SEMIPARAMETRIC METHODS FOR SURVIVAL DATA WITH CLUSTERING, OUTCOME-DEPENDENT SAMPLING, DEPENDENT CENSORING, AND EXTERNAL TIME-DEPENDENT COVARIATE

by

Hui Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2012

Doctoral Committee:

Professor John D. Kalbfleisch, Co-Chair
Associate Professor Douglas E. Schaubel, Co-Chair
Associate Professor Bin Nan
Associate Professor Rajiv Saran

ACKNOWLEDGEMENTS

I am very grateful to all the people who have helped me at different stages during the completion of this thesis. In particular, I would like to thank my advisors, Dr. Douglas E. Schaubel and Dr. John D. Kalbfleisch, who guided me throughout these years of my graduate study and demonstrated how cutting-edge research should be performed. Their tremendous amount of support and encouragement made this dissertation possible. I also want to thank Canadian Organ Replacement Register (CORR) of the Canadian Institute for Health Information, the Dialysis Outcomes and Practice Patterns Study (DOPPS) and the Scientific Registry of Transplant Recipients (SRTR) for providing permission to use their data. Moreover, I thank my committee members, Dr. Bin Nan and Dr. Rajiv Saran, who provided me valuable comments and suggestions about my thesis. Finally, I wish to thank everybody around me, particularly, my parents, my husband Xinghua and son Eric, as well as many close friends including Rui, for providing strong support to my work and much joy to my life at the University of Michigan.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

In this dissertation, I investigate three important problems concerning proportional hazards regression. In two cases, complexities in the data structure require that new methodology be developed; in the third case, existing methods are applied in an interesting and non-standard manner.

In Chapter II, statistical methods are developed for case-cohort designs for clustered failure time data. The case-cohort design is commonly used in large cohort studies. Under this design, covariate data are collected for a random sample (named the subcohort) from the entire cohort, and for any additional subject who experienced the event of interest (cases) outside the subcohort. Therefore, this design is appealing for large cohort studies because of its cost savings, in particular when the disease is rare. It is also useful when multiple disease outcomes are of interest, since the same subcohort can be used as control group for each of the outcomes.

A number of methods have been proposed in the literature for the analysis of case-cohort data. For example, different approaches for estimating the regression parameters and variance estimators have been proposed under the proportional hazards model by Prentice (1986), Self and Prentice (1988), Wacholder et al. (1989), Lin and Ying (1993), Barlow (1994), Chen and Lo (1999), Borgan et al. (2000),

Sorensen and Anderson (2000), Chen (2001b), and Samuelsen, Anestad, and Skrondal (2007). Computation through standard statistical software of the regression parameter and corresponding variance estimators have been described by Therneau and Li (1999) and Langholz and Jiao (2007) in the context of case-cohort data. Many other regression models have also been developed to analyze case-cohort data, including the additive hazards regression model of Kulich and Lin (2000), Sun, Sun, and Flournoy (2004), and Ma (2007); semiparametric transformation models of Chen (2001a), Kong, Cai, and Sen (2004, 2006); and accelerated failure time models of Nan, Yu, and Kalbfleisch (2006) and Nan, Kalbfleisch, and Yu (2009).

Each of the afore-listed methods of analysis for case-cohort data has concerned the analysis of univariate failure time data. However, clustered failure time data are often encountered in public health studies. For example, patients treated at the same center are unlikely to be independent. In general, two approaches are proposed dealing with clustered failure time data. A conditional model is more appropriate when the within-cluster covariate effect is of interest, e.g., Moger, Pawitan, and Borgan (2008). When the population-averaged covariate effect is of interest, a marginal model is appealing. This model leaves the unobservable correlation structure of clustered data unspecified. Examples include Wei, Lin, and Weissfeld (1989); Lee, Wei, and Amato (1992); Cai and Prentice (1995); Spiekerman and Lin (1998); Lu and Wang (2005). Of particular note, Lu and Shih (2006) considered a marginal approach to extend case-cohort designs for clustered failure time data.

In Chapter II, we develop methods based on estimating equations for case-cohort designs for clustered failure time data. We assume a marginal hazards model, with a common baseline hazard and common regression coefficient across clusters. The proposed estimators of the regression parameter and cumulative baseline hazard are

shown to be consistent and asymptotically normal, and consistent estimators of the asymptotic covariance matrices are derived. The regression parameter estimator is easily computed using any standard Cox regression software that allows for offset terms. The proposed estimators are then investigated in simulation studies, and demonstrated empirically to have increased efficiency relative to some existing methods. The proposed methods are applied to a study of mortality among Canadian dialysis patients. Therefore, the methods developed in this chapter will complement, if not substitute, current methods in treating complex case-cohort data that are encountered more commonly.

The case-cohort design is a special case of what is known as an outcome-dependent sampling (ODS) design, wherein subjects are selectively sampled based on the outcomes of interest (e.g., death, survival). Efficient and cost-saving sampling schemes can be derived through ODS. Most methods for analyzing ODS-based data have an underlying assumption that, given covariate information, the censoring and failure times are independent. However, this assumption is sometimes violated in public health studies. For example, wait-listed end-stage liver disease (ESLD) patients may receive a liver transplant and therefore not die on the wait-list, an issue which could produce substantial bias in the estimation of wait-list mortality if treated as independent censoring. The Inverse Probability of Censoring Weighting (IPCW) method has been widely used for the analysis of dependently censored data. There is much literature dealing with the IPCW method; e.g., Robins and Rotnitzky (1992); Robins (1993a); Robins and Finkelstein (2000) ; Scharfstein and Robins (2002a); Matsuyama and Yamaguchi (2008).

In Chapter III, we consider failure time data in the setting with both ODS and dependent censoring. We propose hazard regression methods based on weighted es-

timating equations which employ a double-inverse-weighting scheme. The proposed weights correspond to the probability of being sampled and the probability of remaining uncensored. The proposed estimators of the regression parameter are shown to be consistent and asymptotically normal, and consistent estimators of the asymptotic covariance matrices are derived. Finite sample properties of the proposed estimators are examined through simulation studies. The proposed methods are applied to investigate liver wait-list mortality using data obtained from the Scientific Registry of Transplant Recipients (SRTR).

Chapter IV deals with the challenges of fitting complex models used in the real data analysis in Chapter II to data from the smaller countries participating in the Dialysis Outcomes and Practice Patterns Study (DOPPS), a well-known international cohort study. Hemodialysis (HD) is the most common method of renal replacement therapy. Under a thrice weekly HD schedule (Mon/Wed/Fri, Tue/Thu/Sat) the highest risk of death is thought to be on Mondays and Tuesdays since these days follow the longest intervals without the benefit of dialysis. Many studies have assessed the association between day-of-week-specific mortality risk and dialysis schedule. Examples include Bleyer, Russell, and Satko (1999); Karnik et al. (2001); Bleyer et al. (2006). These studies show that there is an increased risk of sudden death on Monday for MWF schedule patients, and on Tuesday for TTS schedule patients.

However, this phenomenon of highest risk of mortality on Monday for MWF schedule patients, and on Tuesday for TTS schedule patients has rarely been studied in large databases. Moreover, no previous study has addressed whether the day-of-week effect is similar across countries. Most importantly, the logistic regression model, the most commonly used model for the analysis of HD studies, is difficult to adjust for time-dependent covariates or appropriately account for censoring. Survival

analysis (Cox regression) is the natural choice, since the endpoint is time-to-death.

In Chapter IV, we proposed a Cox model with time-dependent covariates to evaluate the association between dialysis schedule and day-of-week-specific mortality for DOPPS patients from the U.S., European countries and Japan. Three models were fitted, distinguished by the factor of interest: (i) day of the week (ii) day of dialysis schedule (iii) days since last dialysis. The models are compared and contrasted, with special attention given to the setting where the sample size is small. We addressed whether the Monday/Tuesday effect is similar across countries. Our results indicate that in all three regions, HD patients have a higher death risk on Mondays if they are on a MWF schedule, or Tuesdays if they are on TTS schedule. Our results imply that there may be advantages to a more frequent dialysis schedule, an idea which has not been evaluated frequently in the nephrology literature.

CHAPTER II

Proportional Hazards Regression for the Analysis of Clustered Survival Data from Case-Cohort Studies

2.1 Introduction

The case-cohort design is commonly used in large cohort studies. The design entails collecting covariate data for all subjects who experienced the event of interest (cases) in the full cohort, and for a random sample (the subcohort) from the entire cohort. Therefore, the most important advantage of this design is cost savings, especially when the disease is rare. A second advantage of the case-cohort design is that the subcohort can be used as the comparison group for multiple disease outcomes. A number of methods have been proposed for regression analysis of case-cohort data under the proportional hazards model. Prentice (1986) proposed a pseudo-likelihood method for estimating the regression parameter. Self and Prentice (1988) and Lin and Ying (1993), using different approaches, derived large sample properties of the pseudo-likelihood related estimators. Wacholder et al. (1989) presented variance estimators for the log relative hazard through a bootstrap resampling plan. Barlow (1994) proposed a computationally convenient robust variance estimator. Chen and Lo (1999) suggested a class of estimating functions which in many cases offered improved efficiency. Therneau and Li (1999) and Langholz and Jiao (2007) described the computation of parameter and variance estimates using common software pack-

ages, such as SAS and R/S-PLUS. Borgan et al. (2000), Chen (2001b) and Samuelsen et al. (2007) obtained more efficient estimators by different approaches. Sorensen and Anderson (2000) considered competing risks analysis of case-cohort data.

The case-cohort design has also been studied in the context of other regression models. For example, Kulich and Lin (2000), Sun et al. (2004) and Ma (2007) studied the case-cohort design under an additive hazards regression model. Chen (2001a) and Kong et al. (2004, 2006) considered semiparametric transformation models in the case-cohort design. Nan et al. (2006) and Nan et al. (2009) considered accelerated failure time models and rank based analyses in case-cohort designs.

Each of the studies in the preceding paragraphs focused on univariate failure time data. However, clustered failure time data are commonly encountered in biomedical research. For example, in a family disease study, members from the same family may be correlated due to shared genetic and/or environmental factors. Similarly, outcomes of patients treated at the same center may be correlated. In these cases, valid statistical inference requires that one account for the intra-cluster dependence. Methods proposed for handling clustered failure time data can generally be categorized into two approaches: conditional models and marginal models. As an example of a conditional approach, frailty models specify the correlation structure by postulating a random effect (frailty) that is common to individuals within the same cluster. The regression parameter for such models is interpreted conditional on the random effect. For example, Moger et al. (2008) proposed frailty based case-cohort methods for analyzing family survival data with families as the sampling unit. If the investigator is interested in population averaged covariate effects, a marginal model is appealing; such a model leaves the dependence structure unspecified in the model formulation, but adjusts for the dependence in the inference. Several methods

have been proposed for fitting marginal proportional hazards models; e.g, Wei et al. (1989); Lee et al. (1992); Cai and Prentice (1995); Spiekerman and Lin (1998); Lu and Wang (2005). Lu and Shih (2006) considered case-cohort designs adapted to clustered failure time data under a marginal model and developed inference procedures.

Our proposed method is motivated by a retrospective cohort study of a possible day-of-week effect on death rates among patients receiving hemodialysis to treat advanced kidney failure. Patients treated at the same renal center are likely to be correlated due to center-specific practice patterns as well as a tendency to share socio-economic and environmental characteristics. The dialysis schedule, Monday/Wednesday/Friday (M/W/F) or Tuesday/Thursday/Saturday (T/T/S), may put patients at higher risk of death on certain days. For example, patients may have higher risk of death on Monday and Tuesday since, on average, these days follow the longest intervals without dialysis.

In this chapter, we propose methods based on estimating equations for three case-cohort designs that are applicable to clustered survival data. We assume a marginal proportional hazards model with a common baseline hazard and common regression coefficient across clusters. The case-cohort sampling designs we consider are similar to those proposed by Lu and Shih (2006). However, the designs we propose feature Bernoulli sampling, which is convenient for establishing theoretical properties. More importantly, we construct the risk sets using not only the information in the subcohort, but also the information collected on future deaths, similar to Chen and Lo (1999). As a result, the proposed estimators have increased efficiency relative to those of Lu and Shih (2006).

The remainder of this chapter is organized as follows. In Section 2, we describe the

proposed estimation procedures. In Section 3, we derive large sample properties for the proposed estimators. We conduct simulation studies in Section 4 to investigate the finite sample properties of the proposed estimators. In Section 5, we apply the proposed methods to a national organ failure database. The chapter concludes with some discussion in Section 6. All proofs are presented in the Web Appendix.

2.2 Proposed Methods

We first describe case-cohort designs with Bernoulli sampling for clustered failure time data. The full cohort consists of n independent clusters, and the ith cluster $(i = 1, \ldots, n)$ has m_i correlated subjects. We assume that subjects within the same cluster are exchangeable. In advance of follow-up, a random sample of the entire cohort, called the subcohort, is selected. Covariate data are then collected from individuals in the subcohort as well as those observed to fail in the entire cohort. Three designs are considered to obtain the subcohort:

- Design A: Randomly sample individuals from each cluster with Bernoulli sampling. That is, each individual in each cluster has an independent fixed probability of being selected to the subcohort.
- Design B: Randomly sample clusters from the full cohort with Bernoulli sampling.
- Design C: Randomly sample clusters from the full cohort with Bernoulli sampling, then randomly sample subjects with Bernoulli sampling from the selected clusters.

These are the same designs proposed by Lu and Shih (2006), except that we consider Bernoulli sampling, which greatly simplifies asymptotic derivations. Note that Design A and Design B are special cases of Design C.

Let T_{ij} and C_{ij} be the failure time and censoring time, where (i, j) represents the jth subject in the ith cluster. Let $\boldsymbol{Z}_{ij}(t)$ be the p-vector of possibly time-dependent covariates; with any time-dependent covariates assumed to be external (Kalbfleisch and Prentice, 2002). We assume that T_{ij} and C_{ij} are independent conditional on the observed covariates. Let $X_{ij} = T_{ij} \wedge C_{ij}$, $Y_{ij}(t) = I\left(X_{ij} \geq t\right)$, $\delta_{ij} = I\left(T_{ij} < C_{ij}\right)$, and $N_{ij}(t) = I\left(T_{ij} \leq C_{ij} \wedge t\right)$, where $I(\cdot)$ is the indicator function and $a \wedge b = \min\{a, b\}$. We assume that $\{N_{ij}(\cdot), Y_{ij}(\cdot), \boldsymbol{Z}_{ij}(\cdot), m_i :$

$j = 1, \ldots, m_i\}$ are independently and identically distributed for $i = 1, \ldots, n$. Let H_i indicate whether or not cluster i is selected into the subcohort, and let H_{ij} be the indicator for subject (i, j) being sampled as a potential individual in the subcohort. Subject (i, j) is selected into the subcohort if and only if $H_i H_{ij} = 1$. The variates H_i and H_{ij} are assumed to be independent of $\{N_{ij}(\cdot), Y_{ij}(\cdot), \boldsymbol{Z}_{ij}(\cdot), m_i : j = 1, \ldots, m_i\}$, for all i, j. Under Design A, B, and C, the H_i's are independent Bernoulli variables with $\mathcal{E}(H_i) = \gamma$ for all $i = 1, \ldots, n$, where $\mathcal{E}(\cdot)$ denotes expectation, and the H_{ij}'s are independent Bernoulli variables with $\mathcal{E}(H_{ij}) = \theta$, for all $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Under Design A, $H_i = 1$, for all $i = 1, \ldots, n$, i.e., $\gamma = 1$. Under Design B, $H_{ij} = 1$, for all $i = 1, \ldots, n$, $j = 1, \ldots, m_i$; i.e., $\theta = 1$.

Let the marginal hazard of failure of individual (i, j) be specified by a proportional hazards model (Cox (1972)),

$$\lambda_{ij}(t) = \lambda_0(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(t)}, \tag{2.1}$$

where $\lambda_0(\cdot)$ is an unspecified marginal baseline hazard function and $\boldsymbol{\beta}_0$ is a p-dimensional regression parameter. Since we are primarily interested in the estimation of $\boldsymbol{\beta}_0$, we leave the dependence structure of individuals within a cluster unspecified.

Many authors have studied the estimation of the regression parameters under model (2.1). Under a working independence assumption, Lee, Wei and Amato (1992)

proposed the estimating function

$$\boldsymbol{U}_{LWA}(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^{\tau}\{\boldsymbol{Z}_{ij}(u) - \boldsymbol{E}_{LWA}(\boldsymbol{\beta},u)\}\,dN_{ij}(u),$$

where $\tau < \infty$ equals the maximum follow-up time, $\boldsymbol{E}_{LWA}(\boldsymbol{\beta},u) = \boldsymbol{S}^{(1)}_{LWA}(\boldsymbol{\beta},u)/S^{(0)}_{LWA}(\boldsymbol{\beta},u)$, $\boldsymbol{S}^{(d)}_{LWA}(\boldsymbol{\beta},u) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}$, with $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Then, $\boldsymbol{\beta}_0$ of model (2.1) can be estimated with $\widehat{\boldsymbol{\beta}}_{LWA}$, the solution to the estimating equation $\boldsymbol{U}_{LWA}(\boldsymbol{\beta}) = 0$. Lu and Shih (2006) considered case-cohort designs for clustered failure time data under model (2.1) and proposed to estimate $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}_{LS}$, the root of the estimating equation $\boldsymbol{U}_{LS}(\boldsymbol{\beta}) = 0$, where

$$\boldsymbol{U}_{LS}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^{\tau}\{\boldsymbol{Z}_{ij}(u) - \boldsymbol{E}_{LS}(\boldsymbol{\beta},u)\}\,dN_{ij}(u),$$

where $\boldsymbol{E}_{LS}(\boldsymbol{\beta},u) = \boldsymbol{S}^{(1)}_{LS}(\boldsymbol{\beta},u)/S^{(0)}_{LS}(\boldsymbol{\beta},u)$ and
$\boldsymbol{S}^{(d)}_{LS}(\boldsymbol{\beta},u) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_iH_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}$.

Lu and Shih (2006) used only subcohort subjects to construct the risk set. Since information on all failures in the full cohort are available, failures outside the subcohort can also contribute to the risk set, as proposed by Chen and Lo (1999) for independent subjects. We propose three procedures to estimate $\boldsymbol{\beta}_0$, the procedures differing with respect to their treatment of the marginal observed-event probability, $Pr(\delta_{ij} = 1)$, which we denote by p_0. In the first proposed procedure, p_0 is assumed known, which follows the Chen and Lo (1999) approach. Usually, p_0 is not known, but this gives a baseline to which other approaches can be compared. We estimate $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_t$, the solution to $\boldsymbol{U}(\boldsymbol{\beta},p_0) = 0$, where

$$(2.2)\quad \boldsymbol{U}(\boldsymbol{\beta},p) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^{\tau}\left\{\boldsymbol{Z}_{ij}(u) - \overline{\boldsymbol{E}}(\boldsymbol{\beta},p,u)\right\}dN_{ij}(u)$$

$$\overline{\boldsymbol{E}}(\boldsymbol{\beta},p,u) = \frac{\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta},p,u)}{\overline{S}^{(0)}(\boldsymbol{\beta},p,u)}$$

$$\overline{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta}, p, u) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\frac{p}{N_1}\delta_{ij} + \frac{1-p}{n_0}(1-\delta_{ij})H_iH_{ij}\right\} Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}$$

with $N_1 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}$, and $n_0 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(1-\delta_{ij})H_iH_{ij}$. The motivation for building estimating equation (2.2) is that $\overline{\boldsymbol{E}}(\boldsymbol{\beta}, p_0, u)$ is a consistent estimator of $\mathcal{E}\left\{\boldsymbol{Z}_{ij}(u)|X_{ij}=u,\ \ \delta_{ij}=1\right\}$, where

$$\begin{aligned}
&\mathcal{E}\left\{\boldsymbol{Z}(u)|X=u,\delta=1\right\}\\
&= \frac{\mathcal{E}\left\{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}\right\}}{\mathcal{E}\left\{Y(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}\right\}}\\
&= \frac{p_0\mathcal{E}\left\{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}|\delta=1\right\} + (1-p_0)\mathcal{E}\left\{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}|\delta=0\right\}}{p_0\mathcal{E}\left\{Y(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}|\delta=1\right\} + (1-p_0)\mathcal{E}\left\{Y(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(u)}|\delta=0\right\}}.
\end{aligned}$$

(2.3)

The first (second) conditional means in numerator and denominator can be estimated by their respective empirical counterparts from all failures in the whole cohort (controls in the subcohort). A derivation of (2.3) is given in the Web Appendix.

In almost all settings, the population failure probability, p_0, is unknown but can be estimated using the subcohort case proportion, $\widehat{p}_s$, or the full cohort case proportion, $\widehat{p}_w$. These give rise to estimating functions $\boldsymbol{U}(\boldsymbol{\beta}, \widehat{p}_s)$ and $\boldsymbol{U}(\boldsymbol{\beta}, \widehat{p}_w)$, with solutions $\widehat{\boldsymbol{\beta}}_s$ and $\widehat{\boldsymbol{\beta}}_w$, respectively. In cases where the study cohort is well defined, $\widehat{p}_w$ can be computed and used to obtain $\widehat{\boldsymbol{\beta}}_w$, which has the most practical value. When the study cohort is less well-defined, $\widehat{\boldsymbol{\beta}}_s$ is a suitable alternative. For example, if the study does not have a roster for the full cohort (such that the cohort size, N, is not known), then $\widehat{\boldsymbol{\beta}}_s$ can still be used.

Some simple algebra shows that

$$\begin{aligned}
\overline{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta}, \widehat{p}_s, u) \;=\; &\frac{\widehat{q}_1}{n_0+n_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\delta_{ij} + \frac{1}{\widehat{q}_1}(1-\delta_{ij})H_iH_{ij}\right\}\\
&\times Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}
\end{aligned}$$

$$\overline{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta}, \widehat{p}_w, u) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\delta_{ij} + \frac{1}{\widehat{q}_0}(1-\delta_{ij})H_iH_{ij}\right\} Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d},$$

where $N = N_0+N_1$, $\widehat{q}_1 = n_1/N_1$, and $\widehat{q}_0 = n_0/N_0$, with $n_1 = \sum_{i=1}^{n}\sum_{j=1}^{m_i} H_iH_{ij}\delta_{ij}$ and $N_0 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(1-\delta_{ij})$. The estimating equations are similar, therefore, to those arising from inverse sampling probability weighting (ISPW), such as that proposed by Kalbfleisch and Lawless (1988) and Borgan et al. (2000) for the Cox model; Kulich and Lin (2000) for the additive hazards model; and Nan, Kalbfleisch and Yu (2009) for the accelerated failure time model. These studies focused on univariate failure time data.

The cumulative baseline hazard function, $\Lambda_0(t) = \int_0^t \lambda_0(u)du$, can be consistently estimated by

$$\widehat{\Lambda}_0(t; \widehat{\boldsymbol{\beta}}, \widehat{p}) = \int_0^t \frac{d\overline{N}(u)}{\widehat{\mu}\overline{S}^{(0)}(\widehat{\boldsymbol{\beta}}, \widehat{p}, u)}, \tag{2.4}$$

where $\overline{N}(u) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} N_{ij}(u)$, $\mu = \mathcal{E}(m_i)$, and $\widehat{\mu} = n^{-1}\sum_{i=1}^{n} m_i$. In (2.4), either $\widehat{p}_s$, $\widehat{p}_w$ or p_0 could be used.

The proportional hazards assumption may be violated for one or more covariates, which could be individual level covariates such as age or time since first dialysis, or cluster level covariates such as center size. Our proposed methods can be extended to allow for stratification on such covariates; details can be found in the Web Appendix Section A.5.

2.3 Asymptotic Properties of the Proposed Estimators

We make the following assumptions:

(*a*) $\{N_{ij}(\cdot), Y_{ij}(\cdot), \boldsymbol{Z}_{ij}(\cdot), m_i : j = 1, \ldots, m_i\}, i = 1, \ldots, n$ are independently and identically distributed.

(*b*) $P\{Y_{ij}(t) = 1\} > 0$ for $t \in (0, \tau]$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, and all m_i .

(c) $|Z_{ijh}(0)| + \int_0^\tau |dZ_{ijh}(t)| < B_{\boldsymbol{Z}} < \infty$ for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, and all m_i, where Z_{ijh} is the hth component of $\boldsymbol{Z}_{ij}$ and $B_{\boldsymbol{Z}}$ is a constant.

(d) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that $\sup_{u\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \|\boldsymbol{S}^{(d)}(\boldsymbol{\beta},u) - \boldsymbol{s}^{(d)}(\boldsymbol{\beta},u)\| \xrightarrow{P} 0$ for $d = 0, 1, 2$, where $\boldsymbol{s}^{(d)}(\boldsymbol{\beta},u) = \mathcal{E}\left\{\boldsymbol{S}^{(d)}(\boldsymbol{\beta},u)\right\}$ is absolutely continuous, for $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $u \in (0,\tau]$. Moreover, $s^{(0)}(\boldsymbol{\beta},u)$ is assumed to be bounded away from zero.

(e) For $d = 0, 1, 2$, $\sup_{u\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \|\overline{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta},p,u) - \mu^{-1}\boldsymbol{s}^{(d)}(\boldsymbol{\beta},u)\| \xrightarrow{P} 0$.

(f) The matrix $\boldsymbol{A}(\boldsymbol{\beta}_0)$ is positive definite, where

$$\boldsymbol{A}(\boldsymbol{\beta}) = \int_0^\tau \left\{\boldsymbol{s}^{(2)}(\boldsymbol{\beta},u)/s^{(0)}(\boldsymbol{\beta},u) - \boldsymbol{e}(\boldsymbol{\beta},u)^{\otimes 2}\right\} dF(u)$$

with $\boldsymbol{e}(\boldsymbol{\beta},u) = \boldsymbol{s}^{(1)}(\boldsymbol{\beta},u)/s^{(0)}(\boldsymbol{\beta},u)$, and $F(u) = \mathcal{E}\left\{\overline{N}(u)\right\}$.

(g) $\Lambda_0(\tau) < \infty$, and $\lambda_0(t)$ is absolutely continuous for $t \in (0,\tau]$.

Our main results are given in Theorems 1 - 4 below, the proofs of which are given in the Web Appendix. We provide only brief summary remarks about the proofs below.

Theorem 1: *Under conditions* $(a)-(g)$, *as* $n \to \infty$, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, p_0)$ *converges to a mean zero Normal distribution with covariance* $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, p_0) = \mathcal{E}\left\{\boldsymbol{W}_1(\boldsymbol{\beta}_0, p_0)^{\otimes 2}\right\}$, *with*

$$\begin{aligned}
\boldsymbol{W}_i(\boldsymbol{\beta},p) &= \sum_{j=1}^{m_i} \int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\} \\
&\quad \times \left[dN_{ij}(u) - \left\{\frac{1}{\mu}\delta_{ij} + \frac{1}{\mu\gamma\theta}(1-\delta_{ij})H_iH_{ij}\right\} Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \right. \\
&\quad \left. \times \left\{\mu^{-1}s^{(0)}(\beta,u)\right\}^{-1} dF(u)\right] + \boldsymbol{D}_1(\boldsymbol{\beta})G_{1i}(p) + \boldsymbol{D}_2(\boldsymbol{\beta})G_{2i}(p) \\
\boldsymbol{D}_1(\boldsymbol{\beta}) &= \mathcal{E}\left[\sum_{j=1}^{m_1} \int_0^\tau \{\boldsymbol{Z}_{1j}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\} \frac{\delta_{1j}}{\mu^2 p} Y_{1j}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1j}(u)}\right.
\end{aligned}$$

$$\times \frac{\sum_{k=1}^{n}\sum_{l=1}^{m_k} dN_{kl}(u)}{\mu^{-1}s^{(0)}(\boldsymbol{\beta},u)}\Bigg]$$

$$\boldsymbol{D}_2(\boldsymbol{\beta}) = \mathcal{E}\left[\sum_{j=1}^{m_1}\int_0^{\tau}\{\boldsymbol{Z}_{1j}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\right.$$
$$\left.\times\frac{(1-\delta_{1j})H_1H_{1j}}{(\mu\gamma\theta)^2(1-p)}Y_{1j}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1j}(u)}\frac{\sum_{k=1}^{n}\sum_{l=1}^{m_k}dN_{kl}(u)}{\mu^{-1}s^{(0)}(\boldsymbol{\beta},u)}\right]$$

$$G_{1i}(p) = n^{-1}\left(\sum_{j=1}^{m_i}\delta_{ij}-\mu p\right)$$

$$G_{2i}(p) = n^{-1}\left\{\sum_{j=1}^{m_i}(1-\delta_{ij})H_iH_{ij}-\mu\gamma\theta(1-p)\right\}.$$

In the Web Appendix, we show that $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0,p_0)=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{W}_i(\boldsymbol{\beta}_0,p_0)+o_p(1)$; hence, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0,p_0)$ is essentially a scaled sum of n independent and identically distributed random quantities with mean zero and finite variance. The proof of asymptotic normality follows from the Multivariate Central Limit Theorem (MCLT) and various results from empirical process theory. The result in Theorem 1 is used to derive the limiting distribution of the proposed estimators.

Theorem 2*: Under conditions* $(a)-(g)$*,* $\widehat{\boldsymbol{\beta}}_t$ *converges in probability to* $\boldsymbol{\beta}_0$ *and* $n^{1/2}(\widehat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0)$ *converges in distribution to a mean zero normal distribution with covariance matrix* $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0,p_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$*.*

The proof of the consistency of $\widehat{\boldsymbol{\beta}}_t$ follows by the Inverse Function Theorem (Foutz, 1977). The proof of asymptotic normality follows from a Taylor series expansion and the Cramèr-Wold device.

Theorem 3*: Under conditions* $(a)-(g)$*, both* $\widehat{\boldsymbol{\beta}}_s$ *and* $\widehat{\boldsymbol{\beta}}_w$ *converge in probability to* $\boldsymbol{\beta}_0$*, and each of* $n^{1/2}(\widehat{\boldsymbol{\beta}}_s-\beta_0)$ *and* $n^{1/2}(\widehat{\boldsymbol{\beta}}_w-\boldsymbol{\beta}_0)$ *converges in distribution to a zero-mean Normal with covariance matrices* $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_s(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ *and* $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_w(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ *respectively, where* $\boldsymbol{\Omega}_a(\beta)=\mathcal{E}\left\{\boldsymbol{\psi}_1^a(\boldsymbol{\beta},p_0)^{\otimes 2}\right\}$ *and* $\boldsymbol{\psi}_i^a(\boldsymbol{\beta},p)$

$= \boldsymbol{W}_i(\boldsymbol{\beta}, p) + \boldsymbol{B}(\boldsymbol{\beta}) Q_i^a(p)$ *for* $a = s$ *or* w, *with* $Q_i^s(p) = \{\mu\gamma\theta\}^{-1} \times \sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p)$, $Q_i^w(p) = \mu^{-1} \sum_{j=1}^{m_i}(\delta_{ij} - p)$, *and*

$$\begin{aligned}
\boldsymbol{B}(\boldsymbol{\beta}) &= \int_0^\tau \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, u) r^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\beta, u)^2} - \frac{\boldsymbol{r}^{(1)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}, u)} \right\} dF(u) \\
\boldsymbol{r}^{(d)}(\boldsymbol{\beta}, u) &= \frac{1}{p_0} \mathcal{E}\left\{ \delta_{11} Y_{11}(u) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{11}(u)} \boldsymbol{Z}_{11}(u)^{\otimes d} \right\} \\
&\quad - \frac{1}{1 - p_0} \mathcal{E}\left\{ (1 - \delta_{11}) Y_{11}(u) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{11}(u)} \boldsymbol{Z}_{11}(u)^{\otimes d} \right\}.
\end{aligned}$$

The results in Theorem 1, combined with two Taylor series expansions, the MCLT and Slutsky's Theorem, conclude the proof of asymptotic normality of $\widehat{\boldsymbol{\beta}}_s$ and $\widehat{\boldsymbol{\beta}}_w$ in Theorem 3. The covariance matrices in Theorems 2 and 3 can be consistently estimated from the observed case-cohort data, as described in the Web Appendix.

We now describe asymptotic results pertaining to the proposed baseline cumulative hazard estimator.

Theorem 4*: Under conditions* $(a) - (g)$, $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t)$ *converges in probability to* $\Lambda_0(t)$ *uniformly in* $t \in [0, \tau]$, *and* $n^{1/2}\left\{ \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \Lambda_0(t) \right\}$ *converges weakly to a Gaussian process with mean zero and covariance function at* (s, t) *given by* $\mathcal{E}\left\{ \boldsymbol{\phi}_1(\boldsymbol{\beta}_0, p_0, s) \boldsymbol{\phi}_1(\boldsymbol{\beta}_0, p_0, t) \right\}$, *where*

$$\begin{aligned}
\boldsymbol{\phi}_i(\boldsymbol{\beta}, p, t) &= k(\boldsymbol{\beta}, p, t) Q_i(p) + \boldsymbol{h}^T(\boldsymbol{\beta}, p, t) \boldsymbol{A}(\boldsymbol{\beta}) \boldsymbol{\psi}_i(\boldsymbol{\beta}, p) + \chi_i(\boldsymbol{\beta}, p, t) \\
\chi_i(\boldsymbol{\beta}, p, t) &= \sum_{j=1}^{m_i} \int_0^t \frac{1}{s^{(0)}(\boldsymbol{\beta}, u)} dM_{ij}(u) \\
&\quad + \sum_{j=1}^{m_i} \int_0^t \frac{1}{s^{(0)}(\boldsymbol{\beta}, u)^2} \left\{ 1 - \delta_{ij} - \frac{1}{\gamma\theta}(1 - \delta_{ij}) H_i H_{ij} \right\} \\
&\qquad \times Y_{ij} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(u)} dF(u) \\
k(\boldsymbol{\beta}, p, t) &= -\int_0^t \frac{\mu r^{(0)}(\boldsymbol{\beta}, p, u)}{s^{(0)}(\boldsymbol{\beta}, u)} d\Lambda_0(u) \\
\boldsymbol{h}(\boldsymbol{\beta}, p, t) &= -\int_0^t \boldsymbol{e}(\boldsymbol{\beta}, u) d\Lambda_0(u).
\end{aligned}$$

A sketch of the proof is given in Web Appendix A.

2.4 Numerical Studies

We conducted simulation studies to investigate the finite sample properties of the estimators proposed in Section 2, and to compare the proposed methods with those of Lu and Shih (2006). We generated clustered failure time data from $n = 100$ clusters. Cluster sizes, m_i, were simulated from a Binomial (50,0.8) distribution for $i = 1, \ldots, n$, with $\mu = \mathcal{E}(m_i) = 40$. The covariate Z_{ij} took values 1 and 0, with probabilities 0.5 and 0.5 respectively. The failure time for the jth subject within the ith cluster was simulated from a distribution with conditional hazard function

$$\lambda_{ij}(t|\, Z_{ij}, Q_i) = Q_i h_0(t) \exp\{\xi_0 Z_{ij}\},$$

where Q_i is a frailty variable following a positive stable distribution with index $\alpha = 0.8$. The variate Q_i is generated following the method in Chambers et al. (1976),

$$Q_i = \frac{\sin(\alpha Q_{1i})}{\{\sin(Q_{1i})\}^{1/\alpha}} \left[\frac{\sin\{(1-\alpha)Q_{1i}\}}{Q_{2i}}\right]^{(1-\alpha)/\alpha},$$

where Q_{1i} follows a $U(0, \pi)$ distribution, Q_{2i} follows an exponential distribution with mean 1, and Q_{1i} and Q_{2i} are independent. The baseline hazard function is given by $h_0(t) = \alpha^{-1} t^{\alpha^{-1}-1}$, with ξ_0 set to $\log(0.5)/\alpha = -0.8664$ or 0. The resulting marginal hazard function is $\lambda_{ij}(t|\, Z_{ij}) = \lambda_0(t) \exp\{\beta_0 Z_{ij}\}$, and the marginal baseline hazard function is given by $\lambda_0(t) = 1, 0 \leq t < \infty$, $\beta_0 = \alpha\xi_0 = \log(0.5)$ or 0. The censoring times C_{ij} were constant and equal to 1, which led to average observed event probabilities of $p = 0.51$ or $p = 0.63$. For each data generation, for Design A, individuals within each cluster were selected into the subcohort by Bernoulli sampling with equal probability 0.2 or 0.15. For Design B, we selected clusters by

Bernoulli sampling with probability 0.2 or 0.15. For Design C, we first sampled clusters by Bernoulli sampling with probability 0.4 or 0.3, then sampled individuals from those selected clusters by Bernoulli sampling with probability 0.5. Therefore for each design, we would expect approximately 800 or 600 individuals in the subcohort. In another data configuration, $\beta_0 = \log(0.5)$, the marginal baseline hazard function is given by $\lambda_0 = 0.2$. The covariates Z_{ij} follows either a Bernoulli distribution, which takes value 1 with probability 0.5, or a Normal distribution with mean 0 and variance 1. The other settings were the same except that only approximately 800 individuals were sampled in the subcohort. In this data configuration, the average observed event probabilities are $p = 0.14$ and $p = 0.21$. Each data configuration was replicated 1000 times. The true case percentage, p_0, would typically be unknown in real world settings; however, it is of course available in our simulation study and is evaluated for comparison purposes.

Tables 2.1 and 2.2 display the results of our proposed estimators and those of Lu and Shih (2006). For each data configuration, we list the empirical bias (BIAS) and standard deviation (ESD), average asymptotic standard error (ASE), asymptotic relative efficiency (ARE) with respect to the full cohort and empirical coverage probability (CP). Each of the estimators is approximately unbiased, and the variance estimators appear to be reasonably accurate. The 95% empirical coverage probabilities are generally close to the nominal value. In Table 2.1, for Design B, slight under-estimation of the standard error and under-coverage occur when $\beta_0 = 0$ and $n_s = 600$. This is due to the small number of clusters in the subcohort. For Design B, clusters are sampled and all individuals in the selected clusters are kept in the subcohort. Little extra information is gained when more subjects in the same cluster are included, since subjects within cluster are correlated. However, more information

is available when the number of sampled clusters increases and, correspondingly, the under-coverage is reduced when n_s is increased to 800.

In Table 2.1, the proposed method appears to be more efficient than that of Lu and Shih (2006), at least for the examples considered. In comparing the proposed sampling designs, for approximately equal subcohort sizes, it appears that Design A is more efficient than Design C, which is more efficient than Design B. This can be attributed to differences in the number of clusters sampled and the resulting differences in the amount of independent information contained in the subcohort. This efficiency gain is more obvious when the covariate is cluster-specific (Web Table 6). In Table 2.2, the efficiency gain of the proposed methods over those of Lu and Shih (2006) is less evident in the presence of a lower event rate. This can be explained by there being fewer failures outside the subcohort to include in the risk sets.

Additional scenarios have been evaluated in order to examine various aspects, such as continuous covariates, stronger correlation among failure times, smaller number of clusters, smaller subcohort size, lower event rate, as well as the performance of the stratified methods. Results of several of these numerical studies are available in the Web Appendix. In the examples we evaluated, the proposed methods generally work well.

Also in the Web Appendix, the estimates based on the proposed methods are compared to those based on simple random sampling (SRS), and to an inverse sampling probability weighting (ISPW) method. The proposed methods do not lose efficiency relative to the SRS or ISPW methods, at least for the set-ups considered.

For the data settings with $\beta_0 = \log(0.5)$, $\lambda_0(t) = 1$ and $n_s = 800$, we calculated the average of the estimate of $\Lambda_0(t)$ at $t = 0.02, t = 0.04, \cdots, t = 1.0$ based on 1000 replications. Figure 2.1 displays the average point estimate for the cumulative

Table 2.1: Simulation results based on 1000 replications: $\beta_0 = \log(0.5)$.

Design & Method		$n = 100, n_s = 800$ Bias	ESD	ASE	ARE	CP	$n = 100, n_s = 600$ Bias	ESD	ASE	ARE	CP
					$\beta_0 = \log(0.5)$						
FC		-0.001	0.053	0.054	1.000	0.959	-0.001	0.053	0.054	1.000	0.959
A	SC	-0.003	0.083	0.082	0.434	0.954	-0.002	0.092	0.091	0.352	0.956
	WC	-0.003	0.082	0.082	0.434	0.953	-0.002	0.091	0.091	0.352	0.958
	T	-0.004	0.080	0.079	0.467	0.951	-0.003	0.089	0.088	0.377	0.950
	LS	-0.003	0.091	0.089	0.368	0.933	-0.003	0.103	0.100	0.292	0.941
B	SC	-0.001	0.083	0.084	0.413	0.946	-0.001	0.094	0.093	0.337	0.943
	WC	-0.003	0.086	0.086	0.394	0.940	-0.004	0.096	0.095	0.323	0.941
	T	-0.004	0.084	0.083	0.423	0.933	-0.004	0.094	0.093	0.337	0.929
	LS	-0.002	0.091	0.091	0.352	0.954	-0.004	0.104	0.102	0.280	0.943
C	SC	0.000	0.084	0.083	0.423	0.955	-0.001	0.093	0.092	0.345	0.941
	WC	-0.001	0.083	0.083	0.423	0.945	-0.002	0.093	0.092	0.345	0.942
	T	-0.002	0.082	0.080	0.456	0.942	-0.003	0.092	0.090	0.360	0.942
	LS	-0.002	0.090	0.090	0.360	0.944	-0.004	0.102	0.101	0.286	0.945
					$\beta_0 = 0$						
FC		0.000	0.040	0.040	1.000	0.942	0.000	0.040	0.040	1.000	0.942
A	SC	0.002	0.035	0.036	0.298	0.954	0.001	0.082	0.082	0.238	0.952
	WC	0.002	0.035	0.036	0.297	0.955	0.001	0.082	0.082	0.238	0.943
	T	0.002	0.035	0.036	0.297	0.952	0.001	0.083	0.082	0.238	0.943
	LS	0.006	0.038	0.041	0.225	0.967	0.003	0.095	0.095	0.177	0.951
B	SC	0.005	0.037	0.036	0.302	0.936	0.001	0.085	0.080	0.250	0.929
	WC	0.005	0.036	0.036	0.292	0.939	0.001	0.087	0.079	0.256	0.915
	T	0.005	0.036	0.036	0.293	0.939	0.001	0.087	0.079	0.256	0.913
	LS	0.002	0.041	0.041	0.231	0.946	0.002	0.099	0.093	0.185	0.922
C	SC	0.008	0.035	0.036	0.297	0.950	0.002	0.086	0.081	0.244	0.928
	WC	0.008	0.035	0.036	0.298	0.947	0.002	0.086	0.081	0.244	0.927
	T	0.008	0.035	0.036	0.298	0.947	0.002	0.086	0.081	0.244	0.922
	LS	0.012	0.040	0.041	0.225	0.939	0.003	0.097	0.095	0.177	0.937

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
The number of clusters $n = 100$, m_i follows a Bin(50,0.8) distribution, α=0.8, λ_0=1, censoring time C=1, Z follows a Bernoulli(0.5) distribution. The number of individuals in the subcohort is either $n_s = 800$ or $n_s = 600$.

Table 2.2: Simulation results with $p_0 = 0.14$ and $p_0 = 0.21$ based on 1000 replications.

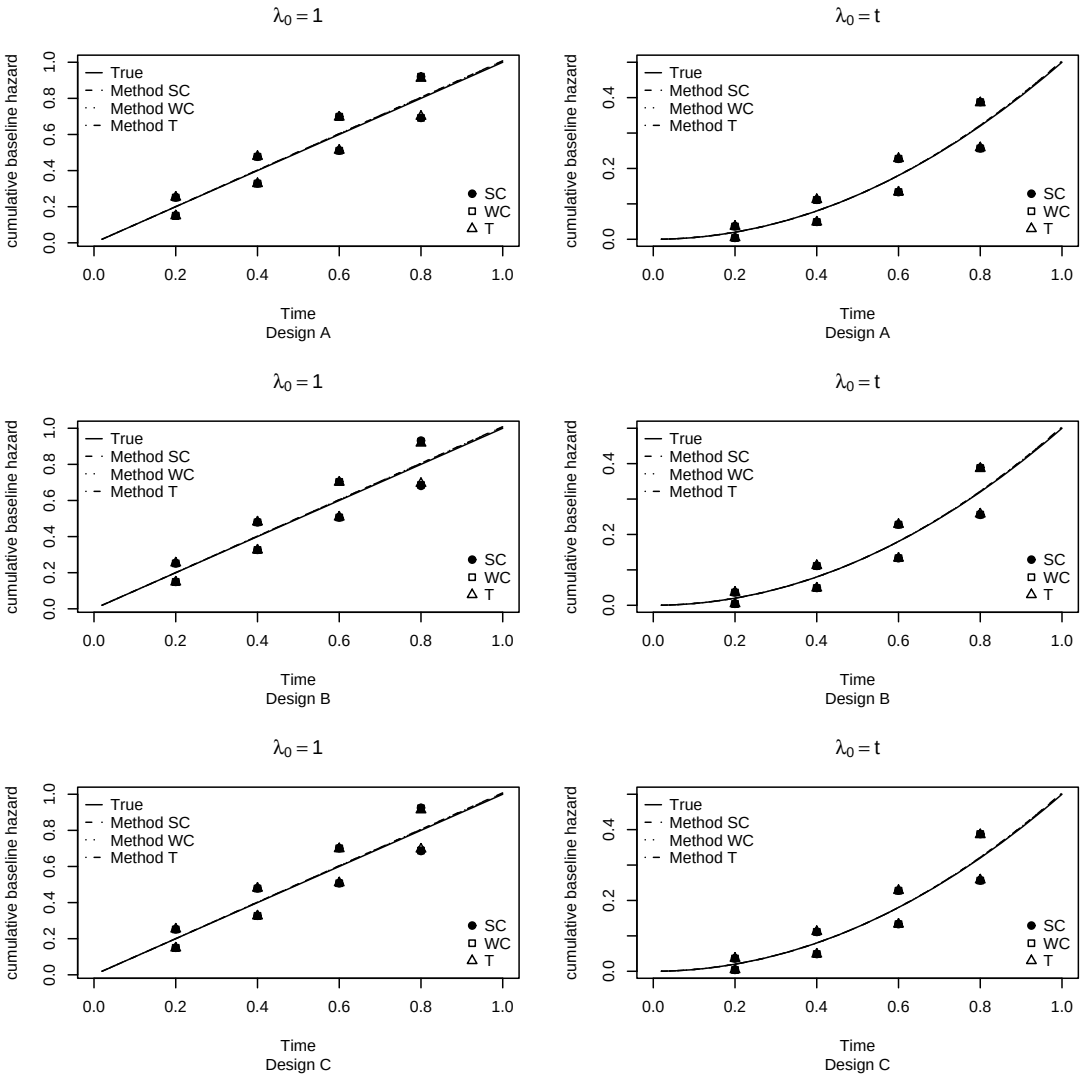
Design & Method		$Z \sim Bernoulli(0.5)$					$Z \sim N(0,1)$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		-0.010	0.111	0.106	1.000	0.934	-0.004	0.055	0.054	1.000	0.932
A	SC	-0.011	0.127	0.123	0.743	0.937	-0.006	0.068	0.068	0.631	0.944
	WC	-0.011	0.127	0.123	0.743	0.935	-0.006	0.068	0.067	0.650	0.940
	T	-0.011	0.125	0.121	0.767	0.935	-0.005	0.063	0.063	0.735	0.944
	LS	-0.011	0.127	0.124	0.731	0.940	-0.008	0.074	0.072	0.563	0.937
B	SC	-0.010	0.129	0.123	0.743	0.928	-0.006	0.069	0.069	0.612	0.943
	WC	-0.011	0.130	0.124	0.731	0.922	-0.007	0.069	0.069	0.612	0.938
	T	-0.011	0.128	0.122	0.755	0.927	-0.006	0.065	0.065	0.690	0.941
	LS	-0.011	0.129	0.124	0.731	0.930	-0.007	0.073	0.074	0.533	0.948
C	SC	-0.011	0.129	0.123	0.743	0.929	-0.004	0.071	0.068	0.631	0.936
	WC	-0.011	0.129	0.124	0.731	0.932	-0.003	0.070	0.068	0.631	0.930
	T	-0.011	0.127	0.122	0.755	0.929	-0.002	0.065	0.064	0.712	0.932
	LS	-0.011	0.130	0.124	0.731	0.937	-0.004	0.077	0.072	0.563	0.930

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
The number of clusters $n = 100$, m_i follows a Bin(50,0.8) distribution, α=0.8, λ_0=0.2, censoring time C=1, β=log(0.5), Z follows either a Bernoulli(0.5) distribution or a N(0,1) distribution, which corresponds to a marginal event rate of $p_0 = 0.14$ or $p_0 = 0.21$, respectively. The number of individuals in the subcohort is $n_s = 800$.

baseline hazards. The true cumulative baseline hazard is also included for comparison purposes. There appears to be no bias for our proposed estimators. We next assumed that the marginal baseline hazard function is given by $\lambda_0(t) = t$. Under this configuration, the proposed estimate is approximately unbiased.

2.5 Application

We applied the proposed methods to the estimation of the day-of-week effect among Canadian hemodialysis (HD) patients. The 1,276 patients who initiated HD between January 1, 1990 and December 31, 1990 were included in the analysis. Patients were followed from the time they first received HD until the time of death caused by cardiovascular disease (CVD), receiving transplantation, switching to peritoneal dialysis, loss to follow up, or last day of observation (December 31, 1998), whichever occurred first. Patients were clustered by center. In total, there were 70 centers yielding clusters with 1 to 75 patients and a mean of 18.2. Design A



Method SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; True = true cumulative baseline hazard function. The lines correspond to the average of $\widehat{\Lambda}_0(t)$ at each time point, and the points denote the confidence intervals of $\Lambda_0(t)$ based on the empirical standard deviation (ESD) at $t = 0.2, t = 0.4, t = 0.6$ and $t = 0.8$.

Figure 2.1: Simulation results to examine the cumulative baseline hazard estimators based on 1000 replications.

was chosen since, all else equal, it is generally at least as efficient as Designs B and C.

The primary outcome of interest is CVD death, and the covariate of interest is day of week (Sunday, Monday, ..., Saturday), which was coded using time-dependent covariates, where $Z_{ij1}(t) = I\{\text{day } t, \text{ for subject } (i,j), \text{ is a Monday}\}$, ...,$Z_{ij6}(t) =$

$I\{\text{day } t\text{, for subject } (i,j)\text{, is a Saturday}\}$, with Sunday chosen as the reference day, where t is the time since initiation of HD for patient (i,j). Adjustment covariates included age, gender, region, comorbid conditions and primary renal diagnosis. Age was categorized into 6 groups: <18, 18-39, 40-49, 50-59, 60-69, and ≥70, and was adjusted for through stratification. Patients from the same renal center may be correlated due to shared practice patterns. Therefore, one needs to account for such intra-cluster dependence for valid statistical inference. In total, there were 249 observed CVD deaths; hence, the event fraction for the full cohort was 0.195. In stratum 1 to 6, the numbers of CVD deaths were 0 (out of 24), 13 (out of 253), 14 (out of 179), 54 (out of 232), 87 (out of 313) and 81 (out of 275), respectively. We analyzed the data using Design A with sampling probability of 0.2. A total of 251 patients was selected into the subcohort. The point estimates were obtained using PROC PHREG in SAS with OFFSET terms, while the variance estimates were calculated using PROC IML. For comparison purposes, we also carried out a full cohort analysis and an analysis with the method of Lu and Shih (2006).

Results of the analysis are shown in Table 2.3. Using $\widehat{p}_s$, patients are estimated to have 1.36 and 1.68 times higher hazards of CVD death on Mondays and Tuesdays, respectively, compared to Sundays. Results based on $\widehat{p}_w$ were similar. Results from the full cohort analysis were close to those from our case-cohort analyses, with smaller standard errors. Results based on the method of Lu and Shih (2006) were also similar to ours, with larger standard errors.

The cumulative baseline hazards for each age-specific stratum are exhibited in Figure 2.2. Each sub-figure contains 3 lines, which correspond to the cumulative baseline hazards for Design A methods SC and WC, as well as the full cohort analysis. Since no CVD deaths occurred in stratum 1, cumulative baseline hazard estimation

Table 2.3: Estimate of day-of-week effect on CVD mortality among dialysis patients.

	Design A					
	SC			WC		
Day	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$
Sunday	0.00	0.00	1.00	0.00	0.00	1.00
Monday	0.31	0.27	1.36	0.33	0.26	1.39
Tuesday	0.52	0.28	1.68	0.51	0.28	1.67
Wednesday	-0.02	0.26	0.98	-0.004	0.25	1.00
Thursday	0.23	0.29	1.26	0.24	0.29	1.27
Friday	0.12	0.27	1.13	0.14	0.26	1.15
Saturday	-0.11	0.30	0.90	-0.09	0.29	0.91

	Full Cohort			LS		
Day	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$
Sunday	0.00	0.00	1.00	0.00	0.00	1.00
Monday	0.39	0.21	1.48	0.33	0.38	1.39
Tuesday	0.55	0.24	1.73	0.79	0.37	2.20
Wednesday	-0.08	0.22	0.92	0.02	0.35	1.02
Thursday	0.27	0.25	1.31	0.23	0.40	1.26
Friday	0.14	0.22	1.15	0.16	0.33	1.17
Saturday	-0.02	0.25	0.98	-0.05	0.37	0.95

is not available for this stratum. In general, the proposed cumulative baseline hazard estimates are close to those for full cohort analysis. The exception was stratum 4, for which the SC and WC estimators are considerably above the full cohort estimator. To examine this phenomenon further, we reanalyzed the data several times (results not shown) which of course involves selecting different subcohorts. Based on this exercise, it appears that the disparity between the SC or WC estimator and the full cohort estimator in any stratum (including stratum 4) is due to sampling variation. In fact, when we drew several bootstrap samples and carried out full-cohort analyses of each, the variability in the estimates of the cumulative hazards was quite large. This suggests that the sampling variation we observed in the case-cohort cumulative hazard estimators was largely inherited from that in the full cohort analysis.

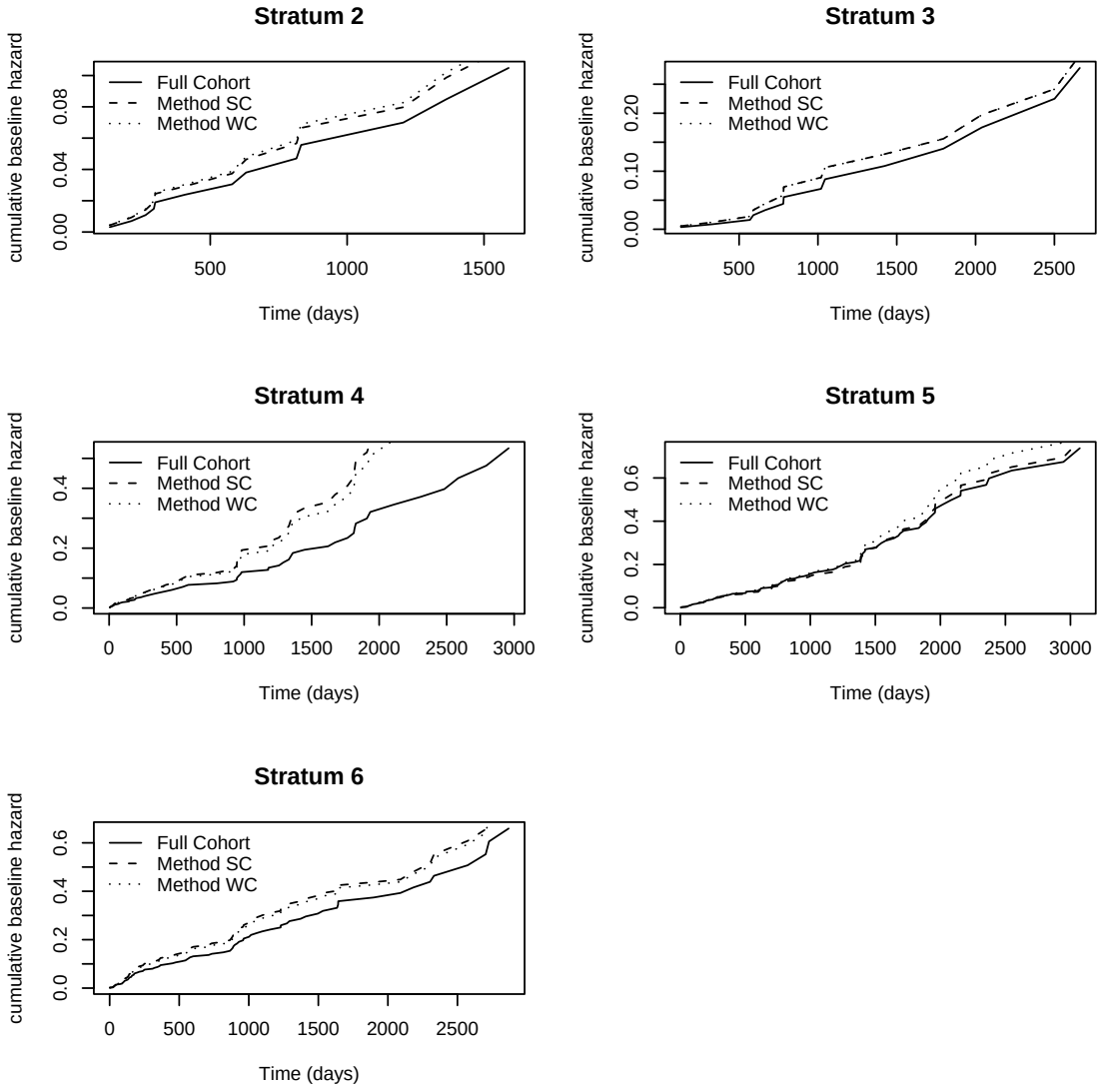


Figure 2.2: Cumulative baseline hazard estimators for the study of CVD mortality among dialysis patients.

Method SC = estimating p_0 using the subcohort, $\widehat{p}_s$, WC = estimating p_0 using whole cohort, $\widehat{p}_w$.

2.6 Discussion

The case-cohort design has been widely studied for univariate failure time data. Lu and Shih (2006) extended the case-cohort design to clustered failure time data. With respect to parameter estimation, compared to Lu and Shih's methods, the methods we propose feature risk sets which use future cases in addition to subcohort

subjects. We demonstrate empirically that the proposed estimators have increased efficiency relative to the methods of Lu and Shih (2006), and that our asymptotic results are applicable to finite samples. The point estimates of our proposed methods are easily computed using standard Cox regression software.

Our simulation results suggest that the proposed methods gain efficiency relative to existing methods (Lu and Shih, 2006) when sampling a smaller number of subjects, or having longer censoring times. This is due to the inclusion of a larger number of failures in the risk sets which are outside the subcohort.

If subcohort sizes are approximately equal, it appears that Design A results in more efficient estimators than Design C, and that Design C has greater efficiency than Design B. This can be attributed to differences in the number of sampled clusters in the subcohort. This trend is stronger when the covariate is cluster-specific. However, the choice between Designs A - C also depends on the cluster size and the availability of data on all clusters.

For each of Designs A - C, we propose three estimation methods which differ based on their treatment of p_0, the marginal probability of the observed event. When β_0 is away from zero, the general superiority of $\widehat{\beta}_t$ over $\widehat{\beta}_w$, and of $\widehat{\beta}_w$ over $\widehat{\beta}_s$, may be explained by the more accurate estimation of the marginal event probability, p_0. Such superiority is more pronounced when β_0 is further from the null (data not shown), as in Chen and Lo (1999). If $\beta_0 = 0$, $\widehat{\beta}_t$ and $\widehat{\beta}_w$ should gain no efficiency over $\widehat{\beta}_s$, since no information about β_0 is provided by p_0. In most real-data applications, the true case percentage p_0 is unknown, and it is not feasible to use $\widehat{\beta}_t$. However, in cases where the study cohort is well-defined, $\widehat{p}_w$ can be computed and used to obtain $\widehat{\beta}_w$, which has the most practical value. In other cases, p_0 can be estimated using the subcohort.

For set-ups with a smaller number of clusters and smaller subcohort size, the proposed methods generally work well, though there is some slight under-coverage for Designs B and C. The asymptotic properties are based on increasing the number of clusters, but Design B samples the smallest number of clusters. Correspondingly, this under-coverage is reduced as the number of clusters increases.

Studies with low event rates often motivate case-cohort sampling. As such, we carried out simulations where the marginal event rate was around $p_0 = 0.03$ (Web Table 4). With a reasonable subcohort sample size, $\widehat{\beta}_s$ appears to work as well as other estimators. In the presence of a very low failure rate, the proposed methods do not gain much efficiency over those of Lu and Shih (2006). This would be expected since, in such settings, the subcohort would tend to contain fewer events; meaning that little efficiency gain would be expected by including future failures in the risk sets. Note that the case-cohort design may still be beneficial for studies with a frequently occurring event. For example, one may need to retrospectively collect additional information from a large database (e.g., disease registry). Case-cohort sampling could then result in substantial cost savings, especially when the collection of detailed covariate information is expensive. The design might also be altered to sample only a fraction of the cases.

The proposed stratified methods appear to perform well with a reasonable number of strata. The baseline cumulative hazard estimator was also examined and performs well.

Point estimates based on simple random samples (SRS) for some non-rare event settings are provided (Web Table 7). It appears that the ESDs of the point estimates based on SRS are very close to those based on Bernoulli sampling. Therefore, one would not gain much efficiency by using SRS, at least for the examples we considered.

Based on our analysis in Section 5, Canadian hemodialysis (HD) patients appear to be at increased risk of cardiovascular disease death on Monday and, in particular Tuesday. Peritoneal dialysis (PD) is an alternative to hemodialysis as a treatment method for kidney failure. A useful follow-up to our analysis would be to study the day-of-week effect on death among PD patients. Unlike HD patients who receive dialysis only 3 days per week, PD patients can get treatments daily at home, at work, or on trips. Therefore, we would expect that the risk of death would be constant from day to day within the week. For HD patients, days on which mortality is increased may depend on schedule (M/W/F or T/T/S), but the dialysis schedule information is not available in the CORR database.

We propose sampling designs which construct the subcohort by independent Bernoulli sampling; in contrast, Lu and Shih (2006) construct the subcohort through sampling without replacement. The subcohort from simple random sampling can only be constructed when accrual into the cohort has ended, while the subcohort from Bernoulli sampling can be formed concurrently. Therefore, case-cohort designs with Bernoulli sampling may be particularly appealing in a prospective study. However, with fixed sample size, case-cohort designs using simple random sampling can improve efficiency, although asymptotic derivations would be more delicate than those in this chapter particularly because of the dependence between sampled clusters induced by Designs B and C.

The proposed methods are based on a marginal proportional hazards model, which does not formulate the within-cluster dependence structure. A proportional hazards frailty model specifies the dependence structure explicitly. Such a model, combined with maximum likelihood estimation, may result in increased efficiency and would be worth investigating.

CHAPTER III

Semiparametric Methods for the Analysis of Failure Time Data with Outcome-Dependent Sampling and Dependent Censoring

3.1 Introduction

Outcome-dependent sampling (ODS) is a cost-saving sampling scheme to enhance study efficiency. In an ODS design, one collects covariate information from a sample by allowing selection probabilities to depend on individuals' outcomes (e.g., death, survival). An ODS design concentrates resources on observations carrying the greatest amount of information. There is a large literature on analyzing data arising from ODS; see for example Breslow and Holubkov (1997a), Zhou et al. (2002), Zhou and You (2007), Schildcrout and Heagerty (2008), Song, Zhou, and Kosorok (2009), and Wang et al. (2009).

The case-control study and case-cohort design are two simple and familiar examples of ODS designs. A number of methods have been proposed for the regression analysis of case-control and case-cohort studies under the proportional hazards model; see Prentice (1986), Breslow and Cain (1988), Self and Prentice (1988), Wacholder et al. (1989), Lin and Ying (1993), Barlow (1994), Breslow and Holubkov (1997b), Chen and Lo (1999), Therneau and Li (1999), Borgan et al. (2000), Langholz and Goldstein (2001), Scheike and Juul (2004) , Scheike and Martinussen (2004), and

Lu and Shih (2006).

Inverse probability of selection weighting (IPSW) is a natural way to generate consistent estimators of population parameters to overcome biased samples, including those generated through ODS designs. In IPSW, each subject is weighted by the inverse of their probability of being sampled. Various authors have proposed IPSW methods for settings in which sampling probability is independent of outcome. For example, for survey data, Binder (1992) proposed an IPSW estimator under Cox's proportional hazards models with weights being treated as fixed; Lin (2000) further studied the case and developed an alternative inference procedure which accounts for the random variation corresponding to the representative population. For two-phase stratified samples, Breslow and Wellner (2007) considered the solution of IPSW likelihood equations with two-phase stratified samples under semiparametric models. For biased samples, Pan and Schaubel (2008) proposed a two-stage weighted method under the proportional hazards model, which estimates the weight using logistic regression at the first stage. For Case-cohort design under Cox's proportional hazards models, Barlow (1994) proposes a pseudolikelihood function with time-dependent weights; Borgan et al. (2000) presented several IPSW estimators for the analysis of exposure stratified case-cohort samples. Kulich and Lin (2000) proposed IPSW estimators for the additive hazards model for case-cohort studies. Nan et al. (2009) presented outcome-dependent weighted estimators for accelerated failure time model in case-cohort studies.

Each of the afore-listed methods for analyzing ODS-based data under the proportional hazards model has an underlying assumption that subjects are censored in a manner independent of the failure rate. However, dependently censored data are commonly encountered in public health studies. For example, wait-listed end-

stage liver disease patients may receive a liver transplant, which censors their wait-list death. In liver transplantation, medical urgency (which is inherently time-dependent) is the criterion by which patients are prioritized for deceased-donor liver transplantation. Therefore, an analysis of baseline factors affecting wait list mortality (i.e., recorded at wait listing and not updated) could result in substantial bias if transplantation were treated as independent censoring. One commonly used method to accommodate dependent censoring is to conduct a weighted analysis, with weights inversely proportional to the probability of remaining uncensored. Such methods have been proposed by Robins and Rotnitzky (1992), Robins (1993b), Robins and Finkelstein (2000), and Scharfstein and Robins (2002b); each of whom showed that Inverse Probability of Censoring Weighting (IPCW) corrects for the dependence between censoring and failure times. Matsuyama and Yamaguchi (2008) applied the IPCW approaches to settings with more than one cause of censoring. Zhang and Schaubel (2011) used IPCW method to estimate group-specific differences in restricted mean lifetime for studies with dependent censoring.

This chapter is motivated by the desire to compare wait-list survival for patients with end-stage liver disease (ESLD). Chronic ESLD patients wait listed for liver transplantation are ordered primarily based on their current (i.e., most recent) Model for End-stage Liver Disease (MELD) score, which is calculated as a log linear combination of bilirubin, creatinine, and international normalized ratio for prothrombin time (Wiesner et al., 2001). As such, the higher a patient's MELD score, the higher their priority to receive a liver transplant. However, higher MELD scores are also associated with an elevated risk of wait-list death, as shown by many previous authors (e.g., Kremers et al., 2004, Huo et al., 2005, Merion et al., 2005, Basto et al., 2008, Subramanian et al., 2010). Thus liver transplantation results in dependent

censoring of wait-list death, due to the correlation between MELD score and both wait-list survival and liver transplant rate. This and other related issues in the liver transplant setting are discussed by Schaubel et al. (2009).

In certain cases, special exceptions are made under which a wait-listed patient may be assigned a MELD score which is higher than that calculated, in an attempt to reflect the patient's actual medical urgency. The most frequent occurrence of such MELD exceptions is for patients with hepatocellular carcinoma (HCC, a form of liver cancer). HCC patients are usually assigned a MELD score of at least 22, which is often considerably higher than the score based on their laboratory measures. To our knowledge, no existing analyses in the liver transplant literature have quantified whether the MELD score of 22 accurately reflects the true wait-list mortality risk faced by HCC patients. As a primary example in this chapter, we carry out such an analysis, with patients classified by their baseline HCC status and MELD scores. Since MELD affects both death and liver transplantation probabilities, liver transplantation is handled as dependent censoring of wait-list death time in this analysis.

In this chapter, we propose methods based on estimating equations for the analysis of failure time data generated by ODS and subject to dependent censoring. We employ a double-inverse-weighting scheme, which combines weights corresponding to the probability of remaining uncensored and the probability of being sampled. A proportional hazards model is assumed for the death process, with the covariate of interest being that observed at baseline (time 0). It is assumed that a longitudinal sequence of measures is observed for each subject, and a proportional hazards model based on such measures is assumed for the dependent censoring process.

The remainder of this chapter is organized as follows. In Section 3.2, we describe

the proposed estimation procedures. In Section 3.3, we derive large sample properties for the proposed estimators. We conduct simulation studies in Section 3.4 to investigate the finite sample properties of the proposed estimators. Section 3.5 provides an application of the methods to wait-list survival data obtained from a national organ failure registry. The chapter concludes with a discussion in Section 3.6.

3.2 Proposed Methods

Let Z_{1i} denote the q_1-vector of time-constant covariates for subject i $(i = 1, \ldots, n)$. Let $Z_{2i}(t)$ be the q_2-vector of time-dependent covariates at time t, $Z_i(t) = \{Z_{1i}^T, Z_{2i}(t)^T\}^T$, and $\widetilde{Z}_i(t) = \{Z_i(u) : 0 \leq u \leq t\}$ denote the history of $Z_i(\cdot)$ up to time t. Let T_i and C_i be the potential failure and censoring times, respectively. We suppose that $C_i = C_{1i} \wedge C_{2i}$, where $a \wedge b = \min\{a, b\}$, C_{1i} is the censoring time due to mechanisms that are independent of T_i given $Z_i(0)$, and C_{2i} denotes the dependent censoring time; that is, C_{2i} is dependent on T_i given $Z_i(0)$. Let $X_i = T_i \wedge C_i$, $Y_i(t) = I(X_i \geq t)$, $\Delta_{1i} = I(T_i \leq C_i)$, $\Delta_{2i} = I(C_{2i} \leq C_{1i}, C_{2i} < T_i)$, $\Delta_{3i} = (1 - \Delta_{1i})(1 - \Delta_{2i})$, $N_i(t) = I(X_i \leq t, \Delta_{1i} = 1)$, and $N_i^C(t) = I(X_i \leq t, \Delta_{2i} = 1)$, where $I(\cdot)$ is the indicator function. The observable data are assumed to be n independently and identically distributed copies of $\{N_i(\cdot), N_i^C(\cdot), Y_i(\cdot), Z_i(\cdot)\}$. Let ξ_i indicate whether or not subject i is sampled. The variate ξ_i is allowed to depend on Δ_{1i}, Δ_{2i} and Δ_{3i} so that the sampling probability can be different for subjects who fail, subjects who are dependent censored and those who are independent censored. Let the cohort be divided into 3 strata according to the outcome $(\Delta_1, \Delta_2, \Delta_3)$ such that $L_k = \{i : \Delta_{ki} = 1\}$, $k = 1, 2, 3$. Let $p_k = \text{pr}(\xi_i = 1 \mid i \in L_k)$, $p = (p_1, p_2, p_3)^T$, and $\rho_i(p) = \sum_{k=1}^3 \Delta_{ki}\xi_i/p_k$. Note that $\rho_i(p)$ weights the ith subject by the inverse

probability that the subject is sampled.

We assume that the hazard of failure of individual i is specified by the following proportional hazards model (Cox, 1972),

$$\lambda_i \{t \mid Z_i(0)\} = \lambda_0(t) \exp\{\beta_0^T Z_i(0)\}, \tag{3.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function for failure time, and β_0 is a $(q_1 + q_2)$-dimensional regression parameter. Note that, we are chiefly interested in inferring the role of $Z_i(0)$ on the death hazard, as opposed to $\{Z_i(t) : t > 0\}$, for reasons of interpretation. For example, it is straightforward to predict survival probability using a pre-specified value of $Z_i(0)$ along with parameter estimates from model (3.1). To do so using a model based on $\widetilde{Z}_i(t)$ would be much more complicated, unless all time-dependent elements are assumed to be external (Kalbfleisch and Prentice, 2002) which is not assumed in the data structure of interest (as previously described).

If it were also the case that C_{2i} was independent of T_i given $Z_i(0)$ (unlike the setting of interest), then β_0 could be consistently estimated by $\widehat{\beta}_{ODS}$, the root of the estimating equation $U_{ODS}(\beta) = 0$, where

$$U_{ODS}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \rho_i(p)\{Z_i(0) - \overline{Z}_{ODS}(\beta, t)\} dN_i(t), \tag{3.2}$$

where $\tau < \infty$ is the maximum follow-up time, $\overline{Z}_{ODS}(\beta, t) = S^{(1)}_{ODS}(\beta, t)/S^{(0)}_{ODS}(\beta, t)$, $S^{(d)}_{ODS}(\beta, t) = \sum_{i=1}^{n} \rho_i(p) Y_i(t) Z_i(0)^{\otimes d} \exp\{\beta^T Z_i(0)\}$, with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$. Estimating equations of the same general structure as (3.2) and arising from IPSW have been proposed by several previous authors; e.g., Kalbfleisch and Lawless (1988), Binder (1992), Borgan et al. (2000) and Lin (2000) for the Cox model; Kulich and Lin (2000) for the additive hazards model; and Nan et al. (2009) for the accelerated failure time model. In IPSW, sampled subjects are weighted by the inverse of their respective probabilities of being selected.

However, since $Z_i(t)$ affects both the event and censoring times, and $Z_i(t)$ is not incorporated into model (3.1), C_{2i} would generally not be independent of T_i given $Z_i(0)$. In this case, the estimate $\widehat{\beta}_{ODS}$ derived from (3.2) could be substantially biased because (3.2) does not accommodate the dependence between C_{2i} and T_i. We assume that conditional on the covariate history $\widetilde{Z}_i(t)$, the hazards of dependent censoring C_{2i} at time t does not further depend on the possibly unobserved failure time T_i; that is,

$$\lambda_i^C\{t \mid \widetilde{Z}_i(t), C_i \geq t, T_i \geq t, T_i\} = \lambda_i^C\{t \mid \widetilde{Z}_i(t), C_i \geq t, T_i \geq t\}. \tag{3.3}$$

This fundamental assumption is called "no unmeasured confounders for censoring" (Rubin, 1977; Robins, 1993). Borrowing terminology from the competing risks literature, assumption (3.3) allows us to identify the cause-specific hazard for C_{2i}. We assume a time-dependent Cox proportional hazards model for the right-hand side of equation (3.3),

$$\lambda_i^C\{t \mid \widetilde{Z}_i(t), X_i \geq t\} = \lambda_0^C(t)\exp\{\alpha_0^T V_i(t)\}, \tag{3.4}$$

where $\lambda_0^C(t)$ is an unspecified baseline hazard function for dependent censoring, $V_i(t)$ is a s-vector consisting of functions of $Z_i(t)$, and α_0 is a s-dimensional regression parameter.

We propose the following estimating function,

$$U(\beta) = \sum_{i=1}^{n} \int_0^\tau R_i(t)\{Z_i(0) - \overline{Z}(\beta, R, t)\}dN_i(t), \tag{3.5}$$

where

$$\begin{aligned} \overline{Z}(\beta, R, t) &= \frac{S^{(1)}(\beta, R, t)}{S^{(0)}(\beta, R, t)} \\ S^{(d)}(\beta, R, t) &= n^{-1}\sum_{i=1}^{n} R_i(t)Y_i(t)Z_i(0)^{\otimes d}\exp\{\beta^T Z_i(0)\} \end{aligned}$$

$$\begin{aligned} R_i(t) &= \rho_i(p) W_i(t) \\ W_i(t) &= e^{\Lambda_i^C(t)} \kappa(t), \end{aligned}$$

where $\Lambda_i^C(t) = \int_0^t \exp\{\alpha^T V_i(u)\} d\Lambda_0^C(u)$ and the function $\kappa(t)$ in the weight $W_i(t)$ is a stabilization factor. We consider three choices of $\kappa(t)$. One choice is $\kappa_1(t) = 1$. However, when the censoring is heavy, $e^{\Lambda_i^C(t)}$ could be quite large and lead to instability in the estimation. In this case, the choice of $\kappa_2(t) = \exp\left[-\int_0^t \exp\{\alpha^T V_i(0)\} \; d\Lambda_0^C(u)\right]$ or $\kappa_3(t) = \exp[-\Lambda_i^\dagger\{t \mid Z_i(0)\}]$ may be more appropriate, where $\Lambda_i^\dagger(t)$ is based on a time-to-censoring model that uses only the baseline covariate values, $Z_i(0)$. Hereafter, we denote $W_{ji}(t) = e^{\Lambda_i^C(t)} \kappa_j(t)$, $j = 1, 2, 3$, and correspondingly estimate β_0 with $\widehat{\beta}_{W_1}$, $\widehat{\beta}_{W_2}$ and $\widehat{\beta}_{W_3}$, the solutions to $U(\beta) = 0$ with weights $W_{1i}(t)$, $W_{2i}(t)$ and $W_{3i}(t)$, respectively.

The weight $W_{1i}(t)$ can be estimated using $\exp\{\widehat{\Lambda}_i^C(t)\}$, where

$$\begin{aligned} \widehat{\Lambda}_i^C(t) &= \int_0^t \exp\{\widehat{\alpha}^T V_i(s)\} d\widehat{\Lambda}_0^C(s, \widehat{\alpha}) \\ \widehat{\Lambda}_0^C(t, \alpha) &= \sum_{i=1}^n \int_0^t \left[\sum_{j=1}^n \rho_j(p) Y_j(s) \exp\{\alpha^T V_j(s)\}\right]^{-1} \rho_i(p) dN_i^C(s), \end{aligned}$$

where $\widehat{\alpha}$ is the partial likelihood estimate of α_0 and is computed under assumption (3.4) as the root of $U^C(\alpha) = 0$; where

$$U^C(\alpha) = \sum_{i=1}^n \int_0^\tau \{V_i(t) - \overline{V}(\alpha, p, t)\} \rho_i(p) dN_i^C(t),$$

is an IPSW-based estimating function, with $\overline{V}(\alpha, p, t) = S_C^{(1)}(\alpha, p, t) / S_C^{(0)}(\alpha, p, t)$ and $S_C^{(d)}(\alpha, p, t) = n^{-1} \sum_{i=1}^n \rho_i(p) Y_i(t) V_i(t)^{\otimes d} e^{\alpha^T V_i(t)}$.

The weight $W_{2i}(t)$ can be estimated using $\widehat{\kappa}_{2i}(t) \exp\{\widehat{\Lambda}_i^C(t)\}$, where $\widehat{\kappa}_{2i}(t) = \exp[-\widehat{\Lambda}_i^C\{t, \widehat{\alpha} \mid Z_i(0)\}]$; i.e., $\kappa_{2i}(t)$ is estimated using the same model (3.4), but only using baseline covariate values,

$$\widehat{\Lambda}_i^C\{t, \widehat{\alpha} \mid Z_i(0)\} = \int_0^t \exp\{\widehat{\alpha}^T V_i(0)\} d\widehat{\Lambda}_0^C(s, \widehat{\alpha}).$$

The weight $W_{3i}(t)$ can be estimated by $\widehat{\kappa}_{3i}(t)\exp\{\widehat{\Lambda}_i^C(t)\}$, where $\widehat{\kappa}_{3i}(t) = \exp\{-\widehat{\Lambda}_i^\dagger(t)\}$, with $\kappa_{3i}(t)$ estimated using an additional baseline model for C_{2i},

$$\lambda_i^\dagger\{t \mid Z_i(0), C_i \geq t, T_i, T_i \geq t\} = \lambda_0^\dagger(t)\exp\{{\alpha^\dagger}^T V_i(0)\},$$

such that we have

$$\begin{aligned}
\widehat{\Lambda}_i^\dagger(t) &= \int_0^t \exp\{\widehat{\alpha}^{\dagger T} V_i(0)\} d\widehat{\Lambda}_0^\dagger(s, \widehat{\alpha}^\dagger), \\
\widehat{\Lambda}_0^\dagger(t, \alpha^\dagger) &= \sum_{i=1}^n \int_0^t \left[\sum_{j=1}^n \rho_j(p) Y_j(s)\exp\{{\alpha^\dagger}^T V_j(0)\}\right]^{-1} \rho_i(p) dN_i^C(s),
\end{aligned}$$

and $\widehat{\alpha}^\dagger$ is the partial likelihood estimate of $\alpha^\dagger$ under the model for dependent censoring with hazard $\lambda_i^\dagger(t)$. Weight stabilizers analogous to $\kappa_{3i}(t)$ have been suggested, for example, by Robins and Finkelstein (2000) and Hernán, Brumback, and Robins (2000). We propose the stabilizer $\kappa_{2i}(t)$ as an alternative. The performance of each of $W_{1i}(t)$, $W_{2i}(t)$ and $W_{3i}(t)$ are compared through simulations studies described in Section 3.4.

3.3 Asymptotic Properties of the Proposed Estimators

The following conditions are assumed throughout this section.

(*a*) $\{N_i(\cdot), N_i^C(\cdot), Y_i(\cdot), Z_i(\cdot)\}, i = 1, \ldots, n$ are independently and identically distributed.

(*b*) $P\{Y_i(\tau) = 1\} > 0$ for $i = 1, \ldots, n$.

(*c*) $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| < B_Z < \infty$ for $i = 1, \ldots, n$, where Z_{ij} is the jth component of Z_i and B_Z is a constant.

(*d*) There exists a neighborhood $\mathcal{B}$ of β_0 such that $\sup_{u\in[0,\tau],\beta\in\mathcal{B}} \|S^{(d)}(\beta, R, u) - s^{(d)}(\beta, R, u)\| \longrightarrow 0$ in probability for $d = 0, 1, 2$, where $s^{(d)}(\beta, R, u) =$

$\mathcal{E}\{S^{(d)}(\beta,R,u)\}$ is absolutely continuous, for $\beta \in \mathcal{B}$, uniformly in $u \in (0,\tau]$, $\mathcal{E}(\cdot)$ denotes expectation. Moreover, $s^{(0)}(\beta,R,u)$ is assumed to be bounded away from zero.

(e) There exists a neighborhood $\mathcal{B}_\mathcal{C}$ of α_0 such that $\sup_{u\in[0,\tau],\alpha\in\mathcal{B}_\mathcal{C}} \|S_C^{(d)}(\alpha,p,u) - s_C^{(d)}(\alpha,u)\| \longrightarrow 0$ in probability for $d=0,1,2$, where for $\alpha \in \mathcal{B}_\mathcal{C}$, $s_C^{(d)}(\alpha,u) = \mathcal{E}\{S_C^{(d)}(\alpha,p,u)\}$ is absolutely continuous, uniformly in $u \in (0,\tau]$. Moreover, $s_C^{(0)}(\alpha,u)$ is assumed to be bounded away from zero.

(f) The matrices $A(\beta_0)$ and $A^C(\alpha_0)$ are positive definite, where

$$\begin{aligned} A(\beta) &= \int_0^\tau \left\{ s^{(2)}(\beta,R,u)/s^{(0)}(\beta,R,u) - \overline{z}(\beta,R,u)^{\otimes 2} \right\} dF(u) \\ A^C(\alpha) &= \int_0^\tau \left\{ s_C^{(2)}(\alpha,u)/s_C^{(0)}(\alpha,u) - \overline{v}(\alpha,u)^{\otimes 2} \right\} dF^C(u) \end{aligned}$$

with $\overline{z}(\beta,R,u) = s^{(1)}(\beta,R,u)/s^{(0)}(\beta,R,u)$, $\overline{v}(\alpha,u) = s_C^{(1)}(\alpha,u)/s_C^{(0)}(\alpha,u)$, $F(u) = \mathcal{E}\{R_i(u)N_i(u)\}$, $F^C(u) = \mathcal{E}\{\rho_i(p_0)N_i^C(u)\}$.

(g) $\Lambda_0(\tau) < \infty$, $\Lambda_0^C(\tau) < \infty$.

We describe the asymptotic properties of the proposed estimators in the following theorems.

Theorem III.1. *Under conditions $(a)-(g)$, as $n\to\infty$, $n^{1/2}(\widehat{\alpha}-\alpha_0)$ converges to a mean zero Normal distribution with covariance $A^C(\alpha_0)^{-1}\Omega(\alpha_0)A^C(\alpha_0)^{-1}$, where*

$$\begin{aligned} \Omega(\alpha) &= \mathcal{E}\left\{\psi_i(\alpha,p)^{\otimes 2}\right\} \\ \psi_i(\alpha,p) &= K_i(\alpha,p) + B^C(\alpha,p)Q_i(p) \\ K_i(\alpha,p) &= \int_0^\tau \left\{V_i(t) - \overline{v}(\alpha,t)\right\}\rho_i(p)dM_i^C(t) \\ B^C(\alpha,p) &= \int_0^\tau \left\{ \frac{s_c^{(1)}(\alpha,p,t)}{s_c^{(0)}(\alpha,p,t)^2} r^{(0)}(\alpha,p,t) - \frac{1}{s_c^{(0)}(\alpha,p,t)} r^{(1)}(\alpha,p,t) \right\} \\ &\quad \times dF^C(t) + d(\alpha,p) \end{aligned}$$

$$
\begin{aligned}
r_k^{(d)}(\alpha,p,t) &= -\frac{1}{p_k}\mathcal{E}\left\{\Delta_{k1}Y_1(t)V_1(t)^{\otimes d}e^{\alpha^T V_1(t)}\right\}, \quad d=0,1 \\
r^{(d)}(\alpha,p,t) &= \left(r_1^{(d)}(\alpha,p,t) \quad r_2^{(d)}(\alpha,p,t) \quad r_3^{(d)}(\alpha,p,t)\right), \quad d=0,1,
\end{aligned}
$$

where we further define

$$
\begin{aligned}
d_k(\alpha,p) &= -\frac{1}{p_k}\mathcal{E}\left[\int_0^\tau \left\{V_1(t)-\overline{v}(\alpha,t)\right\}\Delta_{k1}dN_1^C(t)\right] \\
d(\alpha,p) &= (d_1(\alpha,p) \quad d_2(\alpha,p) \quad d_3(\alpha,p)) \\
Q_{ki}(p) &= \eta_k^{-1}\Delta_{ki}(\xi_i - p_k) \\
\eta_k &= pr(\Delta_k=1), k=1,2,3 \\
Q_i(p) &= (Q_{1i}(p) \quad Q_{2i}(p) \quad Q_{3i}(p))^T,
\end{aligned}
$$

with $dM_i^C(t) = dN_i^C(t) - Y_i(t)d\Lambda_i^C(t)$.

In Web Appendix B.2, we show that $n^{1/2}\left(\widehat{\alpha}-\alpha_0\right) = n^{-1/2}\sum_{i=1}^n \psi_i(\alpha_0,p_0) \times A^C(\alpha_0)^{-1}$ $+o_p(1)$; hence $n^{1/2}\left(\widehat{\alpha}-\alpha_0\right)$ is essentially a scaled sum of n independent and identically distributed random quantities with mean zero and finite variance. By the Multivariate Central Limit Theorem (MCLT) and empirical process theory, one proves the asymptotic normality.

Theorem III.2. *Under conditions* $(a)-(g)$*, as* $n\to\infty$*,* $n^{1/2}\left(\widehat{\beta}_{W_1}-\beta_0\right)$*, converges to a mean zero Normal distribution with covariance* $A(\beta_0)^{-1}\Sigma(\beta_0,R)A(\beta_0)^{-1}$*, where*

$$
\begin{aligned}
A(\beta) &= \int_0^\tau \left\{\frac{s^{(2)}(\beta,R,t)}{s^{(0)}(\beta,R,t)} - \overline{z}(\beta,R,t)^{\otimes 2}\right\}dF(t) \\
\Sigma(\beta,R) &= \mathcal{E}\left\{\Theta_i(\beta,R)^{\otimes 2}\right\} \\
\Theta_i(\beta,R) &= O(\beta,R)Q_i(p_0) \\
&\quad + H(\beta,R)A^C(\alpha_0)^{-1}\psi_i(\alpha_0,p_0) \\
&\quad + \int_0^\tau \chi(u,\tau)d\Phi_i(\alpha_0,p_0,u) \\
O(\beta,R) &= \mathcal{E}\left[\int_0^\tau \left\{Z_i(0)-\overline{z}(\beta,R,t)\right\}\mu_i(p_0)W_{1i}(t)dM_i(t)\right]
\end{aligned}
$$

$$
\begin{aligned}
\mu_{ki}(p) &= \frac{d\rho_i(p)}{dp_k} = -\frac{\Delta_{ki}\xi_i}{p_k^2} \\
\mu_i(p) &= (\mu_{1i}(p) \quad \mu_{2i}(p) \quad \mu_{3i}(p))^T \\
H(\beta, R) &= \mathcal{E}\left[\int_0^\tau \{Z_i(0) - \overline{z}(\beta, R, t)\}\, \Psi_i^T R_i(t) dM_i(t)\right] \\
\Psi_i(t) &= \int_0^t V_i(u) d\Lambda_i^C(u) \\
\chi(t_1, t_2) &= \mathcal{E}\left[e^{\alpha_0^T V_i(t_1)} \int_{t_1}^{t_2} \{Z_i(0) - \overline{z}(\beta, R, t)\}\, R_i(t) dM_i(t)\right]
\end{aligned}
$$

$$
\begin{aligned}
d\Phi_i(\alpha, p, u) &= s_C^{(0)}(\alpha, u)^{-1} \left\{dJ(u) - r^{(0)}(\alpha, p, u) d\Lambda_0^C(u)\right\} Q_i(p) \\
&\quad -\overline{v}^T(\alpha, u) d\Lambda_0^C(u) A^C(\alpha)^{-1} \psi_i(\alpha, p) \\
&\quad +s_C^{(0)}(\alpha, u)^{-1} \rho_i(p) dM_i^C(u) \\
dJ(u) &= \mathcal{E}\left\{\mu_i(p_0)^T dN_i^C(u)\right\},
\end{aligned}
$$

with $dM_i(t) = dN_i(t) - Y_i(t) d\Lambda_i(t)$.

The proof begins by decomposing $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$ into $n^{1/2}\{\widehat{\Lambda}_0^C(t; \widehat{\alpha}, \widehat{p}) - \widehat{\Lambda}_0^C(t; \widehat{\alpha}, p_0)\} + n^{1/2}\{\widehat{\Lambda}_0^C(t; \widehat{\alpha}, p_0) - \widehat{\Lambda}_0^C(t; \alpha_0, p_0)\} + n^{1/2}\{\widehat{\Lambda}_0^C(t; \alpha_0, p_0) - \Lambda_0^C(t)\}$. Then $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$ can be expressed asymptotically as a sum of independent and identically distributed zero-mean variates, as $n \to \infty$. Combining this result and the Functional Delta Method, we can show that $n^{1/2}\{\widehat{R}_i(t) - R_i(t)\}$ can be written asymptotically as a sum of independent and identically distributed zero-mean variates, as $n \to \infty$. Finally, through Functional Delta Method, the asymptotic normality of $n^{1/2}(\widehat{\beta}_{W_1} - \beta_0)$ is demonstrated.

The expression for the asymptotic covariance of $\widehat{\beta}_{W_1}$ is very complicated and difficult to implement numerically. A practical way to estimate the variance of the proposed estimators is to treat the weights $R_i(t)$ as known rather than estimated. Based on results derived in the Web Appendix B that, in the setting where the weight

function is known,

$$n^{1/2}(\widehat{\beta}-\beta_0) = A(\beta_0)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n} U_i^{\ddagger}\{\beta_0, R\} + o_p(1) \tag{3.6}$$

with $U_i^{\ddagger}(\beta_0, R) = \int_0^{\tau}\{Z_i(0) - \overline{z}(\beta, R, t)\}R_i(t)dM_i(t)$; hence, $n^{1/2}(\widehat{\beta}-\beta_0)$ is asymptotically a scaled sum of independent and identically distributed zero-mean random quantities with finite variance. Therefore, the variance of $\widehat{\beta}_{W_1}$ is estimated by $\widehat{A}(\widehat{\beta})^{-1}\widehat{\Sigma}^{\ddagger}(\widehat{\beta}, \widehat{R})\widehat{A}(\widehat{\beta})^{-1}$, where $\Sigma^{\ddagger}(\beta, R) = \mathcal{E}\{U_i^{\ddagger}(\beta, R)^{\otimes 2}\}$, $\widehat{A}(\widehat{\beta})$ and $\widehat{\Sigma}^{\ddagger}(\widehat{\beta}, \widehat{R})$ are calculated by replacing limiting values with their corresponding empirical counterparts.

By similar arguments, the asymptotic normality holds for $n^{1/2}(\widehat{\beta}_{W_2} - \beta_0)$ and $n^{1/2}(\widehat{\beta}_{W_3} - \beta_0)$. However, the covariance will be even more complicated than that of $n^{1/2}(\widehat{\beta}_{W_1} - \beta_0)$. Therefore, similarly, we can treat $R_i(t)$ as fixed to calculate the variance of $\widehat{\beta}_{W_2}$ and $\widehat{\beta}_{W_3}$. Note that (3.6) holds when using W_2 or W_3, such that each of the variance of $\widehat{\beta}_{W_2}$ and $\widehat{\beta}_{W_3}$ is estimated by $\widehat{A}(\widehat{\beta})^{-1}\widehat{\Sigma}^{\ddagger}(\widehat{\beta}, \widehat{R})\widehat{A}(\widehat{\beta})^{-1}$, with $R_i(t)$ being replaced by $\rho_i(\widehat{p})\widehat{W}_{2i}(t)$ and $\rho_i(\widehat{p})\widehat{W}_{3i}(t)$, respectively.

3.4 Numerical Studies

We investigated the finite sample properties of the estimators proposed in Section 2 through a series of simulation studies. We generated failure time data from $n = 2500$ subjects. A treatment group indicator Z_{1i} and baseline time-dependent covariate $Z_{2i}(0)$ were generated as independent Bernoulli variables each with probabilities 0.5. The independent censoring times C_{1i} were constant and equal to 100. After generating a $U(0, 1)$ variable, U_T, the event time T was generated from a Cox model with hazard function

$$\lambda_i(t) = \lambda_0(t)\exp\{\beta_1 Z_{1i} + \beta_2 Z_{2i}(0)\}, \tag{3.7}$$

by solving the equation $\int_0^T \lambda_0(u) \exp\{\beta_1 Z_{1i} + \beta_2 Z_{2i}(0)\}\, du = -\log U_T$ for T, so that U_T corresponds to the survival function at T. The baseline hazard function for event time is given by $\lambda_0(t) = 0.1, 0 \leq t < \infty$ and (β_{10}, β_{20}) is set to $\{\log(1.5), \log(1.5)\} = (0.4055, 0.4055)$. The time-dependent covariate $Z_{2i}(t)$ was generated as

$$\begin{aligned} & Z_{2i}(0) I\left(U_T \leq 0.3\right) + \left\{Z_{2i}(0) + U_T \times \text{int}(t)\right\} I\left(0.3 < U_T \leq 0.6\right) \\ & \qquad + \left\{Z_{2i}(0) + U_T/2 \times \text{int}(t)\right\} I\left(U_T > 0.6\right), \end{aligned} \tag{3.8}$$

where $\text{int}(t)$ is the integer part of t. The dependent censoring time C_{2i} was generated from a Cox model with hazard function

$$\begin{aligned} \lambda_i^C(t) \;\; = \;\; & \lambda_0^C(t) \exp\left[\alpha_1 Z_{1i} + \alpha_2 Z_{2i}(t) I\left\{Z_{2i}(0) = 1\right\} + \right. \\ & \left. \alpha_3 Z_{2i}(t) I\left\{Z_{2i}(0) = 0\right\}\right], \end{aligned} \tag{3.9}$$

where $\lambda_0^C(t) = 0.1, 0 \leq t < \infty$ and $(\alpha_1, \alpha_2, \alpha_3) = \{-0.5, \log(2), \log(1.1)\} = (-0.5, 0.6931, 0.0953)$. In this data configuration, the time-dependent covariate $Z_{2i}(t)$ is correlated with the event time T_i through equation (3.8). In addition, $Z_{2i}(t)$ also affects the censoring time C_{2i} via model (3.9). Since only the baseline value of $Z_{2i}(t)$, i.e. $Z_{2i}(0)$, is adjusted for in the model (3.7), the censoring time C_{2i} is dependent on T_i given Z_{1i} and $Z_{2i}(0)$. In another data configuration, the baseline hazard function for event time is given by $\lambda_0 = 0.2$, with (β_{10}, β_{20}) set to $\{\log(1.5), \log(1.5)\} = (0.4055, 0.4055)$ or $\{0, 0\}$, which leads to a lower dependent censoring rate. All the other settings were the same as those used in the previous configuration.

Additional scenarios have also been evaluated in order to examine the performance of the proposed methods with different dependent censoring models. Specifically, the dependent censoring time C_{2i} was generated from a Cox model with hazard function

$$\lambda_i^C(t) = \lambda_0^C(t) \exp\left\{\alpha_1 Z_{1i} + \alpha_2 Z_{2i}(t)\right\}, \tag{3.10}$$

Table 3.1: Simulation results based on 1000 replications: $\lambda_i^C(t)$ is given by (3.9) and $\lambda_0 = 0.1$.

Weight	Estimator	Bias	ESD	ASE	CP
–	$\widehat{\beta}_1^{ODS}$	0.055	0.173	0.169	0.932
W_1	$\widehat{\beta}_1^{W_1}$	0.040	0.205	0.189	0.927
W_2	$\widehat{\beta}_1^{W_2}$	0.018	0.175	0.171	0.941
W_3	$\widehat{\beta}_1^{W_3}$	0.016	0.173	0.170	0.944
–	$\widehat{\beta}_2^{ODS}$	-0.085	0.172	0.169	0.918
W_1	$\widehat{\beta}_2^{W_1}$	-0.024	0.202	0.190	0.934
W_2	$\widehat{\beta}_2^{W_2}$	-0.007	0.172	0.173	0.958
W_3	$\widehat{\beta}_2^{W_3}$	-0.006	0.172	0.173	0.960

$n = 2500$, $\beta_1 = \beta_2 = \log(1.5)$, Spearman correlation between T and C_2 is 0.14. Approximately 47% of subjects are dependently censored. There were $\approx$ 300 individuals in the subcohort.

in which the censoring model depends on $Z_{2i}(t)$ in the same way for both baseline $Z_2(0)$ groups. Other settings were the same as those used in the first data configuration except that $\lambda_0 = 0.2$. Each data configuration was replicated 1000 times.

Tables 3.1, 3.2 and 3.3 display the results of our proposed estimators and those ignoring the dependent censoring. As expected, the estimators ignoring the dependent censoring are biased in some settings. Each of the proposed estimators is approximately unbiased, and the average asymptotic standard errors (ASEs) are generally close to to the empirical standard deviations (ESDs). Correspondingly, the 95% empirical coverage probabilities (CPs) are generally close to the nominal value. In addition, simulation results suggest that, at least in the examples we evaluated, the stabilized estimators using weights $W_{2i}(t)$ and $W_{3i}(t)$ are more efficient than the unstabilized estimator using weight $W_{1i}(t)$. In general, the performance of the estimators using weights $W_{2i}(t)$ and $W_{3i}(t)$ are comparable. In some data configurations, one may appear to be more efficient than the other. However, differences were usually small and appeared to vary by data set-up.

Table 3.2: Simulation results based on 1000 replications: $\lambda_i^C(t)$ is given by (3.9) and $\lambda_0 = 0.2$.

Weight	Estimator	Bias	ESD	ASE	CP
			$\beta_1 = \beta_2 = 0.4055$		
–	$\widehat{\beta}_1^{ODS}$	0.017	0.146	0.146	0.955
W_1	$\widehat{\beta}_1^{W_1}$	0.004	0.152	0.148	0.944
W_2	$\widehat{\beta}_1^{W_2}$	0.002	0.145	0.145	0.952
W_3	$\widehat{\beta}_1^{W_3}$	0.002	0.145	0.145	0.952
–	$\widehat{\beta}_2^{ODS}$	-0.012	0.152	0.148	0.944
W_1	$\widehat{\beta}_2^{W_1}$	0.009	0.158	0.152	0.933
W_2	$\widehat{\beta}_2^{W_2}$	0.009	0.152	0.148	0.944
W_3	$\widehat{\beta}_2^{W_3}$	0.009	0.152	0.148	0.944
			$\beta_1 = \beta_2 = 0$		
–	$\widehat{\beta}_1^{ODS}$	0.014	0.160	0.156	0.949
W_1	$\widehat{\beta}_1^{W_1}$	0.005	0.180	0.167	0.923
W_2	$\widehat{\beta}_1^{W_2}$	0.002	0.160	0.156	0.946
W_3	$\widehat{\beta}_1^{W_3}$	0.002	0.159	0.156	0.947
–	$\widehat{\beta}_2^{ODS}$	-0.080	0.161	0.157	0.914
W_1	$\widehat{\beta}_2^{W_1}$	0.000	0.172	0.168	0.950
W_2	$\widehat{\beta}_2^{W_2}$	0.002	0.158	0.158	0.951
W_3	$\widehat{\beta}_2^{W_3}$	0.001	0.159	0.158	0.953

$n = 2500$, $\beta_1 = \beta_2 = \log(1.5)$ or $\beta_1 = \beta_2 = 0$, which corresponds to a Spearman correlation between T and C_2 of 0.14 and 0.11. Approximately 29% and 39% of subjects were dependently censored, respectively. Numbers of individuals in the subcohort were ≈ 280 and ≈ 290.

Table 3.3: Simulation results based on 1000 replications: $\lambda_i^C(t)$ is given by (3.10) and $\lambda_0 = 0.2$.

Weight	Estimator	Bias	ESD	ASE	CP
			$\beta_1 = \beta_2 = 0.4055$		
–	$\widehat{\beta}_1^{ODS}$	0.038	0.147	0.147	0.951
W_1	$\widehat{\beta}_1^{W_1}$	0.005	0.153	0.150	0.945
W_2	$\widehat{\beta}_1^{W_2}$	0.002	0.146	0.146	0.949
W_3	$\widehat{\beta}_1^{W_3}$	0.002	0.146	0.147	0.948
–	$\widehat{\beta}_2^{ODS}$	0.015	0.154	0.149	0.940
W_1	$\widehat{\beta}_2^{W_1}$	0.012	0.162	0.153	0.933
W_2	$\widehat{\beta}_2^{W_2}$	0.011	0.154	0.148	0.940
W_3	$\widehat{\beta}_2^{W_3}$	0.011	0.154	0.149	0.942
			$\beta_1 = \beta_2 = 0$		
–	$\widehat{\beta}_1^{ODS}$	0.027	0.163	0.158	0.942
W_1	$\widehat{\beta}_1^{W_1}$	0.004	0.181	0.169	0.929
W_2	$\widehat{\beta}_1^{W_2}$	0.001	0.162	0.158	0.952
W_3	$\widehat{\beta}_1^{W_3}$	0.002	0.161	0.158	0.949
–	$\widehat{\beta}_2^{ODS}$	-0.034	0.165	0.159	0.938
W_1	$\widehat{\beta}_2^{W_1}$	-0.003	0.178	0.170	0.936
W_2	$\widehat{\beta}_2^{W_2}$	0.004	0.163	0.159	0.947
W_3	$\widehat{\beta}_2^{W_3}$	0.003	0.164	0.159	0.949

$n = 2500$, $\beta_1 = \beta_2 = \log(1.5)$ or $\beta_1 = \beta_2 = 0$, which corresponds to a Spearman correlation between T and C_2 of 0.20 and 0.20. Respectively, 30% and 41% of subjects are dependently censored. Numbers of individuals in the subcohort were ≈ 280 and ≈ 290.

3.5 Application

We applied the proposed methods to analyze wait-list mortality for patients with end-stage liver disease (ESLD). Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). The $n = 55,943$ patients who initially wait-listed for liver transplantation in the United States at age ≥ 18 between March 1, 2002 and December 31, 2008 were included in the analysis. Patients were followed from the date of initial wait-listing until the earliest of death, receiving liver transplantation, loss to follow-up, or last day of the observation period (December 31, 2008).

The Model of End-stage Liver Disease (MELD) score is time-dependent and is updated based on a frequency that ranges from weekly to yearly and that may depend on the last reported MELD. In the current liver allocation system, patients are ordered on the wait-list primarily by descending MELD. That is, patients with higher MELD are considered to be at greater medical urgency and, therefore, get higher priority for transplantation. However, for hepatocellular carcinoma (HCC) patients, the calculated MELD based on laboratory measures has generally been considered by the field to understate actual medical urgency. As such, a MELD score of 22 is usually assigned to an HCC patient if the laboratory MELD is less than 22. The primary objective of our analysis is to determine which range of (calculated) MELD score is actually consistent with the HCC wait-list mortality hazard.

In many studies, it has been shown that MELD is the dominant risk factor for liver wait-list mortality. Moreover, as stated in the previous paragraph, MELD also strongly affects the liver transplant hazard. Therefore, unless the death model adjusts for time-dependent MELD, the wait-list mortality and decease-donor liver transplantation will be correlated. However, HCC is a diagnosis category; an under-

lying cause of end-stage liver disease, which is usually recorded at time 0. Therefore, given our analytic objective and in the interests of interpretation, it is appropriate to adjust for other characteristics known at $t = 0$, but not factors realized at $t > 0$. Therefore, we must account for liver transplantation as dependent censoring.

There are additional issues regarding the data structure which must be taken into account. In particular, a patient who is too sick to receive a transplant can be inactivated (usually a temporary measure) or removed (permanent) from the wait-list. During these intervals, the patient is ineligible to receive a transplant. Therefore, an appropriate Cox model in this setting is given by the following,

$$\lambda_i^C(t) = A_i(t)\lambda_0^C(t)\exp\{\alpha^T V_i(t)\}, \tag{3.11}$$

where $A_i(t)$ is an indicator of being active on the wait list (i.e., as opposed to being inactive or having been previously removed) as of time t. When fitting model (3.11), we delete patient subintervals with $A_i(t) = 0$. The time-dependent covariate vector $V_i(t)$ includes MELD at time t (grouped into intervals: [6,8], [9,11], [12,13], [14,15], [16,17], [18,19], [20,22], [23,24], [25,29], [30,39], and 40) with HCC patients chosen as the reference group. The vector $V_i(t)$ also includes the following baseline covariates: age, gender, race and blood type; with age less than 40, Female, Caucasian and blood type O as references, respectively. Note that, for the intervals where the patient was either inactivated or removed, the transplant hazard was treated as 0, as indicated by equation (3.11). However, since the inactivated or removed patients are still at risk of pre-transplant death, such that patient subintervals with $A_i(t) = 0$ are included in the wait-list mortality model. In addition, for both the time-to-death and time-to-transplant models, we adjusted for Organ Procurement Organization (OPO) through stratification, since it may not be appropriate to assume proportionality with respect to the approximately 60 OPO-specific hazard functions, transplant or death.

Table 3.4: ODS design for the analysis of liver wait-list mortality.

Patients	OPO size	p_1	p_2	p_3
HCC	all	1.00	1.00	1.00
Non-HCC				
	≤ 400	1.00	1.00	1.00
	(401, 1300]	0.30	0.10	0.10
	> 1300	0.15	0.10	0.10

The primary outcome of interest is wait-list mortality. Loss to follow up, living-donor transplantation and administrative censoring are considered to be independent censoring. Dependent censoring occurred through deceased-donor liver transplantation. Among the $n = 55,951$ patients wait listed for liver transplantation, a total of 4,475 (8%) were diagnosed with HCC. In term of events, 10,584 (19%) patients died on the wait-list, while 28,621 (51%) received a deceased-donor liver transplant.

To illustrate our methods, we selected an ODS and allowed the sampling rate to depend on the baseline HCC status and OPO size. The sampling rate is shown in Table 3.4. Note that patients who are diagnosed with HCC are selected into the subcohort with probability 1.

Results of the analysis are shown in Table 3.5. Since the IPCW weights could be very large toward the tail of the observation time, we truncated IPCW weights with 10. In general, we use the results based on W_2 or W_3 since they are more stable than those based on W_1. Table 3.5 shows that when the dependent censoring is ignored, MELD group $[16, 17]$ is consistent with the HCC group. By using W_2, MELD group $[14, 15]$ is consistent with the HCC group.

The result based on using W_3 was similar to that based on W_2, except that both MELD groups $[14, 15]$ and $[16, 17]$ are consistent with the HCC group (result not shown). By using W_1, MELD groups $[14, 15]$ and $[16, 17]$ are consistent with the HCC group (result not shown). However, since the unstabilized weight W_1 may be

Table 3.5: Analysis of wait-list mortality by MELD group: HCC group (assigned MELD of 22) is chosen to be the reference.

	Unweighted				Weighted: W_2			
	$\widehat{\beta}$	SE	p-value	$\exp\{\widehat{\beta}\}$	$\widehat{\beta}$	SE	p-value	$\exp\{\widehat{\beta}\}$
HCC	0	–	–	1	0	–	–	1
MELD								
[6, 8]	-1.07	0.12	< 0.0001	0.34	-1.02	0.15	< 0.0001	0.36
[9, 11]	-0.60	0.09	< 0.0001	0.55	-0.48	0.11	< 0.0001	0.62
[12, 13]	-0.53	0.09	< 0.0001	0.59	-0.40	0.11	0.0002	0.67
[14, 15]	-0.25	0.09	0.005	0.78	0.002	0.11	0.98	1.00
[16, 17]	0.10	0.10	0.32	1.10	0.24	0.12	0.0498	1.28
[18, 19]	0.26	0.12	0.02	1.30	0.63	0.15	< 0.0001	1.88
[20, 22]	0.55	0.11	< 0.0001	1.73	0.83	0.12	< 0.0001	2.29
[23, 24]	0.90	0.17	< 0.0001	2.47	1.41	0.21	< 0.0001	4.12
[25, 29]	1.59	0.16	< 0.0001	4.88	1.93	0.19	< 0.0001	6.92
[30, 39]	2.38	0.16	< 0.0001	10.81	2.69	0.17	< 0.0001	14.74
40	3.68	0.29	< 0.0001	39.56	3.65	0.37	< 0.0001	38.67

quite large toward the tail of the observation time, the result from using stabilized weights W_2 or W_3 would be of more interest than those from using W_1. According to the results in Table 3.4, an assigned MELD score between 14 and 15 for HCC patients would be consistent with the wait-list mortality rates for such patients, and therefore may be more appropriate than the MELD exception score of 22 that is currently being used.

3.6 Discussion

In this chapter, we propose methods for analyzing failure time data under an ODS design with dependent censoring. The proposed methods employ a double-inverse-weighting scheme, through which the proposed estimators adjust for the sampling bias and overcome dependent censoring. Simulation studies show that the proposed estimators are approximately unbiased and that our asymptotic results are applicable to finite samples. The proposed estimates can be computed using standard software (e.g., PROC PHREG in SAS with WEIGHT statement) with a counting process input file structure.

We propose three different weights to correct the bias induced by dependent censoring. In general, when the dependent censoring is light or moderate, the unstabilized weight $W_1(t)$ works well. However, when censoring is heavy, $W_1(t)$ may be quite large toward the tail of the observation time resulting in unstable estimates. In this case, stabilized weights, $W_2(t)$ and $W_3(t)$, may be preferable and usually result in more efficient estimator than that from using the unstabilized weight $W_1(t)$. We found little difference in the performance of $W_2(t)$ and $W_3(t)$.

In simulation studies, we treated the IPCW weights and IPSW weights as fixed to simplify the computation, which would result in conservative covariance estimators because those weights are actually estimated as opposed to being known. However, simulation results suggest that the proposed ASEs by treating the IPCW weights and IPSW weights as fixed are quite accurate.

The proposed methods require the consistency of the IPCW weight. Therefore, the proportional hazards model for dependent censoring should be correctly specified. This may be approximately true when a sufficient number of covariates is collected.

In addition to depending on the outcome, $(\Delta_{1i}, \Delta_{2i}, \Delta_{3i})$, we can also allow ξ_i to depend on $Z_i(0)$. For example, in order to obtain reasonably precise estimates of the properties of a rare type, we can oversample subjects of this type.

Applying our methods to ESLD patients wait listed for liver transplantation, we found that (calculated) MELD score group of $[16, 17]$ is consistent with the HCC wait-list mortality hazard if no adjustment was made for dependent censoring. However, the consistent MELD score range changes to $[14, 15]$ after we consider dependent censoring by using weight W_2. Therefore, our results indicate that the current MELD exception score of 22 granted to at wait listing to HCC patients overstates the actual medical urgency; and that an assigned MELD score of 14 or 15 may be more

appropriate.

The proposed methods generally work well for set-ups with light and moderate dependent censoring. However, when the dependent censoring rate is very high, say 60% or more, the proposed estimators may perform more poorly, at least for the data settings we considered. No studies in the literature seem to have considered such high censoring rates and further study of data with heavy dependent censoring would be valuable.

CHAPTER IV

Hazard Regression Models for Estimating the Effect of an External Time-Dependent Covariate

4.1 Introduction

Hemodialysis (HD) is the most common treatment for advanced kidney failure. HD removes harmful waste products such as potassium, urea and free water from the blood, which would normally be eliminated in the urine. Typically, HD patients are required to follow a strict treatment regimen that involves receiving dialysis on either a Monday-Wednesday-Friday (MWF) or a Tuesday-Thursday-Saturday (TTS) schedule. During the interval between dialysis sessions, electrolytes and fluids may accumulate and increase the risk of mortality. Therefore, patients may be at higher risk of death on certain days, due to the intermittent nature of the dialysis schedule. For example, death risk may be elevated on Monday for MWF schedule patients or Tuesday for TTS schedule patients since these days are preceded by the longest intervals without dialysis.

The association between day-of-week-specific mortality risk and dialysis schedule has been investigated in various studies. Using crude death rates, Bleyer et al. (1999) revealed that there was an increased risk of sudden death and cardiac-related death on Monday for MWF schedule patients, and on Tuesday for TTS schedule patients. The authors used logistic models to investigate whether mortality was accentuated

for patients with increased age, coronary artery disease, diabetes mellitus, and/or congestive heart failure. Karnik et al. (2001) performed two-tailed binomial tests to study whether the risk of cardiac arrest was elevated on Monday for MWF schedule patients, and on Tuesday for TTS schedule patients. The authors found that the risk of cardiac arrest on Monday was higher for MWF schedule patients. Bleyer et al. (2006) studied the association between occurrences of sudden death among HD patients and the timing of HD. They performed χ^2 tests to test for differences between observed and expected frequencies in day and timing of deaths of patients. This study also showed that there was an increased risk of sudden death on Monday for MWF schedule patients, and on Tuesday for TTS schedule patients.

However, each of the studies referenced in the preceding paragraph was based on crude death rates and logistic regression models. Such approaches make it difficult (if not impossible) to adjust for time-dependent covariates or appropriately account for censoring. Since the endpoint is time-to-death, survival analysis (e.g., Cox regression) is well-suited for this purpose.

The data studied here were obtained from Dialysis Outcomes and Practice Patterns Study (DOPPS), an international prospective observational study of hemodialysis patients and facilities. DOPPS-I (1996-2001) contained more than 17,000 patients from seven countries including France, Germany, Italy, Japan, Spain, United Kingdom, and the U.S., while DOPPS-II (2002-2004) included more than 12,000 patients from the seven countries above as well as Australia/New Zealand, Belgium, Canada, and Sweden. In each phase, over 300 dialysis facilities were involved. At each facility, a random sample of HD patients was selected into the DOPPS.

In this chapter, we propose Cox models to evaluate the association between Monday/Tuesday mortality and dialysis schedule. Three models were fitted, each dis-

tinguished by the factor of interest: (i) day of the week (ii) day of dialysis schedule (iii) days since last dialysis. In each case, the factor of interest is coded as a time-dependent covariate. The models are compared and contrasted, with special attention given to the setting where the sample size is small. We address whether the Monday/Tuesday effect is similar across countries.

The remainder of this chapter is organized as follows. In Section 4.2, we describe the study population and the proposed methods for analyzing HD patients in the DOPPS. In Section 4.3, we present the results of the proposed models. The chapter concludes with some discussion in Section 4.4.

4.2 Methods

We use data from the DOPPS-I and DOPPS-II with U.S., Japanese and European (Belgium, France, Germany, Italy, Spain, Sweden, United Kingdom) patients composing three regional strata. Details of the DOPPS design have been reported previously; see for example Young et al. (2000), and Pisoni et al. (2004). The total study population consisted of 22,163 patients (9,227 U.S. patients, 4,419 Japanese patients, and 8,517 European patients). Patients were followed from the time they entered DOPPS until the death, receipt of a kidney transplant, loss to follow-up, or the end of the observation period, whichever occurred first.

At the start of participation in the DOPPS, demographic characteristics and comorbid conditions were obtained. Follow-up information was collected every four months. The date dialysis was received is also reported in the four-month period. In the current study, the dialysis schedule was defined to be MWF in the four-month interval if the reported date was a Monday, Wednesday or Friday. The TTS dialysis schedule was defined similarly. If the reported date was missing or a Sunday, the

date from the preceding four-month reporting interval was used to define the dialysis schedule.

The primary outcome of interest is all-cause mortality. We assume the time-dependent Cox proportional hazards model for all-cause mortality,

$$\lambda\left\{t|\boldsymbol{Z}(t)\right\} = \lambda_0(t)\exp\left\{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)\right\}, \tag{4.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard, $\boldsymbol{Z}(t)$ is a p-vector of possibly time-dependent covariates, and $\boldsymbol{\beta}_0$ is a p-dimensional regression parameter. Three models were fitted with a view to assessing the primary questions of interest. In each model, the factor of interest was coded as a time-dependent covariate, and Table 4.1 shows the detailed coding used. In Model 1, the covariate of interest is day of the week (Sunday, Monday, ..., Saturday). Each day was compared to the average of the seven days of the week, where $Z_{i1}(t) = I\,(\text{day } t \text{ for subject } i \text{ is a Monday}) - I\,(\text{day } t \text{ for subject } i \text{ is a Sunday}), \ldots, Z_{i6}(t) = I\,(\text{day } t \text{ for subject } i \text{ is a Saturday}) - I\,(\text{day } t \text{ for subject } i \text{ is a Sunday})$, where t is the time since the first ever HD day for patient i. In addition, $Z_{MWF}(t) = I\,(\text{ Patient receives MWF dialysis schedule at day } t\ )$ was used in this model as a covariate of stratification variable. In Model 2, the covariate of interest is day of dialysis schedule (1st, 2nd, ..., 7th), each was compared to the average of the seven days of the week, where $Z_{i1}(t) = I\,(\text{day } t \text{ for subject } i \text{ is the 1st day of dialysis schedule}) - I\,(\text{day } t \text{ for subject } i \text{ is the 7th day of dialysis schedule}), \ldots, Z_{i6}(t) = I\,(\text{day } t \text{ for subject } i \text{ is the 6th day of dialysis schedule}) - I\,(\text{day } t \text{ for subject } i \text{ is the 7th day of dialysis schedule})$. In Model 3, the covariate of interest is days since last dialysis (1, 2, 3), where $Z_{i1}(t) = I\,(\text{day } t \text{ for subject } i \text{ is 1 day since last dialysis }), \ldots, Z_{i3}(t) = I\,(\text{day } t \text{ for subject } i \text{ is 3 days since last dialysis }).$ Let α_j be the covariate coefficient for Z_{ij}, $j = 1, 2, 3$. Each covariate was compared to the average of the seven days of the week. Therefore, the parameters of interest

in Model 3 are $\beta_j = \alpha_j - 1/7\,(3\alpha_1 + 3\alpha_2 + \alpha_3)$ for $j = 1, 2, 3$. In Models 2 and 3, $Z_{MWF}(t)$ was fitted in the model as a covariate of stratification variable.

Since patients were not under observation until they entered the DOPPS study, our model took patient vintage (previous time on dialysis at entry into DOPPS) into account through left-truncation. Baseline adjustment covariates included the following factors: gender, race, 14 comorbid conditions (coronary heart disease, cancer other than skin, other cardiovascular disease, cerebrovascular disease, congestive heart failure, diabetes, gastrointestinal bleeding, HIV/AIDS, hypertension, lung disease, neurologic disease, psychiatric disorder, peripheral vascular disease, and recurrent cellulitis), body mass index (grouped as $<$20, [20-25), [25-30), $\geq$30 kg/m^2), and vascular access (catheter use). Country, phase, age group (18-29, 30-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, $\geq$70 years) and dialysis schedule were adjusted through stratification. Since patients from the same facility are not likely independent due to shared practice patterns, we accounted for intra-cluster dependence by using a robust ("sandwich") estimator to draw valid statistical inference. Cox models were fitted to each region (U.S., Europe and Japan) separately.

In Model 1, $Z_1 I\{Z_{MWF}(t) = 1\}$, $\ldots$, $Z_6 I\{Z_{MWF}(t) = 1\}$, $Z_1 I\{Z_{MWF}(t) = 0\}$, $\ldots$, and $Z_6 I\{Z_{MWF}(t) = 0\}$ were fitted in the model. With dialysis schedule being adjusted through stratification, this model has the advantage of estimating the specific day-of-week effect on mortality for each dialysis schedule. However, sufficient sample size and events are needed to draw reliable conclusions in fitting this complex model. In particular, in our study, conclusions drawn for Japanese HD patients may be unstable due to the smaller sample size and lower event rate. This limitation of Model 1 motivates the other two models which, although less detailed in nature, retain the ability to look at the primary questions concerning the effect of

Table 4.1: External Time-Dependent Covariates

	MWF patients								TTS patients						
	Sun	Mon	Tue	Wed	Thu	Fri	Sat		Sun	Mon	Tue	Wed	Thu	Fri	Sat
								Model 1							
Z_1	-1	1	0	0	0	0	0	Z_1	-1	1	0	0	0	0	0
Z_2	-1	0	1	0	0	0	0	Z_2	-1	0	1	0	0	0	0
Z_3	-1	0	0	1	0	0	0	Z_3	-1	0	0	1	0	0	0
Z_4	-1	0	0	0	1	0	0	Z_4	-1	0	0	0	1	0	0
Z_5	-1	0	0	0	0	1	0	Z_5	-1	0	0	0	0	1	0
Z_6	-1	0	0	0	0	0	1	Z_6	-1	0	0	0	0	0	1
								Model 2							
Z_1	-1	1	0	0	0	0	0	Z_1	0	-1	1	0	0	0	0
Z_2	-1	0	1	0	0	0	0	Z_2	0	-1	0	1	0	0	0
Z_3	-1	0	0	1	0	0	0	Z_3	0	-1	0	0	1	0	0
Z_4	-1	0	0	0	1	0	0	Z_4	0	-1	0	0	0	1	0
Z_5	-1	0	0	0	0	1	0	Z_5	0	-1	0	0	0	0	1
Z_6	-1	0	0	0	0	0	1	Z_6	1	-1	0	0	0	0	0
								Model 3							
Z_1	0	0	1	0	1	0	1	Z_1	1	0	0	1	0	1	0
Z_2	1	0	0	1	0	1	0	Z_2	0	1	0	0	1	0	1
Z_3	0	1	0	0	0	0	0	Z_3	0	0	1	0	0	0	0

time since last dialysis on mortality. Model 2 provides day-of-week effect on all-cause mortality. In this model, the Monday effect on all-cause mortality for MWF schedule patients is assumed to be identical to the Tuesday effect on all-cause mortality for TTS schedule patients with a similar equivalence for all subsequent days during the schedule. Model 3 specifically emphasizes the time elapsed since last dialysis as a predictor of mortality. In this model, the Monday effect on all-cause mortality for MWF schedule patients and the Tuesday effect on all-cause mortality for TTS schedule patients share the same magnitude, both being three days since the last dialysis treatment. Similarly, only the average effect of Tuesday, Thursday and Saturday from the MWF schedule and Sunday, Wednesday and Friday from the TTS schedule is represented in $Z_{i1}(t) = 1$; the average effect of Sunday, Wednesday and Friday from the MWF schedule and Monday, Thursday and Saturday from the TTS schedule is represented in $Z_{i2}(t) = 1$. This model provides a macroscopic view that facilitates easy comparison of the three regions.

The choice between Models 1-3 can be made according to the availability of data. In general, Model 1 assesses the effect of the day on mortality more precisely than Model 2, and Model 2 provides more detailed interpretation than Model 3. However, achieving readily interpreted parameters could result in loss of stability on the conclusion, especially when the sample size is small and event rate is low. In addition, the likelihood ratio tests are performed to compare the proposed three models in each of the three regions.

4.3 Results

For each region, Cox regression models were used to estimate the covariate-adjusted Monday/Tuesday effect on all-cause mortality. Table 4.2 contains three

sets of covariate-adjusted relative hazards of all-cause mortality for U.S. patients. Model 1 analysis shows all-cause mortality by day-of-week. It indicates that patients from the U.S. are estimated to have significant 1.41 and 1.39 times higher hazards of all-cause death on Mondays with MWF schedule and Tuesdays with TTS schedule, respectively, compared to the average of the seven days of the week ($p < 0.0001$ and $p < 0.0001$, respectively). Model 2 gives all-cause mortality by day of dialysis schedule. Results from the Model 2 analysis show that U.S. patients had a significant 1.40 times higher risk of all-cause death on Mondays with MWF schedule and Tuesdays with TTS schedule compared to overage average ($p < 0.0001$). Model 3 provides all-cause mortality by days since last dialysis. The analysis from Model 3 reveals that, in the U.S., patients experienced a significant 1.40 times higher all-cause mortality hazard on Monday (for MWF schedule) and Tuesdays (for TTS schedule) relative to the overall average ($p < 0.0001$).

In Table 4.3, relative hazards of all-cause mortality for European patients are displayed. Under Model 1, European patients had significant 1.34 and 1.22 times higher hazards of all-cause death on Mondays (for patients on a MWF schedule) and on Tuesdays (for patients on a TTS schedule) ($p = 0.001$ and $p = 0.043$, respectively); in each case the comparison is made with the average over the seven days. The effect of Tuesdays on all-cause mortality with TTS schedule is only marginally significant in this region. Results based on Model 2 show that European patients experienced a significant 1.29 times higher risk of all-cause death on Mondays (with MWF schedule) and Tuesdays (with TTS schedule) when compared to the overall average ($p = 0.0004$). Results from Model 3 show that, in Europe, all-cause mortality risk was significantly 1.28 times higher on Mondays with MWF schedule and Tuesdays with TTS schedule compared to the overall average ($p = 0.0005$). The

Table 4.2: Anslysis of DOPPS data: U.S. patients

Covariates	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$	p-value
		Model 1		
		MWF Schedule Patients		
Sunday	-0.14	0.07	0.87	0.030
Monday	0.34	0.05	1.41	< 0.0001
Tuesday	0.02	0.06	1.02	0.77
Wednesday	0.03	0.06	1.03	0.61
Thursday	-0.22	0.08	0.81	0.004
Friday	0.04	0.07	1.04	0.53
Saturday	-0.07	0.06	0.93	0.26
		TTS Schedule Patients		
Sunday	-0.15	0.08	0.86	0.086
Monday	0.11	0.08	1.12	0.16
Tuesday	0.33	0.07	1.39	< 0.0001
Wednesday	-0.19	0.08	0.83	0.027
Thursday	-0.06	0.07	0.94	0.41
Friday	-0.10	0.08	0.90	0.19
Saturday	0.05	0.08	1.05	0.54
		Model 2		
First	0.34	0.04	1.40	< 0.0001
Second	-0.06	0.05	0.94	0.19
Third	-0.008	0.05	0.99	0.86
Fourth	-0.17	0.05	0.84	0.002
Fifth	0.04	0.05	1.04	0.41
Sixth	-0.10	0.05	0.90	0.051
Seventh	-0.03	0.05	0.97	0.49
		Model 3		
One	-0.11	0.02	0.89	< 0.0001
Two	0.0005	0.02	1.00	0.98
Three	0.34	0.04	1.40	< 0.0001

differences estimated by Models 2 and 3 are more significant than those from Model 1.

Corresponding results for Japanese patients are shown in Table 4.4. The Model 1 analysis shows that in Japan, compared to the overall average, all-cause mortality was estimated to be 1.27 and 1.43 times higher on Mondays with MWF schedule and Tuesdays with TTS schedule, respectively. These differences, however, are not significant or only marginally significant in this region ($p = 0.15$ and $p = 0.044$, respectively). The Model 2 analysis shows that Japanese patients experienced a significant 1.34 times higher hazard of all-cause mortality on Mondays with MWF schedule and Tuesdays with TTS schedule compared with the overall average. The difference here are more significant ($p = 0.017$) compared to the result of Model 1. Model 3, as noted, emphasizes the time since last dialysis and we find that Japanese patients had a 1.31 times higher hazard of all-cause mortality on Mondays with MWF schedule and Tuesdays with TTS schedule compared to the overall average. This result is also more significant ($p = 0.027$) compared to the corresponding result from Model 1.

Our results from DOPPS with U.S., European, and Japanese patients from Models 1-3 indicate that in all three regions, HD patients have a higher hazard of all-cause mortality on Mondays with MWF schedule, or Tuesdays with TTS schedule. In general, Model 1 does not fit the data as well as Models 2 and 3 for European and Japanese patients since, in these regions, the Monday and Tuesday effects under Model 1 may not be statistically significant; such effects may be highly significant in Models 2 and 3. Results of Model 3 are consistent among all three regions, which provide a good macroscopic view of the Monday/Tuesday effect; however, this model

Table 4.3: Anslysis of DOPPS data: European patients

Covariates	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$	p-value
		Model 1		
		MWF Schedule Patients		
Sunday	-0.11	0.10	0.90	0.30
Monday	0.30	0.09	1.34	0.001
Tuesday	0.03	0.11	1.03	0.80
Wednesday	-0.01	0.10	0.99	0.89
Thursday	-0.19	0.10	0.83	0.067
Friday	0.02	0.10	1.02	0.87
Saturday	-0.02	0.09	0.98	0.78
		TTS Schedule Patients		
Sunday	-0.04	0.11	0.96	0.69
Monday	-0.03	0.10	0.97	0.77
Tuesday	0.20	0.10	1.22	0.043
Wednesday	0.03	0.11	1.03	0.80
Thursday	-0.20	0.12	0.82	0.097
Friday	-0.22	0.11	0.80	0.052
Saturday	0.17	0.11	1.30	0.013
		Model 2		
First	0.25	0.07	1.29	0.0004
Second	0.03	0.08	1.03	0.75
Third	-0.09	0.07	0.91	0.21
Fourth	-0.21	0.08	0.81	0.007
Fifth	0.13	0.07	1.14	0.070
Sixth	-0.04	0.06	0.96	0.58
Seventh	-0.07	0.07	0.93	0.28
		Model 3		
One	-0.07	0.03	0.93	0.026
Two	0.01	0.03	0.99	0.74
Three	0.25	0.07	1.28	0.0005

Table 4.4: Anslysis of DOPPS data: Japanese patients

Covariates	$\widehat{\beta}$	SE	$\exp(\widehat{\beta})$	p-value
		Model 1		
		MWF Schedule Patients		
Sunday	-0.09	0.18	0.92	0.63
Monday	0.23	0.16	1.27	0.15
Tuesday	0.09	0.22	1.09	0.69
Wednesday	0.005	0.17	1.01	0.98
Thursday	-0.44	0.22	0.65	0.043
Friday	0.37	0.18	1.44	0.041
Saturday	-0.17	0.20	0.84	0.39
		TTS Schedule Patients		
Sunday	-0.46	0.23	0.63	0.057
Monday	-0.38	0.23	0.68	0.095
Tuesday	0.36	0.18	1.43	0.044
Wednesday	0.19	0.17	1.21	0.27
Thursday	0.13	0.23	1.14	0.57
Friday	-0.19	0.23	0.83	0.41
Saturday	0.36	0.20	1.43	0.075
		Model 2		
First	0.29	0.12	1.34	0.017
Second	0.13	0.14	1.14	0.34
Third	0.06	0.14	1.06	0.66
Fourth	-0.32	0.16	0.73	0.055
Fifth	0.36	0.14	1.43	0.010
Sixth	-0.29	0.15	0.75	0.056
Seventh	-0.24	0.14	0.79	0.10
		Model 3		
One	-0.15	0.07	0.86	0.023
Two	0.06	0.07	1.07	0.36
Three	0.27	0.12	1.31	0.027

does not have enough precision. If the sample size is large and the event rate is high, Model 1 would be preferred since it provides precise Monday/Tuesday effect.

Likelihood ratio tests indicate that, in each of the three regions, Model 1 is not significantly better than Model 2. In addition, the difference between Models 2 and 3 is not significant, except in Japan.

4.4 Discussion

In this chapter, we investigated the association between Monday/Tuesday effect on all-cause mortality and dialysis schedule. In practice, the analysis of such data has been limited to crude death rate or logistic regression models. We used Cox models with the covariate of interest serving as a time-dependent covariate. Three models were fitted, distinguished by the covariate of interest: (i) day of the week (ii) day of dialysis schedule (iii) days since last dialysis. With these models, one gains a thorough understanding of the Monday/Tuesday effect on all-cause mortality. Further, the Cox model appropriately accounts for right censoring, whereas other methods (e.g., such as logistic regression) do not track such events accurately. In particular, Green and Symons (1983) concluded that when the follow-up period is long, the Cox model is superior to the logistic model because the Cox model explains more variability of the data than logistic regression. In addition, Cox models with time-dependent covariates can appropriately use the covariate information that varies over time, whereas logistic regression can not gain such benefit easily.

If the sample size is large, Model 1 results in a more detailed Monday/Tuesday effect on all-cause mortality than Model 2, and Model 2 has more precise Monday/Tuesday effect on all-cause mortality than Model 3. However, if the sample size is small and the event rate is low, inferences drawn from Model 1 are subject to

substantially more uncertainty than Model 2, and Model 2 leads to more unreliable conclusions than Model 3. Therefore, the choice between Models 1-3 depends on the sample size and the precision of the conclusion one intends to gain. In addition, Models 2 and 3 provide an overall insight of the Monday/Tuesday effect on all-cause mortality and make the comparison of the three regions easily. The commonality among the three regions is more pronounced in Model 3 than in Model 2. Though results from Model 2 contains more information than those from Model 3, the likelihood ratio tests show that Model 2 is not significantly better than Model 3 except in Japan. If one needs to get even more detailed Monday/Tuesday effect on all-cause mortality, Model 1 can be used; however, the difference between Models 1 and 2 is not significant in the three regions.

With Models 1-3, our results indicate that in all three regions (U.S., Europe, and Japan), HD patients have a higher hazard of all-cause mortality on Mondays with MWF schedule, or Tuesdays with TTS schedule. This implies that there may be an advantage to a more frequent dialysis schedule in these regions.

In summary, we believe that the time-dependent Cox model provides a useful methodology for estimating the association between the Monday/Tuesday effect on all-cause mortality and dialysis schedule. Our results imply that there may be an advantage to a more frequent dialysis schedule in the U.S., Europe and Japan.

The proposed methods are based on a proportional hazards model. However, in certain situations, the proportional hazards model may not be appropriate. The accelerated failure time (AFT) model is an alternative method and may be appealing since the parameters can be interpreted easily. Such a model with an external time-dependent covariate that rotates regularly would be worth investigating.

CHAPTER V

Conclusion

This dissertation proposes three novel statistical methods for analyzing failure time data, targeting four important issues that frequently arise in observational data: (i) study subjects are clustered, (ii) subjects are sampled in a manner which explicitly depends on the outcome (e.g., death, illness), (iii) subjects are censored in a manner independent of the failure rate, and (iv) covariate of interest is an external time-dependent covariate that rotates regularly. Chapter II proposes methods that are based on estimating equations for case-cohort designs for clustered failure time data. Chapter III considers the setting with outcome-dependent sampling and dependent censoring. Chapter IV estimates the effect of an external time-dependent covariate under a proportional regression model.

Under a marginal hazards model, the methods in Chapter II feature tractable asymptotic derivations. The risk set includes not only subcohort members in the case-cohort design, but future failures outside the subcohort, resulting in potentially increased efficiency relative to some existing methods. Chapter III employs a novel double-inverse-weighting scheme which combines weights corresponding to the probability of remaining uncensored and the probability of being sampled. Chapter IV performs a comprehensive investigation of the association between the day-of-week-

specific death rates and the dialysis schedule in the U.S., several European countries and Japan. The covariate of interest is an external time-dependent covariate that rotates regularly.

The methods developed in Chapters II-IV were motivated by real research questions regarding mortality on dialysis among end-stage renal disease patients and wait-list mortality among patients with end-stage liver disease. Hence, the contribution of this research is both clinical and statistical. Each method proposed in this dissertation was applied to a real medical research question. In Chapter II, by applying the proposed methods to a study of mortality among Canadian dialysis patients, we found that Canadian HD patients appear to be at an increased risk of CVD death on Monday and Tuesday. In Chapter III, the proposed methods were applied to end-stage liver disease data. We observed that MELD score group of $[14, 15]$ is consistent with HCC wait-list mortality. Therefore, the current MELD exception score of 22 assigned to HCC patients overstates the true wait-list mortality. In Chapter IV, we studied the association between day-of-week-specific mortality and dialysis schedule for DOPPS patients from the U.S., European countries and Japan. We found that in the three regions, HD patients have a higher hazard of all-cause mortality on Mondays with MWF schedule, or Tuesdays with TTS schedule.

Several possible extensions to the methods proposed in this dissertation may be worth consideration. For example, the methods in Chapter II are under a marginal proportional hazards model. A proportional hazards frailty model combined with maximum likelihood estimation would be valuable. The performance of the proposed methods in Chapter III may be poor when the dependent censoring rate is very high, implying that future studies of data with heavy dependent censoring would be worth investigating.

APPENDICES

APPENDIX A

Proof of Theorems and addition simulation studies in Chapter II

A.1 Proof of Theorem 1

Evaluated at the true values, the estimating function is given by

$$\boldsymbol{U}(\boldsymbol{\beta}_0, p_0) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \left\{\boldsymbol{Z}_{ij}(u) - \frac{\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}_0, p_0, u)}{\overline{S}^{(0)}(\boldsymbol{\beta_0}, p_0, u)}\right\} dN_{ij}(u).$$

By some simple algebra, we have

$$\begin{aligned} n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}, p_0) &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta}, u)\}\, dN_{ij}(u) \\ &\quad -n^{1/2}\int_0^\tau \left\{\frac{\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, p_0, u)}{\overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)} - \boldsymbol{e}(\boldsymbol{\beta}, u)\right\} dF(u) \\ &\quad -n^{1/2}\int_0^\tau \left\{\frac{\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, p_0, u)}{\overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)} - \boldsymbol{e}(\boldsymbol{\beta}, u)\right\} \left\{d\overline{N}(u) - dF(u)\right\}. \end{aligned}$$

By a functional Taylor expansion of $\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, p_0, u)/\overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)$ with respect to $\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, p_0, u)$ and $\overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)$ around $\mu^{-1}\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, u)$ and $\mu^{-1}s^{(0)}(\boldsymbol{\beta}, u)$, respectively, combined with Conditions (d), (e) and the fact that $n^{1/2}\left\{\overline{N}(u) - F(u)\right\}$ converges in distribution to a zero-mean Gaussian process, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}, p_0)$ can be written as

$$\begin{aligned} &n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta}, u)\}\, dN_{ij}(u) \\ &-n^{1/2}\int_0^\tau \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta}, u)\}\, Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \end{aligned}$$

$$\times \left\{ \frac{p_0}{N_1}\delta_{ij} + \frac{1-p_0}{n_0}(1-\delta_{ij})H_iH_{ij} \right\} \left\{ \frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u) \right\}^{-1} dF(u)$$

$$+o_p(1)$$

$$\begin{aligned} = \; & n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\}\, dN_{ij}(u) \\ & -n^{-1/2}\frac{p_0}{N_1/n}\int_0^\tau \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\}\,\delta_{ij}Y_{ij}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \\ & \qquad \times \left\{ \frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u) \right\}^{-1} dF(u) \qquad \text{(A.1)} \\ & -n^{-1/2}\frac{1-p_0}{n_0/n}\int_0^\tau \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\}\,(1-\delta_{ij})H_iH_{ij}Y_{ij}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \\ & \qquad \times \left\{ \frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u) \right\}^{-1} dF(u) + o_p(1), \qquad \text{(A.2)} \end{aligned}$$

followling a parallel setting described by van der Vaart and Wellner (1996) (example 2.11.16 on p.215). By another functional Taylor expansion, we get

$$\begin{aligned} \frac{p_0}{N_1/n} &= \frac{1}{\mu} - \frac{1}{\mu^2 p_0}\left(\frac{N_1}{n} - \mu p_0\right) + o_p(1) \\ \frac{1-p_0}{n_0/n} &= \frac{1}{\mu\gamma\theta} - \frac{1}{(\mu\gamma\theta)^2(1-p_0)}\left\{\frac{n_0}{n} - \mu\gamma\theta(1-p_0)\right\} + o_p(1), \end{aligned}$$

such that (A.1) can be written as

$$\begin{aligned} & -n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\}\,\frac{\delta_{ij}}{\mu}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \\ & \qquad \times \left\{ \frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u) \right\}^{-1} dF(u) \\ & +n^{-1/2}\left(\frac{N_1}{n} - \mu p_0\right)\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta},u)\}\,\frac{\delta_{ij}}{\mu^2 p_0}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)} \\ & \qquad \times \left\{ \frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u) \right\}^{-1} dF(u) + o_p(1). \end{aligned}$$

It is easy to show that

$$\begin{aligned} n^{-1/2}\left(\frac{N_1}{n} - \mu p_0\right) &= n^{-1/2}\sum_{i=1}^{n}\frac{\sum_{j=1}^{m_i}\delta_{ij} - \mu p_0}{n} \\ &= n^{-1/2}\sum_{i=1}^{n}G_{1i}(p_0), \qquad \text{(A.3)} \end{aligned}$$

where $G_{1i}(p)$ is as defined in Theorem 1, and that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu^2 p_0}\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}dF(u)\\
=\;&\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu^2 p_0}\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}d\overline{N}(u)\\
&-\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu^2 p_0}\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}\\
&\qquad\times d\left\{\overline{N}(u)-F(u)\right\}\\
=\;&n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu^2 p_0}\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}\\
&\qquad\times\sum_{k=1}^{n}\sum_{l=1}^{m_i}dN_{kl}(u)+o_p(1). \qquad \text{(A.4)}
\end{aligned}$$

Combining (A.3), (A.4) and using the fact that (A.4) converges in probability to $\boldsymbol{D}_1(\boldsymbol{\beta})$, (A.1) can be written as

$$\begin{aligned}
&-n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu}\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}\\
&\qquad\times dF(u)+\boldsymbol{D}_1(\boldsymbol{\beta})\times n^{-1/2}\sum_{i=1}^{n}G_{1i}(p_0)+o_p(1).
\end{aligned}$$

Similarly, we can show that (A.2) can be written as

$$\begin{aligned}
&-n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\frac{1}{\mu\gamma\theta}(1-\delta_{ij})Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\\
&\qquad\times\left\{\frac{1}{\mu}s^{(0)}(\boldsymbol{\beta},u)\right\}^{-1}dF(u)+\boldsymbol{D}_2(\boldsymbol{\beta})\times n^{-1/2}\sum_{i=1}^{n}G_{2i}(p_0)+o_p(1).
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
&n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta},p_0)\\
=\;&n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\,dN_{ij}(u)\\
&-n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u)-\boldsymbol{e}(\boldsymbol{\beta},u)\}\left\{\frac{1}{\mu}\delta_{ij}+\frac{1}{\mu\gamma\theta}(1-\delta_{ij})H_iH_{ij}\right\}
\end{aligned}$$

$$\times Y_{ij}(u)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\boldsymbol{\beta}, u) \right\}^{-1} dF(u)$$

$$+\boldsymbol{D}_1(\boldsymbol{\beta}) \times n^{-1/2} \sum_{i=1}^{n} G_{1i}(p_0) + \boldsymbol{D}_2(\boldsymbol{\beta}) \times n^{-1/2} \sum_{i=1}^{n} G_{2i}(p_0) + o_p(1)$$

$$= \quad n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_i(\boldsymbol{\beta}, p_0) + o_p(1),$$

with $\boldsymbol{W}_i(\boldsymbol{\beta}, p)$ as defined in Theorem 1.

The quantity $\boldsymbol{W}_i(\boldsymbol{\beta}_0, p_0)$ can be written as

$$\begin{aligned} \boldsymbol{W}_i(\boldsymbol{\beta}_0, p_0) \quad = \quad & \sum_{j=1}^{m_i} \int_0^{\tau} \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta}_0, u)\} \, dM_{ij}(u) \\ & + \sum_{j=1}^{m_i} \int_0^{\tau} \{\boldsymbol{Z}_{ij}(u) - \boldsymbol{e}(\boldsymbol{\beta}_0, u)\} \, (1 - \delta_{ij})(1 - \frac{1}{\gamma\theta} H_i H_{ij}) Y_{ij}(u) \\ & \times e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(u)} \lambda_0(u) du + \boldsymbol{D}_1(\boldsymbol{\beta}_0) G_{1i}(p_0) + \boldsymbol{D}_2(\boldsymbol{\beta}_0) G_{2i}(p_0), \end{aligned}$$

where $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(u)} \lambda_0(u) du$ is a mean-zero process.

Note that $\mathcal{E}\{1 - (\gamma\theta)^{-1} H_i H_{ij}\} = 0$, $\mathcal{E}\{dM_{ij}(u)\} = 0$, $\mathcal{E}\{G_{1i}(p_0)\} = 0$ and $\mathcal{E}\{G_{2i}(p_0)\} = 0$, such that $\mathcal{E}\{\boldsymbol{W}_i(\boldsymbol{\beta}_0, p_0)\} = 0$, for $i = 1, \ldots, n$. Hence under the assumed conditions, asymptotically, $\{\boldsymbol{W}_i(\boldsymbol{\beta}_0, p_0)\}_{i=1}^n$ are independent and identically distributed random quantities with mean zero and finite variance, $\mathcal{E}\{\boldsymbol{W}_1(\boldsymbol{\beta}_0, p_0)^{\otimes 2}\}$. By the Multivariate Central Limit Theorem (MCLT), $n^{-1/2} \boldsymbol{U}(\boldsymbol{\beta}_0, p_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, p_0))$, where $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, p_0)$ is defined in Theorem 1.

A.2 Proof of Theorem 2

To prove the consistency of $\widehat{\boldsymbol{\beta}}_t$, we use the Inverse Function Theorem (Foutz (1977)) by verifying the following conditions:

(i) $\partial \boldsymbol{U}(\boldsymbol{\beta}, p_0)/\partial \boldsymbol{\beta}^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$.

(ii) $-n^{-1} \partial \boldsymbol{U}(\boldsymbol{\beta}, p_0)/\partial \boldsymbol{\beta}^T \big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is positive definite with probability 1 as $n \to \infty$.

(iii) $-n^{-1}\partial \boldsymbol{U}(\boldsymbol{\beta}, p_0)/\partial \boldsymbol{\beta}^T$ converges in probability to a fixed function, $\boldsymbol{A}(\boldsymbol{\beta})$, uniformly in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$.

(iv) Asymptotic unbiasedness of the estimating function: $-n^{-1}\boldsymbol{U}(\boldsymbol{\beta}_0, p_0) \xrightarrow{P} 0$.

Conditions (i), (ii) and (iii) follow from Conditions (d), (e), (f) and (g). Using the result in the proof of Theorem 1, $n^{-1}\boldsymbol{U}(\boldsymbol{\beta}_0, p_0) \xrightarrow{P} 0$ by Chebyshev's inequality. Then, Condition (iv) holds under the assumed model. Having now verified conditions (i) to (iv), we conclude that $\widehat{\boldsymbol{\beta}}_t$ converges in probability to $\boldsymbol{\beta}_0$.

A.3 Proof of Theorem 3

Here we prove results for $\widehat{\boldsymbol{\beta}}_s$ only, since results for $\widehat{\boldsymbol{\beta}}_w$ can be proved similarly. By a Taylor expansion of the score function $\boldsymbol{U}(\widehat{\boldsymbol{\beta}}_s, \widehat{p}_s)$ with respect to $\boldsymbol{\beta}$ and around $\boldsymbol{\beta}_0$, and by a Taylor expansion of $\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s)$ with respect to p around p_0,

$$\begin{aligned} n^{-1/2}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}_s, \widehat{p}_s) &= n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s) - \widehat{\boldsymbol{A}}(\boldsymbol{\beta}_*, \widehat{p}_s)\, n^{1/2}(\widehat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_0) \\ n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s) &= n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, p_0) + \widehat{\boldsymbol{B}}(\boldsymbol{\beta}_0, p_*)\, n^{1/2}(\widehat{p}_s - p_0), \end{aligned}$$

where $\boldsymbol{\beta}_*$ is on the line segment between $\widehat{\boldsymbol{\beta}}_s$ and $\boldsymbol{\beta}_0$, p_* is on the line segment between $\widehat{p}_s$ and p_0, and

$$\begin{aligned} \widehat{\boldsymbol{A}}(\boldsymbol{\beta}, p) &= -n^{-1}\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{U}(\boldsymbol{\beta}, p) \\ &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \left[\frac{\overline{S}^{(2)}(\boldsymbol{\beta}, p, u)}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)} - \left\{\frac{\overline{S}^{(1)}(\boldsymbol{\beta}, p, u)}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)}\right\}^{\otimes 2}\right] dN_{ij}(u) \end{aligned}$$

$$\begin{aligned} \widehat{\boldsymbol{B}}(\boldsymbol{\beta}, p) &= n^{-1}\frac{\partial}{\partial p}\boldsymbol{U}(\boldsymbol{\beta}, p) \\ &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^\tau \left\{\frac{\overline{S}^{(1)}(\boldsymbol{\beta}, p, u)}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)^2}\frac{\partial}{\partial p}\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)\right. \\ &\quad \left. - \frac{1}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)}\frac{\partial}{\partial p}\overline{S}^{(1)}(\boldsymbol{\beta}, p, u)\right\} dN_{ij}(u). \end{aligned}$$

Since $\widehat{\boldsymbol{\beta}}_s \xrightarrow{P} \boldsymbol{\beta}_0$ and $\|\boldsymbol{\beta}_* - \boldsymbol{\beta}_0\| \leq \|\widehat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_0\|$, $\boldsymbol{\beta}_* \xrightarrow{P} \boldsymbol{\beta}_0$. Using the fact that $\widehat{p}_s \xrightarrow{P} p_0$, Condition (e) and continuity,

$$\begin{aligned}\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_*, \widehat{p}_s) &\xrightarrow{P} \int_0^\tau \left\{ \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} - \boldsymbol{e}(\boldsymbol{\beta}_0, u)^{\otimes 2} \right\} dF(u) \\ &\equiv \boldsymbol{A}(\boldsymbol{\beta}_0).\end{aligned}$$

Since $\widehat{p}_s \xrightarrow{P} p_0$ and $\|p_* - p_0\| \leq \|\widehat{p}_s - p_0\|$, we obtain that $p_* \xrightarrow{P} p_0$. We can express $\boldsymbol{R}^{(d)}(\boldsymbol{\beta}, p, u)$ as follows,

$$\begin{aligned}\boldsymbol{R}^{(d)}(\boldsymbol{\beta}, p, u) &= \frac{\partial}{\partial p} \overline{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta}, p, u) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \frac{1}{N_1} \delta_{ij} - \frac{1}{n_0} (1 - \delta_{ij}) H_i H_{ij} \right\} Y_{ij}(u) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(u)} \boldsymbol{Z}_{ij}(u)^{\otimes d} \qquad \text{(A.5)} \\ &= \frac{1}{N_1/n} \times n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \delta_{ij} Y_{ij}(u) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(u)} \boldsymbol{Z}_{ij}(u)^{\otimes d} \\ &\quad - \frac{1}{n_0/n} \times n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (1 - \delta_{ij}) H_i H_{ij} Y_{ij}(u) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(u)} \boldsymbol{Z}_{ij}(u)^{\otimes d},\end{aligned}$$

such that

$$\begin{aligned}\boldsymbol{R}^{(d)}(\boldsymbol{\beta}_0, p_0, u) &\xrightarrow{P} \frac{1}{p_0} \mathcal{E} \left\{ \delta_{11} Y_{11}(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{11}(u)} \boldsymbol{Z}_{11}(u)^{\otimes d} \right\} \\ &\quad - \frac{1}{1 - p_0} \mathcal{E} \left\{ (1 - \delta_{11}) Y_{11}(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{11}(u)} \boldsymbol{Z}_{11}(u)^{\otimes d} \right\} \\ &= \boldsymbol{r}^{(d)}(\boldsymbol{\beta}_0, u).\end{aligned}$$

Then, by continuous mapping,

$$\begin{aligned}\widehat{\boldsymbol{B}}(\boldsymbol{\beta}_0, p_*) &\xrightarrow{P} \int_0^\tau \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)^{\otimes 2}} r^{(0)}(\boldsymbol{\beta}_0, u) - \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)} \boldsymbol{r}^{(1)}(\boldsymbol{\beta}_0, u) \right\} dF(u) \\ &\equiv \boldsymbol{B}(\boldsymbol{\beta}_0).\end{aligned}$$

Using the fact that

$$\begin{aligned}\widehat{p}_s - p_0 &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}\delta_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}} - p_0 \\ &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p_0)}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}} \\ &= n^{-1}\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p_0)}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}/n},\end{aligned}$$

it follows that

$$\begin{aligned}n^{1/2}(\widehat{p}_s - p_0) &= n^{-1/2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p_0)}{\mu\gamma\theta} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{1}{\mu\gamma\theta}\sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p_0)\right\} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n} Q_i(p_0) + o_p(1).\end{aligned}$$

Note that $\mathcal{E}\left\{H_i H_{ij}(\delta_{ij} - p_0)\right\} = 0$, such that $\mathcal{E}\left\{Q_i(p_0)\right\} = 0$. Therefore,

$$\begin{aligned}n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s) &= n^{-1/2}\sum_{i=1}^{n}\left\{\boldsymbol{W}_i(\boldsymbol{\beta}_0, p_0) + \boldsymbol{B}(\boldsymbol{\beta}_0)Q_i(p_0)\right\} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_i(\boldsymbol{\beta}_0, p_0) + o_p(1),\end{aligned}$$

where $\boldsymbol{\psi}_i(\boldsymbol{\beta}, p)$ is as defined in Theorem 3.

Since $\mathcal{E}\left\{\boldsymbol{\psi}_i(\boldsymbol{\beta}_0, p_0)\right\} = 0$, by the MCLT,

$$n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s) \xrightarrow{D} N\left(0, \boldsymbol{\Omega}(\boldsymbol{\beta}_0)\right),$$

where $\boldsymbol{\Omega}(\boldsymbol{\beta}_0) = \mathcal{E}\left\{\boldsymbol{\psi}_i(\boldsymbol{\beta}_0, p_0)^{\otimes 2}\right\}$. We then have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_0) = \widehat{\boldsymbol{A}}(\boldsymbol{\beta}_*, \widehat{p}_s)^{-1} \times n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}_s),$$

since $\boldsymbol{U}(\widehat{\boldsymbol{\beta}}_s, \widehat{p}_s) = 0$. Note that $\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_*, \widehat{p}_s) \xrightarrow{P} \boldsymbol{A}(\boldsymbol{\beta}_0)$. Therefore by Slutsky's Theorem, $n^{1/2}(\widehat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_0) \xrightarrow{D} N\left(0, \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\right)$, completing the proof.

A.4 Covariance Matrix Estimators

We now describe he consistent estimates of the covariance matrices in Theorems

2 and 3. Let $\widehat{\gamma} = n^{-1}\sum_{i=1}^{n} H_i$, and $\widehat{\theta} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} H_i H_{ij} / \sum_{i=1}^{n} H_i m_i$. The covariance matrices $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, p_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ and $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_a(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ can be consistently estimated by $\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}_t, p_0)^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}_t, p_0)\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}_t, p_0)^{-1}$ and $\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a)^{-1}\widehat{\boldsymbol{\Omega}}_a(\widehat{\boldsymbol{\beta}}_a)$ $\times\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a)^{-1}$, respectively, where $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}_t, p_0) = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{W}}_i(\widehat{\boldsymbol{\beta}}_t, p_0)$, $\widehat{\boldsymbol{\Omega}}_a(\widehat{\boldsymbol{\beta}}_a) = n^{-1}$ $\sum_{i=1}^{n}\widehat{\boldsymbol{\psi}}_i^a(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a)$, $\widehat{\boldsymbol{\psi}}_i^a(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a) = \widehat{\boldsymbol{W}}_i(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a) + \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a)\widehat{Q}_i^a(\widehat{p}_a)$, $\widehat{\boldsymbol{W}}_i(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a) = \sum_{j=1}^{m_i}\widehat{\boldsymbol{W}}_{ij}(\widehat{\boldsymbol{\beta}}_a, \widehat{p}_a)$, and

$$
\begin{aligned}
\widehat{\boldsymbol{W}}_{ij}(\boldsymbol{\beta}, p) &= \left\{\boldsymbol{Z}_{ij}(X_{ij}) - \overline{\boldsymbol{E}}(\boldsymbol{\beta}, p, X_{ij})\right\}\delta_{ij} \\
&\quad -n^{-1}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\left\{\frac{1}{\widehat{\mu}}\delta_{ij} + \frac{1}{\widehat{\mu}\widehat{\gamma}\widehat{\theta}}(1-\delta_{ij})H_i H_{ij}\right\} Y_{ij}(X_{kl}) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(X_{kl})} \\
&\quad \times\left\{\overline{S}^{(0)}(\boldsymbol{\beta}, p, X_{kl})\right\}^{-1}\left\{\boldsymbol{Z}_{ij}(X_{kl}) - \overline{\boldsymbol{E}}(\boldsymbol{\beta}, p, X_{kl})\right\}\delta_{kl} \\
&\quad +\widehat{\boldsymbol{D}}_1(\boldsymbol{\beta})G_{1i}(p) + \widehat{\boldsymbol{D}}_2(\boldsymbol{\beta})G_{2i}(p) \\
\widehat{\boldsymbol{B}}(\boldsymbol{\beta}, p) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\frac{\overline{S}^{(1)}(\boldsymbol{\beta}, p, X_{ij})R^{(0)}(\boldsymbol{\beta}, p, X_{ij})}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, X_{ij})^2} - \frac{\boldsymbol{R}^{(1)}(\boldsymbol{\beta}, p, X_{ij})}{\overline{S}^{(0)}(\boldsymbol{\beta}, p, X_{ij})}\right\}\delta_{ij} \\
\widehat{Q}_i^s(p) &= \frac{1}{\widehat{\mu}\widehat{\gamma}\widehat{\theta}}\sum_{j=1}^{m_i} H_i H_{ij}(\delta_{ij} - p) \\
\widehat{Q}_i^w(p) &= \frac{1}{\widehat{\mu}}\sum_{j=1}^{m_i}(\delta_{ij} - p) \\
\widehat{\boldsymbol{D}}_1(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\left\{\boldsymbol{Z}_{ij}(X_{kl}) - \overline{\boldsymbol{E}}(\boldsymbol{\beta}, p, X_{kl})\right\}\frac{\delta_{ij}}{\widehat{\mu}^2 p} \\
&\quad \times Y_{ij}(X_{kl}) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(X_{kl})}\left\{\frac{1}{\widehat{\mu}}\overline{S}^{(0)}(\boldsymbol{\beta}, p, X_{kl})\right\}^{-1}\delta_{kl} \\
\widehat{\boldsymbol{D}}_2(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\left\{\boldsymbol{Z}_{ij}(X_{kl}) - \overline{\boldsymbol{E}}(\boldsymbol{\beta}, p, X_{kl})\right\}\frac{(1-\delta_{ij})H_i H_{ij}}{(\widehat{\mu}\widehat{\gamma}\widehat{\theta})^2(1-p)} \\
&\quad \times Y_{ij}(X_{kl}) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(X_{kl})}\left\{\frac{1}{\widehat{\mu}}\overline{S}^{(0)}(\boldsymbol{\beta}, p, X_{kl})\right\}^{-1}\delta_{kl},
\end{aligned}
$$

and $\boldsymbol{R}^{(d)}(\boldsymbol{\beta}, p, u)$ is as defined in (A.5).

A.5 Proof of Theorem 4

We can decompose $\alpha_n(t) = \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \Lambda_0(t)$ into three parts, $\alpha_n(t) = \alpha_{1:n}(t) +$

$\alpha_{2:n}(t) + \alpha_{3:n}(t)$, where

$$\begin{aligned} \alpha_{1:n}(t) &= \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, p_0, t) \\ \alpha_{2:n}(t) &= \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, p_0, t) - \widehat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) \\ \alpha_{3:n}(t) &= \widehat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) - \Lambda_0(t). \end{aligned} \tag{A.6}$$

Taking a Taylor expansion of $\alpha_{1:n}(t)$,

$$\begin{aligned} \alpha_{1:n}(t) &= \left. \frac{\partial \widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, p, t)}{\partial p} \right|_{p=p_*} \times (\widehat{p} - p_0) \\ &= -\int_0^t \frac{1}{\mu \overline{S}^{(0)}(\widehat{\boldsymbol{\beta}}, p, u)^2} \frac{\partial}{\partial p} \overline{S}^{(0)}(\widehat{\boldsymbol{\beta}}, p, u) d\overline{N}(u) \Big|_{p=p_*} \times (\widehat{p} - p_0) \\ &= -\int_0^t \frac{R^{(0)}(\widehat{\boldsymbol{\beta}}, p_*, u)}{\mu \overline{S}^{(0)}(\widehat{\boldsymbol{\beta}}, p_*, u)^2} d\overline{N}(u) \times (\widehat{p} - p_0), \end{aligned}$$

where p_* lies between $\widehat{p}$ and p_0, and $R^{(0)}(\boldsymbol{\beta}, p, u)$ is as defined in (A.5). Under assumptions (a)-(g), $\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)$, $R^{(0)}(\boldsymbol{\beta}, p, u)$ and $\overline{N}(u)$ are all bounded and $\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)$ is bounded away from 0. Using the fact that $\widehat{p}$ converges in probability to p_0 implies that $\alpha_{1:n}(t) \xrightarrow{P} 0$.

With respect to the second term of (A.6), applying a Taylor expansion,

$$\begin{aligned} \alpha_{2:n}(t) &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \left. \frac{\partial \widehat{\Lambda}_0(\boldsymbol{\beta}, p_0, t)}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_*} \\ &= -(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \int_0^t \left. \frac{\overline{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, p_0, u)}{\mu \overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)^2} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_*} d\overline{N}(u) \\ &= -(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \int_0^t \frac{\overline{\boldsymbol{E}}(\boldsymbol{\beta}_*, p_0, u)}{\mu \overline{S}^{(0)}(\boldsymbol{\beta}_*, p_0, u)} d\overline{N}(u), \end{aligned}$$

where $\boldsymbol{\beta}_*$ lies between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Since $\overline{\boldsymbol{E}}(\boldsymbol{\beta}, p_0, u)$ and $\overline{N}(u)$ are bounded, $\overline{S}^{(0)}(\boldsymbol{\beta}, p_0, u)$ is bounded away from 0, and $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$, it follows that $\alpha_{2:n}(t) \xrightarrow{P} 0$.

Now, considering the last term in (A.6),

$$\begin{aligned}
\alpha_{3:n}(t) &= \int_0^t \frac{d\overline{N}(u)}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \int_0^t \lambda_0(u)du \\
&= \int_0^t \frac{d\overline{N}(u)}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \int_0^t \frac{n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij}(u)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(u)}\lambda_0(u)}{S^{(0)}(\boldsymbol{\beta}_0, u)}du \\
&= \int_0^t \frac{d\overline{N}(u)}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \{dN_{ij}(u) - dM_{ij}(u)\} \\
&= \int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} d\overline{N}(u) + \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\overline{M}(u) \\
&= \int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} d\left\{\overline{N}(u) - F(u)\right\} \\
&\quad + \int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} dF(u) + \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\overline{M}(u) \\
&= \int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} dF(u) + \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\overline{M}(u) \\
&\quad + o_p(n^{-1/2}).
\end{aligned}$$

Since $\left\{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)\right\}^{-1} - S^{(0)}(\boldsymbol{\beta}_0, u)^{-1} \xrightarrow{P} 0$, with $F(u)$ bounded, $S^{(0)}(\boldsymbol{\beta}_0, u) \xrightarrow{P} s^{(0)}(\boldsymbol{\beta}_0, u)$, which is bounded away from 0, and since $\int_0^t d\overline{M}(u) = n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^t dM_{ij}(u) \xrightarrow{P} 0$ for $t \in [0, \tau]$, it follows that $\alpha_{3:n}(t) \xrightarrow{P} 0$. Combining results for $\alpha_{1:n}(t)$, $\alpha_{2:n}(t)$ and $\alpha_{3:n}(t)$, it follows that $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) \xrightarrow{P} \Lambda_0(t)$.

With respect to convergence to a Gaussian process, note that, by the consistency of $\widehat{\boldsymbol{\beta}}$, p_* and Lemma 1 in the Appendix in Lin et al. (2000), $-\int_0^t R^{(0)}(\widehat{\boldsymbol{\beta}}, p_*, u) \times \left\{\mu\overline{S}^{(0)}(\widehat{\boldsymbol{\beta}}, p_*, u)^2\right\}^{-1} d\overline{N}(u) \xrightarrow{P} k(\boldsymbol{\beta}_0, p_0, t)$, where $k(\boldsymbol{\beta}, p, t) = -\int_0^t \mu r^{(0)}(\boldsymbol{\beta}, p, u) / s^{(0)}(\boldsymbol{\beta}, u) d\Lambda_0(u)$. It then follows that

$$\begin{aligned}
n^{1/2}\alpha_{1:n} &= k(\boldsymbol{\beta}_0, p_0, t)\, n^{1/2}(\widehat{p} - p_0) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n k(\boldsymbol{\beta}_0, p_0, t) Q_i(p_0) + o_p(1).
\end{aligned}$$

Similarly, $-\int_0^t \overline{\boldsymbol{E}}(\boldsymbol{\beta}_*, p_0, u) \left\{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_*, p_0, u)\right\}^{-1} d\overline{N}(u) \xrightarrow{P} \boldsymbol{h}(\boldsymbol{\beta}_0, p_0, t)$, where

$\boldsymbol{h}(\boldsymbol{\beta}_0, p_0, t) = -\int_0^t \boldsymbol{e}(\boldsymbol{\beta}_0, u)/s^{(0)}(\boldsymbol{\beta}_0, u)dF(u)$. We then have that

$$\begin{aligned} n^{1/2}\alpha_{2:n} &= \boldsymbol{h}^T(\boldsymbol{\beta}_0, p_0, t)\; n^{1/2}(\widehat{\boldsymbol{\beta}} - \beta_0) + o_p(1) \\ &= \boldsymbol{h}^T(\boldsymbol{\beta}_0, p_0, t)\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_*, \widehat{p})^{-1}\; n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, \widehat{p}) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n} \boldsymbol{h}^T(\boldsymbol{\beta}_0, p_0, t)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\psi}_i(\boldsymbol{\beta}_0, p_0) + o_p(1). \end{aligned}$$

Considering $n^{1/2}\alpha_{3:n}$,

$$\begin{aligned} n^{1/2}\alpha_{3:n} &= n^{1/2}\int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} dF(u) \\ &\quad + n^{1/2}\int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\overline{M}(u) + o_p(1) \\ &= n^{1/2}\int_0^t \left\{ \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \right\} dF(u) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^t \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)} dM_{ij}(u) + o_p(1). \end{aligned}$$

Applying Taylor expansions of $\left\{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)\right\}^{-1}$ and $S^{(0)}(\boldsymbol{\beta}_0, u)^{-1}$,

$$\begin{aligned} & \frac{1}{\mu\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} \\ = & \left\{ \frac{1}{\mu\mu^{-1}s^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{\overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u) - \mu^{-1}s^{(0)}(\boldsymbol{\beta}_0, u)}{\mu(\mu^{-1}s^{(0)}(\boldsymbol{\beta}_0, u))^2} \right\} \\ & - \left\{ \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{S^{(0)}(\boldsymbol{\beta}_0, u) - s^{(0)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)^2} \right\} + o_p(1) \\ = & \frac{\mu^{-1}S^{(0)}(\boldsymbol{\beta}_0, u) - \overline{S}^{(0)}(\boldsymbol{\beta}_0, p_0, u)}{\mu^{-1}s^{(0)}(\boldsymbol{\beta}_0, u)^2} + o_p(1) \\ = & \frac{1}{\mu^{-1}s^{(0)}(\boldsymbol{\beta}_0, u)^2} \times n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \left\{ \frac{1}{\mu} - \frac{p_0}{N_1/n}\delta_{ij} - \frac{1-p_0}{n_0/n}(1-\delta_{ij})H_iH_{ij} \right\} \\ & \qquad \times Y_{ij}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ij(u)}} + o_p(1) \\ = & \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)^2} \times n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \left\{ 1 - \delta_{ij} - \frac{1}{\gamma\theta}(1-\delta_{ij})H_iH_{ij} \right\} Y_{ij}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ij(u)}} + o_p(1). \end{aligned}$$

It then follows that

$$n^{1/2}\alpha_{3:n} = n^{-1/2}\sum_{i=1}^{n}\vartheta_i(\boldsymbol{\beta}_0, p_0, t) + o_p(1),$$

where

$$\begin{aligned}\vartheta_i(\boldsymbol{\beta}_0, p_0, t) &= \sum_{j=1}^{m_i} \int_0^t \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)} dM_{ij}(u) \\ &\quad + \sum_{j=1}^{m_i} \int_0^t \frac{1}{s^{(0)}(\boldsymbol{\beta}_0, u)^2} \left\{ 1 - \delta_{ij} - \frac{1}{\gamma\theta}(1-\delta_{ij}) H_i H_{ij} \right\} Y_{ij} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(u)} \\ &\quad \times dF(u).\end{aligned}$$

Combining the above results, one obtains $n^{1/2}\left\{\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \Lambda_0(t)\right\} = n^{-1/2} \times$ $\sum_{i=1}^n \phi_i(\boldsymbol{\beta}_0, p_0, t) + o_p(1)$, where $\phi_i(\boldsymbol{\beta}, p, t) = k(\boldsymbol{\beta}, p, t) Q_i(p) + \boldsymbol{h}^T(\boldsymbol{\beta}, p, t) \boldsymbol{A}(\boldsymbol{\beta})\, \boldsymbol{\psi}_i(\boldsymbol{\beta}, p)$ $+ \vartheta_i(\boldsymbol{\beta}, p, t)$. It then follows from the MCLT that $n^{1/2}\left\{\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \Lambda_0(t)\right\}$ converges to a multivariate normal with mean zero and covariance function at (s, t) given by $\mathcal{E}\left\{\phi_1(\boldsymbol{\beta}_0, p_0, s)\phi_1(\boldsymbol{\beta}_0, p_0, t)\right\}$. Using similar arguments to Spiekerman et al. (1998), tightness can be verified. Therefore, by the Functional Central Limit Theorem (Pollard (1990)), $n^{1/2}\left\{\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t) - \Lambda_0(t)\right\}$ converges to a Gaussian process with mean zero and covariance function at (s, t) given by $\mathcal{E}\left\{\phi_1(\boldsymbol{\beta}_0, p_0, s)\phi_1(\boldsymbol{\beta}_0, p_0, t)\right\}$.

A.6 Derivation of Equation (3)

Let $\widetilde{Z}(u) = \{Z(s) : 0 < s \leq u\}$. Equation (3) of the chapter can be derived as follows:

$$\begin{aligned}&\mathcal{E}\left\{ Z(u) \middle| X = u, \delta = 1 \right\} \\ =\ & \int z(u)\, \frac{f_{X=u,\delta=1|\widetilde{z}(u)}}{\int f_{X=u,\delta=1|\widetilde{z}(u)} dF_{\widetilde{z}(u)}} dF_{\widetilde{z}(u)} \\ =\ & \frac{\int z(u)\, \lambda_0(u) e^{\beta^T \widetilde{z}(u)} P(T \geq u | \widetilde{z}(u)) P(C \geq u | \widetilde{z}(u)) dF_{\widetilde{z}(u)}}{\int \lambda_0(u) e^{\beta^T \widetilde{z}(u)} P(T \geq u | \widetilde{z}(u)) P(C \geq u | \widetilde{z}(u)) dF_{\widetilde{z}(u)}} \\ =\ & \frac{\int z(u)\, e^{\beta^T \widetilde{z}(u)} \mathcal{E}\left\{ Y(u) | \widetilde{z}(u) \right\} dF_{\widetilde{z}(u)}}{\int e^{\beta^T \widetilde{z}(u)} \mathcal{E}\left\{ Y(u) | \widetilde{z}(u) \right\} dF_{\widetilde{z}(u)}} \\ =\ & \frac{\mathcal{E}\left\{ Y(u) Z(u) e^{\beta^T Z(u)} \right\}}{\mathcal{E}\left\{ Y(u) e^{\beta^T Z(u)} \right\}}.\end{aligned}$$

A.7 Extension of proposed methods to a stratified model

Let V_{ij} denote the stratum for subject (i, j) and set $V_{ijk} = I\left\{V_{ij} = k\right\}$, $k = 1, \ldots, K$,

where there are K mutually exclusive strata. If subject (i,j) is in the kth stratum, the marginal hazard of failure is specified as

$$\lambda_{ij}(t|\, V_{ijk}=1) = \lambda_{0k}(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ij}(t)}, \tag{A.7}$$

where $\lambda_{0k}(\cdot)$ is an unspecified stratum-specific baseline hazard function. Under model (A.7), let $p_{0k} = Pr(\delta_{ij}=1|\, V_{ij}=k)$ for $k=1,\ldots,K$ and set $\mathbf{p}_0 = (p_{01},\ldots,p_{0K})^T$. Let $N_{1k} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} V_{ijk}\delta_{ij}$ be the total number of failures in stratum k in the full cohort and let $n_{0k} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} V_{ijk}(1-\delta_{ij})H_iH_{ij}$ be the total number of non-failures in stratum k in the subcohort. We assume that $\{N_{ij}(\cdot), Y_{ij}(\cdot), \boldsymbol{Z}_{ij}(\cdot), V_{ij}, m_i : j=1,\ldots,m_i\}, i=1,\ldots,n$ are independent and identically distributed, and for each k, let $m_{ik} = \sum_{j=1}^{m_i} V_{ijk}$, and $\mathcal{E}[m_{ik}] = \mu_k$.

The parameter $\boldsymbol{\beta}_0$ can be estimated by $\widetilde{\boldsymbol{\beta}}$, the solution to $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta},\mathbf{p})=0$, where

$$\widetilde{\boldsymbol{U}}(\boldsymbol{\beta},\mathbf{p}) = \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_0^{\tau} V_{ijk}\left\{\boldsymbol{Z}_{ij}(u) - \widetilde{\boldsymbol{E}}_k(\boldsymbol{\beta},p_k,u)\right\} dN_{ij}(u)$$

with $\widetilde{\boldsymbol{E}}_k(\boldsymbol{\beta},p_k,u) = \widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},p_k,u)/\widetilde{S}_k^{(0)}(\boldsymbol{\beta},p_k,u)$, $\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},p_k,u) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} V_{ijk}\times \left[N_{1k}^{-1}p_k\delta_{ij} + n_{0k}^{-1}(1-p_k)(1-\delta_{ij})H_iH_{ij}\right] Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}$ for $d=0,1,2$. We can estimate p_{0k} by the subcohort case percentage in stratum k, $\widetilde{p}_{ks}$, or by the full cohort case percentage in stratum k, $\widetilde{p}_{kw}$, or p_{0k} itself if it is known. Let $\widetilde{\boldsymbol{\beta}}_s$, $\widetilde{\boldsymbol{\beta}}_w$ and $\widetilde{\boldsymbol{\beta}}_t$ be the solutions of corresponding estimating equations, respectively. The cumulative baseline hazard function, $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u)du$, can be estimated by $\widetilde{\Lambda}_{0k}(t;\widetilde{\boldsymbol{\beta}},\widetilde{p}_k)$, where

$$\widetilde{\Lambda}_{0k}(t;\boldsymbol{\beta},p_k) = \int_0^t \frac{d\overline{N}_k(u)}{\widehat{\mu}_k\widetilde{S}^{(0)}(\boldsymbol{\beta},p_k,u)},$$

with $\overline{N}_k(u) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} V_{ijk}N_{ij}(u)$.

To establish the asymptotic properties of $\widetilde{\boldsymbol{\beta}}_t$, $\widetilde{\boldsymbol{\beta}}_s$ and $\widetilde{\boldsymbol{\beta}}_w$, we need to modify Conditions (a), (e), (f) and (g) as follows:

(a') $\{N_{ij}(\cdot), Y_{ij}(\cdot), \boldsymbol{Z}_{ij}(\cdot), V_{ij}, m_i : j = 1, \ldots, m_i\}, i = 1, \ldots, n$ are independently and identically distributed.

(e') For each k, $\sup_{u\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \|\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, p_k, u) - \mu_k^{-1}\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta}, u)\| \xrightarrow{P} 0$ for $d = 0, 1, 2$, where $\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta}, u)$ is an absolutely continuous function of $\boldsymbol{\beta} \in \mathcal{B}$ and uniformly in $u \in (0, \tau]$, and $s_k^{(0)}(\boldsymbol{\beta}, u)$ is bounded away from zero, $k = 1, \ldots, K$.

(f') $\boldsymbol{I}(\boldsymbol{\beta}_0) = \sum_{k=1}^K \int_0^\tau \left[\boldsymbol{s}_k^{(2)}(\boldsymbol{\beta}_0, u)/s_k^{(0)}(\boldsymbol{\beta}_0, u) - \left\{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}_0, u)/s_k^{(0)}(\boldsymbol{\beta}_0, u)\right\}^{\otimes 2}\right] \times dF_k(u)$ is positive definite, where $F_k(u) = \mathcal{E}\left\{\overline{N}_k(u)\right\}$.

(g') $\Lambda_{0k}(\tau) < \infty$ for each k, and $\lambda_{0k}(t)$ is absolutely continuous for $t \in (0, \tau]$.

Conditions (b'), (c'), and (d') are the same as (b), (c), and (d) respectively.

Theorem A.7.1: *Under conditions* $(a') - (g')$, $\widetilde{\boldsymbol{\beta}}_t$ *converges in probability to* $\boldsymbol{\beta}_0$, *and* $n^{1/2}(\widetilde{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0)$ *converges in distribution to a mean zero Normal with covariance matrix* $\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_t(\boldsymbol{\beta}_0, \boldsymbol{p}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$, *where* $\boldsymbol{\Omega}_t(\boldsymbol{\beta}_0, \boldsymbol{p}_0) = \mathcal{E}\{\boldsymbol{W}_{1\cdot}(\boldsymbol{\beta}_0, \boldsymbol{p}_0)^{\otimes 2}\}$, $\boldsymbol{W}_{i\cdot}(\boldsymbol{\beta}, \boldsymbol{p}) = \sum_{k=1}^K \boldsymbol{W}_{ik}(\boldsymbol{\beta}, p_k)$, *and* $\boldsymbol{W}_{ik}(\boldsymbol{\beta}, p_k)$ *is the same as* $\boldsymbol{W}_i(\boldsymbol{\beta}, \boldsymbol{p})$ *except that* $\boldsymbol{W}_{ik}(\boldsymbol{\beta}, p_k)$ *is calculated within stratum* k.

Theorem A.7.2: *Under conditions* $(a')-(g')$, *both* $\widetilde{\boldsymbol{\beta}}_s$ *and* $\widetilde{\boldsymbol{\beta}}_w$ *converge in probability to* $\boldsymbol{\beta}_0$, *and each of* $n^{1/2}(\widetilde{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_0)$ *and* $n^{1/2}(\widetilde{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0)$ *is asymptotically a zero-mean Normal with covariance matrix* $\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_s(\boldsymbol{\beta}_0, \boldsymbol{p}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$ *and* $\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Omega}_w(\boldsymbol{\beta}_0, \boldsymbol{p}_0)\times \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$, *respectively, where for* $a = s$ *and* w, $\boldsymbol{\Omega}_a(\boldsymbol{\beta}_0, \boldsymbol{p}_0) = \mathcal{E}\{\boldsymbol{\varphi}_1^a(\boldsymbol{\beta}_0, \boldsymbol{p}_0)\}$, $\boldsymbol{\varphi}_i^a(\boldsymbol{\beta}, \boldsymbol{p}) = \boldsymbol{W}_{i\cdot}(\boldsymbol{\beta}, \boldsymbol{p}) + \sum_{k=1}^K \{\boldsymbol{B}_k(\boldsymbol{\beta})Q_{ik}^a(\boldsymbol{p})\}$, $Q_{ik}^s(\boldsymbol{p}) = (\mu_k\gamma\theta)^{-1} \sum_{j=1}^{m_i} V_{ijk}H_iH_{ij}(\delta_{ij} - p_k)$, $Q_{ik}^w(\boldsymbol{p}) = \mu_k^{-1}\sum_{j=1}^{m_i} V_{ijk}(\delta_{ij} - p_k)$, $\boldsymbol{B}_k(\boldsymbol{\beta}) = \int_0^\tau \left\{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}, u) \times r_k^{(0)}(\boldsymbol{\beta}, u)/s_k^{(0)}(\boldsymbol{\beta}, u)^2 - \boldsymbol{r}_k^{(1)}(\boldsymbol{\beta}, u)/s_k^{(0)}(\boldsymbol{\beta}, u)\right\} dF_k(u)$, *with*

$$\begin{aligned}\boldsymbol{r}_k^{(d)}(\boldsymbol{\beta}, u) &= p_k^{-1}\mathcal{E}\left\{\delta_{ij}Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}\middle| V_{ijk} = 1\right\} \\ &\quad - (1 - p_k)^{-1}\mathcal{E}\left\{(1 - \delta_{ij})Y_{ij}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}(u)}\boldsymbol{Z}_{ij}(u)^{\otimes d}\middle| V_{ijk} = 1\right\}.\end{aligned}$$

Table A.1: Simulation results to evaluate the estimate of β_0 with a continuous covariate based on 1000 replications.

Design & Method		$\beta_0 = \log(0.5)$				
		Bias	ESD	ASE	ARE	CP
FC		-0.002	0.033	0.034	1.000	0.947
A	SC	-0.002	0.042	0.044	0.597	0.955
	WC	-0.003	0.041	0.042	0.655	0.943
	T	-0.002	0.037	0.038	0.801	0.935
	LS	-0.005	0.057	0.057	0.356	0.957
B	SC	-0.003	0.049	0.048	0.502	0.940
	WC	-0.004	0.046	0.045	0.571	0.923
	T	-0.004	0.044	0.041	0.688	0.916
	LS	-0.006	0.066	0.063	0.291	0.939
C	SC	-0.003	0.046	0.046	0.546	0.944
	WC	-0.004	0.043	0.043	0.625	0.939
	T	-0.003	0.040	0.040	0.723	0.937
	LS	-0.008	0.061	0.060	0.321	0.944

Estimate of β_0 from 5 methods with a continuous covariate: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.8, λ_0=1, censoring time C=1, β=log(0.5), Z follows a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 800$.

The proofs of Theorems A.7.1 and A.7.2 are very similar to those of Theorems 2 and 3, respectively. The asymptotic properties of $\widetilde{\Lambda}_{0k}(\widetilde{\boldsymbol{\beta}}, \widetilde{p}_k, t)$ and the derivations thereof are analogous to those of $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, \widehat{p}, t)$.

A.8 Additional simulation results

Table A.1 gives some results with a continuous covariate. The proposed methods appear to perform well.

In the chapter, we considered $\alpha = 0.8$, which corresponds to Kendall's τ of 0.2 for weak intracluster association. Here we conducted some simulation studies with $\alpha = 0.5$, which leads to Kendall's τ of 0.5 for fairly strong intracluster association. (Table A.2). The proposed methods still perform well, at least in the examples we considered.

We conducted some simulation studies with smaller number of clusters, smaller number of subjects within clusters, and smaller subcohort size. Table A.3 summa-

Table A.2: Simulation results with $\alpha = 0.5$ based on 1000 replications.

Design & Method		$Z \sim Bernoulli(0.5)$					$Z \sim N(0,1)$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		-0.004	0.068	0.070	1.000	0.949	-0.003	0.049	0.049	1.000	0.949
A	SC	-0.003	0.094	0.093	0.567	0.946	-0.002	0.057	0.057	0.739	0.947
	WC	-0.004	0.093	0.092	0.579	0.950	-0.003	0.056	0.055	0.794	0.944
	T	-0.006	0.090	0.089	0.619	0.947	-0.003	0.053	0.051	0.923	0.939
	LS	-0.005	0.099	0.099	0.500	0.948	-0.004	0.067	0.067	0.535	0.954
B	SC	-0.001	0.102	0.102	0.471	0.937	-0.007	0.066	0.066	0.551	0.941
	WC	-0.006	0.109	0.108	0.420	0.928	-0.009	0.063	0.064	0.586	0.936
	T	-0.008	0.107	0.105	0.444	0.923	-0.010	0.061	0.059	0.690	0.924
	LS	-0.007	0.106	0.107	0.428	0.954	-0.013	0.085	0.083	0.349	0.935
C	SC	-0.005	0.098	0.097	0.521	0.929	-0.002	0.061	0.060	0.667	0.937
	WC	-0.008	0.100	0.099	0.500	0.942	-0.003	0.060	0.059	0.690	0.936
	T	-0.011	0.098	0.095	0.543	0.925	-0.003	0.056	0.055	0.794	0.930
	LS	-0.007	0.104	0.103	0.462	0.946	-0.010	0.076	0.074	0.438	0.943

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.5, λ_0=1, censoring time C=1, β=log(0.5), Z follows either a Bernoulli(0.5) distribution or a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 800$.

rized these results. This illustrates that the proposed method generally works well, though there is some slight under-coverage for Designs B and C, which is reduced as the number of clusters increases.

We also did some simulation studies with smaller marginal event rate of $p_0 = 0.03$ (Table A.4). The results display that even when the event rate is small, $\widehat{\beta}_s$ still performs well.

Next, we examined the stratified method proposed in Section (A.7). As shown in Table A.5, the proposed stratified method appears to perform well with a reasonable small number of strata.

The results in Table A.6 show that the efficiency gain of the proposed method over that of Lu and Shih (2006) is more obvious when the covariate is cluster-specific.

Next, the performance of an inverse sampling probability weighted (ISPW) estimator and the proposed estimator were compared through simulation (Table A.7). The ISPW method used the true sampling probability, while the methods proposed

Table A.3: Simulation results to evaluate the performance of the proposed method with a smaller number of clusters and a smaller cluster size based on 1000 replications.

Design & Method		$Z \sim Bernoulli(0.5)$					$Z \sim N(0,1)$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		-0.005	0.099	0.098	1.000	0.938	-0.006	0.055	0.056	1.000	0.950
A	SC	-0.009	0.161	0.159	0.380	0.943	-0.006	0.077	0.079	0.502	0.951
	WC	-0.009	0.158	0.158	0.385	0.940	-0.007	0.074	0.075	0.558	0.947
	T	-0.010	0.156	0.155	0.400	0.938	-0.006	0.070	0.070	0.640	0.933
	LS	-0.013	0.175	0.173	0.321	0.940	-0.015	0.107	0.108	0.269	0.947
B	SC	-0.003	0.166	0.155	0.400	0.923	-0.009	0.084	0.081	0.478	0.921
	WC	-0.006	0.172	0.156	0.395	0.907	-0.012	0.082	0.076	0.543	0.912
	T	-0.007	0.169	0.153	0.410	0.900	-0.011	0.077	0.071	0.622	0.899
	LS	-0.008	0.185	0.172	0.325	0.929	-0.028	0.127	0.111	0.255	0.910
C	SC	-0.014	0.166	0.156	0.395	0.920	-0.005	0.086	0.079	0.502	0.924
	WC	-0.014	0.168	0.156	0.395	0.913	-0.008	0.079	0.075	0.558	0.929
	T	-0.015	0.165	0.153	0.410	0.911	-0.007	0.075	0.070	0.640	0.918
	LS	-0.013	0.185	0.170	0.332	0.909	-0.021	0.120	0.109	0.264	0.924

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
50 clusters, m_i follows a Bin(25,0.8) distribution, α=0.8, λ_0=1, censoring time C=1, β=log(0.5), Z follows either a Bernoulli(0.5) distribution or a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 200$.

Table A.4: Simulation results with $p_0 = 0.03$ based on 1000 replications.

Design & Method		$Z \sim Bernoulli(0.5)$					$Z \sim N(0,1)$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		-0.041	0.242	0.214	1.000	0.909	-0.017	0.139	0.116	1.000	0.867
A	SC	-0.043	0.248	0.224	0.913	0.903	-0.023	0.153	0.128	0.821	0.883
	WC	-0.042	0.248	0.224	0.913	0.902	-0.023	0.153	0.128	0.821	0.885
	T	-0.042	0.247	0.223	0.921	0.903	-0.022	0.149	0.125	0.861	0.884
	LS	-0.042	0.249	0.224	0.913	0.901	-0.024	0.153	0.128	0.821	0.885
B	SC	-0.041	0.248	0.223	0.921	0.919	-0.021	0.150	0.126	0.848	0.885
	WC	-0.041	0.249	0.224	0.913	0.920	-0.021	0.150	0.127	0.834	0.889
	T	-0.041	0.247	0.223	0.921	0.920	-0.021	0.147	0.125	0.861	0.891
	LS	-0.042	0.249	0.224	0.913	0.919	-0.022	0.151	0.128	0.821	0.888
C	SC	-0.042	0.250	0.224	0.913	0.920	-0.023	0.150	0.127	0.834	0.886
	WC	-0.042	0.250	0.224	0.913	0.919	-0.024	0.150	0.127	0.834	0.886
	T	-0.042	0.248	0.223	0.921	0.920	-0.023	0.147	0.125	0.861	0.886
	LS	-0.042	0.250	0.224	0.913	0.919	-0.024	0.150	0.128	0.821	0.890

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.8, censoring time C=1, β=log(0.5), λ_0=0.04 when Z follows a Bernoulli(0.5) distribution, λ_0=0.0025 when Z follows a N(0,1) distribution, the marginal event rate is $p_0 = 0.03$. The number of individuals in the subcohort is $n_s = 800$.

Table A.5: Simulation results to evaluate the performance of the proposed stratified methods based on 1000 replications.

Design & Method		$Z \sim Bernoulli(0.5)$					$Z \sim N(0,1)$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		-0.014	0.110	0.104	1.000	0.929	-0.004	0.055	0.054	1.000	0.932
A	SC	-0.011	0.133	0.124	0.703	0.924	-0.006	0.068	0.068	0.631	0.944
	WC	-0.012	0.134	0.124	0.703	0.922	-0.006	0.068	0.067	0.650	0.940
	T	-0.011	0.132	0.122	0.727	0.926	-0.005	0.063	0.063	0.735	0.944
	LS	-0.011	0.135	0.125	0.692	0.926	-0.008	0.074	0.072	0.563	0.937
B	SC	-0.016	0.133	0.124	0.703	0.925	-0.006	0.069	0.069	0.612	0.943
	WC	-0.016	0.134	0.125	0.692	0.927	-0.007	0.069	0.069	0.612	0.938
	T	-0.016	0.132	0.123	0.715	0.929	-0.006	0.065	0.065	0.690	0.941
	LS	-0.016	0.134	0.125	0.692	0.930	-0.007	0.073	0.074	0.533	0.948
C	SC	-0.013	0.131	0.124	0.703	0.926	-0.004	0.071	0.068	0.631	0.936
	WC	-0.013	0.131	0.125	0.692	0.929	-0.003	0.070	0.068	0.631	0.930
	T	-0.013	0.129	0.123	0.715	0.930	-0.002	0.065	0.064	0.712	0.932
	LS	-0.014	0.131	0.126	0.681	0.927	-0.004	0.077	0.072	0.563	0.930

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.8, censoring time C=1, β=log(0.5), Z follows either a Bernoulli(0.5) distribution or a N(0,1) distribution, $Z_2 \sim U(0,1)$, stratum k=1,2 or 3 if $Z_2 \leq 0.33$,$0.33 < Z_2 \leq 0.67$ or $Z_2 > 0.67$, respectively. λ_{0k}=0.1×k, for k=1,2,3. The number of individuals in the subcohort is $n_s = 800$.

Table A.6: Simulation results with a cluster-level covariate based on 1000 replications.

Design & Method		$\beta_0 = \log(0.5)$					$\beta_0 = 0$				
		Bias	ESD	ASE	ARE	CP	Bias	ESD	ASE	ARE	CP
FC		0.000	0.145	0.147	1.000	0.939	0.002	0.126	0.127	1.000	0.940
A	SC	0.002	0.160	0.159	0.855	0.944	0.003	0.141	0.141	0.811	0.951
	WC	0.002	0.160	0.159	0.855	0.946	0.003	0.141	0.141	0.811	0.952
	T	0.001	0.159	0.157	0.877	0.948	0.003	0.141	0.141	0.811	0.952
	LS	0.000	0.163	0.163	0.813	0.949	0.002	0.147	0.146	0.757	0.948
B	SC	0.008	0.365	0.350	0.176	0.947	0.005	0.315	0.302	0.177	0.944
	WC	0.013	0.353	0.334	0.194	0.934	0.005	0.315	0.300	0.179	0.929
	T	0.013	0.353	0.333	0.195	0.934	0.005	0.315	0.300	0.179	0.930
	LS	0.015	0.483	0.465	0.100	0.962	0.019	0.483	0.463	0.075	0.962
C	SC	-0.003	0.253	0.247	0.354	0.936	-0.001	0.222	0.215	0.349	0.937
	WC	0.001	0.246	0.240	0.375	0.934	-0.001	0.222	0.215	0.349	0.938
	T	0.000	0.246	0.238	0.381	0.935	-0.001	0.222	0.215	0.349	0.939
	LS	-0.004	0.313	0.306	0.231	0.944	-0.002	0.306	0.299	0.180	0.949

Estimate of β_0 from 5 methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; LS = Lu and Shih (2006) estimator.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.8, λ_0=1, censoring time C=1, β=log(0.5), Z follows a Bernoulli(0.5) distribution. The number of individuals in the subcohort is $n_s = 800$.

Table A.7: Simulation results to compare the proposed methods with the ISPW and SRS methods based on 1000 replications.

		SC		WC		T		
Design	FC	BER	SRS	BER	SRS	BER	SRS	ISPW
$Z \sim Ber(0.5)$								
A	0.053	0.083	0.083	0.082	0.084	0.080	0.082	0.083
B	0.053	0.083	0.083	0.086	0.086	0.084	0.084	0.088
C	0.053	0.084	0.083	0.083	0.084	0.082	0.082	0.085
$Z \sim N(0,1)$								
A	0.033	0.042	0.042	0.041	0.041	0.037	0.038	0.042
B	0.033	0.049	0.049	0.046	0.045	0.044	0.042	0.061
C	0.033	0.046	0.046	0.043	0.043	0.040	0.039	0.050

Empirical standard deviation of the estimate of β_0 from following methods: Method FC = full cohort analysis; SC = estimating p_0 using the subcohort, $\widehat{p}_s$; WC = estimating p_0 using whole cohort, $\widehat{p}_w$; T = using true value, p_0; ISPW = inverse sampling probability method; BER = Bernoulli sampling; SRS = simple random sampling.
100 clusters, m_i follows a Bin(50,0.8) distribution, α=0.8, λ_0=1, censoring time C=1, β=log(0.5), Z follows either a Bernoulli(0.5) distribution or a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 800$.

here used estimates of the sampling probability. The results show that the ESD of the ISPW method is generally comparable to that of our proposed method.

In addition, the point estimates based on simple random samples (SRS) for some non-rare event settings are provided (Table A.7). This investigation showed that the ESDs of the point estimates based on SRS are very close to those based on Bernoulli sampling. Therefore, one does not gain much efficiency by using SRS, at least for the examples we considered.

APPENDIX B

Proof of Theorems in Chapter III

We provide a proof of Theorem 2 and an examination of the corresponding variance estimator. We begin with a review of the notation.

B.0 Notation

$i =$ subject $(i = 1, \ldots, n)$

$T_i =$ failure time

$C_{1i} =$ independent censoring time

$C_{2i} =$ dependent censoring time

$C_i = C_{1i} \wedge C_{2i}$

$X_i = T_i \wedge C_{1i} \wedge C_{2i}$

$Y_i(t) = I\{X_i \geq t\}$

$\Delta_{1i} = I(T_i \leq C_i)$

$\Delta_{2i} = I\{C_{2i} \leq C_{1i}, C_{2i} < T_i)\}$

$\Delta_{3i} = (1 - \Delta_{1i})(1 - \Delta_{2i})$

$N_i(t) = I\{X_i \leq t, \Delta_{1i} = 1\}$

$N_i^C(t) = I\{X_i \leq t, \Delta_{2i} = 1\}$

Z_{1i} = time-constant covariate vector

$Z_{2i}(t)$ = time-dependent covariate vector

$Z_i(t) = \left\{Z_{1i}^T, Z_{2i}^T(t)\right\}^T$

$\overline{Z}_i(t) = \{Z_i(u) : 0 \leq u \leq t\}$

$V_i(t) = \{V_{1i}(t), \ldots, V_{qi}(t)\}$ = functions of $Z_i(t)$

$\lambda_i(t) = \lambda_0(t)e^{\beta^T Z_i(0)}$

$\lambda_i^C(t) = \lambda_0^C(t)e^{\alpha^T V_i(t)}$

$\xi_i = I$ (individual i is selected for the subcohort)

$p_k = Pr(\xi_i = 1 \mid \Delta_{ki} = 1)$, $k = 1, 2, 3$

$\rho_i(p) = \sum_{k=1}^3 \Delta_{ki}\xi_i/p_k$

$dM_i(t) = dN_i(t) - Y_i(t)e^{\beta^T Z_i(0)}\lambda_0(t)dt$

$dM_i^C(t) = dN_i^C(t) - Y_i(t)e^{\alpha^T V_i(t)}\lambda_0^C(t)dt$

The following is a proof of Theorem 2, for the case where $\widehat{W}_{1i}$ is used in the proposed estimators. The proofs for stabilized weights, $\widehat{W}_{2i}$ and $\widehat{W}_{3i}$, proceed through steps analogous to those listed below.

B.1 $n^{-\frac{1}{2}}U^C(\alpha_0, p_0)$

The estimating function for the dependent censoring model is

$$U^C(\alpha_0, p_0) = \sum_{i=1}^n \int_0^\tau \left\{V_i(t) - \overline{V}(\alpha_0, p_0, t)\right\} \rho_i(p_0) dN_i^C(t),$$

where

$$\begin{aligned}\overline{V}(\alpha,p,t) &= \frac{S_C^{(1)}(\alpha,p,t)}{S_C^{(0)}(\alpha,p,t)} \\ S_C^{(d)}(\alpha,p,t) &= n^{-1}\sum_{i=1}^{n}\rho_i(p)Y_i(t)V_i(t)^{\otimes d}e^{\alpha^T V_i(t)}.\end{aligned}$$

Let $s_C^{(d)}(t,\alpha) = \mathcal{E}\{Y_1(t)V_1(t)^{\otimes d}e^{\alpha^T V_1(t)}\}$ and let $\overline{v}(\alpha,t) = s_C^{(1)}(t,\alpha)/s_C^{(0)}(t,\alpha)$. We define $dM_i^C(t) = dN_i^C(t) - Y_i(t)e^{\alpha^T V_i(t)}\lambda_0^C(t)dt$. By van der Vaart & Wellner (1996, Example 2.11.16), $H_n(t) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\rho_i(p_0)M_i^C(t)$ converges weakly to a tight Gaussian process $H(t)$ with continuous sample paths on $[0,\tau]$.

By some simple algebra, we have

$$\begin{aligned}n^{-\frac{1}{2}}U^C(\alpha_0,p_0) &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\{V_i(t)-\overline{V}(\alpha_0,p_0,t)\}\rho_i(p_0)dM_i^C(t) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\{V_i(t)-\overline{v}(\alpha_0,t)\}\rho_i(p_0)dM_i^C(t) \\ &\quad -n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\left\{\overline{V}(\alpha_0,p_0,t)-\overline{v}(\alpha_0,t)\right\}\rho_i(p_0)dM_i^C(t) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^\tau\left\{V_i(t)-\overline{v}(\alpha_0,t)\right\}\rho_i(p_0)dM_i^C(t)+o_p(1) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}K_i(\alpha_0,p_0)+o_p(1),\end{aligned}$$

with $K_i(\alpha,p) = \int_0^\tau\{V_i(t)-\overline{v}(\alpha,t)\}\rho_i(p)dM_i^C(t)$. Note that $\mathcal{E}\{\rho_i(p_0)dM_i^C(t)\} = 0$, such that $\mathcal{E}\{K_i(\alpha_0,p_0)\} = 0$, for $i = 1,\cdots,n$.

B.2 $n^{1/2}(\widehat{\alpha}-\alpha_0)$

Using a Taylor expansion, we can show that

$$n^{-1/2}U^C(\alpha_0,\widehat{p}) = n^{-1/2}U^C(\alpha_0,p_0) + B_n^C(\alpha_0,p_*)n^{1/2}\left(\widehat{p}-p_0\right),$$

where p_* is on the line segment between $\widehat{p}$ and p_0, and

$$\begin{aligned} B_n^C(\alpha,p) &= n^{-1}\frac{\partial}{\partial p}U^C(\alpha,p) \\ &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{ \frac{S_C^{(1)}(\alpha,p,t)}{S_C^{(0)}(\alpha,p,t)^2}\frac{\partial}{\partial p}S_C^{(0)}(\alpha,p,t) \right. \\ &\quad \left. -\frac{1}{S_C^{(0)}(\alpha,p,t)}\frac{\partial}{\partial p}S_C^{(1)}(\alpha,p,t) \right\} \rho_i(p)dN_i^C(t) \\ &\quad +n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{V_i(t)-\overline{V}(\alpha,p,t)\right\}\frac{\partial}{\partial p}\rho_i(p)dN_i^C(t). \end{aligned}$$

We define $R_k^{(d)}(\alpha,p,t)$, $d=0,1$ and $D_k(\alpha,p)$ as follows,

$$\begin{aligned} R_k^{(d)}(\alpha,p,t) &= \frac{\partial}{\partial p_k}S_C^{(d)}(\alpha,p,t) \\ &= n^{-1}\sum_{i=1}^{n}-\frac{\Delta_{ki}\xi_i}{p_k^2}Y_i(t)V_i(t)^{\otimes d}e^{\alpha^T V_i(t)} \end{aligned}$$

such that

$$\begin{aligned} R_k^{(d)}(\alpha,p,t) &\longrightarrow -\frac{1}{p_k}\mathcal{E}\{\Delta_{k1}Y_1(t)V_1(t)^{\otimes d}e^{\alpha^T V_1(t)}\} \\ &\equiv r_k^{(d)}(\alpha,p,t) \end{aligned}$$

in probability. We then have

$$\begin{aligned} D_k(\alpha,p) &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \{V_i(t)-\overline{V}(\alpha,p,t)\}\frac{\partial}{\partial p_k}\rho_i(p)dN_i^C(t) \\ &= -n^{-1}\sum_{i=1}^{n}\int_0^\tau \{V_i(t)-\overline{V}(\alpha,p,t)\}\frac{\Delta_{ki}\xi_i}{p_k^2}dN_i^C(t) \\ &= -n^{-1}\sum_{i=1}^{n}\int_0^\tau \{V_i(t)-\overline{v}(\alpha,t)\}\frac{\Delta_{ki}\xi_i}{p_k^2}dN_i^C(t) \\ &\quad +\int_0^\tau \{\overline{V}(\alpha,p,t)-\overline{v}(\alpha,t)\}n^{-1}\sum_{i=1}^{n}\frac{\Delta_{ki}\xi_i}{p_k^2}dN_i^C(t) \end{aligned}$$

such that

$$\begin{aligned} D_k(\alpha,p) &\longrightarrow -\frac{1}{p_k}\mathcal{E}[\int_0^\tau \{V_1(t)-\overline{v}(\alpha,t)\}\Delta_{k1}dN_1^C(t)] \\ &\equiv d_k(\alpha,p) \end{aligned}$$

in probability. Let $r^{(d)}(\alpha,p,t) = \left[r_1^{(d)}(\alpha,p,t) \quad r_2^{(d)}(\alpha,p,t) \quad r_3^{(d)}(\alpha,p,t)\right]$ and let $d(\alpha,p,t) = [d_1(\alpha,p,t) \quad d_2(\alpha,p,t) \quad d_3(\alpha,p,t)]$. Then by continuous mapping,

$$\begin{aligned} B_n^C(\alpha,p) &\longrightarrow \int_0^\tau \left\{ \frac{s_c^{(1)}(\alpha,p,t)}{s_c^{(0)}(\alpha,p,t)^2} r^{(0)}(\alpha,p,t) - \frac{1}{s_c^{(0)}(\alpha,p,t)} r^{(1)}(\alpha,p,t) \right\} dF^C(t) \\ &\quad + d(\alpha,p) \\ &\equiv B^C(\alpha,p) \end{aligned}$$

in probability, where $F^C(t) = \mathcal{E}\{\rho_i(p) N_i^C(t)\}$. It is easy to show that

$$\begin{aligned} n^{1/2}(\widehat{p}_k - p_{k0}) &= n^{1/2}\left(\frac{\sum_{i=1}^n \Delta_{ki}\xi_i}{\sum_{i=1}^n \Delta_{ki}} - p_{k0}\right) \\ &= n^{1/2}\frac{n^{-1}\sum_{i=1}^n \Delta_{ki}(\xi_i - p_{k0})}{n^{-1}\sum_{i=1}^n \Delta_{ki}} \\ &= n^{-1/2}\sum_{i=1}^n \eta_k^{-1}\Delta_{ki}(\xi_i - p_{k0}) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^n Q_{ki}(p_0) + o_p(1), \end{aligned}$$

where $Q_{ki}(p_0) = \eta_k^{-1}\Delta_{ki}(\xi_i - p_{k0})$, $\eta_k = \mathrm{pr}(\Delta_k = 1), k = 1,2,3$. Let $Q_i(p) = (Q_{1i}(p) \quad Q_{2i}(p) \quad Q_{3i}(p))^T$. Note that $\mathcal{E}\{Q_{ki}(p_0)\} = 0$, therefore,

$$\begin{aligned} n^{-1/2}U^C(\alpha_0,\widehat{p}) &= n^{-1/2}\sum_{i=1}^n \left\{K_i(\alpha_0,p_0) + B^C(\alpha_0,p_0)Q_i(p_0)\right\} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^n \psi_i(\alpha_0,p_0) + o_p(1), \end{aligned}$$

where $\psi_i(\alpha,p) = K_i(\alpha,p) + B^C(\alpha,p)Q_i(p)$. Since $\mathcal{E}\{\psi_i(\alpha_0,p_0)\} = 0$, by the Multivariate Central Limit Theorem (MCLT),

$$n^{-1/2}U^C(\alpha_0,\widehat{p}) \longrightarrow N(0,\Omega(\alpha_0)),$$

in distribution, where $\Omega(\alpha_0) = \mathcal{E}\{\psi_i(\alpha_0,p_0)^{\otimes 2}\}$.

We then have

$$n^{1/2}(\widehat{\alpha} - \alpha_0) = A_n^C(\alpha_*,\widehat{p})^{-1}n^{-1/2}U^C(\alpha_0,\widehat{p}),$$

where $A_n^C(\alpha,p) = n^{-1}\sum_{i=1}^n \int_0^\tau \{S_c^{(2)}(\alpha,p,t)/S_c^{(0)}(\alpha,p,t) - \overline{V}(\alpha,p,t)^{\otimes 2}\}\rho_i(p)dN_i^C(t)$ and α_* is on the line segment between $\widehat{\alpha}$ and α_0. Note the fact that $n^{-1}\sum_{i=1}^n \rho_i(p)$ $\times dN_i^C(t)$ converges to $dF^C(t)$ in probability, such that $A_n^C(\alpha_*,\widehat{p})$ converges in probability to $A^C(\alpha_0)$, with $A^C(\alpha) = \int_0^\tau \{s_c^{(2)}(\alpha,t)/s_c^{(0)}(\alpha,t) - \overline{v}(\alpha,t)^{\otimes 2}\}dF^C(t)$. Therefore, by Slutsky's Theorem, $n^{1/2}(\widehat{\alpha}-\alpha_0)$ converges in distribution to a zero-mean normal variate with covariance $A^C(\alpha_0)^{-1}\Omega(\alpha_0)A^C(\alpha_0)^{-1}$.

B.3 $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$

We can decompose $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$ as follows,

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\} \\
= \; & n^{1/2}\{\widehat{\Lambda}_0^C(t;\widehat{\alpha},\widehat{p}) - \widehat{\Lambda}_0^C(t;\widehat{\alpha},p_0)\} \qquad \text{(B.1)} \\
& + n^{1/2}\{\widehat{\Lambda}_0^C(t;\widehat{\alpha},p_0) - \widehat{\Lambda}_0^C(t;\alpha_0,p_0)\} \qquad \text{(B.2)} \\
& + n^{1/2}\{\widehat{\Lambda}_0^C(t;\alpha_0,p_0) - \Lambda_0^C(t)\}. \qquad \text{(B.3)}
\end{aligned}
$$

Applying a Taylor expansion of $\rho_i(\widehat{p})/S_C^{(0)}(\widehat{\alpha},\widehat{p},t)$ around $\rho_i(p_0)/S_C^{(0)}(\widehat{\alpha},p_0,t)$, we can write (B.1) as

$$
\begin{aligned}
& n^{1/2}\sum_{i=1}^n \int_0^t \left\{ \frac{\rho_i(\widehat{p})}{S_C^{(0)}(\widehat{\alpha},\widehat{p},u)} - \frac{\rho_i(p_0)}{S_C^{(0)}(\widehat{\alpha},p_0,u)} \right\} n^{-1} dN_i^C(u) \\
= \; & n^{-1}\sum_{i=1}^n \int_0^t \left\{ \frac{1}{S_C^{(0)}(\widehat{\alpha},p_0,u)} \frac{d\rho_i(p)}{dp} \mid_{p=p_0} - \frac{\rho_i(p_0)}{S_C^{(0)}(\widehat{\alpha},p_0,u)^2} \frac{\partial}{\partial p} S_C^{(0)}(\widehat{\alpha},p,u) \mid_{p=p_0} \right\} \\
& \times dN_i^C(u) n^{1/2}(\widehat{p}-p_0) + o_p(1) \\
= \; & \int_0^t \left\{ \frac{n^{-1}\sum_{i=1}^n \mu_i^T(p_0) dN_i^C(u)}{s_C^{(0)}(\alpha_0,u)} - \frac{r^{(0)}(\alpha_0,p_0,u) d\widehat{\Lambda}_0^C(u;\widehat{\alpha},p_0)}{s_C^{(0)}(\alpha_0,u)} \right\} \\
& \times n^{-1/2}\sum_{i=1}^n Q_i(p_0) + o_p(1) \\
= \; & L(\alpha_0,p_0,t) n^{-1/2}\sum_{i=1}^n Q_i(p_0) + o_p(1),
\end{aligned}
$$

where

$$\begin{aligned}
\mu_{ki}(p) &= \frac{d\rho_i(p)}{dp_k} = -\frac{\Delta_{ki}\xi_i}{p_k^2} \\
\mu_i(p) &= (\mu_{1i}(p) \quad \mu_{2i}(p) \quad \mu_{3i}(p))^T \\
L(\alpha_0, p_0, t) &= \int_0^t s_C^{(0)}(\alpha_0, u)^{-1} \left\{ dJ(u) - r^{(0)}(\alpha_0, p_0, u) d\Lambda_0^C(u) \right\},
\end{aligned}$$

with $dJ_n(u) = n^{-1} \sum_{i=1}^n \mu_i^T(p_0) dN_i^C(u)$, which converges to $dJ(u)$ in probability.

Considering (B.2),

$$\begin{aligned}
\text{(B.2)} &= n^{1/2} \sum_{i=1}^n \int_0^t \left\{ \frac{1}{S_C^{(0)}(\widehat{\alpha}, p_0, u)} - \frac{1}{S_C^{(0)}(\alpha_0, p_0, u)} \right\} n^{-1} \rho_i(p_0) dN_i^C(u) \\
&= n^{-1} \left\{ \sum_{i=1}^n \int_0^t -\frac{S_C^{(1)}(\alpha_0, p_0, u)}{S_C^{(0)}(\alpha_0, p_0, u)^2} \rho_i(p_0) dN_i^C(u) \right\}^T n^{1/2}(\widehat{\alpha} - \alpha_0) + o_p(1) \\
&= \left\{ -\int_0^t \overline{V}(\alpha_0, p_0, u) d\widehat{\Lambda}_0^C(u; \alpha_0, p_0) \right\}^T A^C(\alpha_0)^{-1} n^{-1/2} \sum_{i=1}^n \psi_i(\alpha_0, p_0) \\
&\quad + o_p(1) \\
&= \widehat{h}_C^T(t; \alpha_0, p_0) A^C(\alpha_0)^{-1} n^{-1/2} \sum_{i=1}^n \psi_i(\alpha_0, p_0) \\
&= h_C^T(t; \alpha_0, p_0) A^C(\alpha_0)^{-1} n^{-1/2} \sum_{i=1}^n \psi_i(\alpha_0, p_0) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{h}_C(t; \alpha_0, p_0) &= -\int_0^t \overline{V}(\alpha_0, p_0, u) d\widehat{\Lambda}_0^C(u; \alpha_0, p_0) \\
h_C(t; \alpha_0, p_0) &= -\int_0^t \overline{v}(\alpha, u) d\Lambda_0^C(u).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{(B.3)} &= n^{-1/2} \sum_{i=1}^n \int_0^t S_C^{(0)}(\alpha_0, p_0, u)^{-1} \rho_i(p_0) dM_i^C(u) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^t s_C^{(0)}(\alpha_0, u)^{-1} \rho_i(p_0) dM_i^C(u) + o_p(1).
\end{aligned}$$

Combining the above results, one obtains

$$\begin{aligned} n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\} &= n^{-1/2}\sum_{i=1}^{n}\Phi_i(\alpha_0, p_0, t) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^t d\Phi_i(\alpha_0, p_0, u) + o_p(1) \end{aligned}$$

where

$$\begin{aligned} \Phi_i(\alpha_0, p_0, t) &= L(\alpha_0, p_0, t)Q_i(p_0) + h_C^T(t; \alpha_0, p_0)A^C(\alpha_0)^{-1}\psi_i(\alpha_0, p_0) \\ &\quad + \int_0^t s_C^{(0)}(\alpha_0, u)^{-1}\rho_i(p_0)dM_i^C(u) \\ d\Phi_i(\alpha_0, p_0, u) &= s_C^{(0)}(\alpha_0, u)^{-1}\left\{dJ(u) - r^{(0)}(\alpha_0, p_0, u)d\Lambda_0^C(u)\right\}Q_i(p_0) \\ &\quad -\overline{v}^T(\alpha_0, u)d\Lambda_0^C(u)A^C(\alpha_0)^{-1}\psi_i(\alpha_0, p_0) \\ &\quad +s_C^{(0)}(\alpha_0, u)^{-1}\rho_i(p_0)dM_i^C(u). \end{aligned}$$

Note that $\mathcal{E}\{Q_i(p_0)\} = 0$, $\mathcal{E}\{\psi_i(\alpha_0, p_0)\} = 0$ and $\mathcal{E}\{\rho_i(p_0)dM_i^C(u)\} = 0$, such that $\mathcal{E}\{d\Phi_i(\alpha_0, p_0, u)\} = 0$.

B.4 $n^{1/2}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\}$

We can decompose $n^{1/2}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\}$ as follows,

$$\begin{aligned} & n^{1/2}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} \\ (\text{B.4}) \quad &= n^{1/2}\{\int_0^t e^{\widehat{\alpha}^T V_i(u)}d\widehat{\Lambda}_0^C(u) - \int_0^t e^{\alpha_0^T V_i(u)}d\widehat{\Lambda}_0^C(u)\} \\ (\text{B.5}) \quad &\quad +n^{1/2}\{\int_0^t e^{\alpha_0^T V_i(u)}d\widehat{\Lambda}_0^C(u) - \int_0^t e^{\alpha_0^T V_i(u)}d\Lambda_0^C(u)\}. \end{aligned}$$

By a Taylor expansion,

$$\begin{aligned} (\text{B.4}) &= n^{1/2}\int_0^t \left\{e^{\widehat{\alpha}^T V_i(u)} - e^{\alpha_0^T V_i(u)}\right\}d\widehat{\Lambda}_0^C(u) \\ &= \int_0^t V_i^T(u)e^{\alpha_0^T V_i(u)}d\widehat{\Lambda}_0^C(u)n^{1/2}(\widehat{\alpha} - \alpha_0) + o_p(1) \\ &= \int_0^t V_i^T(u)d\Lambda_i^C(u)A^C(\alpha_0)^{-1}n^{-1/2}\sum_{l=1}^{n}\psi_l(\alpha_0, p_0) + o_p(1). \end{aligned}$$

Now considering the second term (B.5),

$$\begin{aligned}(\text{B.5}) &= n^{1/2}\int_0^t e^{\alpha_0^T V_i(u)} d\{\widehat{\Lambda}_0^C(u) - \Lambda_0^C(u)\} \\ &= \int_0^t e^{\alpha_0^T V_i(u)} n^{-1/2} \sum_{l=1}^n d\Phi_l(\alpha_0, p_0, u) + o_p(1).\end{aligned}$$

It follows that

$$\begin{aligned}n^{1/2}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} &= \int_0^t V_i^T(u) d\Lambda_i^C(u) A^C(\alpha_0)^{-1} n^{-1/2} \sum_{l=1}^n \psi_l(\alpha_0, p_0) \\ &\quad + \int_0^t e^{\alpha_0^T V_i(u)} n^{-1/2} \sum_{l=1}^n d\Phi_l(\alpha_0, p_0, u) + o_p(1) \\ &= n^{-1/2} \sum_{l=1}^n G_l(t) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}G_l(t) &= \Psi_i^T(t) A^C(\alpha_0)^{-1} \psi_l(\alpha_0, p_0) + \int_0^t e^{\alpha_0^T V_i(u)} d\Phi_l(\alpha_0, p_0, u) \\ \Psi_i(t) &= \int_0^t V_i(u) d\Lambda_i^C(u)\end{aligned}$$

B.5 $n^{1/2}\{\widehat{R}_i(t) - R_i(t)\}$

Letting $R_i(t) = \rho_i(p) W_{1i}(t)$, we have

$$\begin{aligned}n^{1/2}\{\widehat{R}_i(t) - R_i(t)\} &= n^{1/2}\{\rho_i(\widehat{p}) e^{\widehat{\Lambda}_i^C(t)} - \rho_i(p_0) e^{\widehat{\Lambda}_i^C(t)}\} \\ &\quad + n^{1/2}\{\rho_i(p_0) e^{\widehat{\Lambda}_i^C(t)} - \rho_i(p_0) e^{\Lambda_i^C(t)}\} \\ &= \mu_i(p_0)^T e^{\Lambda_i^C(t)} n^{1/2}(\widehat{p} - p_0) \\ &\quad + \rho_i(p_0) e^{\Lambda_i^C(t)} n^{1/2}\{\widehat{\Lambda}_i^C(t) - \Lambda_i^C(t)\} + o_p(1) \\ &= \mu_i(p_0)^T W_{1i}(t) n^{-1/2} \sum_{l=1}^n Q_l(p_0) + R_i(t) n^{-1/2} \sum_{l=1}^n G_l(t) \\ &\quad + o_p(1).\end{aligned}$$

B.6 $n^{1/2}(\widehat{\beta}-\beta_0)$

It is easy to show that

$$n^{1/2}(\widehat{\beta}-\beta_0) = A_n(\beta_0)^{-1}n^{-1/2}\sum_{i=1}^{n}U_i\{\beta_0,\widehat{R}_i(t)\}+o_p(1),$$

where

$$\begin{aligned}
U_i\{\beta,R\} &= \int_0^\tau \left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}R_i(t)dN_i(t)\\
\overline{Z}(\beta,R,t) &= \frac{S^{(1)}(\beta,R,t)}{S^{(0)}(\beta,R,t)}\\
S^{(d)}(\beta,R,t) &= n^{-1}\sum_{i=1}^{n}R_i(t)Y_i(t)Z_i(0)^{\otimes d}e^{\beta^T Z_i(0)}, d=0,1,2\\
s^{(d)}(\beta,R,t) &= \mathcal{E}\left\{R_i(t)Y_i(t)Z_i(0)^{\otimes d}e^{\beta^T Z_i(0)}\right\}\\
\overline{z}(\beta,R,t) &= s^{(1)}(\beta,R,t)/s^{(0)}(\beta,R,t).
\end{aligned}$$

We then write

$$\begin{aligned}
A_n(\beta) &= n^{-1}\sum_{i=1}^{n}\int_0^\tau\left\{\frac{S^{(2)}(\beta,R,t)}{S^{(0)}(\beta,R,t)}-\overline{Z}(\beta,R,t)^{\otimes 2}\right\}R_i(t)dN_i(t)\\
&\longrightarrow \int_0^\tau\left\{\frac{s^{(2)}(\beta,R,t)}{s^{(0)}(\beta,R,t)}-\overline{z}(\beta,R,t)^{\otimes 2}\right\}dF(t)\\
&\equiv A(\beta)
\end{aligned}$$

in probability, with $n^{-1}\sum_{i=1}^{n}R_i(t)dN_i(t)$ converging in probability to $dF(t)$.

We can decompose $n^{-1/2}U(\beta,\widehat{R})$ as follows,

$$\begin{aligned}
& n^{-1/2}U\left(\beta_0,\widehat{R}\right)\\
=\ & n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\left\{Z_i(0)-\overline{Z}(\beta,\widehat{R},t)\right\}\widehat{R}_i(t)dM_i(t)\\
=\ & n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}R_i(t)dM_i(t) \qquad \text{(B.6)}\\
& +n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}\left\{\widehat{R}_i(t)-R_i(t)\right\}dM_i(t) \qquad \text{(B.7)}\\
& -n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\left\{\overline{Z}(\beta,\widehat{R},t)-\overline{Z}(\beta,R,t)\right\}\widehat{R}_i(t)dM_i(t) \qquad \text{(B.8)}
\end{aligned}$$

It is easy to show that (B.6) $= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \{Z_i(0) - \overline{z}(\beta,R,t)\}\, R_i(t)dM_i(t) + o_p(1)$. The third term (B.8) converges in probability to 0. We can express (B.7) as follows,

$$\begin{aligned}
\text{(B.7)} &= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\} \\
&\quad \left\{\mu_i(p_0)^T W_{1i}(t) n^{-1}\sum_{l=1}^{n} Q_l(p_0) + R_i(t) n^{-1}\sum_{l=1}^{n} G_l(t)\right\} dM_i(t) + o_p(1) \\
\text{(B.9)} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\}\mu_i(p_0)^T W_{1i}(t)dM_i(t) n^{-1/2}\sum_{l=1}^{n} Q_l(p_0) \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\} R_i(t) \\
\text{(B.10)} &\quad \times n^{-1}\left\{\sum_{l=1}^{n}\Psi_i^T(t)A^C(\alpha_0)^{-1}\psi_l(\alpha_0,p_0)\right\} dM_i(t) \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\} R_i(t) \\
\text{(B.11)} &\quad \times n^{-1}\sum_{l=1}^{n}\int_0^t e^{\alpha_0^T V_i(u)} d\Phi_l(\alpha_0,p_0,u) dM_i(t).
\end{aligned}$$

We can show that

$$\begin{aligned}
\text{(B.9)} &= \widehat{O}(\beta,R) n^{-1/2}\sum_{i=1}^{n} Q_i(p_0) \\
&= O(\beta,R) n^{-1/2}\sum_{i=1}^{n} Q_i(p_0) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{O}(\beta,R) &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\}\mu_i(p_0)^T W_{1i}(t)dM_i(t) \\
O(\beta,R) &= \mathcal{E}\left[\int_0^\tau \{Z_i(0) - \overline{z}(\beta,R,t)\}\mu_i(p_0)^T W_{1i}(t)dM_i(t)\right].
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{(B.10)} &= n^{-1}\left[\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0) - \overline{Z}(\beta,R,t)\right\} R_i(t)\Psi_i^T(t)dM_i(t)\right] \\
&\quad A^C(\alpha_0)^{-1} n^{-1/2}\sum_{l=1}^{n}\psi_l(\alpha_0,p_0)
\end{aligned}$$

$$
\begin{aligned}
&= \widehat{H}(\beta,R)A^C(\alpha_0)^{-1}n^{-1/2}\sum_{l=1}^{n}\psi_l(\alpha_0,p_0)\\
&= H(\beta,R)A^C(\alpha_0)^{-1}n^{-1/2}\sum_{l=1}^{n}\psi_l(\alpha_0,p_0)+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{H}(\beta,R) &= n^{-1}\left[\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}\Psi_i^T(t)R_i(t)dM_i(t)\right]\\
H(\beta,R) &= \mathcal{E}\left[\int_0^\tau \left\{Z_i(0)-\overline{z}(\beta,R,t)\right\}\Psi_i^T(t)R_i(t)dM_i(t)\right].
\end{aligned}
$$

Changing the orders of integration and summation,

$$
\begin{aligned}
\text{(B.11)} &= n^{-1/2}\sum_{l=1}^{n}\int_0^\tau\left[n^{-1}\sum_{i=1}^{n}e^{\alpha_0^T V_i(u)}\int_u^\tau\left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}R_i(t)dM_i(t)\right]\\
&\qquad\times d\Phi_l(\alpha_0,p_0,u)\\
&= n^{-1/2}\sum_{l=1}^{n}\int_0^\tau\widehat{\chi}(u,\tau)d\Phi_l(\alpha_0,p_0,u)\\
&= n^{-1/2}\sum_{l=1}^{n}\int_0^\tau\chi(u,\tau)d\Phi_l(\alpha_0,p_0,u)+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\chi}(t_1,t_2) &= n^{-1}\sum_{i=1}^{n}e^{\alpha_0^T V_i(t_1)}\int_{t_1}^{t_2}\left\{Z_i(0)-\overline{Z}(\beta,R,t)\right\}R_i(t)dM_i(t)\\
\chi(t_1,t_2) &= \mathcal{E}\left[e^{\alpha_0^T V_i(t_1)}\int_{t_1}^{t_2}\left\{Z_i(0)-\overline{z}(\beta,R,t)\right\}R_i(t)dM_i(t)\right].
\end{aligned}
$$

Combining the above results,

$$
n^{1/2}\left(\widehat{\beta}-\beta_0\right) = A(\beta_0)^{-1}n^{-1/2}\sum_{i=1}^{n}\Theta_i\left(\beta_0,R\right)+o_p(1),
$$

where

$$
\begin{aligned}
\Theta_i\left(\beta,R\right) &= O(\beta,R)Q_i(p_0)\\
&\quad +H(\beta,R)A^C(\alpha_0)^{-1}\psi_i(\alpha_0,p_0)\\
&\quad +\int_0^\tau\chi(u,\tau)d\Phi_i(\alpha_0,p_0,u).
\end{aligned}
$$

Note that $\mathcal{E}\{\Theta_i(\beta, R)\} = 0$. By the MCLT and Slutsky's Theorem, $n^{1/2}(\widehat{\beta} - \beta_0)$ converges in distribution to a $N(0, A(\beta_0)^{-1}\Sigma(\beta, R)A(\beta_0)^{-1})$ variate, where $\Sigma(\beta, R) = \mathcal{E}\{\Theta_i(\beta, R)^{\otimes 2}\}$.

B.7 Estimating var$(\widehat{\beta})$

The variance of $\widehat{\beta}$ can be consistently estimated by $n^{-1}\sum_{i=1}^{n}\widehat{\Theta}_i(\beta, R)$, with $\widehat{\Theta}_i(\beta, R)$ being obtained by substituting limiting values in $\Theta_i(\beta, R)$ with the sample analogs. However, as shown in the Web Appendix, the computation of $\widehat{\Theta}_i(\beta, R)$ is very complicated and difficult to implement numerically. A useful alternative is to estimate the variance of the proposed estimators by treating the weights $R_i(t)$ as known rather than estimated.

By some simple algebra, we have

$$\begin{aligned} n^{-1/2}U(\beta_0, R) &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(0) - \overline{Z}(\beta, R, t)\right\}R_i(t)dN_i(t) \\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(0) - \overline{Z}(\beta, R, t)\right\}R_i(t)dM_i(t) \\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(0) - \overline{z}(\beta, R, t)\right\}R_i(t)dM_i(t) \qquad \text{(B.12)} \\ &\quad -n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\overline{Z}(\beta, R, t) - \overline{z}(\beta, R, t)\right\}R_i(t)dM_i(t). \qquad \text{(B.13)} \end{aligned}$$

Note that $\mathcal{E}\{W_{1i}(t)dM_i(t) \mid Z(0)\} = 0$, such that

$$\begin{aligned} \mathcal{E}\{R_i(t)dM_i(t) \mid Z(0)\} &= \mathcal{E}\{\rho_i(p)W_{1i}(t)dM_i(t) \mid Z(0)\} \\ &= \mathcal{E}[\mathcal{E}\{\rho_i(p)W_{1i}(t)dM_i(t) \mid \Delta_{1i}, \Delta_{2i}, \Delta_{3i}, Z(t)\} \mid Z(0)] \\ &= \mathcal{E}\{W_{1i}(t)dM_i(t) \mid Z(0)\} \\ &= 0. \end{aligned}$$

Therefore, (B.13) converges in probability to 0. It follows that

$$n^{-1/2}U\left(\beta_0, R\right) \;=\; n^{-1/2}\sum_{i=1}^{n} U_i^{\ddagger}(\beta_0, R) + o_p(1),$$

where $U_i^{\ddagger}(\beta_0, R) = \int_0^{\tau}\left\{Z_i(0) - \overline{z}(\beta, R, t)\right\} R_i(t) dM_i(t)$. Hence, under the assumed conditions, $\{U(\beta_0, R)\}$ is asymptotically a sum of independent and identically distributed zero-mean random quantities. By the MCLT, $n^{-1/2}U\left(\beta_0, R\right)$ converges asymptotically to a $N(0, \Sigma^{\ddagger}(\beta_0, R))$ distribution, where $\Sigma^{\ddagger}(\beta, R) = \mathcal{E}\left\{U_i^{\ddagger}(\beta, R)^{\otimes 2}\right\}$. By the Functional Delta methods,

$$n^{1/2}\left(\widehat{\beta} - \beta_0\right) \;=\; A(\beta_0)^{-1} n^{-1/2}\sum_{i=1}^{n} U_i^{\ddagger}\left\{\beta_0, R\right\} + o_p(1),$$

Therefore, the variance of $\widehat{\beta}$ is estimated by $\widehat{A}(\widehat{\beta})^{-1}\widehat{\Sigma}^{\ddagger}(\widehat{\beta}, \widehat{R})\widehat{A}(\widehat{\beta})^{-1}$, where $\widehat{A}(\widehat{\beta})$ and $\widehat{\Sigma}^{\ddagger}(\widehat{\beta}, \widehat{R})$ are calculated by replacing limiting values with their corresponding empirical counterparts.

BIBLIOGRAPHY

BIBLIOGRAPHY

Barlow, W. E. (1994). Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064–1072.

Binder, D. A. (1992). Fitting cox's proportional hazards models from survey data. *Biometrika* **79**, 139–147.

Bleyer, A. J., Russell, G. B., and Satko, S. G. (1999). Sudden and cardiac death rates in hemodialysis patients. *Kidney International* **55**, 1553–1559.

Bleyer, A. J., et al. (2006). Characteristics of sudden death in hemodialysis patients. *Kidney International* **69**, 2268–2273.

Borgan, e., et al. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6**, 39–58.

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

Breslow, N. E. and Holubkov, R. (1997a). Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *Journal of the Royal Statistical Society, B* **59**, 447–461.

Breslow, N. E. and Holubkov, R. (1997b). Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Stat. Med.* **16**, 103–116.

Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics* **34**, 86–102.

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82**, 151–164.

Chen, H. Y. (2001a). Fitting semiparametric transformation regression models to data from a modified case-cohort design. *Biometrika* **88**, 255–268.

Chen, H. Y. (2001b). Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. *Journal of the American Statistical Association* **96**, 1446–1457.

Chen, K. and Lo, S. H. (1999). Case-cohort and case-control analysis with cox's model. *Biometrika* **86**, 755–764.

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

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* **72**, 147–148.

Green, M. S. and Symons, M. J. (1983). A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chronic Dis* **36**, 715–723.

Hernán, M. A., Brumback, B., and Robins, J. M. (2000). Marginal structural models to estimate the causal effect on the survival of hiv-positive men. *Epidemiology* **11**, 561–570.

Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine* **7**, 149–160.

Karnik, J. A., et al. (2001). Cardiac arrest and sudden death in dialysis units. *Kidney International* **60**, 350–357.

Kong, L., Cai, J., and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91**, 305–319.

Kong, L., Cai, J., and Sen, P. K. (2006). Asymptotic results for fitting semiparametric transformation models to failure time data from case-cohort studies. *Statistica Sinica* **16**, 135–151.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87**, 73–87.

Langholz, B. and Goldstein, L. (2001). Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics* **2**, 63–84.

Langholz, B. and Jiao, J. (2007). Computational methods for case-cohort studies. *Computational Statistics & Data Analysis* **51**, 3737–3748.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Art* pages 237–247.

Lin, D. Y. (2000). On fitting cox's proportional hazards models to survey data. *Biometrika* **87**, 37–47.

Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88**, 1341–1349.

Lu, S. E. and Shih, J. H. (2006). Case-cohort designs and analysis for clustered failure time data. *Biometrics* **62**, 1138–1148.

Lu, S. E. and Wang, M. C. (2005). Marginal analysis for clustered failure time data. *Lifetime Data Analysis* **11**, 61–79.

Ma, S. (2007). Additive risk model with case-cohort sampled current status data. *Statistical Papers* **48**, 595–608.

Matsuyama, Y. and Yamaguchi, T. (2008). Estimation of the marginal survival time in the presence of dependent competing risks using inverse probability of censoring weighted (ipcw) methods. *Pharmaceutical Statistics* **7**, 202–214.

Moger, T. A., Pawitan, Y., and Borgan, e. (2008). Case-cohort methods for survival data on families from routine registers. *Statistics in Medicine* **27**, 1062–1074.

Nan, B., Kalbfleisch, J. D., and Yu, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *Annals of Statistics* **37**, 2351–2376.

Nan, B., Yu, M., and Kalbfleisch, J. D. (2006). Censored linear regression for case-cohort studies. *Biometrika* **93**, 747–762.

Pan, Q. and Schaubel, D. E. (2008). Proportional hazards models based on biased samples and estimated selection probabilities. *Can. J. Statist.* **36**, 111–127.

Pisoni, R. L., et al. (2004). The dialysis outcomes and practice patterns study (dopps): Design, data elements, and methodology. *Am J Kidney Dis* **44**, S7–S15.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics: Hayward, CA.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

Robins, J. M. (1993a). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceurical Section, American Statistical Association* pages 24–33.

Robins, J. M. (1993b). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceurical Section, American Statistical Association* pages 24–33.

Robins, J. M. and Finkelstein, D. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics* **56**, 779–788.

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

Samuelsen, S. O., Anestad, H., and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scandinavian Journal of Statistics* **34**, 103–119.

Scharfstein, D. O. and Robins, J. M. (2002a). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89**, 617–634.

Scharfstein, D. O. and Robins, J. M. (2002b). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89**, 617–634.

Schaubel, D. E., et al. (2009). Estimating the effect of a time-dependent treatment in by levels of an internal time-dependent covariate. *J. Am. Stat. Assoc.* **104**, 49–59.

Scheike, T. H. and Juul, A. (2004). Maximum likelihood estimation for cox's regression model under nested case-control sampling. *Biostatistics* **5(2)**, 193–206.

Scheike, T. H. and Martinussen, T. (2004). Maximum likelihood estimation for coxs regression model under caseccohort sampling. *Scandinavian Journal of Statistics* **31**, 283–293.

Schildcrout, J. S. and Heagerty, P. J. (2008). On outcome-dependent sampling designs for longitudinal binary response data with time-varying covariates. *Biostatistics* **9**, 735–749.

Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Annals of Statistics* **16**, 64–81.

Song, R., Zhou, H., and Kosorok, M. R. (2009). On semiparametric efficient inference for two-stage outcome-dependent-sampling with a continuous outcome. *Biometrika* **96**, 221–228.

Sorensen, P. and Anderson, P. K. (2000). Competing risks analysis of the case-cohort design. *Biometrika* **87**, 49–59.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93**, 1164–1175.

Sun, J., Sun, L., and Flournoy, N. (2004). Additive hazards model for competing risks analysis of the case-cohort design. *Communications in statistics: Theory and methods* **33**, 351–366.

Therneau, T. M. and Li, H. (1999). Computing the cox model for case cohort designs. *Lifetime Data Analysis* **5**, 99–112.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.

Wacholder, S., et al. (1989). Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* **76**, 117–123.

Wang, W., et al. (2009). Causal effects in outcome-dependent two-phase sampling designs. *Journal of the Royal Statistical Society, B* **71**, 947–969.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distrubutions. *Journal of the American Statistical Association* **84**, 1065–1073.

Young, E. W., et al. (2000). The dialysis outcomes and practice patterns study (dopps): an international hemodialysis study. *Kidney Int* **57**, S74–S81.

Zhang, M. and Schaubel, D. E. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics* **67**, 740–749.

Zhou, H. and You, J. H. (2007). Semiparametric methods for data from an outcome-dependent sampling scheme. In Hong, D. and Shyr, Y., editors, *Quantitative Medical Data Analysis Using Mathematical Tools and Statistical Techniques*, pages 93–110. World Scientific Publications, Singapore.

Zhou, H., et al. (2002). A semiparametric empirical likelihood method for data from an outcome-dependent sampling design with a continuous outcome. *Biometrics* **58**, 413–421.