the average experience that would have been observed if that particular OPO performed at the national average. We are specifically focused on the probability that a patient receives a transplant. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR), a national population-based organ transplant registry.

There are two frameworks for casting the event times in a competing risks setting. The first method is based on latent failure times (*Gail*, 2001, 1975; *Crowder*, 2001), in which there exists a latent event time for every cause, but only the minimum of the latent event times is observed (*Cox*, 1959; *Moeschberger and David*, 1971). For marginal quantities (such as the failure-type-specific survival functions) to be identifiable under the latent failure times set up, each of the latent failure times must act independently.

The second framework assumes there is only one event time for each subject, with the event occurring from one of two or more causes. Under this framework, there only exists one failure time, due to one cause. For example, if a particular type of event occurs, then the event times due to other causes are undefined. The key functions that arise from this framework are the crude functions known as cause-specific hazard (CSH) and the cumulative incidence function (CIF) (*Anderson et al.*, 1993; *Chiang*, 1968; *Kalbfleisch and Prentice*, 2002). These models do not require the assumption that the causes of death are independent (*Tsiatis*, 1975).

In this chapter, we propose methods that contrast centers with respect to cumulative incidence. We model the cause-specific hazards explicitly through *Cox* (1972) regression, with the CSHs then combined and integrated, then transformed to obtain the CIF. The effect measure, or, the effect of a particular OPO on a specific cause, is obtained through appropriately averaging fitted values. The event times are subject to right censoring, which is assumed to be independent of the event time given the covariates.

In Section 2.2, we introduce the notation and present the proposed method and estimation procedures. Section 2.3 describes the asymptotic properties of the proposed estimators. In Section 2.4 we evaluate the finite sample performance of the proposed estimators in simulation studies. We return to the motivating example in Section 2.5, applying the proposed methods to data from the SRTR to compare OPOs across the United States. Section 2.6 concludes with a discussion.

2.2 Proposed Methods

Let T_i and C_i be the event and censoring times, respectively, for individual i ($i = 1, \dots, n$). Center and cause will be denoted by j ($j = 1, \dots, J$) and k ($k = 1, \dots, K$), respectively. There are n individuals in the entire sample, with n_j individuals in center j . For concreteness, we refer to the factor of interest as center, although in practice this factor could be any categorical covariate defining subgroups of individuals. Let $\Delta_i = kI(T_i < C_i)$, where $I(\cdot)$ is the indicator function. A_i ($A_i = 1, \dots, J$) is the center that subject i belongs to and $A_{ij} = I(A_i = j)$ is equal to 1 if subject i is in center j . The observation time is defined as $X_i = T_i \wedge C_i$. Thus, the observed data consist of $\{X_i, \Delta_i, \mathbf{Z}_i, A_i\}$, where \mathbf{Z}_i is the vector of covariates which is assumed to be time constant. The at-risk indicator is given by $Y_{ij}(t) = I(X_i \geq t, A_i = j)$ and the counting process for subject i in center j , cause k is denoted by $N_{ijk}(t) = I(X_i \leq t, A_i = j, \Delta_i = k)$. Let the CIF for cause k for individual i , at center j be denoted by

$$F_{ijk}(t) \equiv P(T_i \leq t, \Delta_i = k | A_i = j, \mathbf{Z}_i),$$

interpreted as the probability that individual i in center j experiences an event of type k by time t .

Our goal is to contrast the average CIF for center j with an appropriate average across all centers. Along those lines, we define the effect of center j on cause k

as

$$\delta_{jk}(t) = E_{Z|A}[F_{ijk}(t)] - E_{Z|A}[E_A[F_{iAk}(t)]], \quad (2.1)$$

which is intended to reflect the impact of center j on the CIF of cause k . The effect measure $\delta_{jk}(t)$ compares, through the CIF, two scenarios: (i) the observed reality: all individuals assigned to center j are treated at center j and (ii) all individuals assigned to center j are instead treated at a hypothetical center with cumulative incidence equal to those of the national average.

The outer expectation in both the first and second terms on the right side of (2.1) are evaluated with respect to the distribution of $[Z|A = j]$. The quantity $E_{Z|A}[F_{ijk}(t)]$ is the CIF for group j , averaged over the covariate distribution of individuals in center j . The inner expectation in the second term is evaluated with respect to the marginal distribution of A . The term $E_A[F_{iAk}(t)]$ is the CIF if individual i , who is actually at center j , were at the hypothetical national center. This national average center does not actually exist, but individual i 's experience at this hypothetical center can be obtained by considering the CIFs for individual i at all of the centers and the likelihood of being in a particular center. The CIF if individual i , known to be in center j , were actually at center l is denoted by $F_{ilk}(t)$. By considering $F_{ilk}(t)$ for $l = 1, \dots, J$, we consider what would have happened to individual i at each of the actual centers in the study sample. An individual is defined by his/her covariate pattern.

One option for obtaining the CIFs is to use the cause-specific hazards. The cause k

CIF for individual i at center j is

$$F_{ijk}(t) = \int_0^t S_{ij}(s) \lambda_{ijk}^{\#}(s) ds, \quad (2.2)$$

where $S_{ij}(t) = P(T_i > t | \mathbf{Z}_i, A_i = j)$ is the survival function for individual i at center j , and the cause-specific hazards are given by

$$\lambda_{ijk}^{\#}(t) = \lim_{\epsilon \downarrow 0} P(t \leq T_i < t + \epsilon, \Delta_i = k | T_i \geq t, \mathbf{Z}_i, A_i = j).$$

Note that $F_{ijk}(t)$ can be written entirely in terms of the cause-specific hazard, since $S_{ij}(t) = \exp\{-\sum_{k=1}^K \Lambda_{ijk}^{\#}(t)\}$, where the cumulative CSH for individual i at center j is $\Lambda_{ijk}^{\#}(t) = \int_0^t \lambda_{ijk}^{\#}(s) ds$. Also called the subdistribution function, $F_{ijk}(t)$ is the probability that subject i experiences an event of type j by time t , acknowledging that he/she could experience another event first, which would preclude event k from happening. From (2.2), it can be seen that the CIF is a function of all of the CSHs, so that the CIF acknowledges the presence of other types of events. As $t \rightarrow \infty$, $F_{ijk}(t) \rightarrow P(\Delta_i = k | \mathbf{Z}_i, A_i = j)$. Therefore, unlike the cumulative distribution function in the single cause setting, the CIF will generally not approach 1 if the competing causes have non-zero probabilities.

In this chapter, Cox regression (Cox, 1972) stratified on center is assumed to relate the covariates to the cause-specific hazards,

$$\lambda_{ijk}^{\#}(t) = \lambda_{0jk}(t) \exp\{\boldsymbol{\beta}_k^T \mathbf{Z}_i\}. \quad (2.3)$$

The Cox model is selected due to its flexibility and popularity. Stratifying on center adjusts for center effects non-parametrically. As such, we are assuming proportionality with respect to the cause k hazard at time t among subjects alive at time t with respect to the adjustment covariates, \mathbf{Z}_i , although not the centers.

A few comments on model (2.2) are in order. First, the stratification by center would make it difficult to estimate the effect of center j on $\lambda_{ijk}^\#(t)$ based on model (2.2) alone. However, as defined previously, the cause-specific hazards are not of primary interest per se, and are useful only through their connection to $F_{ijk}(t)$. Second, the covariate vector is assumed to be constant. If not, accurate estimation of $F_{ijk}(t)$ would be substantially more complicated, and the methods proposed in this chapter would not be recommended. Third, covariate effects, β_k , are allowed to vary by cause but are assumed to be constant over time. This could be relaxed through additional stratification, or by parametrically modeling covariate effects that are time-dependent; e.g., through the Cox non-proportional hazards model; see *Klein and Moeschberger* (2003).

We estimate $\delta_{jk}(t)$ by using the finite sample estimators for the quantities involved. The covariate effects β_k are estimated through partial likelihood (*Cox*, 1975), while the cumulative baseline cause-specific hazards $\Lambda_{0jk}^\#(t)$ are estimated by the Breslow estimator (*Breslow*, 1972). Referring back to (2.1), the quantity $E_{Z|A}[F_{ijk}(t)]$ can be estimated by taking the average of $\hat{F}_{ijk}(t)$ with respect to the empirical distribution of $[\mathbf{Z}_i|A_i = j]$. The CIF for individual i at the national average center is estimated by taking a weighted average of the CIFs for individual i across centers, expressed

as $\sum_{l=1}^J \hat{F}_{ilk}(t)\hat{p}_l$, with $\hat{p}_l = n_l/n$ an estimator for $p_l = P(A_i = l)$. Combining these estimators, we then estimate the proposed center effect measure by

$$\hat{\delta}_{jk}(t) = \frac{1}{n_j} \sum_{i=1}^n I(A_i = j) \hat{F}_{ijk}(t) - \frac{1}{n_j} \sum_{i=1}^n I(A_i = j) \left\{ \sum_{l=1}^J \hat{F}_{ilk}(t) \hat{p}_l \right\}, \quad (2.4)$$

where $\hat{F}_{ijk}(t) = \int_0^t \exp\{-\sum_{k=1}^K \hat{\Lambda}_{ijk}^\#(s)\} d\hat{\Lambda}_{ijk}^\#(s)$ and $\hat{\Lambda}_{ijk}^\#(t) = \int_0^t \exp(\hat{\beta}_k^T \mathbf{Z}_i) d\hat{\Lambda}_{0jk}^\#(s)$. Note that, in cases where $A_i \neq l$, the cause-specific hazard function of individual i at center l is estimated using the baseline hazard from center l with the covariates of individual i ; i.e., $\hat{\Lambda}_{ilk}^\#(t) = \int_0^t d\hat{\Lambda}_{0lk}^\#(s) \exp(\hat{\beta}_k^T \mathbf{Z}_i)$. The center to which a subject actually belongs is accounted for in the model fitting. Once the regression parameters and cause-specific cumulative baseline hazards are estimated, subject i 's actual center is not factored into the averaging in the second term on the right side of (2.4). These methods are valid under independent censoring, which can be formally written as

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} P(t \leq T_i < t + \epsilon, \Delta_i = k | T_i > t, C_i > t, \mathbf{Z}_i, A_i) \\ & = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} P(t \leq T_i < t + \epsilon, \Delta_i = k | T_i > t, \mathbf{Z}_i, A_i) \end{aligned}$$

for $k = 1, \dots, K$.

We now describe the properties of the proposed estimators when n becomes large.

2.3 Asymptotic Properties

We begin by listing the assumed regularity conditions for $i = 1, \dots, n$, $j = 1, \dots, J$, and $k = 1, \dots, K$.

- (a) $\{X_i, \Delta_i, \mathbf{Z}_i, A_i\}$ are independent and identically distributed.
- (b) $P(Y_{ij}(\tau) = 1) > 0$.
- (c) $P(A_i = j | \mathbf{Z}_i) > 0$.
- (d) $|Z_{iq}| \leq K$, where \mathbf{Z}_{iq} is the q th element of \mathbf{Z}_i and K is a constant.
- (e) $\int_0^\tau \lambda_{0jk}^\#(t) dt < \infty$.
- (f) Continuity of the following functions:

$$\mathbf{r}_{jk}^{(d)}(t; \boldsymbol{\beta}) = E \left[Y_{ij}(t) \mathbf{Z}_i^{\otimes d} \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \right], \quad d = 0, 1, 2.$$

with $\mathbf{r}_{jk}^{(d)}(t; \boldsymbol{\beta})$ bounded away from 0 for $t \in (0, \tau]$.

- (g) Positive-definiteness of the following matrices:

$$\mathbf{I}_k(\boldsymbol{\beta}_k) = E \left[\sum_{j=1}^J \int_0^\tau \left\{ \frac{\mathbf{r}_j^{(2)}(t; \boldsymbol{\beta}_k)}{r_j^{(0)}(t; \boldsymbol{\beta}_k)} - \bar{\mathbf{z}}_j(t; \boldsymbol{\beta}_k)^{\otimes 2} \right\} dN_{ijk}(t) \right],$$

where $\bar{\mathbf{z}}_j(t; \boldsymbol{\beta}) = \mathbf{r}_{jk}^{(1)}(t; \boldsymbol{\beta})^{-1} r_{jk}^{(0)}(t; \boldsymbol{\beta})$.

Condition (a) permits fairly standard applications of the Central Limit Theorem. Conditions (b) and (c) ensure identifiability, with (c) being the positivity assumption familiar to causal inference methodology. Condition (e) ensures boundedness of many integrals arising in the asymptotic development. The framework we work with has $n \rightarrow \infty$ with the number of centers J remaining constant. In this case, $n_j \rightarrow \infty$ for all j as $n \rightarrow \infty$.

Theorem 2.1: *Under Conditions (a) through (g), $\hat{F}_{ijk}(t)$ converges almost surely to $F_{ijk}(t)$ and $\hat{\delta}_{jk}(t)$ converges almost surely to $\delta_{jk}(t)$.*

$\hat{F}_{ijk}(t)$ is a functional of consistent estimators $\hat{\beta}_k$ and $\hat{\Lambda}_{0jk}^\#(t)$. The proof of consistency then follows from successive applications of the Continuous Mapping Theorem.

Theorem 2.2: *Under conditions (a) through (g), $n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \}$ converges asymptotically to a zero-mean Gaussian process with variance function $\sigma_{jk}(t) = E [\psi_{mjk}(t; \beta)^2]$ where $\psi_{mjk}(t; \beta) = \sum_{d=1}^4 \{ \phi_{ijk}^d(t; \beta) - \phi_{ik}^d(t; \beta) \} + o_p(1)$, with $\beta^T =$*

$[\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_k^T]$ and $\phi_{lk}^d(t; \boldsymbol{\beta}) = E_A [\phi_{ljk}^d(t; \boldsymbol{\beta})]$ and where we define

$$\begin{aligned}\phi_{ljk}^1(t; \boldsymbol{\beta}) &= -\frac{1}{p_j} E \left[\sum_{m=1}^K A_{ij} \int_0^t \mathbf{D}_{ijm}^T(s; \boldsymbol{\beta}_m) dF_{ijk}(s) \right] \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} \mathbf{U}_{lm}(\boldsymbol{\beta}_m), \\ \phi_{ljk}^2(t; \boldsymbol{\beta}) &= -\sum_{m=1}^K \int_0^t \frac{1}{p_j} E \left[A_{ij} \{F_{ijk}(t) - F_{ijk}(u)\} Y_{ij}(u) \exp(\boldsymbol{\beta}_m^T \mathbf{Z}_i) \mathbf{r}_{jk}^{(0)}(u; \boldsymbol{\beta}_m)^{-1} \right] \\ &\quad dM_{ijm}(u; \boldsymbol{\beta}_m), \\ \phi_{ljk}^3(t; \boldsymbol{\beta}) &= \frac{1}{p_j} E \left[A_{ij} \int_0^t [\mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta})]^T dF_{ijk}(s) \right] \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} \mathbf{U}_{lk}(\boldsymbol{\beta}_k), \\ \phi_{ljk}^4(t; \boldsymbol{\beta}) &= \int_0^t \frac{1}{p_j} E \left[A_{ij} S_{ij}(s; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \mathbf{r}_{jk}^{(0)}(s; \boldsymbol{\beta}_2)^{-1} \right] dM_{ljk}(s; \boldsymbol{\beta}_k),\end{aligned}$$

where $\mathcal{I}_k(\boldsymbol{\beta}_k)$ and $\bar{\mathbf{z}}_j(t; \boldsymbol{\beta})$ are defined in Condition (g), $\mathbf{r}_{jk}^{(d)}$ defined in Condition (f), and

$$\begin{aligned}\mathbf{D}_{ijk}(t; \boldsymbol{\beta}_k) &= \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k)\} d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k), \\ \mathbf{U}_{ik}(\boldsymbol{\beta}_k) &= \sum_{j=1}^J \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(t; \boldsymbol{\beta}_k)\} dM_{ijk}(t), \\ dM_{ijk}(s; \boldsymbol{\beta}_k) &= dN_{ijk}(s) - Y_{ij}(s) d\Lambda_{ijk}^\#(s).\end{aligned}$$

The martingales $M_{ijk}(t; \boldsymbol{\beta}_k)$ are defined with respect to the joint filtration $\mathcal{F}_{ij}(t) = \sigma \{N_{ij}(s), Y_{ij}(s), \mathbf{Z}_i; s \in (0, t)\}$, where $N_{ij}(s) = I(X_i \leq s, A_i = j)$

The asymptotic variance can be estimated by:

$$\widehat{Var} \left[n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\} \right] = \frac{1}{n} \sum_{l=1}^n \left[\sum_{d=1}^4 \left\{ \hat{\phi}_{ljk}^d(t; \boldsymbol{\beta}) - \hat{\phi}_{lk}^d(t; \boldsymbol{\beta}) \right\} \right]^2,$$

where the $\hat{\phi}_{ljk}(t)$ are obtained by replacing limiting values in $\phi_{ljk}(t)$ with their empirical counterparts.

As implied by the formulas in Theorem 2.2, calculation of the asymptotic variance is cumbersome computationally. As a result, we propose the bootstrap, which we evaluate through simulation in the next section.

2.4 Simulation Studies

We conducted simulation studies to evaluate the finite sample performance of the proposed estimator $\hat{\delta}_{jk}(t)$. In the first simulation study, there are $K = 2$ causes, with cause $k = 1$ being of primary interest. There are $J = 5$ centers, with varying center sizes of $n_j = 100, 125, \text{ and } 150$ individuals. There are three covariates, all of which are binary and hierarchically dependent. The covariate Z_{i1} follows a Bernoulli distribution with $p = \theta_{1j}$, $Z_{i2}|Z_{i1}$ is distributed as a Bernoulli with $P(Z_{i2} = 1|Z_{i1}) = Z_{i1}\theta_{21} + (1 - Z_{i1})\theta_{22}$, and $Z_{i3}|Z_{i2}$ follows a Bernoulli distribution with $P(Z_{i3} = 1|Z_{i2}) = Z_{i2}\theta_{31} + (1 - Z_{i2})\theta_{32}$. The covariate patterns varied for each center due to center-specific θ_{1j} . In these set of simulations, $\theta_{1j}^T = [0.55, 0.75, 0.6, 0.65, 0.5]$, $\theta_{21} = 0.55$, $\theta_{22} = 0.45$, $\theta_{31} = 0.45$, and $\theta_{32} = 0.65$.

The event time for cause one T_{ij1} and the event time for cause two T_{ij2} follow exponential distributions, with

$$\lambda_{ijk}^{\#}(t) = \lambda_{0jk} \exp\{\beta_{k1}Z_{i1} + \beta_{k2}Z_{i2} + \beta_{k3}Z_{i3}\}$$

for $k = 1, 2$. The baseline hazards $\lambda_{0jk}(t)$ vary across centers and causes. For both configurations, the covariate effect for cause 1 is $\beta_1^T = [0.4, 0.5, 0.6]$ and for cause 2 is $\beta_2^T = [-0.1, 0.3, -0.2]$. The baseline cause-specific hazards are shown in Table 2.1.

Table 2.1: Baseline cause-specific hazards for the simulation study

| Configuration 1 | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ |
|-----------------|---------|---------|---------|---------|---------|
| λ_{0j1} | 0.1 | 0.15 | 0.2 | 0.22 | 0.7 |
| λ_{0j2} | 0.12 | 0.1 | 0.08 | 0.09 | 0.08 |
| Configuration 2 | | | | | |
| λ_{0j1} | 0.1 | 0.8 | 0.2 | 0.76 | 0.4 |
| λ_{0j2} | 0.12 | 0.1 | 0.08 | 0.09 | 0.08 |

The censoring distribution C_i is also exponential and also depends on the covariates, with

$$\lambda_i^C(t) = \lambda_0^C \exp\{\alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \alpha_3 Z_{i3}\}.$$

The distribution of C_i differs for each center due to the different covariate patterns. There is administrative censoring at $t = 10$ to reflect the real life scenario of a finite observation period. The censoring baseline hazard is $\lambda_0^C = 0.02$.

The observation time, X_i , is defined as the minimum of T_{i1} , T_{i2} , and C_i and whichever is the minimum will determine the value of Δ_i . Although we are using the latent failure time model to generate the data, analysis of the simulated data will not take that perspective. For example, the analysis stills assumes that T_{i1} does not exist if

$$T_{i1} > T_{i2}.$$

We used the bootstrapping method to estimate the variability of the effect, with standard errors calculated using 25 bootstrap samples. The results based on 500 replicates of Configuration 1 and 2 are shown in Tables 2.2 and 2.3, respectively. We calculated the estimated effect at the times $t = 1, t = 3$, and $t = 5$ after the start of follow-up. In Configuration 1, the true effect varied greatly from having a large negative effect, -0.243 to a large positive effect, 0.369. Given the inherent bounds of 0 and 1 on cumulative incidence, and thus bounds of -1 and 1 on the difference of cumulative incidence, a 30 percent displacement represents a rather large effect. Centers $j = 1$ and $j = 2$ are smaller centers that perform worse than expected and center $j = 4$ is a larger center that performs better than expected.

The proposed estimators $\hat{\delta}_{jk}(t)$ have small bias, even when the size of the center is small. As expected, the larger the center, in general the smaller the bias of $\hat{\delta}_{jk}(t)$. Within a center, the bias tends to be smaller at the earlier times than at later times. This could be due to the fact that there are more individuals at risk at the earlier times. The bootstrap standard errors (BSE) were close to the empirical standard deviations (ESD). Coverage probabilities (CP) are mostly just under 95 percent. Larger centers do not necessarily have higher CPs than smaller centers, which is somewhat counter-intuitive.

In Configuration 2, we investigated the behavior of the proposed estimator in the presence of smaller center sizes. In Table 2.3, center $j = 1$ is a small center that has negative true effects of high magnitude, both factors that may hinder the estimator

Table 2.2: Performance of $\hat{\delta}_{j1}(t)$ based on 500 simulations of Configuration 1, with bias, empirical standard deviation (ESD), the bootstrap standard error (BSE), and the 95% confidence interval coverage probabilities (CP).

| j | n_j | t | $\delta_{j1}(t)$ | Bias | ESD | BSE | CP |
|-----|-------|-----|------------------|--------|-------|-------|------|
| 1 | 100 | 1 | -0.205 | 0.015 | 0.037 | 0.037 | 0.93 |
| | | 3 | -0.243 | 0.016 | 0.041 | 0.044 | 0.94 |
| | | 5 | -0.221 | 0.018 | 0.040 | 0.045 | 0.94 |
| 2 | 100 | 1 | -0.101 | -0.009 | 0.043 | 0.043 | 0.92 |
| | | 3 | -0.078 | -0.015 | 0.049 | 0.050 | 0.93 |
| | | 5 | -0.056 | -0.022 | 0.047 | 0.052 | 0.93 |
| 3 | 125 | 1 | -0.040 | 0.007 | 0.036 | 0.037 | 0.95 |
| | | 3 | 0.004 | 0.008 | 0.035 | 0.036 | 0.93 |
| | | 5 | 0.021 | 0.009 | 0.031 | 0.033 | 0.94 |
| 4 | 150 | 1 | -0.009 | -0.000 | 0.032 | 0.033 | 0.95 |
| | | 3 | 0.032 | 0.003 | 0.031 | 0.031 | 0.93 |
| | | 5 | 0.038 | 0.004 | 0.028 | 0.029 | 0.95 |
| 5 | 100 | 1 | 0.369 | -0.011 | 0.035 | 0.038 | 0.96 |
| | | 3 | 0.268 | -0.019 | 0.027 | 0.029 | 0.92 |
| | | 5 | 0.194 | -0.019 | 0.026 | 0.028 | 0.92 |

Table 2.3: Performance of $\hat{\delta}_{j1}(t)$ based on 500 simulations of Configuration 2, with bias, empirical standard deviation (ESD), the bootstrap standard error (BSE), and the 95% confidence interval coverage probabilities (CP).

| j | n_j | t | $\delta_{j1}(t)$ | Bias | ESD | BSE | CP |
|-----|-------|-----|------------------|--------|-------|-------|------|
| 1 | 50 | 1 | -0.379 | 0.014 | 0.051 | 0.053 | 0.95 |
| | | 3 | -0.379 | 0.008 | 0.063 | 0.064 | 0.94 |
| | | 5 | -0.323 | 0.003 | 0.059 | 0.066 | 0.97 |
| 2 | 100 | 1 | 0.231 | -0.016 | 0.034 | 0.037 | 0.94 |
| | | 3 | 0.137 | -0.024 | 0.023 | 0.028 | 0.92 |
| | | 5 | 0.093 | -0.021 | 0.022 | 0.027 | 0.93 |
| 3 | 125 | 1 | -0.215 | 0.012 | 0.036 | 0.035 | 0.93 |
| | | 3 | -0.133 | 0.014 | 0.032 | 0.034 | 0.92 |
| | | 5 | -0.080 | 0.013 | 0.027 | 0.031 | 0.94 |
| 4 | 100 | 1 | 0.207 | -0.002 | 0.036 | 0.037 | 0.95 |
| | | 3 | 0.134 | -0.008 | 0.025 | 0.027 | 0.96 |
| | | 5 | 0.093 | -0.007 | 0.023 | 0.025 | 0.96 |
| 5 | 150 | 1 | 0.014 | -0.001 | 0.032 | 0.031 | 0.94 |
| | | 3 | 0.057 | 0.001 | 0.023 | 0.024 | 0.96 |
| | | 5 | 0.050 | 0.001 | 0.021 | 0.022 | 0.95 |

by reducing the number of type 1 events. Having only $n_j = 50$ individuals while the other four centers have at least twice as many individuals, center $j = 1$ not only has a small center in absolute terms, but also in relative terms. The bias of center 1 is still quite small at all of the three studied time points and its coverage probability is close to the nominal value. Thus, even for smaller centers that perform worse than expected, the proposed estimator performs quite well.

2.5 Application

We applied the proposed methods to compare, by OPO, the probability of receiving a kidney transplant among patients wait-listed for kidney transplantation. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). We selected patients from OPOs that make up Region 10 and who were wait-listed between January 1, 2000 and December 31, 2009, with a resulting sample size of $n = 23,490$ from $J = 6$ centers. Each patient's follow-up started when the patient was put on the wait-list. Follow-up ended at the earliest of receiving a transplant, death on wait-list, removal from wait-list, or loss to follow-up. Since we were interested in evaluating the OPOs based on their ability to ensure as many patients as possible receive the preferred treatment of receiving a deceased-donor transplant, transplantation was the cause of interest, with deaths and removals treated as competing risks. We focused on three time points, years 1, 3, and 5. This reflects current practice since survival statistics are usually reported at chosen year intervals rather than on a daily or a monthly basis.

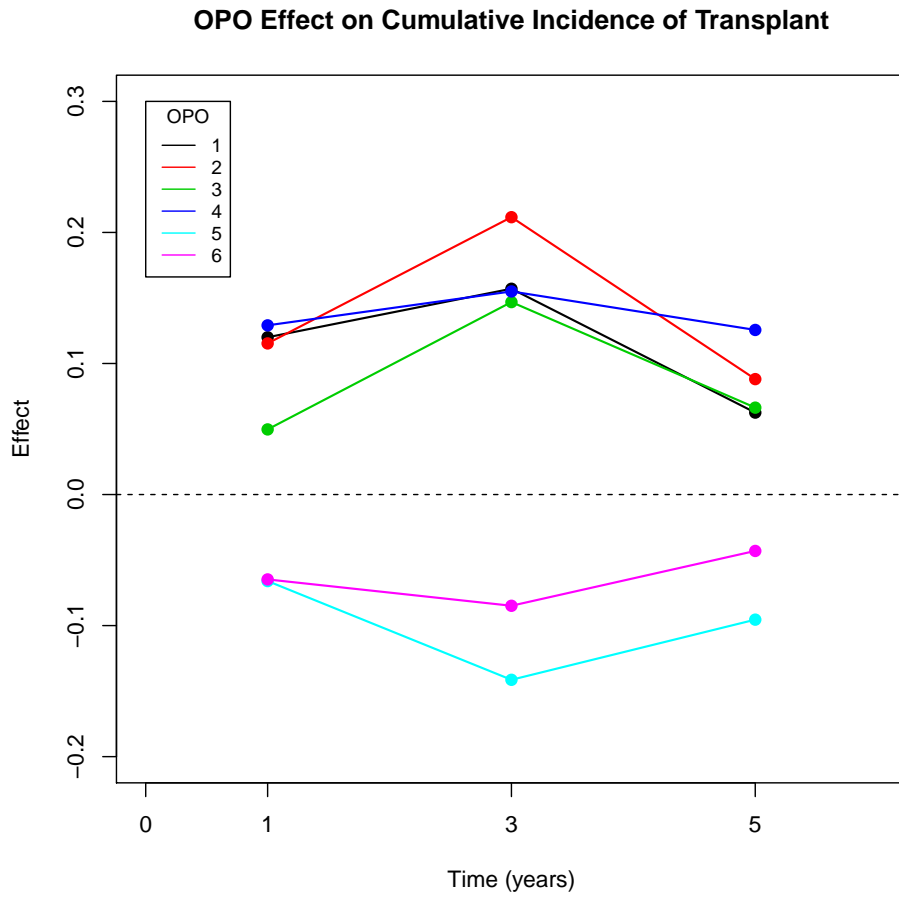


Figure 2.1: Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ for 1, 3, and 5 years post wait-listing, for $j = 1, \dots, 6$.

Table 2.4: Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ with 95% confidence limits for 1, 3, and 5 years post wait-listing, for $j = 1, \dots, 6$.

| year | OPO | $\hat{\delta}_{j1}(t)$ | upper 95% confidence limit | lower 95% confidence limit |
|------|-----|------------------------|-------------------------------|-------------------------------|
| 1 | 1 | 0.12 | 0.09 | 0.15 |
| 3 | 1 | 0.16 | 0.12 | 0.20 |
| 5 | 1 | 0.06 | 0.02 | 0.11 |
| 1 | 2 | 0.12 | 0.09 | 0.14 |
| 3 | 2 | 0.21 | 0.18 | 0.24 |
| 5 | 2 | 0.09 | 0.05 | 0.12 |
| 1 | 3 | 0.05 | 0.03 | 0.07 |
| 3 | 3 | 0.15 | 0.12 | 0.17 |
| 5 | 3 | 0.07 | 0.04 | 0.09 |
| 1 | 4 | 0.13 | 0.11 | 0.15 |
| 3 | 4 | 0.15 | 0.13 | 0.18 |
| 5 | 4 | 0.13 | 0.10 | 0.15 |
| 1 | 5 | -0.07 | -0.08 | -0.06 |
| 3 | 5 | -0.14 | -0.16 | -0.12 |
| 5 | 5 | -0.10 | -0.12 | -0.07 |
| 1 | 6 | -0.06 | -0.07 | -0.06 |
| 3 | 6 | -0.08 | -0.10 | -0.07 |
| 5 | 6 | -0.04 | -0.06 | -0.03 |

A figure of the estimated effects $\hat{\delta}_{jk}(t)$ for $k = 1$ at the three time points is shown in Figure 2.1. We would have expected to see that as follow-up time increased, the spread of the estimated effects increased as well. This is because we believe that earlier differences in performance are less likely to be attributable to the OPO than are later differences. In the earlier period of follow-up, other factors independent of the OPO, such as the patient's inherent overall health may affect his/her chance of surviving until a transplant becomes available. It appears that even one year after the start of follow-up, there is differentiation between the OPOs that perform better than expected and ones that perform worse than expected. The differences between the OPOs widen by year three, but decrease by year five. The 95% confidence intervals for $\hat{\delta}_{jk}(t)$ are shown in Table 2.4.

2.6 Discussion

We have proposed a summary measure that quantifies the center effect in terms of CIF. By averaging over transformed fitted values obtained by Cox models and transformations, we compare the patient experience under two scenarios, one actual and one hypothetical. The proposed method would allow one to determine which groups of patients are at greater or lesser probability of experiencing the event of interest. In the context of evaluating centers, a center could compare its actual performance to that if it were performing at the overall average. The proposed effect measure has negligible bias. Although calculating the theoretical asymptotic standard error is cumbersome, the bootstrap standard error appeared to be fairly

accurate, as evidenced by comparison to the empirical standard errors and coverage probabilities.

The proposed methods were applied to national transplant registry data to evaluate OPOs with respect to average probability of receiving a kidney transplant. From the perspective of each individual OPO, an estimate of the cumulative incidence of transplantation is a meaningful metric of quality of delivered care. It answers the question, how would these same patients have done elsewhere, on average? It gives an OPO valuable information since the evaluation is done based on that particular OPO's case mix. Applied to the country, this method would allow us to see which parts of the country are under-served or well-served, taking into account that the profiles of wait-listed patients of each OPO can be different. Each effect estimate is specific to an OPO's demographics, so the effect estimates from different OPOs generally cannot be compared to each other meaningfully.

The OPO is responsible for allocating organs to patients on a wait-list for organ transplants. It would be considered optimal if the greatest number of patients eventually receive an organ before an event can occur that prevents a transplant. Thus, the performance of an OPO is crucial for the patients whose health depends on these time sensitive transplants.

We constructed a measure that uses the CIF to quantify the center effect. An alternative to modeling the CIF is to base the effect measure on the cumulative cause-specific hazard. The CIF incorporates information from all causes to give a natural interpretation in the competing risks setting. In our motivating example,

we want to take into account the entire patient experience, and not just his/her transplant experience, to determine the probability of receiving a transplant. For example, patients in a particular OPO may be getting transplanted at a faster rate while alive, but are also dying on the wait-list at a faster rate. We would want to use both pieces of information in the comparison. However, if interest lies in estimating the event rate of one specific cause without the input of the other causes which are not of direct interest, then the CSH is more apt. For example, the cumulative CSH would be more appropriate for comparing the rate of transplants among surviving patients.

CHAPTER III

Comparing Cumulative Incidence Functions using Weighted Counting Processes

3.1 Introduction

In observational studies, the goal is often to compare groups with respect to a certain outcome or measure. Such groups can be defined by any categorical variable such as sex, age group, race, and geographic area. Since these groups are not assigned using a randomization scheme, but rather are observed, adjustments need to be made to account for the potentially disparate covariate distributions across groups.

Time-to-event data are often of interest in observational studies and, within this framework, competing risks survival analysis is often appropriate. In the competing risks setting, a subject may experience one of many outcomes, where the occurrence of any one type of event prevents all others from happening. In practice, investigators may be only interested in the relationship between the group and one particular

health outcome, even in settings where the event of interest is just one of many for which each subject is at risk. A function which is often of primary interest in the competing risks setting is the cumulative incidence function (CIF), which reflects the cumulative probability that the event of interest occurs, explicitly acknowledging that the occurrence of one of the competing causes precludes subsequent occurrence of the cause of interest.

The motivating example concerns the evaluation of organ procurement organizations (OPO) with respect to kidney transplantation. A patient in need of a kidney transplant and deemed medically suitable will typically have to wait until an organ becomes available by being on a wait-list. There are 58 OPOs in the United States, with each responsible for administering a wait-list and executing the allocation of available deceased-donor organs procured from a geographically defined donation service area. An OPO usually serves a state or an area smaller than a state. A patient on the wait-list can receive a transplant, or die, or be removed from the waiting list. The latter two outcomes are the competing risks because, if death or removal occurs, the patient cannot subsequently receive a transplant. Patients are removed from the wait-list often because their general health has declined to the extent that they are not likely to survive the transplantation surgery. The distribution of patient characteristics across OPOs may differ considerably due to differences in the underlying demographics across the United States. Since different demographics can contribute to differences in death rates while on the wait-list, confounding is likely to be present unless differences in the covariate distribution among OPOs are taken into account.

Ultimately, an OPO aims to maximize the number of transplants among patients on its wait-list, with the underlying goal of minimizing the number of deaths among patients in its donation service area. A meaningful comparison may be how each OPO performs relative to the national average, with performance quantified by the likelihood that a patient's eventual outcome is receiving a transplant. The CIF captures the probability that a transplant occurs, accounting for the possibility of death and removal from the wait-list. This chapter develops methods for comparing centers, but using center as a grouping variable is only for concreteness. Any categorical variable can be used to create groups within the study population.

Competing risks and the cumulative incidence function have been the focus of a large number of previous reports. For example, *Cheng et al.* (1998) developed methods to predict the CIF in the case of dependent competing risks and proportional hazards. Various methods for competing risks are explored by *Pepe and Mori* (1993), who make a strong case for using the CIF to make treatment decisions. *Zhang and Fine* (2008) presented non-parametric inference procedures for medically useful functions of the CIF, such as the difference, relative risk, and odds ratio between two groups. *Scheike et al.* (2008) developed a method that allows the effect of covariates on the CIF to be directly assessed. Their semi-parametric model also incorporates time dependent effects.

There are few methods for comparing groups or centers in the presence of competing risks, adjusted for covariates. *Gray* (1988) developed a non-parametric test for group differences in CIFs for one specific cause of failure. Gray makes the comparison

between the CIF for a particular group and an unspecified distribution. *Lin* (1997) developed non-parametric methods for estimating CIFs and proposed resampling-based procedures to construct confidence bands and tests of comparison. *Fine and Gray* (1999) introduced regression methods for the hazard function corresponding to the CIF. Previously, most of the competing risks methodology was focused on the cause-specific hazards and, if ultimate interest lies in the CIFs, then inference would be carried out by integrating the cause-specific hazard functions and the survival function. There is also some existing work comparing the CIFs of different causes, such as tests by *Aly et al.* (1994). However, in this chapter the objective is to estimate group effects with respect to CIFs of the same cause; hence, the nature of the comparison is quite different from that of interest to *Aly et al.* (1994).

This chapter proposes an effect estimate to contrast the population average cumulative incidence under two scenarios: (i) subjects are distributed across groups as per the existing population and (ii) all subjects are members of a particular group. In terms of the target quantity, the methods are akin to direct standardization. The CIF is estimated non-parametrically, using counting processes instead of via modeling the cause-specific hazards. Inverse Probability of Treatment Weighting (IPTW) and Inverse Probability of Censoring Weighting (IPCW) are employed to create comparable empirical covariate distributions across the centers and to account for censoring, respectively.

The remainder of the chapter consists of the formalization of the afore-listed data elements and description of the proposed method in Section 3.2. Results from simu-

lation studies are presented in Section 3.3 to evaluate the finite-sample performance. Section 3.4 includes results from the application of the proposed method to address the motivating question.

3.2 Proposed Methods

Let T_i and C_i be the event time and censoring time, respectively. The minimum of T_i and C_i is the observation time, X_i . Individuals are denoted by $i = 1, \dots, n$, centers by $j = 1, \dots, J$, and cause by $k = 1, \dots, K$. The observed event indicator is $\Delta_i = kI(T_i < C_i)$. The variable for center is A_i ($A_i = 1, \dots, J$), so that if individual i is in center j , then $A_i = j$ and the center j indicator is $A_{ij} = I(A_i = j) = 1$, where $I(\cdot)$ is the indicator function. The covariate vector is denoted by \mathbf{Z}_i . A random sample of size n consists of data $\{X_i, \Delta_i, \mathbf{Z}_i, A_i\}$ for $i = 1, \dots, n$. The unobserved counting process for individual i for cause k is $N_{ik}^*(t) = I(T_i \leq t, \Delta_i = k)$, while the observed counting process is $N_{ik}(t) = \int_0^t I(C_i > s) dN_{ik}^*(s)$ for cause k . Further, we define $N_{ijk}(t) = A_{ij}N_{ik}(t)$. We assume the T_i and C_i are conditionally independent, given \mathbf{Z}_i and A_i , such that

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P(t \leq T_i < t + \epsilon | T_i \geq t, C_i > t, \mathbf{Z}_i) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P(t \leq T_i < t + \epsilon | T_i \geq t, \mathbf{Z}_i). \quad (3.1)$$

The cumulative incidence function for cause k is given by

$$F_{ijk}(t) = P(T_i \leq t, \Delta_i = k | A_i = j, \mathbf{Z}_i) = E \left[\int_0^t dN_{ijk}^*(s) \right], \quad (3.2)$$

which can be interpreted as the cumulative probability that an event of type k happens by time t , taking into account of other events that are competing to occur. We define the effect of center j on the CIF of cause k as follows,

$$\delta_{jk}(t) = F_{jk}(t) - F_k(t), \quad (3.3)$$

where $F_{jk}(t) = E_Z[F_{ijk}(t)]$ and $F_k(t) = E_Z[E_A[F_{iAk}(t)]]$. The effect measure is analogous to direct standardization, where both expectations are taken with respect to the marginal distribution of the covariates \mathbf{Z}_i , instead of the conditional distribution of the covariates given center j , $[\mathbf{Z}_i|A_i = j]$. The effect can be seen as terms representing an expected value subtracting an observed value, with expectation taken over the distribution of the centers. The first term is center-specific and represents what the national average would have been if all patients were subjected to a center with CIF equal to that of center j . For patients who are not actually observed to be at center j , this is contrary to fact. The second term is not center-specific, since we have already taken the expectation with respect to center. It can be interpreted as the national or overall average CIF.

We now describe how to estimate $\delta_{jk}(t)$ from (3.3), using the relationship set out in (3.2). To begin, we consider the case where covariate adjustment does not need to be considered. Since $dN_{ijk}^*(s)$ is not observable, it is useful to develop an estimator of $\delta_{jk}(t)$ in terms of the observable quantity, $N_{ijk}(t) = \int_0^t I(C_i > s) dN_{ijk}^*(s)$. In the case of right-censored survival data, the censoring times are unknown if an event of any type is observed. The often used Inverse Probability of Censoring Weighting (IPCW)

technique (*Robins and Rotnitzky, 1992; Robins, 1993*) can be applied to obtain the unobserved counting process from the observed counting process via weighting. The IPCW technique weights each observation by the inverse of the survival function for C_i , conditional on the covariates $P(C_i > t|A_i = j, \mathbf{Z}_i)$. We can write $E[dN_{ijk}^*(s)] = E[dN_{ijk}(s)P(C_i > t|A_i = j, \mathbf{Z}_i)^{-1}]$ due to the independent censoring assumption (3.1) stated above and, integrating both sides over $(0, t]$, we obtain

$$F_{ijk}(t) = E \left[\int_0^t \frac{dN_{ijk}(s)}{P(C_i > t|A_i = j, \mathbf{Z}_i)} \right]. \quad (3.4)$$

An estimator for the CIF for type k in center j is then given by

$$\hat{F}_{ijk}(t) = \int_0^t \frac{dN_{ijk}(s)}{\hat{G}_{ij}(s)} = A_{ij} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s)},$$

where $G_{ij}(t) = P(C_i > t|A_i = j, \mathbf{Z}_i)$ with estimator $\hat{G}_{ij}(t) = \exp\{-\hat{\Lambda}_{ij}^C(t)\}$. The estimated cumulative hazard for censoring $\hat{\Lambda}_{ij}^C(t)$ is obtained by the proportional hazards *Cox* (1972) model stratified on center, $\lambda_{ij}^C(t) = \lambda_{0j}^C(t) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i)$. The estimate for the cumulative incidence is based on the observed counting process for events of type k in center j .

Given the focus on observational studies, confounding is likely to be an issue and the estimator of the effect may be biased if the differences in covariate distributions are not taken into account. Inverse Probability of Treatment Weighting (*Robins et al., 2000; Hernan et al., 2000*) is a technique that allows valid comparisons to be made across groups (or treatments) in observational studies by eliminating any group-

specific covariate patterns in the analysis step. For each individual, IPTW assigns a weight that is inversely related to the probability that the individual belongs to his/her actual observed group given his/her covariates values. Thus, if a patient, given his/her covariate pattern, is unlikely to be in the particular center that he/she is observed to be in, then that patient's experience gets more weight. The denominator of this weight, $P(A_i = j | \mathbf{Z}_i) \equiv p_{ij}(\boldsymbol{\theta})$, is assumed to follow the generalized logit model,

$$\log \left\{ \frac{p_{ij}(\boldsymbol{\theta})}{p_{iJ}(\boldsymbol{\theta})} \right\} = \boldsymbol{\theta}_j^T \mathbf{Z}_i,$$

for $j = 1, \dots, J - 1$. The proposed estimator of $\delta_{jk}(t)$ is then given by

$$\hat{\delta}_{jk}(t) = \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s)} - \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^J A_{il} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{il}(s)}. \quad (3.5)$$

The first term only uses information from subjects in center j due to the term $I(A_i = j)$. The subject-specific observed counting process for events of type k , $dN_{ik}(s)$, is weighted by the probability $p_{ij}(\hat{\boldsymbol{\theta}})$ so that the resulting empirical covariate distribution that is used for center j is the same as that of the entire study sample. The contribution of each individual in center j is also weighted by the inverse of the estimated conditional censoring survival distribution given the covariate $\hat{G}_{ij}(t)$, which handles the censoring. Thus the first term is doubly weighted by IPTW and IPCW.

The estimator is non-parametric in the sense that no models are assumed for the

cause-specific hazards or subdistribution. However, modeling is involved due to the IPCW and IPTW methods.

The second term utilizes the entire sample by summing through everyone in each center, and then summing through all of the centers. Each individual's contribution to the CIF is weighted by the quantity $\hat{G}_{ij}(t)$ that is also used in the first term. Thus the second term is an overall average of the observed cause k counting processes of all subjects in the study weighted by IPCW.

A direct consequence of applying IPTW to the first term is that the resulting empirical covariate distribution used for the first term is the same as the (unweighted) empirical covariate distribution of the second term. The center-specific first term can be interpreted as the estimated average CIF if everyone in the sample were at center j . The second term, which is not center-specific, can be thought of as an estimator of the overall average CIF.

By (3.4), we can show that

$$E \left[\frac{I(A_i = j)}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} \right] = E_Z[F_{ijk}(t)].$$

By known properties of Cox regression, $\hat{G}_{ij}(t)$ converges in probability to $G_{ij}(t)$ (*Anderson and Gill*, 1982). Assuming that the center assignment model is correct, established properties of maximum likelihood imply that $p_{ij}(\hat{\boldsymbol{\theta}})$ is a consistent estimator for $p_{ij}(\boldsymbol{\theta})$. Since the summands are all individuals from center j , they are independent and identically distributed (i.i.d.) and applying the Weak Law of Large

Numbers gives

$$\frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s)} \xrightarrow{p} E_Z[F_{ijk}(t)],$$

which is the first term of $\delta_{jk}(t)$ in (3.3).

Similarly, we can show that

$$E \left[\sum_{l=1}^J A_{il} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{il}(s)} \right] = E_Z[E_A[F_{iAk}(t)]]. \quad (3.6)$$

In (3.6), there is an additional step where the expectation with respect to the centers is taken.

3.3 Asymptotic Properties

We describe the asymptotic distribution of $\hat{\delta}_{jk}(t)$ by the following result.

Theorem 3.1: *As $n \rightarrow \infty$, $\hat{\delta}_{jk}(t)$ converges in probability to $\delta_{jk}(t)$, for $t \in [0, \tau]$. In addition, $n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \}$ converges to a Normal distribution with asymptotic variance $E \left[\{ \phi_{\ell jk}(t) \}^2 \right]$, where we define $\phi_{\ell jk}(t) = \phi_{\ell jk}^{(1)}(t) + \phi_{\ell jk}^{(2)}(t) + \phi_{\ell jk}^{(3)}(t) + \phi_{\ell jk}^{(4)}(t) -$*

$\{\phi_{\ell k}^{(5)}(t) + \phi_{\ell k}^{(6)}(t) + \phi_{\ell k}^{(7)}(t)\}$ with

$$\begin{aligned}
\phi_{\ell j k}^{(1)}(t) &= \mathbf{d}_{jk}^T(t) \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} U_{\ell}^{IPTW}(\boldsymbol{\theta}) \\
\phi_{\ell j k}^{(2)}(t) &= E \left[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ \int_0^s \{\mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha})\}^T d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \\
&\quad \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_{\ell}^C(\boldsymbol{\alpha}) \\
\phi_{\ell j k}^{(3)}(t) &= \int_0^t E [F_{jk}(t) - F_{jk}(u)] Y_{ij}^C(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i s_j^{(0)}(u; \boldsymbol{\alpha}))^{-1} dM_{\ell j}^C(u; \boldsymbol{\alpha}) \\
\phi_{\ell j k}^{(4)}(t) &= \frac{A_{\ell j}}{p_{\ell j}(\boldsymbol{\theta})} \int_0^t \frac{dN_{\ell k}(s)}{G_{\ell j}(s)} - F_{jk}(t) \\
\phi_{\ell k}^{(5)}(t) &= E \left[\sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ \int_0^s \{\mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha})\}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \\
&\quad \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_{\ell}^C(\boldsymbol{\alpha}) \\
\phi_{\ell k}^{(6)}(t) &= \sum_{m=1}^J \int_0^t E \left[A_{im} \int_u^t \exp(\Lambda_{im}^C(s)) dN_{ik}(s) Y_{im}(s) \right] \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i s_m^{(0)}(u; \boldsymbol{\alpha}))^{-1} \\
&\quad dM_{\ell m}^C(u; \boldsymbol{\alpha}),
\end{aligned}$$

where we define

$$\begin{aligned}
\mathbf{d}_{jk}^T(t) &= E \left[-A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \mathbf{Z}_i \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s; \hat{\boldsymbol{\alpha}})} \Delta_{ij}^T(\boldsymbol{\theta}) \right] \\
U_i^{IPTW}(\boldsymbol{\theta}) &= \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial l_i(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial l_i(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial l_i(\boldsymbol{\theta})}{\partial \theta_{J-1}} \end{bmatrix} = \begin{bmatrix} U_{i1}(\boldsymbol{\theta}) \\ U_{i2}(\boldsymbol{\theta}) \\ \vdots \\ U_{i,J-1}(\boldsymbol{\theta}) \end{bmatrix}
\end{aligned}$$

for $j = 1, \dots, J - 1$, with

$$\begin{aligned}
U_{ij}(\boldsymbol{\theta}) &= \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l} = \mathbf{Z}_i (A_{ij} - p_{ij}(\boldsymbol{\theta})) . \\
\mathcal{I}^{IPTW}(\boldsymbol{\theta}) &= -E \left[\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \\
&= -E \begin{bmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}) & \mathcal{I}_{12}(\boldsymbol{\theta}) & \cdots & \mathcal{I}_{1,J-1}(\boldsymbol{\theta}) \\ \mathcal{I}_{12}(\boldsymbol{\theta}) & \mathcal{I}_{22}(\boldsymbol{\theta}) & & \\ \vdots & & \ddots & \\ \mathcal{I}_{1,J-1}(\boldsymbol{\theta}) & & & \mathcal{I}_{J-1,J-1}(\boldsymbol{\theta}) \end{bmatrix}
\end{aligned}$$

where for $j = 1, 2, \dots, J - 1$ and $l = 1, 2, \dots, J - 1$ and $j \neq l$ we define the following:

$$\begin{aligned}
\mathcal{I}_{ll}(\boldsymbol{\theta}) &= -\frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_l^T} = \mathbf{Z}_i^{\otimes 2} \left[\frac{\exp(\mathbf{Z}_i^T \boldsymbol{\theta}_l)}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)} - \frac{\exp(2\mathbf{Z}_i^T \boldsymbol{\theta}_l)}{\left(1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)\right)^2} \right] \\
&= \mathbf{Z}_i^{\otimes 2} p_{il}(\boldsymbol{\theta}) \{1 - p_{il}(\boldsymbol{\theta})\} \\
\mathcal{I}_{jl}(\boldsymbol{\theta}) &= -\frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_l^T} = \mathbf{Z}_i^{\otimes 2} \frac{\exp\{\mathbf{Z}_i^T (\boldsymbol{\theta}_l + \boldsymbol{\theta}_j)\}}{\left\{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)\right\}^2} \\
&= -\mathbf{Z}_i^{\otimes 2} p_{il}(\boldsymbol{\theta}) p_{ij}(\boldsymbol{\theta}) \\
dM_{ij}^C(t) &= dN_{ij}^C(t) - Y_{ij}(t) d\Lambda_{ij}^C(t) \\
s_j^{(d)}(t; \boldsymbol{\beta}_k) &= E \left[Y_{ij}(t) \mathbf{Z}_i^{\otimes d} \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \right] \quad d = 0, 1, 2.
\end{aligned}$$

$$\Delta_{ij}(\boldsymbol{\theta}) = -\mathbf{Z}_i \begin{bmatrix} p_{ij}(\boldsymbol{\theta})p_{i1}(\boldsymbol{\theta}) \\ p_{ij}(\boldsymbol{\theta})p_{i2}(\boldsymbol{\theta}) \\ \vdots \\ p_{ij}(\boldsymbol{\theta})(1 - p_{ij}(\boldsymbol{\theta})) \\ \vdots \\ p_{ij}(\boldsymbol{\theta})p_{i,J-1}(\boldsymbol{\theta}) \end{bmatrix}.$$

We can write $n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \} = n^{-\frac{1}{2}} \sum_{\ell=1}^n \phi_{\ell jk}(t)$, where the $\phi_{\ell jk}(t)$ are independent and identically distributed zero-mean variates. Then, $\text{Var} \left(n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \} \right) = E \left[\{ \phi_{\ell jk}(t) \}^2 \right]$ and an estimator for $\text{Var}(\hat{\delta}_{jk}(t))$ is $\widehat{\text{Var}}(\hat{\delta}_{jk}(t)) = n^{-2} \sum_{\ell=1}^n \{ \hat{\phi}_{\ell jk}(t) \}^2$.

As introduced in Theorem 3.1, the expression for the variance is very involved and, thus, calculating the estimated variance is quite demanding in terms of coding and computation. However, the estimator can be calculated very fast using standard software such as SAS, and so the bootstrap method is suggested for estimating the variance of the estimate.

3.4 Simulation Studies

We studied the performance of the proposed estimator through simulations from the competing risks setting with two causes, with cause $k = 1$ being of interest. The failure times T_{ij1} and T_{ij2} follow exponential distributions with $\lambda_{ijk}(t) = \lambda_{0jk} \exp\{ \boldsymbol{\beta}_k^T \mathbf{Z}_i \}$, $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ where $\lambda_{0j1} = 0.08$ and $\lambda_{0j2} = 0.01$. The base-

line hazards are center- and cause-specific. The censoring was also exponential with $\lambda^C(t) = \lambda_0^C \exp\{\boldsymbol{\alpha}^T \mathbf{Z}_i\}$, with administrative censoring at $t = 10$ and baseline censoring hazard $\lambda_0^C = 0.011$. There were three covariates: $Z_{i1} \sim \text{Ber}(\boldsymbol{\theta}_{1j})$, $Z_{i2}|Z_{i1} \sim \text{Ber}(Z_{i1}\theta_{21} + (1 - Z_{i1})\theta_{22})$, and $Z_{i3}|Z_{i2} \sim \text{Ber}(Z_{i2}\theta_{31} + (1 - Z_{i2})\theta_{32})$, where Ber denotes Bernoulli distribution with covariate effects implied by $\boldsymbol{\beta}_1^T = [0.4; 0.5; 0.6]$, and $\boldsymbol{\beta}_2^T = [-0.1; 0.3; 0.2]$. The values of the parameters for generating the covariates were $\boldsymbol{\theta}_{1j}^T = [0.55, 0.75, 0.6, 0.65, 0.5]$, $\theta_{21} = 0.55$, $\theta_{22} = 0.45$, $\theta_{31} = 0.45$, and $\theta_{32} = 0.65$. There were $J = 5$ centers, each with 200 individuals. The bootstrap method was employed to estimate variability, with 25 bootstrap samples used per replicate. We evaluated the performance of the estimator at times $t = 1, t = 3$, and $t = 5$. Simulation results in Table 3.1 are obtained by replicating each data structure 500 times. At $t = 5$, on average across the iterations, there were still between 23% and 40% of subjects at risk across the five centers.

In the first set of simulations, the estimator $\hat{\delta}_{j1}(t)$ performed quite well and has relatively small bias, as seen in Table 3.1. Although bias was quite low uniformly, it tended to be larger at later time points ($t = 5$) relative to earlier time points. This is likely due to the fact that there are more data at the earlier stages of follow-up. The value of $\delta_{j1}(t)$ also seems to affect the bias of $\hat{\delta}_{j1}(t)$. The larger the displacement from having no effect, or in other words, the larger the absolute value of $\delta_{j1}(t)$, the larger the bias. For example, center $j = 1$ performs much better than the overall average, and the bias is larger compared to that of center $j = 3$ which performs at a level appropriately equal to the overall average.

Table 3.1: Simulation results using Configuration 1: estimate for effect of center j on CIF of cause $k = 1$, with bias, empirical standard deviation (ESD), bootstrap standard error (BSE), and 95% confidence interval coverage probabilities (CP).

| j | t | $\delta_{j1}(t)$ | BIAS | ESD | BSE | CP |
|-----|-----|------------------|--------|-------|-------|------|
| 1 | 1 | 0.07 | -0.001 | 0.024 | 0.023 | 0.94 |
| | 3 | 0.16 | -0.008 | 0.030 | 0.031 | 0.93 |
| | 5 | 0.21 | -0.018 | 0.033 | 0.032 | 0.90 |
| 2 | 1 | 0.04 | -0.000 | 0.023 | 0.023 | 0.93 |
| | 3 | 0.08 | -0.005 | 0.031 | 0.032 | 0.94 |
| | 5 | 0.10 | -0.015 | 0.036 | 0.035 | 0.91 |
| 3 | 1 | 0.02 | 0.000 | 0.022 | 0.021 | 0.93 |
| | 3 | 0.04 | 0.000 | 0.031 | 0.030 | 0.92 |
| | 5 | 0.06 | 0.002 | 0.031 | 0.032 | 0.94 |
| 4 | 1 | -0.06 | 0.000 | 0.015 | 0.015 | 0.93 |
| | 3 | -0.14 | 0.009 | 0.023 | 0.023 | 0.92 |
| | 5 | -0.18 | 0.019 | 0.028 | 0.027 | 0.89 |
| 5 | 1 | -0.06 | 0.000 | 0.016 | 0.016 | 0.93 |
| | 3 | -0.14 | 0.005 | 0.026 | 0.024 | 0.93 |
| | 5 | -0.19 | 0.013 | 0.030 | 0.029 | 0.91 |

Table 3.2: Simulation results using Configuration 2: estimate for effect of center j on CIF of cause $k = 1$, with bias, empirical standard deviation (ESD), bootstrap standard error (BSE), and 95% confidence interval coverage probabilities (CP).

| j | t | $\delta_{j1}(t)$ | BIAS | ESD | BSE | CP |
|-----|-----|------------------|--------|-------|-------|------|
| 1 | 1 | 0.01 | -0.000 | 0.02 | 0.021 | 0.95 |
| | 3 | 0.04 | -0.003 | 0.031 | 0.029 | 0.92 |
| | 5 | 0.06 | -0.006 | 0.031 | 0.032 | 0.94 |
| 2 | 1 | 0.02 | 0.001 | 0.021 | 0.031 | 0.93 |
| | 3 | 0.04 | 0.000 | 0.029 | 0.021 | 0.94 |
| | 5 | 0.04 | -0.001 | 0.032 | 0.031 | 0.94 |
| 3 | 1 | 0.001 | -0.000 | 0.020 | 0.020 | 0.92 |
| | 3 | 0.002 | -0.001 | 0.028 | 0.028 | 0.93 |
| | 5 | 0.002 | -0.001 | 0.032 | 0.031 | 0.94 |
| 4 | 1 | -0.02 | 0.000 | 0.018 | 0.019 | 0.95 |
| | 3 | -0.04 | 0.003 | 0.028 | 0.027 | 0.93 |
| | 5 | -0.06 | 0.001 | 0.028 | 0.03 | 0.93 |
| 5 | 1 | -0.01 | 0.003 | 0.031 | 0.03 | 0.93 |
| | 3 | -0.04 | 0.005 | 0.019 | 0.018 | 0.93 |
| | 5 | -0.05 | 0.006 | 0.030 | 0.031 | 0.94 |

The second set of simulations, shown in Table 3.2, contain center effects of smaller magnitudes compared to those of the first set. As can be seen, when the effects are closer to 0, the bias tends to be smaller, consistent with Table 3.1. The coverage probabilities are also closer to the nominal value of 0.95, though there does not seem to be a trend of lower coverage probabilities at later follow-up times.

3.5 Application

Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). We applied our methods to assess each (OPO) with respect to the average probability that a patient received a kidney transplant, relative to the regional average. For organ allocation purposes, the 58 OPOs are grouped into 11 regions. We selected the $J = 6$ OPOs that comprise Region 10, which contains Michigan. Patients wait-listed during the last decade (i.e., between January 1, 2000 and December 31, 2009) were included in the study. The resulting sample consisted of $n = 22,685$ subjects from $J = 6$ OPOs. The observation period ended on December 31, 2009.

Wait-listed patients leave the wait-list due to death, receipt of a kidney transplant, or being removed. The cause of interest is transplant with a deceased donor, and the competing risks are removal and death. The adjustment covariates were age, sex, body mass index (BMI), race, hypertension status, blood type, diabetes, and panel reactive antibodies. Living donor transplantation was treated as (independent) censoring, an issue to which we return in Section 3.6. Figure 3.1 shows the trend of

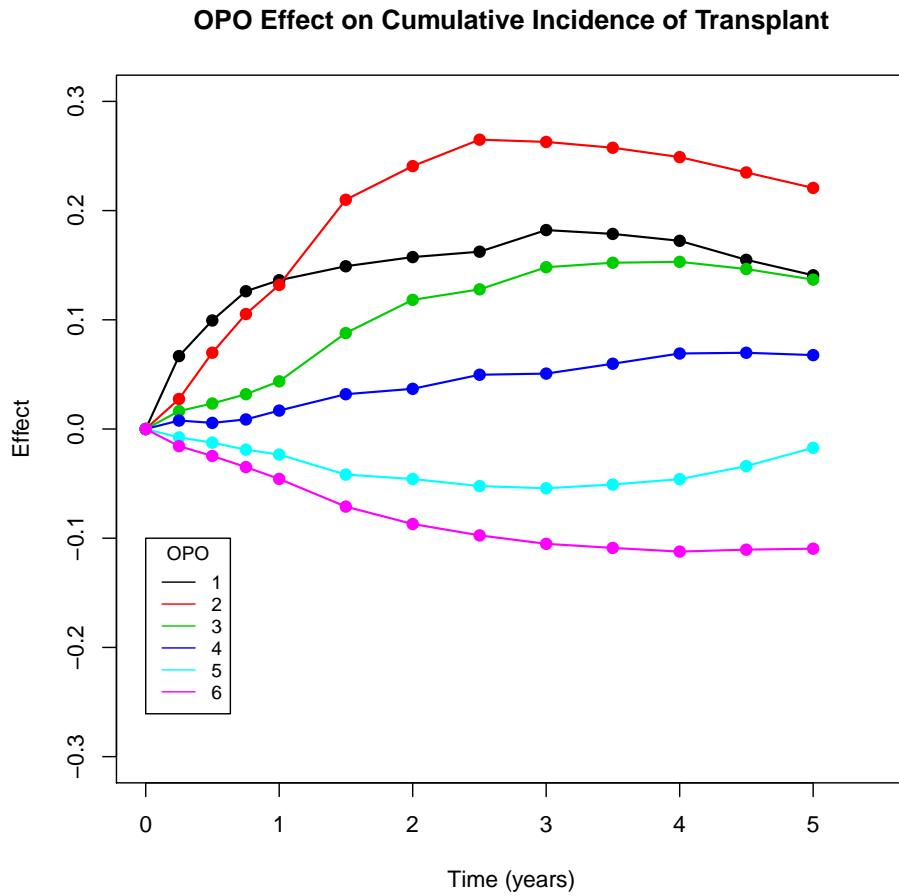


Figure 3.1: Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ for various $t \in [0, 5]$ years post wait-listing, for $j = 1, \dots, 6$.

$\hat{\delta}_{j1}(t)$ over 5 years of follow-up for $j = 1, \dots, 6$. OPOs 1, 2, and 3 perform better than expected, while OPOs 5 and 6 perform worse than expected. Interestingly, for the OPOs with elevated cumulative incidence of transplantation, $\hat{\delta}_{j1}(t)$ increases early; usually peaking between $t = 2$ and $t = 4$ years before dipping slightly during the last couple years of follow-up. In general, the differences in performance among OPOs first become larger as follow-up time increases, and then decrease slightly after year 3, where many of the estimated effects stabilize. The OPOs are sorted in ascending sample size, with 1,434 and 9,269 patients on the wait-list in OPO $j = 1$ and OPO $j = 6$, respectively. There emerged a pattern where the larger OPOs performed worse than the smaller OPOs. This could be due to disparities in organ donation rates across the OPOs. Since donations reflect one aspect of the OPO's efforts and performance in allocating organs, adjusting for it would attenuate, perhaps inappropriately, some of the OPO effect. Figures 3.2 and 3.3 are similar to Figure 3.1, but with a 95% confidence limits for $\hat{\delta}_{j1}(t)$ at the selected time points. OPO 1 was put into a separate figure because its confidence limits overlap with those of OPO 2. As expected, the confidence limits are narrower at the start of follow-up and are wider at the end of the follow-up period, due to sparser data at the later times. In addition, larger OPOs have narrower confidence limits. Figure 3.4 shows the change in $\hat{\delta}_{j1}(t)$ for each time interval for OPO 1. The bars represent the difference over time in the difference between OPOs. From the start of follow-up to 3 months after start of follow-up, $\hat{\delta}_{11}(t)$ increased by 0.07. From the start of follow-up until year 1, $\hat{\delta}_{11}(t)$ is increasing at a decreasing rate. Starting from end of year 3, $\hat{\delta}_{11}(t)$ decreases, though still remains greater than 0. Figures 3.5 and 3.6 are similar

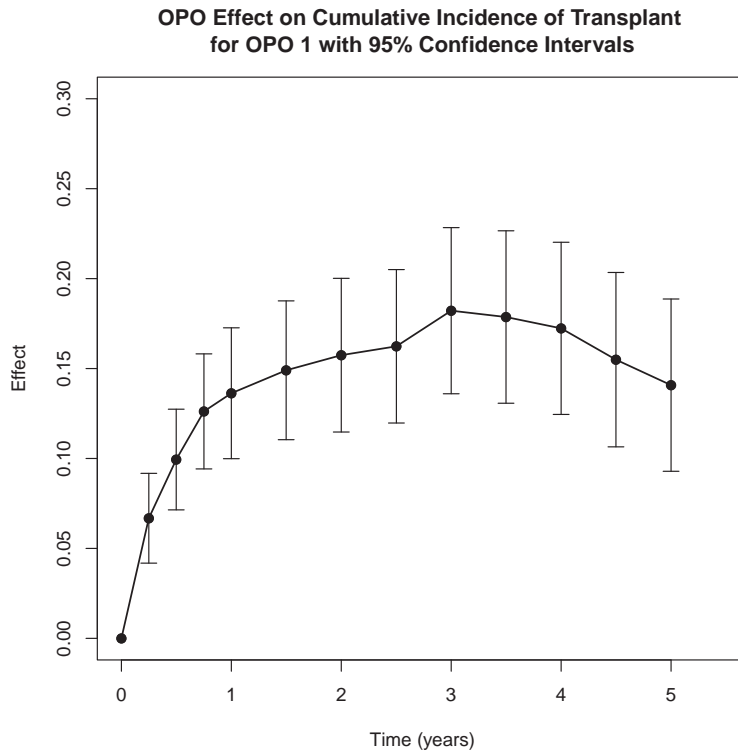


Figure 3.2: Analysis of SRTR data: $\hat{\delta}_{11}(t)$ with 95% confidence interval for various $t \in [0, 5]$ years post wait-listing.

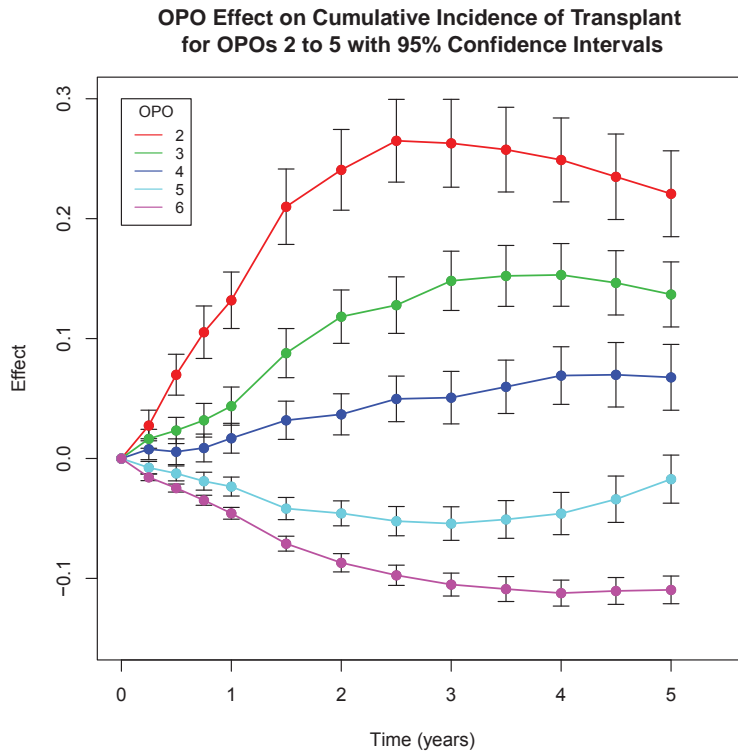


Figure 3.3: Analysis of SRTR data: $\hat{\delta}_{j1}(t)$ with 95% confidence interval of for various $t \in [0, 5]$ years post wait-listing, for $j = 2, \dots, 6$.

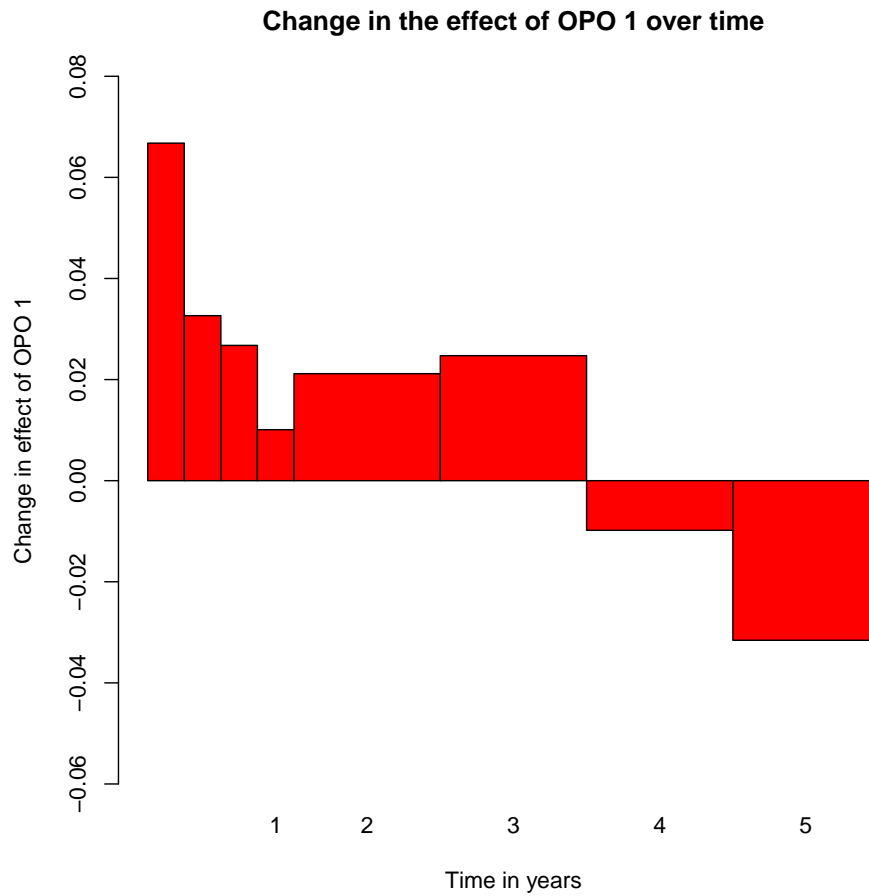


Figure 3.4: Analysis of SRTR data: Change in $\hat{\delta}_{11}(t)$ over various time intervals, for $t \in [0, 5]$ years post wait-listing.

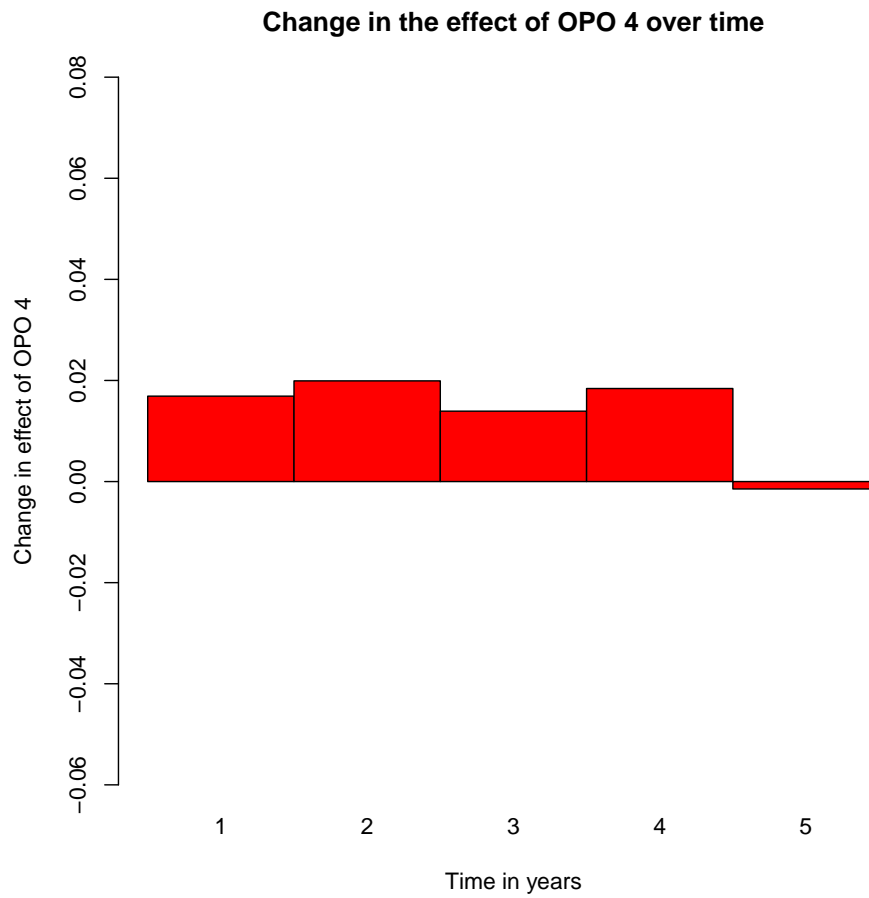


Figure 3.5: Analysis of SRTR data: Change in $\hat{\delta}_{41}(t)$ for various $t \in [0, 5]$ years post wait-listing.

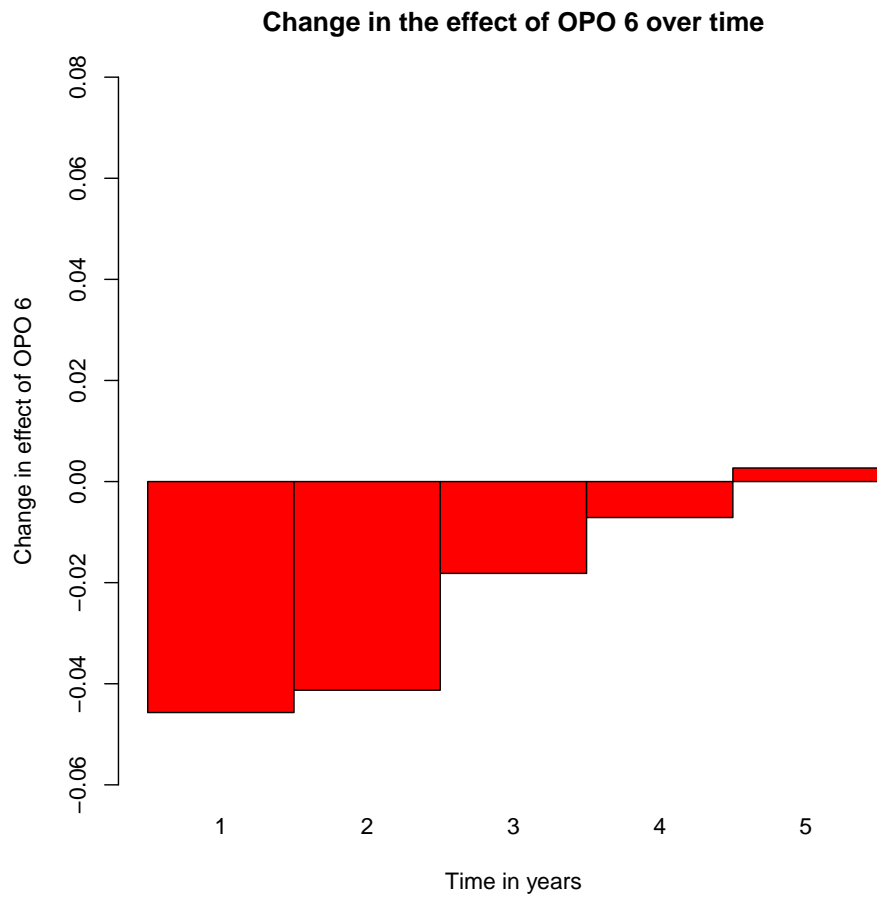


Figure 3.6: Analysis of SRTR data: Change in $\hat{\delta}_{61}(t)$ for various $t \in [0, 5]$ years post wait-listing.

to Figure 3.4, but for OPOs 4 and 6, respectively. The $\hat{\delta}_{41}(t)$ increases at an almost constant rate from start of follow-up until end of year 4. Then, during year 5 there is almost no change. From Figure 3.6, the effect of OPO 6 decreases at a decreasing rate until end of year 4.

Another analysis was also carried out using all $J = 58$ OPOs in the SRTR. The results, shown via two figures, are included in Appendix B.

3.6 Discussion

In this chapter, we develop methods for comparing subgroups to the overall average based on cumulative incidence. The cumulative incidence itself is estimated non-parametrically, while being weighted by IPCW and IPTW semi-parametrically and parametrically, respectively. The effect of each subgroup on cumulative incidence is quantified as the difference between the subgroup-specific cumulative incidence and the overall average. For each subgroup, patients are inverse weighted to reflect the marginal covariate distribution, while the marginal covariate distribution is used explicitly in the computation of the overall average CIF. Simulation results indicate that the proposed method has negligible bias with empirical coverage probability, based on bootstrapped standard errors, being approximately equal to (although, persistently, slightly below) the nominal value. The method is applied to registry data to identify Donation Service Areas in Region 10 that are under-served with respect to deceased-donor kidney transplantation.

Similar to direct standardization, methods proposed in this chapter use the marginal covariate distribution in measuring the group effect. Therefore, it is valid to compare $\hat{\delta}_{j1}(t)$ across different values of j . Comparing $\hat{\delta}_{j1}(t)$ across OPOs would not be valid if indirect standardization was used, which involves the covariate distribution given a particular subgroup. The method described can be useful to a governing or regulatory body that wants to make direct comparisons between groups. In applying the proposed methods, an issue is the possibility of unstable weights. This was more of a concern for the IPTW than for IPCW, since we do not expect the covariates to have very strong effects on censoring. A simple stabilizer based on the proportion of total subjects in each OPO did not improve the estimation and precision of the estimator noticeably. The simulation results indicated that the bias and confidence interval coverage probabilities are acceptable without any stabilization for the IPTW.

The IPCW is semi-parametric efficient since it is based on the partial likelihood. The IPTW is based on the maximum likelihood, and is therefore efficient. Therefore, this method should not lose much efficiency compared to using the cause-specific hazards (which is also estimated efficiently by Cox regression) to model the CIF.

The methods outlined in *Fine and Gray* (1999) can be applied to compare subgroups by using indicator functions to represent each subgroup, giving the subdistribution hazard ratios. In the example of the SRTR data used in this chapter, one OPO would have to be selected as the reference, lending the coefficients to meaningful interpretation. Using this approach, proportionality of hazards across the OPOs also needs to be plausible. Thus, the methods proposed in this chapter are more

flexible in that no proportionality needs to be assumed. Another possible approach to apply *Fine and Gray* (1999) to the set up of interest in this chapter would be to stratify by group, then average the resulting fitted values. An advantage of *Fine and Gray* (1999) may be efficiency gains due to more parametrization compared to the methods in this chapter.

We considered living donor transplantation as independent censoring since it can be argued that living donors do not reflect either positively nor negatively on the effort of the OPO to obtain and allocate kidneys. This is because most living organ donors know their recipients, being either a relative or friend, and thus are specifically donating the kidney for that recipient. Therefore, the recipient of a living donor transplant is not determined by the OPO. That said, it is possible that, in OPOs with low rates of deceased-donor transplantation, patients have increased motivation to seek out a living donor. If this were the reality, it may be more appropriate to treat living donor transplantation as a competing risk. Another possibility would be to treat living donor transplantation as dependent censoring, which would fall outside the scope of the methods proposed in this chapter.

CHAPTER IV

Semi-Parametric Methods for Modeling the Subdistribution Hazard using Multiple Imputation

4.1 Introduction

In many biomedical studies, interest lies in the relationship between certain factors and the time until a particular event. Often, the subject is at risk for several mutually exclusive health outcomes, one of which is of chief interest. For example, consider an observational study that tries to identify factors correlated with cancer recurrence. Patients in that study can also experience death and, if so, cancer recurrence cannot subsequently occur. Since recurrence is of chief interest, death serves as a competing risk, which may be of at most secondary interest to this study. In these cases, standard survival analysis methods are not appropriate and methods that account for competing risks are required for valid inference. A popular quantity in the competing risks landscape is the cumulative incidence function (CIF), which

is the probability that a particular outcome occurs by a certain time, in the presence of other events which compete to occur (*Kalbfleisch and Prentice*, 2002). Another important quantity in the competing risks setting is the cause-specific hazard (CSH), which is the hazard of a particular event occurring at a certain time t , conditional on being at risk (i.e., have not experienced any cause) as of time t .

A frequently employed method of estimating the CIF involves estimating the cause-specific hazard for each cause, then using the relationship between the CSHs and the CIF to obtain an estimate of the latter. Nonetheless, there has been some work in the competing risks literature where the focus was on estimation of the CIF without explicitly estimating the CSHs. For example, *Gray* (1988) proposed tests comparing group-specific cumulative incidence functions via the subdistribution hazard. *Pepe* (1991) developed methods for the estimation and inference of the Kaplan-Meier survival function (*Kaplan and Meier*, 1958) and cumulative incidence function as sums of independent and identically distributed random variables. Two sample tests were also introduced, along with the asymptotic null distribution of the test statistics. *Fine and Gray* (1999) developed methods that involve applying a regression model to the subdistribution hazard for the cause of interest. This approach allows the covariate effects on the cause of interest to be evaluated without having to consider models for the causes not-of-interest. The method involves a variant of Inverse Probability of Censoring Weighting (IPCW; *Robins and Rotnitzky* (1992)), which may be computationally intensive using standard software. In particular, fitting the model using standard Cox regression (*Cox*, 1972) software would often involve expanding the original data set to include separate records for each subject at each time the

weight changes. In addition, the estimates may be unstable due to the possibility of large weights, a phenomenon familiar to IPCW.

In addition to the method of *Fine and Gray* (1999), other methods have been developed for subdistribution regression. *Scheike et al.* (2008) proposed a binomial regression method that directly models the CIF. Procedures for estimation and inference were presented for time-dependent effects. Recently, *Klein and Andersen* (2005) also developed methods for direct regression on the CIF using pseudovalues based on a jackknife statistic.

Censoring-complete data are observed when the event times are subject to censoring, but the censoring times are all known. In such cases, even when a subject's failure time is observed (and, hence not censored), the potential censoring time is still known for that subject. Censoring-complete survival data arise when there is only administrative censoring and no random censoring caused by loss to follow-up. An example of a setting which would generate such data would be a clinical trial where patients are followed very closely, so that no patients are lost to follow-up. For instance, the trial may be designed so that the patients can enter at different chronological times, but the trial ends on a pre-determined specific calendar date upon which follow-up ends for all patients. Thus, even if an event is observed, the potential censoring time is still known through its correspondence to the planned end date of the trial. As another example, consider a study that follows patients who are hospitalized, with death while hospitalized or discharge from the hospital being the two events that compete to occur. Since the patients are hospitalized, they will

not be lost to follow-up. Censoring-complete data arise from this setting if the end of study date has already been decided at the start of the study. Even if an event were observed, the potential censoring time (i.e., the planned end of study) is still known.

Although censoring-complete data are more convenient, most real-world applications involve random loss to follow-up. *Fine and Gray* (1999) framed their inverse weighting method as the replacement of an unobserved indicator function by a weighted quantity with the same expectation. Although the method proposed by *Fine and Gray* (1999) involves IPCW, the use of the inverse weight is motivated by considerations very different from the IPCW methods used for dependent censoring (*Robins and Rotnitzky*, 1992). However, the inverse weighting used in subdistribution hazard modeling shares the considerable added computational burden with other time-dependent weighting methods. For this reason, it would be desirable to have an alternative to inverse weighting.

In the context of proportional hazards modeling of the subdistribution function, one advantage of censoring-complete data over ordinary censored data is that standard software can be applied without a weight function. Multiple imputation is a technique that can be used when the analysis dataset contains missing, incomplete data (*Little and Rubin*, 2002; *Rubin*, 1987). The imputation procedure essentially replaces the missing data with imputed values so that the resulting dataset, called an augmented dataset, is complete. The missing data are generated independently using the same procedure multiple times, creating a set of augmented datasets. Usually

between five and ten augmented datasets are enough to provide acceptable results. The analysis is done on each augmented dataset separately. Valid parameter estimates and corresponding standard errors can be calculated by combining results across augmented datasets, with an additional step for the variance estimator due to variability introduced by the multiple imputes (*Little and Rubin, 2002*). The imputes can be drawn from distributions that are estimated by parametric or non-parametric methods; for example, *Rubin (1987)* uses imputation to analyze data obtained from surveys, which often contain missing data.

In standard univariate time-to-event data, multiple imputation has been used in cases with missing failure times due to censoring. For observations that are censored, the imputed value replaces the censoring time. Thus, the imputed value represents what failure time would have been observed if censoring did not occur. *Taylor et al. (2002)* developed non-parametric multiple imputation procedures with the end goal of estimating and testing survival functions. The imputed time replaces the censored time, creating a new dataset where failures are observed for all subjects.

The motivating example for this chapter is based on kidney transplantation. In the United States, a patient who is in need of a kidney transplant is included on a waiting list that is maintained by an organ procurement organization (OPO). The patient, while on the wait-list, also faces death and being removed from the wait-list for being of very ill-health. If either removal or death occurs, the patient will not receive a transplant. Thus, in the context of our motivating example, the cause of interest is receiving a kidney transplant while on the wait-list, with the competing risks being

pre-transplant death and removal from the wait-list.

Our interest in this chapter primarily lies in determining the covariates that correlate with receiving a kidney transplant. Specifically, since blood type affects the likelihood of being matched with a donor kidney, it has been suggested that certain blood types place a patient at a disadvantage in terms of the probability of receiving a transplant. This has been referred to as the type O problem (*Barone et al.*, 2008; *Stanford et al.*, 2008; *Glander et al.*, 2010) since patients with type O blood can be donors for all other blood types but can only receive an organ with blood type O. Note that the set up in this chapter is different from those mentioned above, in that we are working with individuals who are exposed to more than one cause of failure and we will not observe the censoring time if an event (of any type) has happened. Thus, from the perspective of censoring-completeness, the would-be censoring time for an observation can be considered as missing data and one option is to impute the unobserved censoring times. In this case, failure times for causes not-of-interest would be replaced by an imputed censoring time. Thus, because of the competing risks setting, the censoring time is considered missing in this chapter; whereas in the examples two paragraphs ago, the failure times were the missing data.

Our proposed method has similarities to the imputation algorithm developed by *Schaubel and Zhang* (2010), where imputation aided in the estimation of treatment effects on the marginal recurrent event mean. Values of unobserved censoring times were imputed for the patients who were observed to die. *Ruan and Gray* (2008) developed a similar method to the one proposed in this chapter, specifically, imputing

censoring times for observations that experienced a competing event. However, the censoring distribution was estimated via Kaplan-Meier methods (*Kaplan and Meier*, 1958), thus assuming that censoring is not affected by the covariates. A more flexible approach would be to let censoring depend on the covariates.

In this chapter, we propose methods that use imputation to create censoring-complete data from a standard right-censored survival setting with competing risks. For observations that experience a competing risk, a time is imputed which represents the would-be censoring time if the competing risk had not been observed. The imputation will incorporate information from the covariates with a Cox model assumed for censoring hazard. We also use an interpolation method to reflect the continuity of the assumed baseline hazard for the censoring variable.

The remainder of the chapter is structured as follows. Section 4.2 sets up the notation and introduces the imputation method and interpolation technique. Section 4.3 presents simulation results to assess the performance of the imputation procedure in finite samples. Section 4.4 presents results of the proposed method applied to the motivating example and Section 4.5 completes the chapter with a discussion.

4.2 Proposed Methods

4.2.1 Set-up and Models

We begin by formalizing the data structure and assumed models. Let T_i and C_i be the event and censoring times, respectively, for individual i ($i = 1, \dots, n$). Cause will be denoted by k ($k = 1, \dots, K$), and let $\Delta_i = kI(T_i < C_i)$, where $I(\cdot)$ is the indicator function. The observation time is defined as $X_i = T_i \wedge C_i$. Thus, the observed data consist of $\{X_i, \Delta_i, \mathbf{Z}_i\}$, where \mathbf{Z}_i is a vector of covariates which is assumed to be time constant. The at-risk indicator is given by $Y_i(t) = I(X_i \geq t)$ and the counting process for subject i cause k is denoted by $N_{ik}(t) = I(X_i \leq t, \Delta_i = k)$. For concreteness, the rest of this chapter we will take k as the cause of interest. We assume that T_i and C_i are conditionally independent given \mathbf{Z}_i , and that the time scale is continuous. We also assume proportional hazards for the censoring variable,

$$\lambda_i^C(t) = \lambda_0^C(t) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i). \quad (4.1)$$

The proposed method is an alternative to inverse weighting for fitting the proportional subdistribution hazards model proposed by *Fine and Gray* (1999), who developed methods for directly testing the effect of covariates on the subdistribution hazard. *Fine and Gray* (1999) posit that, in the case of censoring-complete competing risks data, unweighted methods can be applied to fit their model. With standard right censored data, the censoring times are not known if a competing event has

occurred. To estimate the proportional subdistribution hazards model proposed by *Fine and Gray* (1999), we create imputed datasets that are then censoring-complete. We consider the censoring times for the subjects that were observed to experience one of the competing risks as missing data and impute potential censoring times.

The CIF is defined as $F_{ik}(t) = P(T_i \leq t, \Delta_i = k | \mathbf{Z}_i)$ and typically expressed as $F_{ik}(t) = \int_0^t \lambda_{ik}^\#(s) S_i(s) ds$, where $S_i(t) = P(T_i > t | \mathbf{Z}_i)$. The quantity $F_{ik}(t)$ represents the probability that an event of type k has occurred by time t , allowing for the presence of competing events (*Kalbfleisch and Prentice*, 2002); while $\lambda_{ik}^\#(t)$ is the cause-specific hazard (CSH),

$$\lambda_{ik}^\#(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P(t \leq T_i < t + \epsilon, \Delta_i = k | X_i \geq t, \mathbf{Z}_i). \quad (4.2)$$

Gray (1988) defined the improper random variable

$$T_i^* = T_i I(\Delta_i = k) + \infty I(\Delta_i \neq k),$$

which has the associated risk set $I(T_i \geq t \cup \{T_i \geq t, \Delta_i \neq k\})$. The variate T_i^* is not proper because if a subject experiences a competing event, the failure time due to the event of interest is infinity. *Fine and Gray* (1999) extended the risk set to incorporate the presence of censoring,

$$R_i(t) = I(X_i \geq t \cup \{T_i \leq t, \Delta_i \neq k, C_i \geq t\}).$$

The subdistribution hazard as defined by *Fine and Gray* (1999) is given by

$$\begin{aligned}
\lambda_{ik}(t) &= -\frac{d}{dt} \log \{1 - F_{ik}(t)\} \\
&= \frac{\lambda_{ik}^{\#}(t) S_{ik}(t)}{1 - F_{ik}(t)} \\
&= \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq T_i < t + dt, \Delta_i = k | X_i \geq t \cup \{T_i \leq t, \Delta_i \neq k, C_i \geq t\}, \mathbf{Z}_i).
\end{aligned}$$

The improper random variable creates an improper risk set in the subdistribution hazard since if a competing risk has already occurred, it is still considered at risk for the event of interest. In the standard survival set up, $I(X_i \geq t)$ is the risk-set indicator. Therefore, the individual is defined to be at risk until the occurrence of the earliest of any event or censoring. In this improper risk set, the at-risk indicator $R_i(t)$ considers the individual to be at risk until the event of interest or censoring happens, so that being censored removes an individual from the risk set for experiencing the event of interest, but experiencing a competing event does not. Thus, if a competing event has occurred, the individual is defined to be at risk even though the event of interest will never happen.

The proportional subdistribution hazard model for the event of interest k proposed by *Fine and Gray* (1999) is

$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i), \tag{4.3}$$

where $\boldsymbol{\beta}_k$ serves as the parameter of interest and $\lambda_{0k}(t)$ is an unspecified baseline subdistribution hazard. From *Fine and Gray* (1999), the estimating function for

censoring-complete data is

$$\mathbf{U}_k(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n Y_{jk}(s) \mathbf{Z}_j \exp\{\beta_k^T \mathbf{Z}_j\}}{\sum_{j=1}^n Y_{jk}(s) \exp\{\beta_k^T \mathbf{Z}_j\}} \right] dN_{ik}(s), \quad (4.4)$$

where $Y_{ik}(t) = I(C_i \geq t)\{1 - N_{ik}(t-)\}$. The equation in (4.4) can be used only when C_i is known. If not, a variant of IPCW has been suggested (*Fine and Gray, 1999*). As an alternative to inverse weighting, we propose imputing the unknown C_i through a procedure that we now describe.

4.2.2 Imputing Censoring Times

The censoring times can be considered as a special case of missing data. One often-used method for dealing with missing data is multiple imputation (*Rubin, 1987*), where procedures impute censoring times for the observations which experienced a competing event (i.e., an event not-of-interest). In the context of an observational study, it is quite possible that covariates affect censoring. Therefore, we incorporate covariate information when imputing censoring times.

The censoring-complete data consist of $\{\mathbf{Z}_i, X_i^m, \Delta_i\}$, for $i = 1, \dots, n$ and $m = 1, \dots, M$, with

$$X_i^m = I(\Delta_i = k) \cdot X_i + I(\Delta_i = 0) \cdot X_i + I(\Delta_i \neq k, \Delta_i \neq 0) \cdot C_i^m,$$

where C_i^m is the m th imputed censoring time for subject i and M indexes the aug-

mented dataset. Thus a censoring time is imputed when the subject experiences a competing risk. If the subject experienced the event of interest or was censored, then no values are imputed for that subject since the observed data are already considered complete in such cases. From the censoring-complete data perspective, a competing event is classified as censored. If an individual experiences an event $l \neq k$ at time T_i , then that individual is still considered to be at risk for the event of interest until its imputed censoring time, C_i^m .

The following steps detail the imputation procedure.

1. Take a bootstrap sample, with replacement, of the original observed data. Estimate the distribution of the censoring variable by fitting a *Cox* (1972) proportional hazards model on the censoring times, $\lambda_i^C(t) = \lambda_0^C(t) \exp(\alpha^T \mathbf{Z}_i)$. In this censoring model, all observed failure events are considered to be ‘censored’, since the time of interest is considered to be C_i .
2. Suppose subject i is observed to experience a competing event ($\Delta_i = l$, where $l \neq k$ and $l \neq 0$) at time T_i . Randomly draw C_i^m from $\hat{G}_i(t; T_i)$, where

$$G_i(t; T_i) \equiv P(C_i > t | C_i > T_i, \mathbf{Z}_i, T_i) = \frac{G_i(t) I(t > T_i)}{G_i(T_i)}$$

with $G_i(t) = P(C_i > t | \mathbf{Z}_i) = \exp\{-\Lambda_i^C(t)\}$. The quantity $G_i(t; T_i)$ is the conditional survival function for censoring given the censoring time is greater than T_i and $\Lambda_i^C(t) = \int_0^t \lambda_i^C(s) ds$. We use the conditional survival distribution instead of the marginal survival distribution to take into account of the fact that, although the exact value

of the censoring time is unobserved, C_i is known to be greater than T_i , the observed failure time.

Instead of using a step function to estimate the baseline cumulative censoring hazard function, we use a linear interpolation based on the step function. This approach is thought to more accurately reflect the true baseline hazard function, since it is unlikely that the true censoring hazard function is a series of non-zero values at the observed censoring times and zero elsewhere.

The following steps describe the linear interpolation:

2A. Let s_1 and s_2 be censoring times such that $s_1 < T_i < s_2$ and no other censoring occurs between s_1 and s_2 . Then, calculate $\hat{\Lambda}_0^C(T_i)$ as follows,

$$\hat{\Lambda}_0^C(T_i) = \hat{\Lambda}_0^C(s_1) + (T_i - s_1) \frac{\hat{\Lambda}_0^C(s_2) - \hat{\Lambda}_0^C(s_1)}{s_2 - s_1}.$$

2B. Compute

$$\hat{\Lambda}_0(t_i^*) = \frac{-\log \left\{ [1 - U_i] \exp \left(-\hat{\Lambda}_0^C(T_i) \exp(\hat{\alpha}^T \mathbf{Z}_i) \right) \right\}}{\exp(\hat{\alpha}^T \mathbf{Z}_i)},$$

where U_i is a random draw from a Uniform(0,1) distribution.

2C. Suppose that $\hat{\Lambda}_0^C(t_1) < \hat{\Lambda}_0^C(t) < \hat{\Lambda}_0^C(t_2)$ where t_1 and t_2 are times when censoring has occurred and that no censoring occurs between t_1 and t_2 . Then the imputed

censoring time for subject i is

$$C_i^m = t_1 + (t_2 - t_1) \frac{\hat{\Lambda}_0^C(t_i^*) - \hat{\Lambda}_0^C(t_1)}{\hat{\Lambda}_0^C(t_2) - \hat{\Lambda}_0^C(t_1)}.$$

3. To generate one imputed dataset consisting of $\{\mathbf{Z}_i, X_i^m, \Delta_i\}, i = 1, \dots, n$, repeat Step 2 for all subjects who experience a competing event. This dataset only contains the event of interest and censoring, resembling the univariate survival setting with no competing risks.

4. Apply methods for subdistribution hazard regression to the resulting dataset, $\{\mathbf{Z}_i, X_i^m, \Delta_i\}, i = 1, \dots, n$. We fit the model proposed by *Fine and Gray* (1999) by using standard software for fitting Cox proportional hazard models to estimate β_k^m as the solution to $U_k^m(\beta) = \mathbf{0}$, where

$$\mathbf{U}_k^m(\beta) = \sum_{i=1}^n \int_0^\tau \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n Y_{jk}^m(s) \mathbf{Z}_j \exp\{\beta^T \mathbf{Z}_j\}}{\sum_{j=1}^n Y_{jk}^m(s) \exp\{\beta^T \mathbf{Z}_j\}} \right] dN_{ik}(s),$$

with $Y_{ik}^m(t) = I(C_i^m \geq t)(1 - N_{ik}(t-))$. An estimator of $V(\hat{\beta}_k^m)$ is given by the estimated covariance matrix

$$\hat{\Sigma}_k^m = \left[\sum_{i=1}^n \int_0^\tau \left\{ \frac{\sum_{j=1}^n \mathbf{Z}_j^{\otimes 2} Y_{jk}^m(t) \exp(\hat{\beta}_k^m)^T \mathbf{Z}_j}{\sum_{j=1}^n Y_{jk}^m(t) \exp(\hat{\beta}_k^m)^T \mathbf{Z}_j} - \bar{\mathbf{Z}}^{\otimes 2}(t; \hat{\beta}_k^m) \right\} dN_{ik}(t) \right]^{-1},$$

where

$$\bar{\mathbf{Z}}(t; \beta_k) = \frac{\sum_{i=1}^n \mathbf{Z}_i Y_{ik}^m(t) \exp(\beta_k^T \mathbf{Z}_i)}{\sum_{i=1}^n Y_{ik}^m(t) \exp(\beta_k^T \mathbf{Z}_i)}.$$

The event of interest is $I(\Delta_i = k)$.

5. Repeat Steps 2 through 4 a total of M times, to produce M sets of parameter estimates $\{\hat{\beta}_k^1, \dots, \hat{\beta}_k^M\}$ and covariance estimators $\{\hat{\Sigma}_k^1, \dots, \hat{\Sigma}_k^M\}$.

4.2.3 Inference Procedures

Results from the M imputed datasets are then combined according to appropriate procedures for imputed data to provide the relevant estimates and inference. Specifically, the well-established variance formula of *Rubin* (1987) was developed for proper imputation and can then be applied in our set up. The parameter estimate will be the mean of the parameter estimates from Step 1 across the M augmented datasets,

$$\hat{\beta}_k = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_k^m.$$

The variance estimator is calculated by combining the within- and between-imputation variation as follows,

$$\frac{1}{M} \sum_{m=1}^M \hat{\Sigma}_k^m + \left(1 + \frac{1}{M}\right) \left[\frac{1}{m-1} \sum_{m=1}^M (\hat{\beta}_k^m - \hat{\beta}_k)^{\otimes 2} \right].$$

Note that the imputation steps proposed in this chapter randomly draw from the estimated censoring distribution, \hat{G}_i . Standard imputation requires that the imputed censoring times are drawn from the asymptotic distribution of the censoring distri-

bution. Our imputation scheme introduces additional variability, which comes from the estimation of the censoring distribution. Therefore it is necessary to develop a variance estimator that takes into account this additional variability. Instead of drawing from the distribution of the censoring variable, we bypass the parametrization aspect by using the bootstrap method by drawing with replacement from the observed data. The parameter coefficients resulting from the bootstrap datasets are used as the draws in a proper imputation.

4.3 Simulation Studies

Simulation studies were carried out to evaluate the performance of the imputation procedure in finite samples. Data were generated according to the structure described in the simulation study of *Fine and Gray* (1999), except instead of using two covariates, we used three binary covariates. There are two causes of failure, with cause $k = 1$ serving as the cause of interest. The cumulative incidence function for cause $k = 1$ is given by

$$P(T_i \leq t, \Delta_i = 1 | \mathbf{Z}_i) = 1 - \{1 - p [1 - \exp(-t)]\}^{\exp(\beta_1^T \mathbf{Z}_i)},$$

where $p \in (0, 1)$ and with $p = P(\Delta_i = 1 | \mathbf{Z}_i = \mathbf{0})$. For events of cause $k = 2$, the conditional cumulative incidence function is given by $P(T_i \leq t | \Delta_i = 2, \mathbf{Z}_i) = 1 - \exp^{-\exp(\beta_1^T \mathbf{Z}_i) t}$, which is an exponential distribution with rate $\exp(\beta_2^T \mathbf{Z}_i)$. The covariate vector was specified as $\mathbf{Z}_i = [Z_{i1}, Z_{i2}, Z_{i3}]^T$, where $Z_{i1} \sim \text{Ber}(\theta_1)$, $Z_{i2} | Z_{i1} \sim$

$\text{Ber}(Z_{i1}\theta_{21} + (1 - Z_{i1})\theta_{22})$, and $Z_{i3}|Z_{i2} \sim \text{Ber}(Z_{i2}\theta_{31} + (1 - Z_{i2})\theta_{32})$. The values of the parameters are $\theta_1 = 0.05, \theta_{21} = 0.55, \theta_{22} = 0.45, \theta_{31} = 0.45$, and $\theta_{32} = 0.65$. Censoring is dependent on the covariates and follows an exponential distribution with $\lambda_i^C(t) = \lambda_0^C \exp(\alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3)$, with $\lambda_0^C = 0.09$. The administrative censoring is set at $t = 10$ and $M = 10$ imputes were used. Several configurations were included, where the extent of the censoring and cause $k = 2$ events were varied.

Table 4.1: Simulation results for proposed estimator for five configurations, with bias, asymptotic standard error (ASE), empirical standard deviation (ESD), and 95% confidence interval coverage probabilities (CP).

| Configuration | | | True value | BIAS | ASE | ESD | CP | |
|---------------|-----------|--------|--------------|------|--------|------|------|------|
| 1 | Censored: | 15.3 % | β_{11} | 0.2 | 0.032 | 0.49 | 0.46 | 0.95 |
| | Event 1: | 57.6 % | β_{12} | 0.3 | 0.014 | 0.21 | 0.19 | 0.96 |
| | Event 2: | 27.1 % | β_{13} | 0.4 | -0.006 | 0.22 | 0.20 | 0.98 |
| 2 | Censored: | 15.0 % | β_{11} | 0.4 | 0.018 | 0.45 | 0.46 | 0.94 |
| | Event 1: | 59.4 % | β_{12} | 0.3 | 0.009 | 0.21 | 0.18 | 0.97 |
| | Event 2: | 25.5 % | β_{13} | 0.46 | 0.003 | 0.22 | 0.20 | 0.96 |
| 3 | Censored: | 14.1 % | β_{11} | 0.5 | -0.000 | 0.43 | 0.42 | 0.95 |
| | Event 1: | 64.2 % | β_{12} | 0.4 | -0.001 | 0.20 | 0.19 | 0.96 |
| | Event 2: | 21.7 % | β_{13} | 0.65 | 0.017 | 0.21 | 0.19 | 0.96 |
| 4 | Censored: | 32.2 % | β_{11} | 0.5 | -0.040 | 0.47 | 0.45 | 0.96 |
| | Event 1: | 50.8 % | β_{12} | 0.4 | 0.004 | 0.22 | 0.21 | 0.97 |
| | Event 2: | 17.1 % | β_{13} | 0.65 | 0.019 | 0.23 | 0.22 | 0.95 |
| 5 | Censored: | 9.1 % | β_{11} | 0.5 | 0.023 | 0.41 | 0.41 | 0.95 |
| | Event 1: | 67.9 % | β_{12} | 0.4 | 0.017 | 0.19 | 0.20 | 0.95 |
| | Event 2: | 23.0 % | β_{13} | 0.65 | 0.005 | 0.20 | 0.19 | 0.96 |

Table 4.1 summarizes the performance of the proposed parameter estimators. The

bias of the estimated coefficients is generally quite low. As the percent of observations that were censored increased, the bias increased slightly. The coverage probabilities generally attain and sometimes exceed the nominal value. The percent censored may affect the imputation since the censoring distribution is estimated from the observations that were censored. In addition, the percent of observations that were of cause $k = 2$ also affected the bias. The more cause $k = 2$ events, the more imputations need to be made, which could affect the accuracy of the resulting estimate.

4.4 Application

We applied the proposed method to kidney transplant data from the Scientific Registry of Transplant Recipients (SRTR). Patients who were waitlisted at any of the 6 OPOs in Region 10, which contains Michigan, were eligible to be included. The final study sample consisted of the $n = 22,685$ patients who were wait-listed for a kidney transplant between January 1, 2000 and December 31, 2009. We are primarily interested in the probability of receiving a transplant from a deceased donor, with removal from the wait-list and death on the wait-list as competing risks. The covariate vector contained terms representing OPO, age, sex, body mass index (BMI), race, hypertension status, blood type, diabetes, and panel reactive antibody. Since there were six OPOs, five indicator variables were used to estimate the OPO effect.

The reference categories are OPO 6 for location, glomerulonephritis for the primary renal disease diagnosis, between 25 and 30 for BMI, female for sex, Caucasian for

Table 4.2: Analysis of SRTR data using proposed method with $M = 10$.

| Covariate | Mean, n, % | Parameter estimate | P-value | Hazard Ratio |
|----------------------------------|---------------|-----------------------|----------------|--------------|
| Age at listing | 48.1 | -0.008 | < 0.001 | 0.99 |
| Panel reactive antibody | 17.9 | -0.008 | < 0.001 | 0.99 |
| Location: | | | | |
| OPO 1 | 1434 | 0.43 | < 0.001 | 1.53 |
| OPO 2 | 1551 | 0.70 | < 0.001 | 1.99 |
| OPO 3 | 2420 | 0.27 | < 0.001 | 1.32 |
| OPO 4 | 3462 | 0.60 | < 0.001 | 1.81 |
| OPO 5 | 4549 | -0.02 | 0.54 | 0.97 |
| OPO 6 | 9269 | 0 | - | 1 |
| BMI | | | | |
| Less than 20 | 5.2 % | 0.01 | 0.059 | 1.11 |
| 20 to 25 | 26.5 % | 0.03 | 0.30 | 1.03 |
| 25 to 30 | 32.5 % | 0 | - | 1 |
| 30 to 35 | 23.2 % | -0.12 | < 0.001 | 0.89 |
| Greater than 35 | 12.6 % | -0.34 | < 0.001 | 0.71 |
| Primary renal disease diagnosis: | | | | |
| Polycystic kidney disease | 7.7 % | 0.08 | 0.22 | 1.09 |
| Diabetes | 14.7 % | 0.17 | 0.004 | 1.19 |
| Hypertension | 17.9 % | 0.05 | 0.43 | 1.05 |
| Other diagnosis | 54.4 % | 0.08 | 0.16 | 1.08 |
| Glomerulonephritis | 5.28 % | 0 | - | 1 |
| Sex: | | | | |
| Male | 59.5 % | -0.02 | 0.44 | 0.98 |
| Female | 40.5% | 0 | - | 1 |
| Race: | | | | |
| African American | 30.4 % | -0.23 | < 0.001 | 0.79 |
| Hispanic | 2.6 % | -0.08 | 0.28 | 0.93 |
| Asian | 1.6 % | 0.01 | 0.90 | 1.02 |
| Caucasian | 64.9 % | 0 | - | 1 |
| Blood type: | | | | |
| Blood type A | 30.8 % | 0.33 | < 0.001 | 1.38 |
| Blood type AB | 3.7 % | 0.71 | < 0.001 | 2.02 |
| Blood type B | 13.8 % | -0.06 | 0.67 | 0.94 |
| Blood type O | 46.7 % | 0 | - | 1 |

race, and type O for blood type. From Table 4.2, age at listing, BMI greater than 30, panel reactive antibody, and being African American have significant negative effects on the overall probability of receiving a kidney transplant. Being in OPO 1 through OPO 4, being diagnosed with diabetes, and having blood type A or AB have significant positive effect on the overall probability of receiving a kidney transplant. The effects of being in OPO 5, male, being Hispanic, and having blood type B, although negative, are not significant, while the effects of being Asian, BMI less than 25, being diagnosed either with polycystic kidney disease, hypertension, or any other primary renal disease diagnosis are positive but not significant.

To revisit the motivating question, compared to individuals with blood type O, those with blood types A or AB are significantly more likely to receive a kidney transplant, after adjusting for relevant covariates. Interestingly, those with blood type B actually have a lower probability of receiving a transplant compared to those with type O, though this advantage is not significant. Thus, we do have evidence that patients with blood type O are disadvantaged in terms of kidney transplantation, relative to types A and AB.

4.5 Discussion

In this chapter we present a method of creating censoring-complete data from right-censored survival data by imputing potential censoring times for the subjects observed to experience a competing risk. Censoring-complete data do not occur very

frequently in biomedical data since random loss to follow-up is often present. However, by generating censoring-complete data, the subdistribution hazard models proposed by *Fine and Gray* (1999) can be fitted without inverse weighting. The imputed time is drawn from the conditional survival function of the censoring random variable given the censoring time is greater than the observed time of the competing risk. The cumulative baseline hazard function is estimated by interpolating the values at which events of interest occurs, presumably more realistically reflecting the true form than a step function.

The procedures outlined in this chapter allow for the imputation step to incorporate the effect of the covariates. The variance is estimated by Rubin's method (*Rubin*, 1987), with its application justified by obtaining parameter estimates by bootstrapping the observed data. The simulation results demonstrated that the point estimator has relatively small bias and that the variance estimator provides confidence intervals with satisfactory coverage. The performance of the proposed method was found to be consistent across various percentages of censoring and percentages of events not-of-interest.

The imputation procedures described in this chapter consider the would-be censoring time as missing data and impute censoring times for subjects observed to experience a competing risk. In most cases where multiple imputation techniques are applied to univariate survival data, event times are imputed for censored observations and the potential event times are considered to be missing data.

The methods presented in this chapter can be thought of as a generalization of

methods developed by *Ruan and Gray* (2008). A difference between the two methods is the way in which the censoring distribution is estimated. This chapter proposes a Cox regression model, hence allowing censoring to depend on covariates, whereas *Ruan and Gray* (2008) uses the Kaplan-Meier estimator.

CHAPTER V

Conclusion

This dissertation develops three methods for the competing risks data structure, which arises naturally in organ transplantation as well as numerous other biomedical settings. The patient experience while on the wait-list for kidney transplantation provides the motivating medical application area for this dissertation, and we apply each of the proposed methods to data from the Scientific Registry of Transplant Recipients (SRTR). In this context, a patient is entered into a wait-list when he/she has a need for a kidney transplant and is determined to be healthy enough to undergo the transplant surgery. While on the wait-list, a patient can experience death, removal (due to the deterioration of health), or receipt of a transplant. If either death or removal from the wait-list occurs, then a transplant cannot occur.

This dissertation focuses on the cumulative incidence function (CIF), which has received a fair amount of attention in survival literature in recent years. The cumulative incidence can be thought of as the probability of a particular event occurring after taking into account the possibility of competing events. The first two methods aim

to compare subgroups with respect to the cumulative incidence of the event of chief interest. The methods are designed for observational studies, in which the factor defining subgroups is not randomized. The third method imputes censoring times, so that unweighted methods can then be applied for parameter estimation. Large sample properties are derived and simulation results are carried out for each of the proposed methods.

In Chapter II, we established a method that compares group-specific CIFs. The cause-specific hazard of each cause is modeled through Cox regression, stratified by group. Subsequently, the CSHs are combined to provide an estimated CIF for the cause of interest. Averaging the CIF fitted values provides an estimate of the average cumulative incidence given the patient profile of a particular group. Analogous to indirect standardization, the effect estimate for a particular group captures the difference between the scenarios where patients from that group are (i) actually members of that group, versus (ii) members of a hypothetical group that reflects the overall average cumulative incidence.

Chapter III developed a method with a similar goal to Chapter II, but employing direct standardization. Thus, the key comparison is between the current overall average, and what the average would be if all subjects were in fact members of the index subgroup. Estimators of group effects on cumulative incidence are nonparametric and based on a weighted version of the counting process for the event of interest. In particular, Inverse Probability of Censoring Weighting (IPCW) is employed due to the presence of right censoring, while Inverse Probability of Treatment

Weighting (IPTW) is used to account for differences in covariate distributions across groups.

In Chapter IV, we developed a multiple imputation approach to analyzing competing risks data, through which factors affecting cumulative incidence of the cause of interest can be identified. A censoring-complete dataset is created by imputing censoring times for subjects who were observed to experience a competing event. The Cox proportional hazards model is then used to estimate the censoring distribution that provides the potential censoring times. Methods developed for the subdistribution hazard can be applied, with a specific example being the proportional subdistribution hazard model developed by Fine and Gray (1999).

Chapters II and III propose approaches to estimate group effects. There are not many existing methods that compare covariate-adjusted group-specific CIFs, and the methods in these two chapters make a contribution in such settings with direct and indirect standardization. In Chapter IV, we have developed a covariate-adjusted computational technique for competing risks data that may be preferable to other methods involving inverse weighting. The inverse weighting procedures are computationally intensive and the weight changes as a function of time for each individual.

In Chapters II and III, we have used the bootstrap method to estimate the variability of the point estimators. Another option would be to program the respective theoretical variances. Since this would be very computationally intensive, it may be possible to ignore some terms that can be empirically demonstrated to be negligible. For example, the terms associated with the variability of the IPTW and IPCW

in Chapter III may be relatively minor compared to the variability of the counting process elements. Additionally, in Chapter IV, it would be desirable to devise a technique to sample from the joint distribution of $\hat{\beta}$ and $\hat{\Lambda}_0^C$ in the imputation step. This may save computation time compared to the currently proposed bootstrap (the latter being infinite dimensional).

We used the Cox proportional hazards model to model the cause-specific hazards in Chapter II, and to model the censoring distribution in Chapters III and IV. The Cox model was chosen due to its flexibility and widespread popularity. However, it is possible that the covariates act additively, as opposed to multiplicatively, in which case the additive hazards models of

APPENDICES

APPENDIX A

Proof of Theorem 2.1

To begin, we review the essential notation. The models assumed for the cause-specific hazard are given by $\lambda_{ijk}(t) = \lambda_{0jk} \exp\{\boldsymbol{\beta}_k^T \mathbf{Z}_i\}$ and the risk-set and counting process-related quantities are as follows:

$$\begin{aligned}
Y_i(t) &= I(X_i \geq t) \\
Y_{ij}(t) &= Y_i(t) \cdot I(A_i = j) \\
\Delta_i &= kI(T_i < C_i) \\
\Delta_{ik} &= I(\Delta_i = k) \\
N_{ik}(t) &= I(X_i \leq t)\Delta_{ik} \\
N_{ijk}(t) &= N_{ik}(t)I(A_i = j) \\
R_{jk}^{(d)}(t; \boldsymbol{\beta}_k) &= \frac{1}{n} \sum_{i=1}^n Y_{ij}(t) \mathbf{Z}_i^{\otimes d} \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \\
r_{jk}^{(d)}(t; \boldsymbol{\beta}_k) &= E \left[Y_{ij}(t) \mathbf{Z}_i^{\otimes d} \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \right] \\
\bar{\mathbf{Z}}_{jk}(t; \boldsymbol{\beta}_k) &= \frac{R_{jk}^{(1)}(t; \boldsymbol{\beta}_k)}{R_{jk}^{(0)}(t; \boldsymbol{\beta}_k)} \\
\bar{\mathbf{z}}_{jk}(t; \boldsymbol{\beta}_k) &= \frac{r_{jk}^{(1)}(t; \boldsymbol{\beta}_k)}{r_{jk}^{(0)}(t; \boldsymbol{\beta}_k)} \\
dM_{ijk}(t) &= dN_{ijk}(t) - Y_{ij}(t) d\Lambda_{ijk}^{\#}(t).
\end{aligned}$$

The processes $M_{ijk}(t; \boldsymbol{\beta}_1)$, $k = 1, 2$ are martingales with respect to the filtration

$$\mathcal{F}_{ij}(t) = \sigma \{Y_{ij}(s), \mathbf{Z}_i; s \in (0, t]\}.$$

The proof of Theorem 2.1 revolves around asymptotic expansions of the following quantities.

1. $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$

$$2. \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^{\#}(t) - \Lambda_{0jk}^{\#}(t) \right\}$$

$$3. \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{ijk}^{\#}(t) - \Lambda_{ijk}^{\#}(t) \right\}$$

$$4. \quad n^{\frac{1}{2}} \left\{ \hat{S}_{ij}(t) - S_{ij}(t) \right\}$$

$$5. \quad n^{\frac{1}{2}} \left\{ \hat{F}_{ijk}(t) - F_{ijk}(t) \right\}$$

$$6. \quad n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\}$$

For the remainder of the proof, we assume k is the cause of interest.

$$[1.] \quad n^{\frac{1}{2}}(\hat{\beta}_k - \beta_k)$$

By a Taylor expansion of $U_k(\beta)$ around β_k ,

$$\begin{aligned} U_k(\hat{\beta}_k) - U_k(\beta_k) &= \frac{\partial}{\partial \beta_k^T} [U_k(\beta)]_{\beta_k} (\hat{\beta}_k - \beta_k) \\ n^{\frac{1}{2}}(\hat{\beta}_k - \beta_k) &= \mathcal{I}_k^{-1}(\beta_k) n^{-\frac{1}{2}} U_k(\beta_k) \\ &= \mathcal{I}_k^{-1}(\beta_k) n^{-\frac{1}{2}} \sum_{i=1}^n U_{ik}(\beta_k) + o_p(1), \end{aligned}$$

where

$$U_{ik}(\beta_k) = \sum_{j=1}^J \int_0^{\tau} \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j(t; \beta_k) \} dM_{ijk}(t)$$

and

$$\mathcal{I}_k(\boldsymbol{\beta}_k) = E \left[\sum_{j=1}^J \int_0^\tau \left\{ \frac{r_j^{(2)}(t; \boldsymbol{\beta}_k)}{r_j^{(0)}(t; \boldsymbol{\beta}_k)} - \bar{\mathbf{z}}_j(t; \boldsymbol{\beta}_k)^{\otimes 2} \right\} dN_{ijk}(t) \right].$$

The result follows from standard Martingale theory; e.g., as described in Fleming and Harrington (1991).

$$[2.] \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t) - \Lambda_{0jk}^\#(t) \right\}$$

We decompose the quantity as follows,

$$\begin{aligned} n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t) - \Lambda_{0jk}^\#(t) \right\} &= n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t; \hat{\boldsymbol{\beta}}_k) - \Lambda_{0jk}^\#(t; \boldsymbol{\beta}_k) \right\} \\ &= n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t; \hat{\boldsymbol{\beta}}_k) - \hat{\Lambda}_{0jk}^\#(t; \boldsymbol{\beta}_k) \right\} \end{aligned} \quad (\text{A.1})$$

$$+ n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t; \boldsymbol{\beta}_k) - \Lambda_{0jk}^\#(t; \boldsymbol{\beta}_k) \right\}. \quad (\text{A.2})$$

Since $\hat{\Lambda}_{0jk}^\#(t)$ is the Breslow-Aalen analog of $\Lambda_{0jk}^\#(t)$, we adapt results derived for the Breslow-Aalen estimator (Fleming and Harrington, 1991). From this perspective,

$$(\text{A.1}) = n^{\frac{1}{2}} n^{-1} \sum_{i=1}^n \int_0^t \left[R_{jk}^{(0)}(s; \hat{\boldsymbol{\beta}})^{-1} - R_{jk}^{(0)}(s; \boldsymbol{\beta})^{-1} \right] dN_{ijk}(s).$$

Taylor expanding the function $\hat{\Lambda}_{0jk}^\#(t; \hat{\beta}_k)$ around β_k yields

$$\begin{aligned}\hat{\Lambda}_{0jk}^\#(t; \hat{\beta}_k) &= \hat{\Lambda}_{0jk}^\#(t; \beta_k) + \left. \frac{\partial}{\partial \beta_k} \hat{\Lambda}_{0jk}^\#(t; \beta_k) \right|_{\beta_k} (\hat{\beta}_k - \beta_k) \\ n^{\frac{1}{2}} [\hat{\Lambda}_{0jk}^\#(t; \hat{\beta}_k) - \hat{\Lambda}_{0jk}^\#(t; \beta_k)] &= \left. \frac{\partial}{\partial \beta_k} \hat{\Lambda}_{0jk}^\#(t; \beta_k) \right|_{\beta_k} n^{\frac{1}{2}} (\hat{\beta}_k - \beta_k) + o_p(1). \quad (\text{A.3})\end{aligned}$$

The required derivative is as follows:

$$\begin{aligned}\left. \frac{\partial}{\partial \beta_k} \hat{\Lambda}_{0jk}^\#(t; \beta_k) \right|_{\beta_k} &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\partial}{\partial \beta_k} R_{jk}^{(0)}(s; \beta_k)^{-1} dN_{ijk}(s) \\ &= -\frac{1}{n} \sum_{i=1}^n \int_0^t R_{jk}^{(0)}(s; \beta_k)^{-2} R_{jk}^{(1)}(s; \beta_k) dN_{ijk}(s) \\ &= -\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{R_{jk}^{(1)}(s; \beta_k)}{R_{jk}^{(0)}(s; \beta_k)} R_{jk}^{(0)}(s; \beta_k)^{-1} dN_{ijk}(s) \\ &= -\int_0^t \frac{R_{jk}^{(1)}(s; \beta_k)}{R_{jk}^{(0)}(s; \beta_k)} \frac{1}{n} \sum_{i=1}^n \frac{dN_{ijk}(s)}{R_{jk}^{(0)}(s; \beta_k)} \\ &= -\int_0^t \bar{\mathbf{Z}}_j(s; \beta_k) d\hat{\Lambda}_{0jk}(t; \beta_k) \\ &\equiv \hat{\mathbf{h}}_{jk}(t; \beta_k).\end{aligned}$$

The vector $\bar{\mathbf{Z}}_j(s; \beta_k) \xrightarrow{p} \bar{\mathbf{z}}_j(s; \beta_k)$ by repeated applications of the Continuous Mapping Theorem (CMT). In addition, $\hat{\Lambda}_{0jk}^\#(t; \beta_k) \xrightarrow{p} \Lambda_{0jk}^\#(t; \beta_k)$ due to established properties of the Breslow-Aalen estimator (e.g., Andersen and Gill, 1982). Thus, by the CMT and continuity:

$$\hat{\mathbf{h}}_{jk}(t; \beta_k) \xrightarrow{p} \mathbf{h}_{jk}(t; \beta_k),$$

where $\mathbf{h}_{jk}(t; \boldsymbol{\beta}_k) = -\int_0^t \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k) \cdot d\Lambda_{0jk}(t; \boldsymbol{\beta}_k)$. For the second term on the RHS of (A.3) we exploit Result [1] such that,

$$\begin{aligned} \text{(A.1)} &= \hat{\mathbf{h}}_{jk}^T(t; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{ik}(\boldsymbol{\beta}_k) + o_p(1) \\ &= \mathbf{h}_{jk}^T(t; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{ik}(\boldsymbol{\beta}_k) + o_p(1), \end{aligned}$$

where the last equality holds by Slutsky's Theorem. With respect to (A.2), we can write

$$\begin{aligned} \text{(A.2)} &= n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t; \boldsymbol{\beta}_k) - \Lambda_{0jk}^\#(t; \boldsymbol{\beta}_k) \right\} \\ &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \int_0^t R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dN_{ijk}(s) - \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)}{R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)} d\Lambda_{0jk}^\#(s; \boldsymbol{\beta}_k) \right\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \int_0^t R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dN_{ijk}(s) \times \right. \\ &\quad \left. - \int_0^t R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) d\Lambda_{0jk}^\#(s; \boldsymbol{\beta}_k) \right\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \int_0^t R_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ijk}(s; \boldsymbol{\beta}_k) \right\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ijk}(s; \boldsymbol{\beta}_k) + o_p(1), \end{aligned}$$

where the last equality holds because $R_{jk}^{(0)}(t) \xrightarrow{p} r_{jk}^{(0)}(t)$. Putting (A.1) and (A.2) together, we get:

$$n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0jk}^\#(t) - \Lambda_{0jk}^\#(t) \right\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_{ijk}(t; \boldsymbol{\beta}_k) + o_p(1),$$

where $\Phi_{ijk}(t; \boldsymbol{\beta}_k) = -\mathbf{h}_{jk}^T(t; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} U_{ik}(\boldsymbol{\beta}_k) + \int_0^t r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ijk}(s; \boldsymbol{\beta}_k)$.

$$[3.] \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{ijk}^{\#}(t) - \Lambda_{ijk}^{\#}(t) \right\}$$

We can write the following,

$$[3.] = n^{\frac{1}{2}} \left\{ \int_0^t Y_{ij}(s) \exp(\hat{\boldsymbol{\beta}}_k^T \mathbf{Z}_i) d\hat{\Lambda}_{0jk}^{\#}(s; \hat{\boldsymbol{\beta}}_k) - \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) d\hat{\Lambda}_{0jk}^{\#}(s; \hat{\boldsymbol{\beta}}_k) \right\} \quad (\text{A.4})$$

$$+ n^{\frac{1}{2}} \left\{ \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) d\hat{\Lambda}_{0jk}^{\#}(s; \hat{\boldsymbol{\beta}}_k) - \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) d\Lambda_{0jk}^{\#}(s; \boldsymbol{\beta}_k) \right\}. \quad (\text{A.5})$$

We can express the first term as follows,

$$(\text{A.4}) = n^{\frac{1}{2}} \int_0^t \left\{ \exp(\hat{\boldsymbol{\beta}}_k^T \mathbf{Z}_i) - \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \right\} Y_{ij}(s) d\hat{\Lambda}_{0jk}^{\#}(s; \hat{\boldsymbol{\beta}}_k).$$

Doing a Taylor expansion on the function $\exp(\hat{\boldsymbol{\beta}}_k^T \mathbf{Z}_i)$ around the value $\boldsymbol{\beta}_k$,

$$\exp(\hat{\boldsymbol{\beta}}_k^T \mathbf{Z}_i) = \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) + \mathbf{Z}_i \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \left[\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k \right] + o_p(1).$$

Then, using Result [1],

$$n^{\frac{1}{2}} \left\{ \exp(\hat{\boldsymbol{\beta}}_k^T \mathbf{Z}_i) - \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \right\} = \mathbf{Z}_i^T \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_{ik}(\boldsymbol{\beta}_k) + o_p(1).$$

We can then write

$$\begin{aligned} \text{(A.4)} &= \int_0^t \mathbf{Z}_i^T \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) Y_{ij}(s) d\hat{\Lambda}_{0jk}^\#(s; \hat{\boldsymbol{\beta}}_k) \\ &= \int_0^t Y_{ij}(s) \mathbf{Z}_i^T d\hat{\Lambda}_{ijjk}^\#(s; \hat{\boldsymbol{\beta}}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \\ &= \int_0^t \mathbf{Z}_i^T Y_{ij}(s) d\Lambda_{ijjk}^\#(s; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) + o_p(1), \end{aligned}$$

where the last equality holds by the convergence in probability of $\hat{\Lambda}_{ijjk}^\#(t)$ to $\Lambda_{ijjk}^\#(t)$.

We re-express (A.5) as

$$\begin{aligned} \text{(A.5)} &= n^{\frac{1}{2}} \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \left\{ d\hat{\Lambda}_{0jk}^\#(s; \hat{\boldsymbol{\beta}}_k) - d\Lambda_{0jk}^\#(s; \boldsymbol{\beta}_k) \right\} \\ &= n^{\frac{1}{2}} \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) d \left\{ \hat{\Lambda}_{0jk}^\#(s; \hat{\boldsymbol{\beta}}_k) - \Lambda_{0jk}^\#(s; \boldsymbol{\beta}_k) \right\} \\ &= \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n d\Phi_{ljk}(t; \boldsymbol{\beta}_k) \right\} + o_p(1), \end{aligned}$$

where the last line involves use of Result [2], and we define

$$d\Phi_{ijk}(t; \boldsymbol{\beta}_k) = [-\bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k) d\Lambda_{0jk}(s; \boldsymbol{\beta}_k)]^T \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} U_{ik}(\boldsymbol{\beta}_k) + r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ijk}(s; \boldsymbol{\beta}_k).$$

Combining (A.4) and (A.5), we get:

$$\begin{aligned}
[3] &= \int_0^t \mathbf{Z}_i Y_{ij}(s) d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \\
&\quad + \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) n^{-\frac{1}{2}} \left\{ \sum_{l=1}^n d\Phi_{ljk}(s; \boldsymbol{\beta}_k) \right\} + o_p(1) \\
&= \int_0^t \mathbf{Z}_i Y_{ij}(s) d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \\
&\quad + \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) n^{-\frac{1}{2}} \\
&\quad - \sum_{l=1}^n \left\{ \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k) d\Lambda_{0jk}(s; \boldsymbol{\beta}_k) \mathcal{I}_n(\boldsymbol{\beta}_k)^{-1} U_{lk}(\boldsymbol{\beta}_k) \right. \\
&\quad \left. + r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k) \right\} \\
&= \int_0^t \mathbf{Z}_i Y_{ij}(s) d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \\
&\quad - n^{-\frac{1}{2}} \int_0^t d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k) \sum_{l=1}^n \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} U_{lk}(\boldsymbol{\beta}_k) \\
&\quad + n^{-\frac{1}{2}} \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) \left\{ \sum_{l=1}^n r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k) \right\} \\
&= \int_0^t \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k) \}^T d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \\
&\quad + n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k) \\
&= \mathbf{D}_{ijk}^T(t; \boldsymbol{\beta}_k) \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{iljk}(t; \boldsymbol{\beta}_k),
\end{aligned}$$

where we let

$$\begin{aligned}\mathbf{D}_{ijk}(t; \boldsymbol{\beta}_k) &= \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k)\} d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \\ J_{iljk}(t; \boldsymbol{\beta}_k) &= \int_0^t Y_{ij}(s) \cdot \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k).\end{aligned}$$

$$[4.] \quad n^{\frac{1}{2}} \left\{ \hat{S}_{ij}(t) - S_{ij}(t) \right\}$$

We decompose [4] as follows,

$$[4] = - \sum_{m=1}^K S_{ij}(t) n^{-1/2} \left\{ \hat{\Lambda}_{ijm}^\#(t) - \Lambda_{ijm}^\#(t) \right\},$$

due to the Functional Delta Method, combined with the convergence, $e^{-\hat{\Lambda}_{ijm}^\#(s; \hat{\boldsymbol{\beta}}_m)} \xrightarrow{p} e^{-\Lambda_{ijm}^\#(s; \boldsymbol{\beta}_m)}$, by the CMT. Using Result [3], we obtain

$$[4] = -S_{ij}(t) \sum_{m=1}^K \left\{ \mathbf{D}_{ijm}^T(t; \boldsymbol{\beta}_m) \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lm}(\boldsymbol{\beta}_m) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{iljm}(t; \boldsymbol{\beta}_m) \right\}.$$

$$[5.] \quad n^{\frac{1}{2}} \left\{ \hat{F}_{ijk}(t) - F_{ijk}(t) \right\}$$

We can write

$$\begin{aligned}
[5] &= n^{\frac{1}{2}} \left\{ \int_0^t \hat{S}_{ij}(s; \hat{\beta}) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) - \int_0^t S_{ij}(s; \beta) d\Lambda_{ijk}^{\#}(s; \beta_k) \right\} \\
&= n^{\frac{1}{2}} \left\{ \int_0^t \hat{S}_{ij}(s; \hat{\beta}) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) - \int_0^t S_{ij}(s; \beta) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) \right\} \quad (\text{A.6})
\end{aligned}$$

$$+ n^{\frac{1}{2}} \left\{ \int_0^t S_{ij}(s; \beta) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) - \int_0^t S_{ij}(s; \beta) d\Lambda_{ijk}^{\#}(s; \beta_k) \right\}, \quad (\text{A.7})$$

where $\beta^T = [\beta_1^T, \dots, \beta_K^T]$. Note that equation (A.6) will eventually give rise to $\phi_{ilk}^1(t, \beta)$, $\phi_{ilk}^2(t, \beta)$ as defined in Theorem 2.1, while (A.7) will give rise to $\phi_{ilk}^3(t, \beta)$ and $\phi_{ilk}^4(t, \beta)$. We can write (A.6) as

$$\begin{aligned}
(\text{A.6}) &= \int_0^t n^{\frac{1}{2}} [\hat{S}_{ij}(s; \hat{\beta}) - S_{ij}(s; \beta)] d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) \\
&= - \sum_{m=1}^K \int_0^t S_{ij}(s) \mathbf{D}_{ijm}^T(s; \beta_m) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k) \times \\
&\quad \mathcal{I}_m(\beta_m)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lm}(\beta_m) \quad (\text{A.8})
\end{aligned}$$

$$- n^{-\frac{1}{2}} \sum_{m=1}^K \int_0^t S_{ij}(s) \sum_{l=1}^n J_{ilm}(s; \beta_m) d\hat{\Lambda}_{ijk}^{\#}(s; \hat{\beta}_k), \quad (\text{A.9})$$

where we have used Result [4]. Focusing on (A.8), we have,

$$(\text{A.8}) = - \sum_{m=1}^K \int_0^t \mathbf{D}_{ijm}^T(s; \beta_m) dF_{ijk}(s) \mathcal{I}_m(\beta_m)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lm}(\beta_m) + o_p(1),$$

by the fact that $\hat{\Lambda}_{ijk}^{\#}(s; \beta_k) \xrightarrow{p} \Lambda_{ijk}^{\#}(s; \beta_k)$, and $\hat{\beta}_k \xrightarrow{p} \beta_k$ and so, by the CMT,

$\hat{\Lambda}_{ijk}^\#(s; \hat{\boldsymbol{\beta}}_k) \xrightarrow{p} \Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k)$. Therefore, from (A.8), define

$$\begin{aligned}
\phi_{ijk}^1(t, \boldsymbol{\beta}) &= - \sum_{m=1}^K \int_0^t \mathbf{D}_{ijm}^T(s; \boldsymbol{\beta}_m) dF_{ijk}(s) \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_1) + o_p(1) \\
&= - \sum_{m=1}^K \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(u; \boldsymbol{\beta}_k)\}^T dF_{ijk}(s) \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_m) \\
&= - \sum_{m=1}^K \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(u; \boldsymbol{\beta}_k)\}^T \{F_{ijk}(t) - F_{ijk}(u)\} \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_m) \\
&= \sum_{m=1}^K \left\{ \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(u; \boldsymbol{\beta}_k)\}^T F_{ijk}(u) \right. \\
&\quad \left. - F_{ijk}(t) \int_0^t \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(u; \boldsymbol{\beta}_k)\}^T d\Lambda_{ijk}^\#(u; \boldsymbol{\beta}_k) \right\} \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_m).
\end{aligned}$$

$$\begin{aligned}
(A.9) &= - \sum_{m=1}^K n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t e^{-\sum_{k=1}^2 \Lambda_{ijk}(s; \beta_k)} J_{iljm}(s; \beta_m) d\hat{\Lambda}_{ijk}^\#(s; \hat{\beta}_k) \\
&= - \sum_{m=1}^K n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t e^{-\sum_{k=1}^2 \Lambda_{ijk}(s; \beta_k)} \left\{ \int_0^s Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m) \right\} \\
&\quad d\hat{\Lambda}_{ijk}^\#(s; \hat{\beta}_k) + o_p(1) \tag{A.10} \\
&= - \sum_{m=1}^K n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t \left\{ \int_u^t e^{-\sum_{k=1}^2 \Lambda_{ijk}(s; \beta_k)} d\hat{\Lambda}_{ijk}^\#(s; \hat{\beta}_k) \right\} \times \\
&\quad Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m) \\
&= - \sum_{m=1}^K n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t \{F_{ijk}(t) - F_{ijk}(u)\} Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m) \\
&= - \sum_{m=1}^K n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ F_{ijk}(t) \int_0^t Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m) \right. \\
&\quad \left. - \int_0^t F_{ijk}(u) Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m) \right\},
\end{aligned}$$

where (A.10) holds by the properties of $\hat{\Lambda}_{ijk}^\#(s; \beta_k)$ and $\hat{\beta}_k$. Thus, from (A.9) define:

$$\phi_{iljk}^2(t, \beta_1) = - \sum_{m=1}^K \int_0^t \{F_{ijk}(t) - F_{ijk}(u)\} Y_{ij}(u) \exp(\beta_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \beta_m)^{-1} dM_{ljm}(u; \beta_m).$$

Now shifting to (A.7):

$$\begin{aligned}
(A.7) &= n^{\frac{1}{2}} \int_0^t S_{ij}(s; \boldsymbol{\beta}) \left\{ d\hat{\Lambda}_{ijk}^\#(s; \hat{\boldsymbol{\beta}}_k) - d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \right\} \\
&= \int_0^t S_{ij}(s; \boldsymbol{\beta}) d \left[n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{ijk}^\#(s; \hat{\boldsymbol{\beta}}_k) - \Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \right\} \right] \int_0^t S_{ij}(s; \boldsymbol{\beta}) \\
&\quad \left[\mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) dD_{ijk}^T(s; \boldsymbol{\beta}_k) \right. \\
&\quad \left. + n^{-\frac{1}{2}} \sum_{l=1}^n dJ_{iljk}(s; \boldsymbol{\beta}_k) \right] \tag{A.11}
\end{aligned}$$

$$= \left\{ \int_0^t S_{ij}(s; \boldsymbol{\beta}) dD_{ijk}^T(s; \boldsymbol{\beta}_k) \right\} \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k) \tag{A.12}$$

$$+ n^{-\frac{1}{2}} \int_0^t S_{ij}(s; \boldsymbol{\beta}) \sum_{l=1}^n dJ_{iljk}(s; \boldsymbol{\beta}_k). \tag{A.13}$$

For equation (A.11) we have used Result [3]. We can then write

$$(A.12) = \left\{ \int_0^t S_{ij}(s; \boldsymbol{\beta}) [\mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k)]^T d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \right\} \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_{lk}(\boldsymbol{\beta}_k).$$

Corresponding to (A.12), we define:

$$\phi_{iljk}^3(t, \boldsymbol{\beta}) = \left\{ \int_0^t S_{ij}(s; \boldsymbol{\beta}) [\mathbf{Z}_i - \bar{\mathbf{z}}_j(s; \boldsymbol{\beta}_k)]^T d\Lambda_{ijk}^\#(s; \boldsymbol{\beta}_k) \right\} \mathcal{I}_k(\boldsymbol{\beta}_k)^{-1} U_{lk}(\boldsymbol{\beta}_k).$$

We can re-express (A.13) as follows,

$$(A.13) = n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t S_{ij}(s; \boldsymbol{\beta}) Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k),$$

and, correspondingly redefine

$$\phi_{iljk}^4(t, \boldsymbol{\beta}) = \int_0^t S_{ij}(s; \boldsymbol{\beta}_1) Y_{ij}(s) \exp(\boldsymbol{\beta}_k^T \mathbf{Z}_i) r_{jk}^{(0)}(s; \boldsymbol{\beta}_k)^{-1} dM_{ljk}(s; \boldsymbol{\beta}_k).$$

Combining results derived in this subsection, we obtain,

$$n^{\frac{1}{2}} \left\{ \hat{F}_{ijk}(t) - F_{ijk}(t) \right\} = n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \sum_{d=1}^4 \phi_{iljk}^d(t, \boldsymbol{\beta}) \right\} + o_p(1),$$

where the $\sum_{d=1}^4 \phi_{iljk}^d(t, \boldsymbol{\beta})$ are asymptotically independent and identically distributed variates with mean 0.

$$[6.] \quad n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\}$$

We now complete the proof by averaging over $i = 1, \dots, n$ to obtain the limiting distribution of the proposed estimator. Setting $n_j = \sum_{i=1}^n A_{ij}$ and $p_j = E[A_{ij}]$, we

have

$$\begin{aligned}
[6] &= n^{\frac{1}{2}} \left[\frac{1}{n_j} \sum_{i=1}^n A_{ij} \hat{F}_{ijk}(t) - \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left\{ \frac{1}{n} \sum_{r=1}^J \hat{F}_{irk}(t) n_r \right\} \right. \\
&\quad \left. - \left\{ \frac{1}{n_j} \sum_{i=1}^n A_{ij} F_{ijk}(t) - \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left[\frac{1}{n} \sum_{r=1}^J F_{irk}(t) n_r \right] \right\} \right] \\
&= \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left[n^{\frac{1}{2}} \left\{ \hat{F}_{ijk}(t) - F_{ijk}(t) \right\} \right. \\
&\quad \left. - \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left\{ \frac{1}{n} \sum_{r=1}^J n^{\frac{1}{2}} \left[\hat{F}_{irk}(t) - F_{irk}(t) \right] n_r \right\} \right] \\
&= \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left[n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \sum_{d=1}^4 \phi_{iljk}^d(t, \boldsymbol{\beta}) \right\} \right. \\
&\quad \left. - \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n \left[\frac{1}{n} \sum_{r=1}^J \left\{ \sum_{d=1}^4 \phi_{ilrk}^d(t, \boldsymbol{\beta}) \right\} n_r \right] \right\} \right] \\
&= \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n \left[\sum_{d=1}^4 \phi_{iljk}^d(t, \boldsymbol{\beta}) - \frac{1}{n} \sum_{r=1}^J \left\{ \sum_{d=1}^4 \phi_{ilrk}^d(t, \boldsymbol{\beta}) \right\} n_r \right] \right\} \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \frac{1}{n_j} \sum_{i=1}^n A_{ij} \left[\sum_{d=1}^4 \phi_{iljk}^d(t, \boldsymbol{\beta}) - \frac{1}{n} \sum_{r=1}^J \left\{ \sum_{d=1}^4 \phi_{ilrk}^d(t, \boldsymbol{\beta}) \right\} n_r \right] \right\} \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \sum_{d=1}^4 \left[\frac{1}{n_j} \sum_{i=1}^n A_{ij} \phi_{iljk}^d(t, \boldsymbol{\beta}) \right. \right. \\
&\quad \left. \left. - \frac{1}{n_j} \sum_{i=1}^n A_{ij} \frac{1}{n} \sum_{r=1}^J \phi_{ilrk}^d(t, \boldsymbol{\beta}) n_r \right] \right\}. \tag{A.14}
\end{aligned}$$

Focusing on each component in (A.14), we have the following for the expression

involving $\phi_{iljk}^1(t; \boldsymbol{\beta})$,

$$\begin{aligned}
& \frac{1}{n_j} \sum_{i=1}^n A_{ij} \phi_{iljk}^1(t; \boldsymbol{\beta}) \\
&= -\frac{1}{p_j} E \left[\sum_{m=1}^K A_{ij} \int_0^t D_{ijm}^T(s; \boldsymbol{\beta}_m) dF_{ijk}(s; \boldsymbol{\beta}_k) \right] \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_m) \\
&= \phi_{ljk}^1(t; \boldsymbol{\beta}),
\end{aligned}$$

since $n_j/n \xrightarrow{p} p_j$ by the Weak Law of Large Numbers (WLLN), continuity, and Slutsky's Theorem. The simplification for the term involving $\phi_{iljk}^3(t; \boldsymbol{\beta})$ unfold in a similar way. The term involving $\phi_{iljk}^2(t; \boldsymbol{\beta})$ can be written as the following:

$$\begin{aligned}
& \frac{1}{n_j} \sum_{i=1}^n A_{ij} \phi_{iljk}^2(t; \boldsymbol{\beta}) \\
&= -\sum_{m=1}^K \int_0^t \left[\frac{1}{p_j} \frac{1}{n} \sum_{i=1}^n A_{ij} \{F_{ijk}(t) - F_{ijk}(u)\} Y_{ij}(u) \exp(\boldsymbol{\beta}_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \boldsymbol{\beta}_m)^{-1} \right] dM_{ljm}(u; \boldsymbol{\beta}_m) \\
&= -\sum_{m=1}^K \int_0^t \frac{1}{p_j} E \left[\{F_{ijk}(t) - F_{ijk}(u)\} Y_{ij}(u) \exp(\boldsymbol{\beta}_m^T \mathbf{Z}_i) r_{jk}^{(0)}(u; \boldsymbol{\beta}_m)^{-1} \right] dM_{ljm}(u; \boldsymbol{\beta}_m) \\
&= \phi_{ljk}^2(t; \boldsymbol{\beta}),
\end{aligned}$$

by the WLLN, continuity, and Slutsky's Theorem. The term involving $\phi_{iljk}^4(t; \boldsymbol{\beta})$ unfold in a similar way. The term involving $\phi_{ilrk}^1(t; \boldsymbol{\beta})$ can be written as the follow-

ing:

$$\begin{aligned}
& \frac{1}{n_j} \sum_{i=1}^n \sum_{m=1}^K A_{ij} \left\{ \frac{1}{n} \sum_{r=1}^J \phi_{ilrk}^1(t; \boldsymbol{\beta}) n_r \right\} \\
&= -\frac{1}{p_j} \sum_{r=1}^J p_r E \left[\sum_{m=1}^K A_{ij} \int_0^t D_{irm}^T(s; \boldsymbol{\beta}_m) dF_{irk}(s; \boldsymbol{\beta}_k) \right] \mathcal{I}_m(\boldsymbol{\beta}_m)^{-1} U_{lm}(\boldsymbol{\beta}_m) \\
&= \phi_{lk}^1(t; \boldsymbol{\beta}).
\end{aligned}$$

The term involving $\phi_{ilrk}^3(t; \boldsymbol{\beta})$ can be expressed in a similar way. The term involving $\phi_{ilrk}^2(t; \boldsymbol{\beta})$ can be written as,

$$\begin{aligned}
& \frac{1}{n_j} \sum_{i=1}^n \sum_{m=1}^K A_{ij} \left\{ \frac{1}{n} \sum_{r=1}^J \phi_{ilrk}^3(t; \boldsymbol{\beta}) \cdot n_r \right\} \\
&= \sum_{m=1}^K \int_0^t \sum_{r=1}^J \frac{1}{p_j} p_r E \left[A_{ij} \{F_{irk}(t) - F_{irk}(u)\} Y_{ir}(u) \cdot \exp(\boldsymbol{\beta}_m^T \mathbf{Z}_i) r_{rk}^{(0)}(u; \boldsymbol{\beta}_m)^{-1} \right] dM_{lrm}(u; \boldsymbol{\beta}_m) \\
&= \phi_{lk}^2(t; \boldsymbol{\beta}).
\end{aligned}$$

The term involving $\phi_{ilrk}^4(t; \boldsymbol{\beta})$ can be expressed analogously. Therefore, we can write

$$[6] = n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \sum_{d=1}^4 \left(\phi_{ljk}^d(t; \boldsymbol{\beta}) + \phi_{lk}^d(t; \boldsymbol{\beta}) \right) \right\}. \quad (\text{A.15})$$

All summands across l have mean 0 since the ϕ 's have mean 0. If we apply the Functional Central Limit Theorem to [6], where each component is independent

across l , we have

$$Var \left(n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\} \right) = E \left[\left\{ \sum_{d=1}^4 \left(\phi_{ljk}^d(t; \boldsymbol{\beta}) - \phi_{lk}^d(t; \boldsymbol{\beta}) \right) \right\}^2 \right], \quad (\text{A.16})$$

with the corresponding estimator given by

$$\widehat{Var} \left(n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\} \right) = \frac{1}{n} \sum_{l=1}^n \left\{ \sum_{d=1}^4 \left(\hat{\phi}_{ljk}^d(t; \boldsymbol{\beta}) - \hat{\phi}_{lk}^d(t; \boldsymbol{\beta}) \right) \right\}^2. \quad (\text{A.17})$$

APPENDIX B

Analysis of SRTR data utilizing $J = 58$ OPOs for Chapter III

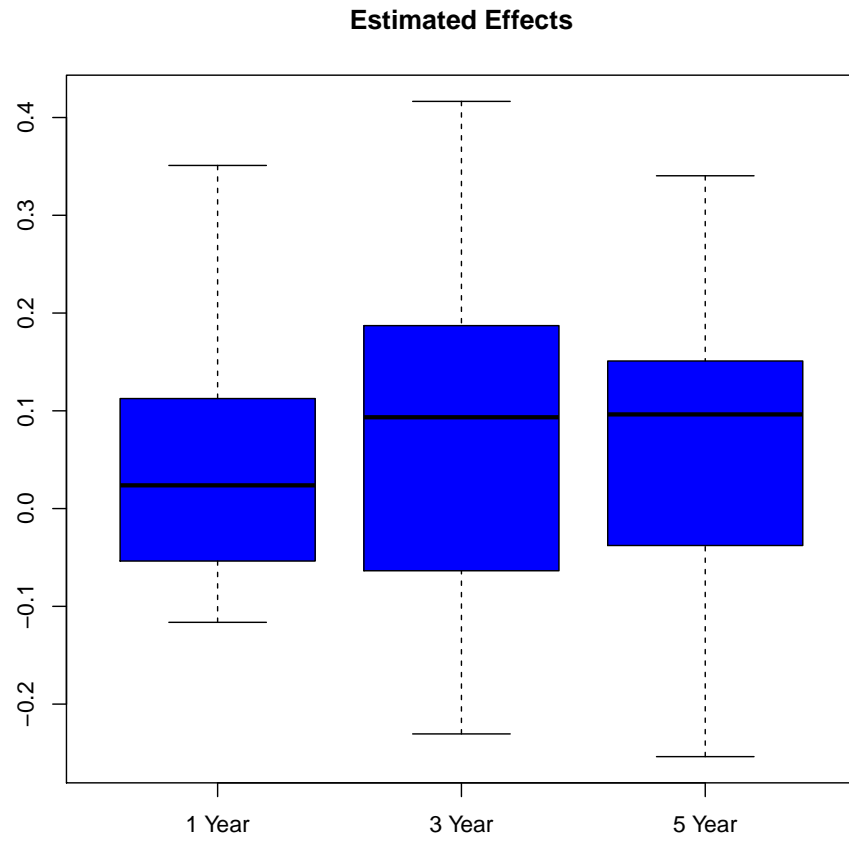


Figure B.1: Analysis of SRTR data: Boxplot of $\hat{\delta}_{j1}(t)$ for $t = 1$, $t = 3$, and $t = 5$ years. Whiskers represent minimum and maximum.

Estimated Effects at 5 Years

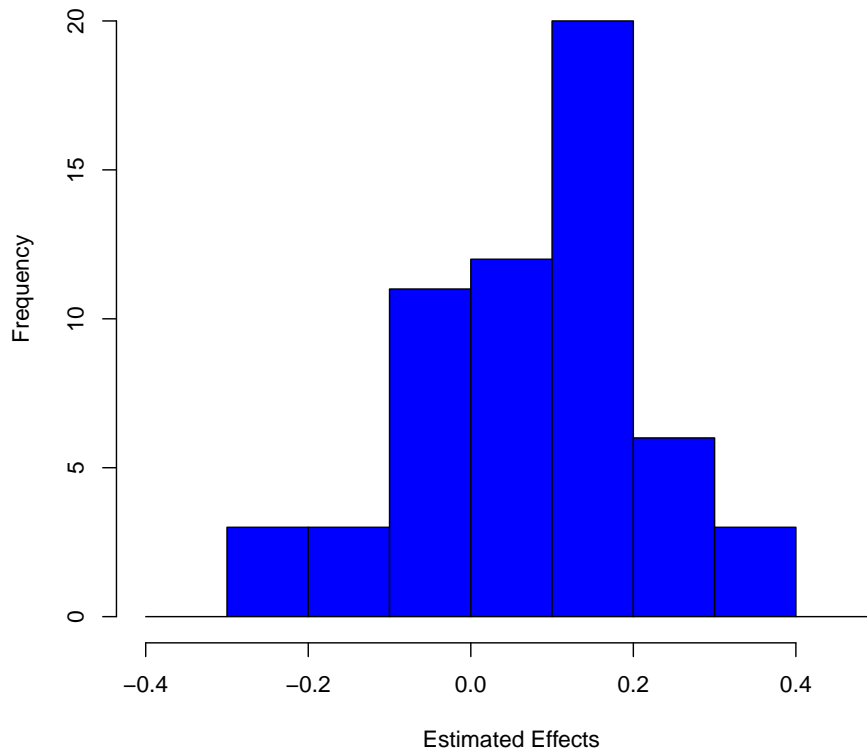


Figure B.2: Analysis of SRTR data: Histogram of $\hat{\delta}_{j1}(t)$ at $t = 5$ years post wait-listing.

APPENDIX C

Proof of Theorem 3.1

Recall that the effect of group j on the cumulative incidence for cause k is given by

$$\delta_{jk}(t) = F_{jk}(t) - F_k(t),$$

where $F_{jk}(t) = E_Z[F_{ijk}(t)]$ and $F_k(t) = E_Z[E_A[F_{iAk}(t)]]$. We estimate δ_{jk} through

$$\begin{aligned} \hat{\delta}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) &= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} - \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{im}(s; \hat{\boldsymbol{\alpha}})} \\ &= \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - \hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}). \end{aligned}$$

In order to prove Theorem 3.1, we derive the influence function for $\hat{\delta}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})$

through Taylor series expansions of the following quantity,

$$n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - \delta_{jk}(t) \right\} \quad (\text{C.1})$$

$$= n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - F_{jk}(t) \right. \quad (\text{C.2})$$

$$\left. - \left[\hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}) - F_k(t) \right] \right\}. \quad (\text{C.3})$$

We will focus on the first and second terms separately, with the first term expressed more explicitly by

$$(\text{C.2}) = n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - F_{jk}(t) \right\} \quad (\text{C.4})$$

$$= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} - E_Z[F_{ijk}(t)] \right\},$$

and the second term by

$$(\text{C.3}) = n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}) - F_k(t) \right\} \quad (\text{C.5})$$

$$= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{im}(s; \hat{\boldsymbol{\alpha}})} - E_Z[E_A[F_{iAk}(t)]] \right\}.$$

Recall that, based on the assumed generalized logit model (multinomial logistic regression), with center J as the reference, we have

$$p_{ij}(\boldsymbol{\theta}) = \frac{\exp(\mathbf{Z}_i^T \boldsymbol{\theta}_j)}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)}$$

$$p_{iJ}(\boldsymbol{\theta}) = \frac{1}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)},$$

for $j = 1, \dots, J - 1$ and where we define

$$\mathbf{Z}_i = \begin{bmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{ip} \end{bmatrix}, \boldsymbol{\theta}_j = \begin{bmatrix} \theta_{1j} \\ \theta_{2j} \\ \vdots \\ \theta_{pj} \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_{J-1} \end{bmatrix}.$$

Under the assumption that C_i follows a proportional hazards model stratified on center, we can write

$$G_{ij}(t) \equiv P(C_i > t | A_i = j, \mathbf{Z}_i) = \exp(-\Lambda_{ij}^C(t)) = \exp(-\Lambda_{0j}^C(t) \exp(\mathbf{Z}_i^T \boldsymbol{\alpha})).$$

In linearizing (C.1), we derive iid expansions for the following quantities:

1. $n^{\frac{1}{2}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$
2. $n^{\frac{1}{2}} \{p_{ij}^{-1}(\hat{\boldsymbol{\theta}}) - p_{ij}^{-1}(\boldsymbol{\theta})\}$
3. $n^{\frac{1}{2}} \{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\}$
4. $n^{\frac{1}{2}} \{\hat{\Lambda}_{0j}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{0j}^C(t)\}$
5. $n^{\frac{1}{2}} \{\hat{\Lambda}_{ij}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{ij}^C(t)\}$
6. $n^{\frac{1}{2}} \{\hat{G}_{ij}^{-1}(t; \hat{\boldsymbol{\alpha}}) - G_{ij}^{-1}(t)\}$
7. $n^{\frac{1}{2}} \{\hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta})\}$

$$8. \ n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta}) - \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) \right\}$$

$$9. \ n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) - F_{jk}(t) \right\}$$

$$10. \ n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}) - \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) \right\}$$

$$11. \ n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) - F_k(t) \right\}$$

$$12. \ n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\}$$

We can write (C.2) as

$$(C.2) = n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta}) \right\} \quad (C.6)$$

$$+ n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta}) - \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) \right\} \quad (C.7)$$

$$+ n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) - F_{jk}(t) \right\}, \quad (C.8)$$

while (C.3) can be expressed as

$$(C.3) = n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}) - \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) \right\} \quad (C.9)$$

$$+ n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) - F_k(t) \right\}. \quad (C.10)$$

$$[1.] \ n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

By a Taylor expansion,

$$[1] = \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^{IPTW}(\boldsymbol{\theta}) + o_p(1), \quad (C.11)$$

which can be derived by first considering the likelihood and log-likelihood contributions,

$$\begin{aligned}
L_i(\boldsymbol{\theta}) &= \prod_{j=1}^J p_{ij}^{A_{ij}}(\boldsymbol{\theta}) \\
l_i(\boldsymbol{\theta}) &= \sum_{j=1}^J A_{ij} \log p_{ij}(\boldsymbol{\theta}) \\
&= \sum_{j=1}^{J-1} A_{ij} \log \left[\frac{\exp(\mathbf{Z}_i^T \boldsymbol{\theta}_j)}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)} \right] + A_{iJ} \log \left[\frac{1}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)} \right] \\
&= \left(\sum_{j=1}^{J-1} A_{ij} \mathbf{Z}_i^T \boldsymbol{\theta}_j \right) - \log \left(1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m) \right).
\end{aligned}$$

The score function contribution for subject i is then given by

$$U_i^{IPTW}(\boldsymbol{\theta}) = \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} \\ \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} \\ \vdots \\ \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{J-1}} \end{bmatrix} = \begin{bmatrix} U_{i1}(\boldsymbol{\theta}) \\ U_{i2}(\boldsymbol{\theta}) \\ \vdots \\ U_{i,J-1}(\boldsymbol{\theta}) \end{bmatrix}$$

for $j = 1, \dots, J - 1$, with

$$U_{ij}(\boldsymbol{\theta}) = \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l} = \mathbf{Z}_i (A_{ij} - p_{ij}(\boldsymbol{\theta})).$$

The information matrix is given by

$$\begin{aligned} \mathcal{I}^{IPTW}(\boldsymbol{\theta}) &= -E \left[\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \\ &= -E \begin{bmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}) & \mathcal{I}_{12}(\boldsymbol{\theta}) & \cdots & \mathcal{I}_{1,J-1}(\boldsymbol{\theta}) \\ \mathcal{I}_{12}(\boldsymbol{\theta}) & \mathcal{I}_{22}(\boldsymbol{\theta}) & & \\ \vdots & & \ddots & \\ \mathcal{I}_{1,J-1}(\boldsymbol{\theta}) & & & \mathcal{I}_{J-1,J-1}(\boldsymbol{\theta}) \end{bmatrix}, \end{aligned}$$

where for $j = 1, 2, \dots, J-1$ and $l = 1, 2, \dots, J-1$ and $j \neq l$:

$$\begin{aligned} \mathcal{I}_{ll}(\boldsymbol{\theta}) &= -\frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_l^T} = \mathbf{Z}_i^{\otimes 2} \left[\frac{\exp(\mathbf{Z}_i^T \boldsymbol{\theta}_l)}{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)} - \frac{\exp(2\mathbf{Z}_i^T \boldsymbol{\theta}_l)}{\left(1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)\right)^2} \right] \\ &= \mathbf{Z}_i^{\otimes 2} p_{il}(\boldsymbol{\theta}) \{1 - p_{il}(\boldsymbol{\theta})\} \\ \mathcal{I}_{jl}(\boldsymbol{\theta}) &= -\frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_l^T} = -\mathbf{Z}_i^{\otimes 2} \frac{\exp\{\mathbf{Z}_i^T(\boldsymbol{\theta}_l + \boldsymbol{\theta}_j)\}}{\left\{1 + \sum_{m=1}^{J-1} \exp(\mathbf{Z}_i^T \boldsymbol{\theta}_m)\right\}^2} \\ &= -\mathbf{Z}_i^{\otimes 2} p_{il}(\boldsymbol{\theta}) p_{ij}(\boldsymbol{\theta}). \end{aligned}$$

$$[2.] \quad n^{\frac{1}{2}} \{w_{ij}(\hat{\boldsymbol{\theta}}) - w_{ij}(\boldsymbol{\theta})\}$$

Let the IPTW component of the weight be denoted by $w_{ij}(\boldsymbol{\theta}) = A_{ij}/p_{ij}(\boldsymbol{\theta})$. By the Delta method,

$$[2] = \frac{\partial}{\partial \boldsymbol{\theta}} [w_{ij}(\boldsymbol{\theta})] \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^{IPTW}(\boldsymbol{\theta}),$$

where we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} [w_{ij}(\boldsymbol{\theta})] = \frac{\partial}{\partial \boldsymbol{\theta}} \left[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \right] = -A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} [p_{ij}(\boldsymbol{\theta})],$$

with

$$\frac{\partial}{\partial \boldsymbol{\theta}} [p_{ij}(\boldsymbol{\theta})] = \begin{bmatrix} \frac{\partial}{\partial \theta_1} [p_{ij}(\boldsymbol{\theta})] \\ \frac{\partial}{\partial \theta_2} [p_{ij}(\boldsymbol{\theta})] \\ \vdots \\ \frac{\partial}{\partial \theta_{J-1}} [p_{ij}(\boldsymbol{\theta})] \end{bmatrix} = \begin{bmatrix} -\mathbf{Z}_i p_{ij}(\boldsymbol{\theta}) p_{i1}(\boldsymbol{\theta}) \\ -\mathbf{Z}_i p_{ij}(\boldsymbol{\theta}) p_{i2}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{Z}_i p_{ij}(\boldsymbol{\theta}) (1 - p_{ij}(\boldsymbol{\theta})) \\ \vdots \\ -\mathbf{Z}_i p_{ij}(\boldsymbol{\theta}) p_{i,J-1}(\boldsymbol{\theta}) \end{bmatrix} \equiv \Delta_{ij}(\boldsymbol{\theta}).$$

So $n^{\frac{1}{2}} \{w_{ij}(\hat{\boldsymbol{\theta}}) - w_{ij}(\boldsymbol{\theta})\}$ can be written as

$$[2] = -A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \Delta_{ij}^T(\boldsymbol{\theta}) \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^{IPTW}(\boldsymbol{\theta}) + o_p(1).$$

$$[3.] \quad n^{\frac{1}{2}} \{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\}$$

By a Taylor expansion of $U^C(\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}$:

$$\begin{aligned} U^C(\hat{\boldsymbol{\alpha}}) - U^C(\boldsymbol{\alpha}) &= \frac{\partial}{\partial \boldsymbol{\alpha}'} [U^C(\boldsymbol{\alpha})]_{\boldsymbol{\alpha}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ n^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} U^C(\boldsymbol{\alpha}) \\ &= \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\boldsymbol{\alpha}), \end{aligned}$$

where $U_i^C(\boldsymbol{\alpha}) \xrightarrow{p} \sum_{j=1}^J \int_0^\tau \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(t; \boldsymbol{\alpha})\} dM_{ij}^C(t)$ and $dM_{ij}^C(t) = dN_{ij}^C(t) - Y_{ij}(t)d\Lambda_{ij}^C(t)$.

The information matrix is given by

$$\mathcal{I}^C(\boldsymbol{\alpha}) = E \left[\sum_{j=1}^J \int_0^\tau \left\{ \frac{s_j^{(2)}(t; \boldsymbol{\alpha})}{s_j^{(0)}(t; \boldsymbol{\alpha})} - \bar{\mathbf{z}}_j(t; \boldsymbol{\alpha})^{\otimes 2} \right\} dN_{ij}^C(t) \right].$$

Properties of [3] as $n \rightarrow \infty$ are based on standard martingale theory (e.g., as described in Fleming and Harrington, 1991).

$$[4.] \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t) - \Lambda_{0j}^C(t) \right\}$$

$$\begin{aligned} [4] &= n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{0j}^C(t; \boldsymbol{\alpha}) \right\} \\ &= n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t; \hat{\boldsymbol{\alpha}}) - \hat{\Lambda}_{0j}^C(t; \boldsymbol{\alpha}) \right\} \end{aligned} \quad (\text{C.12})$$

$$+ n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t; \boldsymbol{\alpha}) - \Lambda_{0j}^C(t) \right\}. \quad (\text{C.13})$$

Applying a Taylor expansion and the Continuous Mapping Theorem (CMT),

$$\begin{aligned} (\text{C.12}) &= \hat{h}_j^C(t; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\boldsymbol{\alpha}) + o_p(1) \\ &\xrightarrow{p} h_j^C(t; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n U_i^C(\boldsymbol{\alpha}) + o_p(1), \end{aligned}$$

with

$$h_j^C(t; \boldsymbol{\alpha}) = - \int_0^t \bar{\mathbf{z}}_j^C(s; \boldsymbol{\alpha}) \cdot d\Lambda_{0j}^C(t; \boldsymbol{\alpha}), \quad (\text{C.14})$$

which holds by Slutsky's Theorem. In addition, $n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t; \boldsymbol{\alpha}) - \Lambda_{0j}^C(t) \right\}$ can be writ-

ten as

$$(C.13) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t s_j^{(0)}(s; \boldsymbol{\alpha})^{-1} dM_{ij}^C(s; \boldsymbol{\alpha}) + o_p(1). \quad (C.15)$$

Putting (C.14) and (C.15) together, we get:

$$n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{0j}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{0j}^C(t) \right\} = n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_{ij}^C(t; \boldsymbol{\alpha}) + o_p(1),$$

where $\Phi_{ij}^C(t; \boldsymbol{\alpha}) = -h_j^C(t; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_i^C(\boldsymbol{\alpha}) + \int_0^t s_j^{(0)}(s; \boldsymbol{\alpha})^{-1} dM_{ij}^C(s; \boldsymbol{\alpha})$.

$$[5.] \quad n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{ij}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{ij}^C(t) \right\}$$

$$[5] = n^{\frac{1}{2}} \left\{ \int_0^t Y_{ij}(s) \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{Z}_i) d\hat{\Lambda}_{0j}^C(s; \hat{\boldsymbol{\alpha}}) - \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) d\hat{\Lambda}_{0j}^C(s; \hat{\boldsymbol{\alpha}}) \right\} \quad (C.16)$$

$$+ n^{\frac{1}{2}} \left\{ \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) d\hat{\Lambda}_{0j}^C(s; \hat{\boldsymbol{\alpha}}) - \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) d\Lambda_{0j}^C(s) \right\}. \quad (C.17)$$

We can express the first component as

$$\begin{aligned}
\text{(C.16)} &= n^{\frac{1}{2}} \int_0^t Y_{ij}(s) \left\{ \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{Z}_i) - \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) \right\} d\hat{\Lambda}_{0j}^C(s; \hat{\boldsymbol{\alpha}}) \\
&= \int_0^t \mathbf{Z}_i d\Lambda_{ij}^C(s; \boldsymbol{\alpha}) \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) + o_p(1), \tag{C.18}
\end{aligned}$$

obtained by a Taylor expansion and Result [3]. Using Result [4], the second expression can be written as follows,

$$\begin{aligned}
\text{(C.17)} &= n^{\frac{1}{2}} \left\{ \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) d\hat{\Lambda}_{0j}^C(s; \hat{\boldsymbol{\alpha}}) - \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) d\Lambda_{0j}^C(s; \boldsymbol{\alpha}) \right\} \\
&= \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n d\Phi_{ljk}^C(t; \boldsymbol{\alpha}) \right\} + o_p(1). \tag{C.19}
\end{aligned}$$

Combining (C.18) and (C.19), we get:

$$n^{\frac{1}{2}} \left\{ \hat{\Lambda}_{ij}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{ij}^C(t; \boldsymbol{\alpha}) \right\} = D_{ij}^C(t; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilj}^C(t; \boldsymbol{\alpha}),$$

where $D_{ij}^C(t; \boldsymbol{\alpha}) = \int_0^t \left\{ \mathbf{Z}_i - \bar{z}_j^C(s; \boldsymbol{\alpha}) \right\} d\Lambda_{ij}^C(s; \boldsymbol{\alpha})$

and $J_{ilj}^C(t; \boldsymbol{\alpha}) = \int_0^t Y_{ij}(s) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(s; \boldsymbol{\alpha})^{-1} dM_{ij}^C(s; \boldsymbol{\alpha})$.

[6.] $n^{\frac{1}{2}} \left\{ \hat{G}_{ij}^{-1}(t; \hat{\boldsymbol{\alpha}}) - G_{ij}^{-1}(t) \right\}$

We can write

$$\begin{aligned}
[6] &= n^{\frac{1}{2}} \left[\exp \left\{ \hat{\Lambda}_{ij}^C(t; \hat{\boldsymbol{\alpha}}) \right\} - \exp \left\{ \Lambda_{ij}^C(t) \right\} \right] \\
&= n^{\frac{1}{2}} \exp \left\{ \Lambda_{ij}^C(t) \right\} \left\{ \hat{\Lambda}_{ij}^C(t; \hat{\boldsymbol{\alpha}}) - \Lambda_{ij}^C(t) \right\} \\
&= \exp(\Lambda_{ij}^C(t)) \left\{ D_{ij}^C(t; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilj}^C(t; \boldsymbol{\alpha}) \right\},
\end{aligned}$$

where we have used the Delta method as well as Result [5].

$$[7.] \quad n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) - \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta}) \right\}$$

We carry out the following calculations,

$$\begin{aligned}
[7] &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} - \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \right\} \\
&= \frac{1}{n} \sum_{i=1}^n n^{\frac{1}{2}} \left\{ w_{ij}(\hat{\boldsymbol{\theta}}) - w_{ij}(\boldsymbol{\theta}) \right\} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \\
&= -\frac{1}{n} \sum_{i=1}^n \left\{ A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \mathbf{Z}_i \Delta_{ij}^T(\boldsymbol{\theta}) \left[\mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^{IPTW}(\boldsymbol{\theta}) \right] \right\} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \\
&= -\left[\frac{1}{n} \sum_{i=1}^n \left\{ A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \mathbf{Z}_i \Delta_{ij}^T(\boldsymbol{\theta}) \right\} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \right] \left[\mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^{IPTW}(\boldsymbol{\theta}) \right],
\end{aligned}$$

where we define

$$\mathbf{d}_{jk}^T(t) = -E \left[A_{ij} p_{ij}^{-2}(\boldsymbol{\theta}) \mathbf{Z}_i \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \Delta_{ij}^T(\boldsymbol{\theta}) \right].$$

After using the Weak Law of Large Numbers (WLLN), continuity, and the fact that

$\hat{G}_{ij}(t; \hat{\boldsymbol{\alpha}}) \xrightarrow{p} G_{ij}(t)$, we can write

$$[7] = \mathbf{d}_{jk}^T(t) \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^{IPTW}(\boldsymbol{\theta}).$$

We then let $\phi_{ljk}^1(t) = \mathbf{d}_{jk}^T(t) \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} U_l^{IPTW}(\boldsymbol{\theta})$, an element quoted in Theorem 3.1.

$$[8.] n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta}) - \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) \right\}$$

We start by writing

$$\begin{aligned} [8] &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} - \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t n^{\frac{1}{2}} \left\{ \hat{G}_{ij}^{-1}(s; \hat{\boldsymbol{\alpha}}) - G_{ij}^{-1}(s) \right\} dN_{ik}(s) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ D_{ij}^C(s; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right. \\ &\quad \left. + n^{-\frac{1}{2}} \left[\sum_{l=1}^n J_{ilj}^C(s; \boldsymbol{\alpha}) \right] \right\} dN_{ik}(s). \end{aligned} \tag{C.21}$$

Now, the first line of the last equation can be decomposed as follows:

$$\begin{aligned}
\text{(C.20)} &= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ D_{ij}^C(s; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \times \\
&\quad \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \times \\
&\quad \left\{ \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right\} \\
&\xrightarrow{p} E \left[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \times \\
&\quad \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \\
&= E \left[\int_0^t \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T \{ F_{ijk}(t) - F_{ijk}(u) \} d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right] \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}).
\end{aligned}$$

We then define

$$\phi_{ijk}^2(t) = E \left[\int_0^t \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T \{ F_{ijk}(t) - F_{ijk}(u) \} d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right] \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_l^C(\boldsymbol{\alpha}),$$

which is among the non-zero elements listed in Theorem 3.1. The second line, (C.21),

can be written as

$$\begin{aligned}
\text{(C.21)} &= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilj}^C(s; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \exp(\Lambda_{ij}^C(s)) \times \\
&\quad \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^s Y_{ij}(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t \left[\frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_u^t \exp(\Lambda_{ij}^C(s)) dN_{ik}(s) \right] \times \\
&\quad Y_{ij}(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha}) \\
&\xrightarrow{p} n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t E \left[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_u^t \exp(\Lambda_{ij}^C(s)) dN_{ik}(s) \right] \times \\
&\quad Y_{ij}(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha}) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^t E [F_{ijk}(t) - F_{ijk}(u)] Y_{ij}(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha}).
\end{aligned}$$

In arriving at the last expression, we have switched the order of summation and order of integration so that the outer integral is a martingale transform. Then, we used the consistency of the estimators and replaced them with the limiting values, $F_{ijk}(t)$. We then let $\phi_{ijk}^3(t) = \int_0^t E [F_{ijk}(t) - F_{ijk}(u)] Y_{ij}(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha})$, which also appears in Theorem 3.1. Combining expressions for (C.20) and (C.21) then yields

$$[8] = n^{-\frac{1}{2}} \left\{ \phi_{ijk}^2(t) + \phi_{ijk}^3(t) \right\}.$$

$$[9.] \quad n^{\frac{1}{2}} \left\{ \hat{F}_{jk}(t; \Lambda_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) - F_{jk}(t) \right\}$$

$$\begin{aligned} [9] &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}_j)} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} - E_Z[F_{ijk}(t)] \right\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} - F_{jk}(t) \right\}. \end{aligned}$$

Let $\phi_{ijk}^A(t) = \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} - F_{jk}(t)$, which has mean 0 since $E \left[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta})} \int_0^t \frac{dN_{ik}(s)}{G_{ij}(s)} \right] = F_{jk}(t)$.

$$[10.] \quad n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \hat{\Lambda}_0^C, \hat{\boldsymbol{\alpha}}) - \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) \right\}$$

We start by writing

$$\begin{aligned} [10] &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{im}(s; \hat{\boldsymbol{\alpha}})} - \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{G_{im}(s)} \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t n^{\frac{1}{2}} \left\{ \hat{G}_{im}^{-1}(s; \hat{\boldsymbol{\alpha}}) - G_{im}^{-1}(s) \right\} dN_{ik}(s) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \times \\ &\quad \left\{ D_{im}^C(s; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) + n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilm}^C(s; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \times \\ &\quad \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ D_{im}^C(s; \boldsymbol{\alpha})^T \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right\} dN_{ik}(s) \quad (\text{C.22}) \end{aligned}$$

$$+ \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n J_{ilm}^C(s; \boldsymbol{\alpha}) \right\} dN_{ik}(s). \quad (\text{C.23})$$

(C.22) can be further written as

$$\begin{aligned}
(C.22) &= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \times \\
&\quad \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \times \\
&\quad \left[\mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \right] \\
&\xrightarrow{p} E \left[\sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \times \\
&\quad \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}).
\end{aligned}$$

Let

$$\phi_{ik}^{\bar{5}}(t) = E \left[\sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_l^C(\boldsymbol{\alpha}).$$

(C.23) can be written as

$$\begin{aligned}
(C.23) &= \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \times \\
&\quad \left\{ n^{-\frac{1}{2}} \sum_{l=1}^n \int_0^s Y_{im}^C(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lm}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \\
&= n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{m=1}^J \int_0^t \left\{ \frac{1}{n} \sum_{i=1}^n A_{im} \int_u^t \exp(\Lambda_{im}^C(s)) dN_{ik}(s) \right\} \times \\
&\quad Y_{im}^C(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lm}^C(u; \boldsymbol{\alpha}) \\
&\xrightarrow{p} n^{-\frac{1}{2}} \sum_{l=1}^n \sum_{m=1}^J \int_0^t E \left[A_{im} \int_u^t \exp(\Lambda_{im}^C(s)) dN_{ik}(s) \right] \times \\
&\quad Y_{im}^C(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lm}^C(u; \boldsymbol{\alpha}).
\end{aligned}$$

We then define

$$\phi_{ik}^6(t) = \sum_{m=1}^J \int_0^t E \left[A_{im} \int_u^t \exp(\Lambda_{im}^C(s)) dN_{ik}(s) \right] Y_{im}^C(u) \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lm}^C(u; \boldsymbol{\alpha}),$$

which is quoted in Theorem 3.1. So

$$[10] = n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \phi_{lk}^5(t) + \phi_{lk}^6(t) \right\}.$$

$$[11.] \quad n^{\frac{1}{2}} \left\{ \hat{F}_k(t; \Lambda_0^C, \boldsymbol{\alpha}) - F_k(t) \right\}$$

We can write, asymptotically, that

$$\begin{aligned}
[11] &= n^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{G_{im}(s)} - E_Z [E_A [F_{imk}(t)]] \right\} \\
&= n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \sum_{m=1}^J A_{im} \int_0^t \frac{dN_{ik}(s)}{G_{im}(s)} - F_k(t) \right\},
\end{aligned}$$

which motivates us to define

$$\phi_{lk}^7(t) = \sum_{m=1}^J A_{lm} \int_0^t \frac{dN_{lk}(s)}{G_{lm}(s)} - F_k(t),$$

used in Theorem 3.1.

$$[12.] \quad n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\}$$

As the culmination of Results [1] to [11], we set $\phi_{ljk}(t) = \phi_{ljk}^1(t) + \phi_{ljk}^2(t) + \phi_{ljk}^3(t) + \phi_{ljk}^4(t) - \left\{ \phi_{lk}^5(t) + \phi_{lk}^6(t) + \phi_{lk}^7(t) \right\}$, which yields the final step in the proof of Theorem 3.1,

$$n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t) - \delta_{jk}(t) \right\} = n^{-\frac{1}{2}} \sum_{l=1}^n \phi_{ljk}(t),$$

where we have defined

$$\begin{aligned}
\phi_{ljk}^1(t) &= \mathbf{d}_{jk}^T(t) \mathcal{I}^{IPTW}(\boldsymbol{\theta})^{-1} U_l^{IPTW}(\boldsymbol{\theta}) \\
\phi_{ljk}^2(t) &= E \left[\int_0^t \{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \boldsymbol{\alpha}) \}^T \{ F_{ijk}(t) - F_{ijk}(u) \} d\Lambda_{ij}^C(u; \boldsymbol{\alpha}) \right] \mathcal{I}^C(\boldsymbol{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\boldsymbol{\alpha}) \\
\phi_{ljk}^3(t) &= \int_0^t E [F_{jk}(t) - F_{jk}(u)] Y_{ij}(u) \cdot \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_j^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lj}^C(u; \boldsymbol{\alpha}) \\
\phi_{ljk}^4(t) &= \frac{A_{lj}}{p_{lj}(\boldsymbol{\theta})} \int_0^t \frac{dN_{lk}(s)}{G_{lj}(s)} - F_{jk}(t) \\
\phi_{lk}^5(t) &= E \left[\sum_{m=1}^J A_{im} \int_0^t \exp(\Lambda_{im}^C(s)) \left\{ \int_0^s \{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \boldsymbol{\alpha}) \}^T d\Lambda_{im}^C(u; \boldsymbol{\alpha}) \right\} dN_{ik}(s) \right] \times \\
&\quad \mathcal{I}^C(\boldsymbol{\alpha})^{-1} U_l^C(\boldsymbol{\alpha}) \\
\phi_{lk}^6(t) &= \sum_{m=1}^J \int_0^t E \left[A_{im} \int_u^t \exp(\Lambda_{im}^C(s)) dN_{ik}(s) \right] Y_{im}(u) \cdot \exp(\boldsymbol{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \boldsymbol{\alpha})^{-1} dM_{lm}^C(u; \boldsymbol{\alpha}) \\
\phi_{lk}^7(t) &= \left[\sum_{m=1}^J A_{lm} \int_0^t \frac{dN_{lk}(s)}{G_{lm}(s)} \right] - F_k(t).
\end{aligned}$$

Therefore, we obtain

$$\text{Var} \left(n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t; \hat{G}_{il}(t; \hat{\boldsymbol{\alpha}}), \hat{\boldsymbol{\theta}}) - \delta_{jk}(t) \} \right) = E \left[\{ \phi_{ljk}(t) \}^2 \right],$$

with corresponding estimator

$$\widehat{\text{Var}} \left(n^{\frac{1}{2}} \{ \hat{\delta}_{jk}(t; \hat{G}_{il}(t; \hat{\boldsymbol{\alpha}}), \hat{\boldsymbol{\theta}}) - \delta_{jk}(t) \} \right) = n^{-1} \sum_{l=1}^n \{ \hat{\phi}_{ljk}(t) \}^2,$$

such that

$$\widehat{Var} \left\{ \hat{\delta}_{jk}(t; \hat{G}_{il}(t; \hat{\alpha}), \hat{\theta}) \right\} = n^{-2} \sum_{l=1}^n \left\{ \hat{\phi}_{ljk}(t) \right\}^2.$$

Thus, the proposed variance estimators are:

$$\hat{\phi}_{ljk}^1(t) = \hat{\mathbf{d}}_{jk}^T(t) \mathcal{I}^{IPTW}(\hat{\theta})^{-1} U_l^{IPTW}(\hat{\theta})$$

$$\hat{\phi}_{ljk}^2(t) = \frac{1}{n} \sum_{i=1}^n \left[\int_0^t \left\{ \mathbf{Z}_i - \bar{\mathbf{z}}_j^C(u; \hat{\alpha}) \right\}^T \left\{ \hat{F}_{ijk}(t) - \hat{F}_{ijk}(u) \right\} d\hat{\Lambda}_{ij}^C(u; \hat{\alpha}) \right] \times$$

$$\mathcal{I}^C(\hat{\alpha})^{-1} n^{-\frac{1}{2}} \sum_{l=1}^n U_l^C(\hat{\alpha})$$

$$\hat{\phi}_{ljk}^3(t) = \int_0^t \frac{1}{n} \sum_{i=1}^n \left[\hat{F}_{ijk}(t) - \hat{F}_{ijk}(u) \right] Y_{ij}(u) \exp(\hat{\alpha}^T \mathbf{Z}_i) \hat{s}_j^{(0)}(u; \hat{\alpha})^{-1} d\hat{M}_{lj}^C(u; \hat{\alpha})$$

$$\hat{\phi}_{ljk}^4(t) = \frac{A_{lj}}{p_{lj}(\hat{\theta})} \int_0^t \frac{dN_{lk}(s)}{\hat{G}_{lj}(s; \hat{\alpha})} - \hat{F}_{jk}(t)$$

$$\hat{\phi}_{lk}^5(t) = \frac{1}{n} \sum_{i=1}^n \left[\sum_{m=1}^J A_{im} \int_0^t \exp(\hat{\Lambda}_{im}^C(s; \hat{\alpha})) \left\{ \int_0^s \left\{ \mathbf{Z}_i - \bar{\mathbf{z}}_m^C(u; \hat{\alpha}) \right\}^T d\hat{\Lambda}_{im}^C(u; \hat{\alpha}) \right\} dN_{ik}(s) \right] \times$$

$$\mathcal{I}^C(\hat{\alpha})^{-1} U_l^C(\hat{\alpha})$$

$$\hat{\phi}_{lk}^6(t) = \sum_{m=1}^J \int_0^t \frac{1}{n} \sum_{i=1}^n \left[A_{im} \int_u^t \exp(\hat{\Lambda}_{im}^C(s; \hat{\alpha})) dN_{ik}(s) \right] \times$$

$$Y_{im}(u) \times \exp(\hat{\alpha}^T \mathbf{Z}_i) s_m^{(0)}(u; \hat{\alpha})^{-1} dM_{lm}^C(u; \hat{\alpha})$$

$$\hat{\phi}_{lk}^7(t) = \left[\sum_{m=1}^J A_{lm} \int_0^t \frac{dN_{lk}(s)}{\hat{G}_{lm}(s; \hat{\alpha})} \right] - \hat{F}_k(t),$$

where

$$\begin{aligned}
\hat{F}_{jk}(t) &= \hat{E}_Z[\hat{F}_{ijk}(t)] \\
&= \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{\hat{p}_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \\
\hat{F}_{ijk}(t) &= A_{ij} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \\
d\hat{M}_{lj}^C(u; \hat{\boldsymbol{\alpha}}) &= dN_{ijk}(t) - Y_{ij}(t) d\hat{\Lambda}_{ijk}^\#(t) \\
\hat{F}_k(t) &= \sum_{m=1}^J \frac{1}{n} \sum_{l=1}^n A_{lm} \int_0^t \frac{dN_{lk}(s)}{\hat{G}_{lm}(s; \hat{\boldsymbol{\alpha}})}.
\end{aligned}$$

For the purposes of estimating $Var(\hat{\delta}_{jk}(t))$, we will treat the weights as fixed and, hence, ignore the variability due to the use of IPCW and IPTW. As such, the pertinent components are as follows:

$$\begin{aligned}
n^{\frac{1}{2}} \left\{ \hat{\delta}_{jk}(t; \hat{\Lambda}_0^C, \boldsymbol{\alpha}, \boldsymbol{\theta}) - \delta_{jk}(t) \right\} &= n^{-\frac{1}{2}} \sum_{l=1}^n \left\{ \phi_{ljk}^4(t) - \phi_{lk}^7(t) \right\} \\
\widetilde{Var}(\hat{\delta}_{jk}(t)) &= \frac{1}{n} E \left[\left\{ \phi_{ljk}^4(t) - \phi_{lk}^7(t) \right\}^2 \right] \\
\widehat{Var}(\hat{\delta}_{jk}(t)) &= n^{-2} \sum_{l=1}^n \left\{ \hat{\phi}_{ljk}^4(t) - \hat{\phi}_{lk}^7(t) \right\}^2,
\end{aligned}$$

where

$$\begin{aligned}
\hat{\phi}_{ljk}^4(t) &= \frac{A_{lj}}{\hat{p}_{lj}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{lk}(s)}{\hat{G}_{lj}(s; \hat{\boldsymbol{\alpha}})} - \frac{1}{n} \sum_{i=1}^n \frac{A_{ij}}{\hat{p}_{ij}(\hat{\boldsymbol{\theta}})} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{ij}(s; \hat{\boldsymbol{\alpha}})} \\
\hat{\phi}_{lk}^7(t) &= \left[\sum_{m=1}^J A_{lm} \int_0^t \frac{dN_{lk}(s)}{\hat{G}_{lm}(s; \hat{\boldsymbol{\alpha}})} \right] - \sum_{m=1}^J \frac{1}{n} \sum_{i=1}^n A_{im} \int_0^t \frac{dN_{ik}(s)}{\hat{G}_{im}(s; \hat{\boldsymbol{\alpha}})}.
\end{aligned}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

- Aly, E. A. A., S. C. Kochar, and I. W. McKeague (1994), Some tests for comparing cumulative incidence functions and cause-specific hazard rates, *Journal of the American Statistical Association*, *89*, 994–999.
- Anderson, P. K., and R. D. Gill (1982), Cox’s regression model for counting processes: a large sample study, *Annals of Statistics*, *4*, 1100–1120.
- Anderson, P. K., A. Borgan, R. D. Gill, and N. Keiding (1993), Statistical Models Based on Counting Processes, *Biometrics*, *24*, 100–101.
- Barone, M., A. W. Avolio, A. DiLeo, P. Burra, and A. Francavilla (2008), ABO blood group-related waiting list disparities in liver transplant candidates: Effect of the MELD adoption, *Transplantation*, *85*, 844–849.
- Benichou, J., and M. H. Gail (1990), Estimates of absolute cause-specific risk in cohort studies, *Biometrics*, *46*, 813–826.
- Breslow, N. (1972), Contribution to the discussion on the paper by D. R. Cox, regression and life tables, *Journal of the Royal Statistical Society, Series B*, *34*, 216–217.
- Cheng, S. C., J. P. Fine, and L. J. Wei (1998), Prediction of cumulative incidence function under the proportional hazards model, *Biometrics*, *54*, 219–228.
- Chiang, C. L. (1968), *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York.
- Cox, D. R. (1959), The analysis of exponentially distributed lifetimes with two types of failure, *Journal of the Royal Statistical Society, Series B*, *21*, 411–421.
- Cox, D. R. (1972), Regression models and life tables (with Discussion), *Journal of the Royal Statistical Society, Series B*, *34*, 187–200.

- Cox, D. R. (1975), Partial likelihood, *Biometrika*, *62*, 269–275.
- Crowder, M. J. (2001), *Classical competing risks*, Chapman and Hall/CRC Press, London.
- Dabrowska, D. M., and K. A. Doksum (1988), Estimates and testing in a two-sample generalized odds-rate model, *Journal of the American Statistical Association*, *83*, 744–749.
- Fine, J. P., and R. J. Gray (1999), A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, *94*, 496–509.
- Gail, M. H. (1975), A review and critique of some models used in competing risk analysis, *Biometrics*, *31*, 209–222.
- Gail, M. H. (2001), *Competing Risks*, encyclopedia, London.
- Glander, P., K. Budde, D. Schmidt, T. F. Fuller, M. Giessing, H. H. Neumayer, and L. Liefeldt (2010), The 'blood group O problem' in kidney transplantation - time to change?, *Nephrology Dialysis Transplantation*, *25*, 1998–2004.
- Gray, R. J. (1988), A class of K-sample tests for comparing the cumulative incidence of a competing risk, *Annals of Statistics*, *16*, 1141–1154.
- Harrington, D. P., and T. R. Fleming (1982), A class of rank test procedures for censored survival data, *Biometrika*, *69*, 553–566.
- Hernan, M. A., B. Brumback, and J. M. Robins (2000), Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men, *Epidemiology*, *11*, 561–570.
- Kalbfleisch, J. D., and R. L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, 2nd ed., 251-252 pp., Wiley, New York.
- Kaplan, E. L., and P. Meier (1958), Nonparametric Estimation From Incomplete Observations, *Journal of the American Statistical Association*, *53*, 457–481.
- Klein, J. P., and P. K. Andersen (2005), Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function, *Biometrics*, *61*, 223–229.

- Klein, J. P., and M. L. Moeschberger (2003), *Survival Analysis*, 2nd ed., Springer, New York.
- Lin, D. Y. (1997), Non-parametric inference for cumulative incidence functions in competing risks studies, *Statistics in Medicine*, *16*, 901–910.
- Little, R. J. A., and D. B. Rubin (2002), *Statistical Analysis with Missing Data*, Wiley, New York.
- Moeschberger, M. L., and H. A. David (1971), Life tests under competing causes of failure and the theory of competing risks, *Biometrics*, *27*, 909–923.
- Pepe, M. S. (1991), Inference for events with dependent risks in multiple endpoint studies, *Journal of the American Statistical Association*, *86*, 770–778.
- Pepe, M. S., and M. Mori (1993), Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data?, *Statistics in Medicine*, *12*, 737–751.
- Robins, J. M. (1993), Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers, *Proceedings of the Biopharmaceutical Section, American Statistical Association, Alexandria, VA*, pp. 24–33.
- Robins, J. M., and A. Rotnitzky (1992), Recovery of information and adjustment for dependent censoring using surrogate markers, *AIDS Epidemiology-Methodological Issues*, *1*, 24–33.
- Robins, J. M., M. A. Hernan, and B. Brumback (2000), Marginal structural models and causal inference in epidemiology, *Epidemiology*, *11*, 550–560.
- Ruan, P. K., and R. J. Gray (2008), Analyses of cumulative incidence functions via non-parametric multiple imputation, *Statistics in Medicine*, *27*, 5709–5724.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, Wiley & Sons, New York.
- Schaubel, D. E., and M. Zhang (2010), Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event, *Lifetime Data Analysis*, *16*, 451–477.
- Scheike, T. H., M. J. Zhang, and T. A. Gerds (2008), Predicting cumulative incidence probability by direct binomial regression, *Biometrika*, *95*, 205–220.

- Stanford, D. A., E. M. Renouf, and V. C. McAlister (2008), Waiting for liver transplantation in Canada: Waitlist history 2000-04 and sensitivity analysis for the future, *Health Care Management Science*, 11, 184–195.
- Taylor, J. M. G., S. Murray, and C. H. Hsu (2002), Survival estimation and testing via multiple imputation, *Statistics and Probability Letters*, 58, 221–232.
- Tsiatis, A. A. (1975), A nonidentifiability aspect of the problem of competing risks, *Proceedings of the National Academy of the Sciences*, 72, 20–22.
- Zhang, M. J., and J. P. Fine (2008), Summarizing differences in cumulative incidence functions, *Statistics in Medicine*, 27, 4939–4949.