STATISTICAL METHODS FOR CASE-CONTROL AND CASE-COHORT STUDIES WITH POSSIBLY CORRELATED FAILURE TIME DATA

by
Sangwook Kang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, School of Public Health.

Chapel Hill
2007

Approved by:

Dr. Jianwen Cai, Advisor
Dr. Daniel Caplan, Reader
Dr. Amy Herring, Reader
Dr. Danyu Lin, Reader
Dr. Chirayath Suchindran, Reader

ABSTRACT

SANGWOOK KANG: STATISTICAL METHODS FOR CASE-CONTROL AND CASE-COHORT STUDIES WITH POSSIBLY CORRELATED FAILURE TIME DATA.
(Under the direction of Dr. Jianwen Cai.)

In large cohort studies, the major effort and cost typically arise from the assembling of covariate measurements. Case-control and case-cohort study designs are widely used ones to reduce the cost and achieve the same goals in such studies, especially when the disease rate is low. In this dissertation, we consider analyzing the multivariate failure time data arising from case-control and case-cohort studies.

First, we consider a case-control within cohort study with correlated failure times. A retrospective dental study was conducted to evaluate the effect of pulpal involvement on tooth survival (Caplan and Weintraub, 1997; Caplan et al., 2005). Due to the clustering of teeth, the survival times of the matched teeth within subjects could be correlated and thus the statistical methods for conventional case-control studies cannot not be directly applied. We study the marginal proportional hazards regression model for data from this type studies.

Second, we consider a case-cohort study with multiple disease outcomes. A case-cohort design was implemented in the Busselton Health Study (Cullen, 1972) and it was of interest to study the relationship between serum ferritin and coronary heart disease and stroke events. Since times to coronary heart disease and stroke events observed from the same subject could be correlated, valid statistical method needs to take it into consideration. To this end, we consider marginal proportional hazards regression model.

Third, we consider marginal additive hazards regression model for case-cohort studies with multiple disease outcomes. Most modern analyses of survival data focus on multiplicative models for relative risk using proportional hazards models. The additive hazards model, which model the risk differences has often been suggested as an alternative to the proportional hazards model.

In each of the three cases, we propose a weighted estimating equation approach for model parameter estimation, with different types weights to enhance the efficiency. The asymptotic properties of the proposed estimators are derived and their finite sample properties are assessed via simulation studies. The proposed method are applied to the aforementioned dental study and the Busselton Health Study for illustration.

ACKNOWLEDGMENTS

I would like to give my deepest gratitude to my dissertation advisor, Dr. Jianwen Cai for her guidance, insights and encouragement throughout my dissertation research process. The lessons I have learned under her direction and the experience I have gained as a graduate research assistant during last four years are invaluable. I am deeply grateful for her financial support as well.

I also would like to thank the committee members, Drs. Daniel Caplan, Amy Herring, Danyu Lin and Chirayath Suchindran for their constructive comments and valuable suggestions. Especially, I want to thank Dr. Daniel Caplan and the Kaiser Permanente Dental Care Program for providing KPDCP data. My thanks also go to Dr. Donglin Zeng who led a workshop on empirical processes and semiparametric theory during my last year at UNC-CH.

In addition, I owe my thanks to my friends, wife and family. I would like to thank my friend, Jacob Vacarro, who have helped me improve my english in every way. I also would like to thank my parents and brother who have provided endless support and encouragement. Last, but not least, my special thanks are reserved for my wife, Hee young Seok. She has always been there for me when I needed her the most.

CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Epidemiologic cohort studies and disease prevention trials often need the follow-up of several thousand subjects or more for many years and thus can be prohibitively expensive (Prentice, 1986). The major effort and cost typically arise from the assembling of covariate measurements, such as expensive analysis of biological specimens, or assessment of exposure from the raw covariate histories. When the disease rate is low, which is often the case, much of the covariate information on disease-free subjects is largely redundant. To reduce the cost in such studies and achieve the same goals as a cohort study, several study designs have been proposed. Case-control and case-cohort study designs are the two most widely used ones among them. In this dissertation, we develop statistical methods for these two study designs with multivariate failure time data.

1.1 Marginal Hazards Regression for Case-Control within Cohort Studies with Possibly Correlated Failure Time Data

Case-control within cohort study involves the independent random sampling of subjects with disease(cases) and without disease(controls) from a cohort that has already been enumerated. The covariate measurements are only assembled for the case-control samples. An important assumption for these conventional case-control studies is the statistical independence among subjects. However, in many biomedical studies, this assumption might not hold. For example, in a retrospective cohort dental study (Caplan and Weintraub, 1997; Caplan et al., 2005), it was of interest to evaluate the effect of pulpal involvement on tooth survival. Root canal filled (RCF) teeth were used as an indicator of pulpal involvement. After cases and controls were sampled, a non-RCF tooth was matched to the RCF tooth within each subject. Here cases were defined as those who lost the RCF tooth, while controls were defined as those who did not lose the RCF tooth during the study period. The survival times of the two teeth within the same subject could be correlated and thus the independence assumption might not be valid. The primary goal of the study is to evaluate the effect of pulpal involvement on tooth survival. The fact that the survival times of the teeth from the same individual are correlated is considered as a nuisance. In such case, a marginal model approach is appealing.

Failure time models from such retrospective case-control studies have been studied in the literature. However, all these methods assume independent failure times and cannot be directly applied to multivariate failure time data. There is an extensive literature of statistical methods for correlated failure time data from prospective cohort studies. However, these methods cannot be directly applied to correlated failure time data from case-control. Work for correlated failure time data from case-control studies has been limited. Some efforts have been made to analyze failure time data from case-control family studies where the investigators are usually interested in estimating the strength of dependence of failure times within family. Consequently, most of the methods concentrated on frailty models or parametric approach. When the correlation of the failure times is not of interest, as in the aforementioned dental study, statistical inference procedure that is easy to conduct and has nice asymptotic properties remains to be developed. It is desirable to develop hazard regression models for the correlated failure time data from case-control within cohort studies which account for the possible correlation within subject while avoiding the specification of the correlation structure.

1.2 Marginal Hazards Model for Case-cohort Studies with Multiple Disease Outcomes

The case-cohort study design was originally proposed by Prentice (1986). Under the case-cohort design, a random sample called subcohort is selected from the entire cohort. The covariate measurements are only assembled for the subjects in the subcohort and all the cases (failures) who experience the disease of interest regardless of whether or not they are in the subcohort. A key advantage of the case-cohort design is its ability to use the same subcohort for several diseases or for subtypes of disease. For example, the case-cohort design was implemented in the Busselton Health Study (Cullen, 1972). This study was conducted every 3 years from 1966 to 1981 and general health information for adult participants were collected by means of questionnaire and clinical visit. It was of interest to study the relationship between serum ferritin and coronary heart disease and stroke events. To reduce costs and preserve stored serum, case-cohort sampling was used. In order to compare the effect of serum ferritin on coronary heart disease and stroke, times to coronary heart disease and stroke events need to be modeled simultaneously. Since times to coronary heart disease and stroke events observed from the same subject could be correlated, valid statistical method needs to take it into consideration.

The additive and multiplicative risk models provide the two principal frameworks for studying the association between risk factors and disease occurrence or death. Most modern analyses of survival data focus on multiplicative models for relative risk using proportional hazards models, mostly due to desirable theoretical properties along with the simple interpretation of the results and the wide availability of computer programs. However, epidemiologists often are interested in the risk difference attributed to the exposure, and the risk difference is known to be more relevant to public health because it translates directly into the number of disease cases that would be avoided by eliminating a particular exposure (Kulich and Lin, 2000). Consequently, the additive hazards model, which model the risk differences, has often been suggested as an alternative to the proportional hazards model.

For data from case-cohort study for a single disease outcome, estimating procedures have been proposed in the literature for various models. However, methodologies to address analysis of case-cohort data with multiple diseases outcomes have been limited. A commonly used method for dealing with multiple diseases is to analyze each disease separately. This approach does not allow comparison of the risk factors for different diseases, because it does not account for the induced correlation between outcomes (Langholz and Thomas, 1990). Statistical methods which account for the correlation between outcomes is needed.

Motivated by these needs, we propose statistical methods for modeling correlated failure time data from case-control within cohort studies and modeling multiple disease outcomes with data from case-cohort studies. We will consider both the multiplicative as well as additive models.

In the next chapter, we will review the relevant literature in these areas.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we review the literature on statistical methods for : 1) univariate failure time data arising from case-control within cohort and case-cohort studies, 2) correlated failure time data from prospective studies assuming random samples, and 3) failure time data from case-control-family studies. The organization of the rest of this section is as following. We review literature on statistical methods for univariate failure time data from case-control and case-cohort studies in section 2.1, for correlated failure time data from prospective studies assuming random samples in section 2.2, and for correlated failure time data from the so-called case-control family studies in section 2.3. In section 2.4, we review the literature on statistical methods for additive hazards models.

2.1 Univariate failure time models from cohort studies

The Cox proportional hazards model (Cox, 1972) has been the most widely used procedure to study the effects of covariates on a failure time. The Cox model assumes that the hazard function for the failure time T associated with a covariate vector $\boldsymbol{Z}$ is given by

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)\}, \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters.

Let C denote the potential censoring time, $X = \min(T, C)$ denote the observed time. Let $N(t)$ denote the counting process, $Y(t) = I(X \leq t)$ be an 'at-risk' indicator process and $\Delta = I(T \leq C)$ be an indicator for failure, where $I(.)$ is an indicator function. The failure time is assumed to be subject to independent right censorship. Let $(T_i, C_i, \boldsymbol{Z}_i)(i = 1, \dots, n)$ be n independent replicates of $(T, C, \boldsymbol{Z})$ and τ denote the study end time.

The regression parameter $\boldsymbol{\beta}_0$ can be estimated by the partial likelihood score function introduced by Cox (1975)

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Delta_i \left\{ \boldsymbol{Z}_i(X_i) - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, X_i)}{S^{(0)}(\boldsymbol{\beta}, X_i)} \right\},$$

where

$$S^{(0)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_i(t)\}, \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_i(t)\} \boldsymbol{Z}_i(t)$$

The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$, defined as the solution to the score equation $\boldsymbol{U}(\boldsymbol{\beta}) = \mathbf{0}$, has been shown to be approximately normal in large samples with mean $\boldsymbol{\beta}_0$ and with a covariance matrix that can be consistently estimated by $-\{\partial \boldsymbol{U}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\}^{-1}$ (Andersen and Gill, 1982; Tsiatis, 1981).

2.1.1 Nested case-control studies

The nested case-control study design was originally suggested by Thomas (1977). Prentice and Breslow (1978) clarified the conceptual foundations of the nested case-control study and formally derived the conditional likelihood. The study consists of selecting a random sample of controls at each distinguished failure time either without replacement or with replacement independently across time. Specifically, suppose that at each observed failure time, or age, $t_i (i = 1, \dots, L)$ exactly $\tilde{n}_{i1} = \tilde{n}_{i1}(t_i)$ cases with exposure variables $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_{\tilde{n}_{i1}}$ and $\tilde{n}_{i0} = \tilde{n}_{i0}(t_i)$ controls with exposures $\boldsymbol{Z}_{\tilde{n}_{i1}+1}, \dots, \boldsymbol{Z}_{\tilde{n}_{i1}+\tilde{n}_{i0}}$ are sampled. Given that the sample consists of individuals with these $\tilde{n}_{i1} + \tilde{n}_{i0}$ risk vectors, the probability that the first $\tilde{n}_{i1}$ such vectors actually correspond to the cases as observed and the remainder to controls can be written, under a Cox proportional hazard model, as

$$\exp\{\boldsymbol{\beta}_0^T \sum_{j=1}^{\tilde{n}_{i1}} \boldsymbol{Z}_j\} / \sum_{l \in R(\tilde{n}_{i1}, \tilde{n}_{i0})} \exp\{\boldsymbol{\beta}_0^T \sum_{j=1}^{\tilde{n}_{i1}} \boldsymbol{Z}_{l_j}\} \tag{2.2}$$

where $R(\tilde{n}_{i1}, \tilde{n}_{i0})$ is the set of all subsets of size $\tilde{n}_{i1}$ from $\{1, \ldots, \tilde{n}_{i1} + \tilde{n}_{i0}\}$ and $l = (l_1, \ldots, l_{\tilde{n}_{i1}})$

A conditional likelihood for $\boldsymbol{\beta}_0$ based on case and control samples of respective sizes $\tilde{n}_{i1}$ and $\tilde{n}_{i0}$ at the observed failure times, or ages, $t_i (i = 1, \ldots, L)$ is simply the product of terms (2.2) over the L distinct times

$$\prod_{i=1}^{L} \left(\exp\{\boldsymbol{\beta}_0^T \sum_{j=1}^{\tilde{n}_{i1}} \boldsymbol{Z}_j\} / \sum_{l \in R(\tilde{n}_{i1}, \tilde{n}_{i0})} \exp\{\boldsymbol{\beta}_0^T \sum_{j=1}^{\tilde{n}_{i1}} \boldsymbol{Z}_{l_j}\} \right) \tag{2.3}$$

Note that (2.3) is of precisely the same form as the partial likelihood for prospective data. However, the 'risk-sets' contributing to (2.3), instead of including all individuals known to be at risk at time t_i, includes only those actually sampled for the retrospective study at t_i.

The asymptotic properties of the estimator have been formally derived (Goldstein and Langholz, 1992; Borgan et al., 1995) using counting process and martingale theory (Andersen et al., 1993). Several authors proposed improved estimators. Langholz and Thomas (1991) proposed some sample reuse methods while Samuelsen (1997) used weighted estimating equations via inclusion probabilities. Chen and Lo (1999) and Chen (2001) made further improvements by using modified weights.

2.1.2 Classical case-control studies

The classical case-control study design was mostly restricted to dose-response models rather than dose-time-response models (Chen, 2001). There has been little development in the literature on survival analysis for classical case-control sampling. Binder (1992) described a procedure for fitting proportional hazards models to survey data with complex sampling designs from the finite population including the classical case-control study.

In population-based surveys, $\{X_i, \Delta_i, \boldsymbol{Z}_i(.)\}(i = 1, \ldots, n)$ are treated as fixed and $\boldsymbol{B}$, which is the solution to $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$, is the finite-population parameter of interest. Suppose that a sample of size $\tilde{n}$ is drawn via a complex design, such as case-control sampling for our case. Binder (1992) proposed to estimate $\boldsymbol{B}$ by the estimating function

$$\hat{\boldsymbol{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i \Delta_i \{\boldsymbol{Z}_i(X_i) - \frac{\hat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, X_i)}{\hat{S}^{(0)}(\boldsymbol{\beta}, X_i)}\},$$

where $w_i = \frac{1}{\pi_i}$ if the ith member is selected in the sample and 0 otherwise, π_i is the inclusion probability of the ith member, $\hat{S}^{(0)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} w_i Y_i(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}$ and $\hat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} w_i Y_i(t) \boldsymbol{Z}_i(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}$. Then $n^{-\frac{1}{2}} \hat{\boldsymbol{U}}(\boldsymbol{B})$ is shown to asymptotically follow a zero-mean normal distribution with covariance matrix $\boldsymbol{V}(\boldsymbol{B})$, and $n^{\frac{1}{2}}(\hat{\boldsymbol{B}} - \boldsymbol{B})$ is shown to be asymptotically normal with mean zero and covariance matrix $\boldsymbol{D}^{-1}(\boldsymbol{B})\boldsymbol{V}(\beta)\boldsymbol{D}^{-1}(\boldsymbol{B})$, where $\boldsymbol{D}(\boldsymbol{\beta}) = \lim_{n\to\infty} -n^{-1} \frac{\partial \boldsymbol{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{B}}$ and

$$\boldsymbol{V}(\boldsymbol{B}) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \boldsymbol{U}_i(\boldsymbol{B}) \boldsymbol{U}_j(\boldsymbol{B})^T,$$

where π_{ij} is the probability of both the ith and the jth members being sampled and

$$\begin{aligned} \boldsymbol{U}_i(\boldsymbol{\beta}) &= \int_0^{\tau} \left\{ \boldsymbol{Z}_i(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} \right\} \left\{ dG_i(t) - \frac{Y_i(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)} dg(t)}{s^{(0)}(\boldsymbol{\beta}, t)} \right\}, \\ \boldsymbol{S}^{(l)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^{n} Y_i(t) \boldsymbol{Z}_i(t)^{l-1} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}, \ \boldsymbol{s}^{(l)} = \lim_{n\to\infty} \boldsymbol{S}^{(l)}(\boldsymbol{\beta}, t), \ l = 1, 2, \\ G_i(t) &= \Delta_i I(X_i \le t), \ G(t) = n^{-1} \sum_{i=1}^{n} G_i(t) \text{ and } g(t) = \lim_{n\to\infty} G(t). \end{aligned}$$

Lin (2000) provided a formal justification of Binder's method and also presented an alternative approach which regards the survey population as a random sample from an infinite universe and accounts for this randomness in the statistical inference. Under superpopulation approach, the survey population are not treated as fixed quantities, but rather as a random sample from the joint distribution of $\{X, \Delta, \boldsymbol{Z}(\cdot)\}$. Then the inclusion probabilities are allowed to depend on $\mathcal{F}$ where $\mathcal{F}$ is the sigma-field generated by $\{X_i, \Delta_i, \boldsymbol{Z}_i(\cdot)\}(i = 1, \ldots, n)$, i.e.,

$$\pi_i = \Pr(\xi_i = 1 | \mathcal{F}), \ i = 1, \ldots, n$$

Then $n^{-\frac{1}{2}}\hat{\boldsymbol{U}}(\boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean zero and covariance matrix $\boldsymbol{D}(\boldsymbol{\beta}_0)+\boldsymbol{V}(\boldsymbol{\beta}_0)$ and, thus, $n^{\frac{1}{2}}(\hat{\boldsymbol{B}}-\boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0) + \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{V}(\boldsymbol{\beta}_0)\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)$$

where $\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the survey population from the super-population and $\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{V}(\boldsymbol{\beta}_0)\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the survey sample from the survey population which is the variance under the finite-population approach.

The sampling weights used in Binder (1992) and Lin (2000) are proportional to the inverse of the sampling probability. Chen and Lo (1999) proposed similar type of weight for classical case-control studies. Chen (2001) proposed a more efficient estimator by using local average type of weights. Specifically, let $0 = t_0 \le t_1 \le \cdots \le t_{a_n} = \tau$ and $0 = s_0 \le s_1 \le \cdots \le s_{b_n} = \tau$ be two partitions of $[0, \tau)$. Use $\pi_i = r_n(X_i, \Delta_i)$, where

$$r_n(t,d) = \begin{cases} \frac{\sum_{l=1}^n \Delta_l \xi_l I\{X_l \in [t_{i-1}, t_i)\}}{\sum_{l=1}^n \Delta_l I\{X_l \in [t_{i-1}, t_i)\}} & \text{if } d = 1 \text{ and } t \in [t_{i-1}, t_i)\,, \\ \frac{\sum_{l=1}^n (1-\Delta_l) \xi_l I\{X_l \in [s_{j-1}, s_j)\}}{\sum_{l=1}^n (1-\Delta_l) I\{X_l \in [s_{j-1}, s_j)\}} & \text{if } d = 0 \text{ and } t \in [s_{j-1}, s_j)\,. \end{cases}$$

This is based on a simple idea of estimating each missing covariate by a local average. In univariate failure time setting, with additional assumptions, the estimator using this weight function is more efficient than the previous one using the inclusion probabilities. Specifically, $n^{-\frac{1}{2}}\hat{\boldsymbol{U}}(\boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean zero and covariance matrix $\boldsymbol{D}(\boldsymbol{\beta}_0) + \boldsymbol{V}(\boldsymbol{\beta}_0) - \boldsymbol{\Gamma}(\boldsymbol{\beta}_0)$ where

$$\boldsymbol{\Gamma}(\boldsymbol{\beta}_0) = \lim_{n\to\infty} N^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \mathrm{E}(\boldsymbol{U}_i(\boldsymbol{\beta}_0)|X_i, \Delta_i)\mathrm{E}(\boldsymbol{U}_i(\boldsymbol{\beta}_0)|X_i, \Delta_i)^T,$$

and $n^{\frac{1}{2}}(\hat{\boldsymbol{B}} - \boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0) + \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)\{\boldsymbol{V}(\boldsymbol{\beta}_0) - \boldsymbol{\Gamma}(\boldsymbol{\beta}_0)\}\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0).$$

Clearly, the asymptotic variance using local average type estimator of Chen (2001) is smaller than the one using the inclusion probability. However, this approach requires the knowledge of the failure or censoring times ($X_i, i = 1, \ldots, n$) of all the members of the cohort which might not be always available.

2.1.3 Case-cohort studies

As an alternative to the nested case-control design to reduce the number of subjects for whom covariate data are required, Prentice (1986) proposed a case-cohort design. This design involves the selection of a random sample, or a stratified random sample, of the entire cohort, and the assembly of covariate histories only for this random subcohort and for all cases. The subcohort in a given stratum constitutes the comparison set of cases occurring at a range of failure times. The subcohort also provides a basis for covariate monitoring during the course of cohort follow-up. The hazard function of the ith subject at time t assumes a relative risk regression model (Cox, 1972) which has the form :

$$\lambda_i(t|\boldsymbol{Z}(t)) = \lambda_0(t) r\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)\}, \tag{2.4}$$

where $r(x)$ is a fixed function with $r(0) = 1$. The pseudolikelihood function for the estimation of $\boldsymbol{\beta}_0$ is given by

$$\tilde{L}(\boldsymbol{\beta}_0) = \prod_{i=1}^{n} \left(r_{ii} / \sum_{l \in \tilde{R}(t_i)} r_{li} \right)^{\Delta_i}, \tag{2.5}$$

where $r_{li} = Y_l(t_i) r\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_l(t_i)\}$, $D(t) = \{i | N_i(t) \neq N_i(t^-)\}$, $\tilde{R}(t) = D(t) \cup C$ and C is a random subcohort. The maximum pseudolikelihood estimate $\tilde{\boldsymbol{\beta}}_p$ is defined by $\boldsymbol{U}(\tilde{\boldsymbol{\beta}}_p) = \boldsymbol{0}$, where

$$\boldsymbol{U}(\boldsymbol{\beta}) = \partial \log \tilde{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = \sum_{i=1}^{n} U_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Delta_i \left(c_{ii} - \sum_{l \in \tilde{R}(t_i)} b_{li} / \sum_{l \in \tilde{R}(t_i)} r_{li} \right) \tag{2.6}$$

and where $b_{li} = Y_l(t_i) \boldsymbol{Z}_l(t_i) r'\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(t_i)\}$, $c_{li} = b_{li} \left(r\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(t_i)\}\right)^{-1}$ and $r'(u) = dr(u)/du$. Under some regularity conditions, $n^{-\frac{1}{2}} \boldsymbol{U}(\boldsymbol{\beta})$ can be shown to converge weakly to a normal variate with mean zero and a variance matrix $\boldsymbol{A}$. Therefore, $n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0)$ is shown to converge in distribution to a normal variate with mean zero and a sandwich type of variance matrix $S = \Omega^{-1}\boldsymbol{A}\Omega^{-1}$. The asymptotic variance matrix can be consistently estimated by $nI(\tilde{\boldsymbol{\beta}}_p)^{-1}\tilde{V}(\tilde{\boldsymbol{\beta}}_p)I(\tilde{\boldsymbol{\beta}}_p)^{-1}$ where

$$\tilde{\boldsymbol{V}}(\boldsymbol{\beta}) = I(\boldsymbol{\beta}) + 2\sum_{j=1}^{n}\Delta_j\tilde{\Delta}(t_j)\sum_{\{k|t_k<t_j\}}\Delta_k v_{kj}, \tag{2.7}$$

with

$$\begin{aligned} I(\boldsymbol{\beta}) &= -\frac{\partial^2 \log \tilde{L}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \\ v_{kj} &= -\sum_{i\in\tilde{R}(t_j)}\left(\frac{B_k + b_{jk} - b_{ik}}{R_k + r_{jk} - r_{ik}}\right)^T\left(c_{ij} - \frac{B_j}{R_j}\right)r_{ij}R_j^{-1} \\ R_j &= \sum_{l\in\tilde{R}(t_j)} r_{lj},\ B_j = \sum_{l\in\tilde{R}(t_j)} b_{lj} \text{ and } \tilde{\Delta}(t) = 1 \text{ if } \tilde{R}(t) \neq C, \text{ and 0 otherwise} \end{aligned}$$

A natural estimator for the cumulative baseline failure rate, $\Lambda_0(\cdot)$ is proposed as

$$\hat{\Lambda}_0(t) = \tilde{n}n^{-1}\int_0^t [\sum_{l\in C} Y_l(w)r\{Z_l(w)\tilde{\boldsymbol{\beta}}\}]^{-1}d\bar{N}(w),$$

where $\bar{N} = N_1 + \ldots + N_n$.

This was shown to be able to be extended to a stratified model. Suppose that baseline data available for the entire cohort are used to partition the cohort into Q strata, and that a relative risk regression model

$$\lambda_q\{t|\boldsymbol{Z}(t)\} = \lambda_{0q}(t)r\{\boldsymbol{\beta}_q^T\boldsymbol{Z}(t)\},\ q = 1, \ldots, Q,$$

is specified for the disease incidence rate in each stratum. A case-cohort approach to the estimation of $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_Q^T)^T$ would involve the selection of a subcohort from each stratum and the assembly of covariate histories for cases and subcohort members. A pseudolikelihood function for $\boldsymbol{\beta}_0$ can be written as a product of terms (2.5) over strata :

$$\tilde{L}_{st}(\boldsymbol{\beta}_0) = \prod_{q=1}^{Q} \prod_{i=1}^{n_q} \left(r_{iiq} / \sum_{l \in \tilde{R}_q(t_{iq})} r_{liq} \right)^{\Delta_{iq}}$$

where $r_{liq} = Y_{lq}(t_{iq}) r\{\boldsymbol{\beta}_q^T \boldsymbol{Z}_{lq}(t_{iq})\}$, $D_q(t) = \{i, q | N_{iq}(t) \neq N_{iq}(t^-)\}$, $\tilde{R}_q(t) = D_q(t) \cup C_q$ and C_q is a random subcohort from qth stratum. Thus, the corresponding score statistic is also a sum of (2.6) over strata and has mean zero and variance can be estimated by the sum over strata of matrices (2.7).

Self and Prentice (1988) developed asymptotic distribution theory for the case-cohort maximum pseudolikelihood estimator and related quantities via using a combination of martingale and finite population convergence results using a slightly different pseudolikelihood and variance estimator. In their formulation of the risk set, only members in the subcohort were included while, in Prentice (1986), a nonsubcohort case that fails at time t_i would be considered at risk and was included in the risk set.

They considered the same type of relative risk regression model (2.4) for the hazard function. The maximum pseudolikelihood estimator, $\tilde{\boldsymbol{\beta}}_{sp}$, is defined as a solution to $\partial \log \tilde{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = 0$, where

$$\log \tilde{L}(\boldsymbol{\beta}) = \sum_{i \in C} \int_0^{\tau} \log r\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(u)\} dN_i(u) - \int_0^{t} \log \left[\sum_{i \in \tilde{C}} Y_l(u) r\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(u)\} \right] d\bar{N}(u)$$

and where $\tilde{C}$ is a random subcohort of size $\tilde{n}$. Under some regularity conditions, $\tilde{\boldsymbol{\beta}}_{sp}$ is shown to converge to $\boldsymbol{\beta}_0$ in probability and $n^{-\frac{1}{2}} \tilde{U}(\boldsymbol{\beta}_0)$ is shown to converge in distribution to a Gaussian random variable with mean zero and covariance matrix given by $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0) = \boldsymbol{D}(\boldsymbol{\beta}_0) + \boldsymbol{A}(\boldsymbol{\beta}_0)$ where $\boldsymbol{D}(\boldsymbol{\beta}) = -\lim_{n \to \infty} \partial^2 \log \tilde{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^2$ and $\boldsymbol{A}(\boldsymbol{\beta}_0)$, which reflects the contribution of the covariance among score components induced by sampling, has very complicated expressions.

Thus, $n^{1/2} \left(\tilde{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0 \right)$ is shown to converge in distribution to a Gaussian random variable with mean zero and covariance matrix given by $\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0) + \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0) \boldsymbol{A}(\boldsymbol{\beta}_0) \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)$ by the usual Taylor series expansions. Consistent estimators for $\boldsymbol{D}(\boldsymbol{\beta}_0)$ and $\boldsymbol{A}(\boldsymbol{\beta}_0)$ have been proposed. For the cumulative hazard function, $\Lambda_0(t)$, $\tilde{\Lambda}_0(t)$ has been proposed as a natural estimator and is given by

$$\tilde{\Lambda}(t) = \tilde{n}n^{-1}\int_0^t \left[\sum_{i\in\tilde{C}} Y_i(u)r\{\tilde{\boldsymbol{\beta}}_{sp}^T\boldsymbol{Z}_i(u)\}\right]^{-1} d\bar{N}(u), \tag{2.8}$$

$n^{1/2}(\tilde{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0)$ and $n^{1/2}(\tilde{\Lambda}(\cdot) - \Lambda_0(\cdot))$ are shown to converge weakly and jointly to Gaussian random variables with mean zero and the appropriate limiting covariance functions are provided.

It is also shown that Prentice (1986)'s estimator, $\tilde{\boldsymbol{\beta}}_p$, and $\tilde{\boldsymbol{\beta}}_{sp}$ are asymptotically equivalent provided an individual's contributions to $\boldsymbol{S}^{(1)}$ and $S^{(0)}$ are asymptotically negligible. The variance estimator proposed by Prentice (1986) is somewhat different than the estimator proposed by Self and Prentice (1988), however, it is shown to converge to $\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0) + \boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)\boldsymbol{D}^{-1}(\boldsymbol{\beta}_0)$.

The variance estimators proposed by Prentice (1986) and Self and Prentice (1988) are very complicated. Wacholder et al. (1989) proposed a bootstrap estimate of the variance of $\tilde{\boldsymbol{\beta}}_p$ to avoid the direct estimation. However, this method is computationally very intensive so it might be very time-consuming for large studies. Different ways of obtaining easily computed variances estimators were proposed by Barlow (1994) and Lin and Ying (1993).

Barlow (1994) proposed a robust estimator of the variance based on the influence of an individual observation on the overall score. He assumed a standard Cox proportional hazard regression model for the relative risk

$$\lambda_i(t|\boldsymbol{Z}(t)) = \lambda_0(t)r(t),$$

where $r(x) = \exp\{\boldsymbol{\beta}_0^T\boldsymbol{Z}(x)\}$. He proposed a slightly different pseudolikelihood function than those of Prentice (1986) or Self and Prentice (1988). The conditional probability of failure at time t_j is given by

$$p_i(t_j) = \frac{Y_i(t_j)w_i(t_j)r_i(t_j)}{\sum_{k=1}^n Y_k(t_j)w_k(t_j)r_k(t_j)}$$

where the weight of the ith subject at time t, $w_i(t)$, is defined as

$$w_i(t) = \begin{cases} 1 & \text{if } dN_i(t) = 1, \\ m(t)/\tilde{m}(t) & \text{if } dN_i(t) = 0 \text{ and } i \in C, \\ 0 & \text{if } dN_i(t) = 0 \text{ and } i \notin C. \end{cases}$$

where $m(t)$ is the number of disease-free individuals at risk at time t in the cohort, $\tilde{m}(t)$ is the number of disease-free individuals at risk at time t in the subcohort, and $r_i(t) = \exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)\}$. Note that the Prentice (1986)'s likelihood used an indicator function as a weight, i.e., $w_i(t) = 1$ if $dN_i(t) = 1$ or $i \in C$, otherwise the weight is zero. The Self and Prentice (1988)'s likelihood used a denominator summed over subcohort members only. In Barlow (1994), estimation of $\boldsymbol{\beta}_0$ follows directly from the logarithm of the pseudolikelihood function, $\sum_t \sum_i dN_i(t) \log(p_i(t))$. The robust variance estimator was proposed using the infinitesimal jacknife estimator and is given by :

$$\hat{\text{Var}}(\tilde{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \hat{e}_i',$$

where $\hat{e}_i = \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{(-i)} = I^{-1}(\tilde{\boldsymbol{\beta}})\hat{c}_i(t_0)$ is the change in $\tilde{\boldsymbol{\beta}}$ if the ith observation is deleted. Let $c_i(t_0)$ denote the influence of an individual observation on the overall score for person i at time t_0 and it is given by

$$c_i(t_0) = \int_0^{t_0} Y_i(t)[dN_i(t) - \lambda_i(t)][\boldsymbol{Z}_i(t) - \text{E}(t)]d\bar{N}(t),$$

where $\text{E}(t) = \sum_{k=1}^{n} p_k(t)\boldsymbol{Z}_k(t)$. Let $I^{-1}(\tilde{\boldsymbol{\beta}})$ denote the inverse of the information matrix generated by the pseudolikelihood. Then $\hat{e}_i$ can be approximated by $I^{-1}(\tilde{\boldsymbol{\beta}})\hat{c}_i(t_0)$, where

$$\hat{c}_i(t_0) = \int_0^{t_0} Y_i(t)[dN_i(t) - \hat{p}_i(t)][\boldsymbol{Z}_i(t) - \hat{\text{E}}(t)]d\bar{N}(t)$$

is an estimate of $c_i(t_0)$ and $\hat{p}_i$ and $\hat{\text{E}}(t)$ are the corresponding estimates of $p_i(t)$ and $E(t)$ replacing $\boldsymbol{\beta}$ by $\tilde{\boldsymbol{\beta}}$.

Lin and Ying (1993) proposed a general solution to the problem of missing covariate data under the Cox regression model and the case-cohort designs were considered as a special case. An approximated partial-likelihood score function was proposed for the estimation of the regression parameters. A new variance-covariance estimator which is much easier to calculate than that of Prentice (1986) or Self and Prentice (1988) has been also proposed.

A standard Cox proportional hazard regression model was assumed for the relative risk :

$$\lambda_i(t|\boldsymbol{Z}(t)) = \lambda_0(t)\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)\}$$

Suppose that the data consist of iid random quintuplets $\{X_i, \Delta_i, \boldsymbol{Z}_i(\cdot), H_{0i}(\cdot), \boldsymbol{H}_i(\cdot)\}$ where $\boldsymbol{Z}_i(\cdot) = \{Z_{1i}(\cdot), \ldots, Z_{pi}(\cdot)\}^T$ may not be completely observed, $H_{0i}(\cdot)$ is an indicator function, and $\boldsymbol{H}_i(\cdot)$ is a $p \times p$ diagonal matrix with indicator functions $\{H_{1i}(\cdot), \ldots, H_{pi}(\cdot)\}$ as the diagonal elements. For the original case-cohort design, $\boldsymbol{H}_i(\cdot) = \boldsymbol{I}_p$, the $p \times p$ identity matrix, and $H_{0i}(t) = 1$ if and only if the ith subject belongs to the subcohort at time t. The approximate partial-likelihood score function for estimating $\boldsymbol{\beta}_0$ can be written as

$$\tilde{\boldsymbol{U}}_H(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Delta_i \boldsymbol{H}_i(X_i)\{\boldsymbol{Z}_i(X_i) - \boldsymbol{E}_H(\boldsymbol{\beta}, X_i)\} \quad \text{where}$$

$\boldsymbol{E}_H(\boldsymbol{\beta}, t) = \boldsymbol{S}_H^{(1)}(\boldsymbol{\beta}, t)/S_H^{(0)}(\boldsymbol{\beta}, t)$ and $\boldsymbol{S}_H^{(r)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^{n} H_{0i}(t)Y_i(t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}\boldsymbol{Z}_i(t)^{\otimes d}$, $d = 0, 1$. $\tilde{\boldsymbol{\beta}}_H$ is the root to the estimating equation $\{\tilde{\boldsymbol{U}}_H(\boldsymbol{\beta}) = \boldsymbol{0}\}$. Under certain regularity conditions, $n^{1/2}(\tilde{\boldsymbol{\beta}}_H - \boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean 0 and covariance matrix $\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{B}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)^T$ where

$$\begin{aligned}
\boldsymbol{A}_n(\boldsymbol{\beta}) &= -n^{-1}\partial\tilde{\boldsymbol{U}}_H(\boldsymbol{\beta})/\partial\boldsymbol{\beta},\ \boldsymbol{A}(\boldsymbol{\beta}) = \lim_{n\to\infty}\boldsymbol{A}_n(\boldsymbol{\beta}),\\
\boldsymbol{B}(\boldsymbol{\beta}) &= \mathrm{E}\{\boldsymbol{W}_1(\boldsymbol{\beta})^{\otimes 2}\},\\
\boldsymbol{W}_i(\boldsymbol{\beta}) &= \Delta_i \boldsymbol{H}_i(X_i)\{\boldsymbol{Z}_i(X_i) - \boldsymbol{e}_H(\boldsymbol{\beta}, X_i)\}\\
&\quad - \int_0^{X_i}\{\boldsymbol{h}(t)/h_0(t)\}H_{0i}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}\{\boldsymbol{Z}_i(t) - \boldsymbol{e}_H(\boldsymbol{\beta}, t)\}\lambda_0(t)dt\\
\boldsymbol{e}_H(\boldsymbol{\beta}, t) &= \boldsymbol{s}_H^{(1)}(\boldsymbol{\beta}, t)/s_H^{(0)}(\boldsymbol{\beta}, t),\ \boldsymbol{s}_H^{(r)}(\boldsymbol{\beta}, t) = \mathrm{E}\{\boldsymbol{S}_H^{(r)}(\boldsymbol{\beta}, t)\},\ \boldsymbol{h}(t) = \mathrm{E}\{\boldsymbol{H}_1(t)\} \text{ and}\\
h_j(t) &= \mathrm{E}\{H_{j1}(t)\}(j = 0, 1, \ldots, p,\ r = 0, 1)
\end{aligned}$$

The covariance matrix can be consistently estimated by $\boldsymbol{A}_n^{-1}(\widetilde{\boldsymbol{\beta}}_H)\boldsymbol{B}_n(\widetilde{\boldsymbol{\beta}}_H)\boldsymbol{A}_n^{-1}(\widetilde{\boldsymbol{\beta}}_H)$ where

$$\boldsymbol{B}_n(\tilde{\boldsymbol{\beta}}) = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{W}}_i(\tilde{\boldsymbol{\beta}}),\ \ \widehat{\boldsymbol{W}}_i(\boldsymbol{\beta}) = \Delta_i\boldsymbol{H}_i(X_i)\{\boldsymbol{Z}_i(X_i) - \boldsymbol{E}_H(\boldsymbol{\beta}, X_i)\}$$
$$- \quad n^{-1}\sum_{l=1}^{n}\Delta_l Y_i(X_l)H_{0i}(X_l)\boldsymbol{H}_l(X_l)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(X_l)\}\{\boldsymbol{Z}_i(X_l) - \boldsymbol{E}_H(\boldsymbol{\beta}, X_l)\}/S_H^{(0)}(\boldsymbol{\beta}, X_l)$$

For the case-cohort design, the variance estimator $\boldsymbol{A}_n^{-1}(\tilde{\boldsymbol{\beta}}_H)\boldsymbol{B}_n(\tilde{\boldsymbol{\beta}}_H)\boldsymbol{A}_n^{-1}(\tilde{\boldsymbol{\beta}}_H)$ is much easier to calculate than the estimators of Prentice (1986) and Self and Prentice (1988), especially in the presence of time-dependent covariates. Another advantage of the estimator by Lin and Ying (1993) is that its form remains unchanged under multiple subcohort augmentations. Furthermore, incomplete covariate measurements on the cases are allowed. A natural estimator of the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ has been proposed and is given by

$$\tilde{\Lambda}(\tilde{\boldsymbol{\beta}}_H, t) = \sum_{i=1}^{n}\frac{I(X_i \le t)\Delta_i H_{0i}(X_i)}{nS_H^{(0)}(\tilde{\boldsymbol{\beta}}, X_i)} = \sum_{i=1}^{n}\int_0^t \frac{H_{0i}(s)dN_i(s)}{nS_H^{(0)}(\tilde{\boldsymbol{\beta}}, s)}$$

The process $n^{1/2}\{\tilde{\Lambda}(\tilde{\boldsymbol{\beta}}_H, \cdot) - \Lambda_0(\cdot)\}$ is shown to converge weakly to a Gaussian process with mean 0 and covariance function

$$\begin{aligned}\boldsymbol{\psi}(t,s) \quad &= \quad \int_0^{\min(t,s)}\frac{d\Lambda_0(u)}{s_H^{(0)}(\boldsymbol{\beta}_0, u)} + \boldsymbol{J}'(t)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{B}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{J}(s) \\ &- \quad \boldsymbol{J}'(t)\boldsymbol{A}^{-1}\boldsymbol{G}(t) - \boldsymbol{J}'(t)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{G}(s) \end{aligned} \tag{2.9}$$

where

$$\begin{aligned}\boldsymbol{J}(t) \quad &= \quad \int_0^t \frac{\boldsymbol{s}_H^{(1)}(\boldsymbol{\beta}_0, u)d\Lambda_0(u)}{s_H^{(0)}(\boldsymbol{\beta}_0, u)}, \\ \boldsymbol{G}(t) \quad &= \quad \mathrm{E}\Bigg[\int_0^\infty\int_0^{\min(t,v)}\frac{H_{01}(u)\exp\{\boldsymbol{\beta}_0'\boldsymbol{Z}_1(u)\}d\Lambda_0(u)}{s_H^{(0)}(\boldsymbol{\beta}_0, u)} \\ &\times\left\{\boldsymbol{H}_1(v) - \frac{H_{01}(v)}{h_0(v)}\boldsymbol{h}(v)\right\}\{\boldsymbol{Z}_1(v) - \boldsymbol{e}_H(\boldsymbol{\beta}_0, v)\} \\ &\times Y_1(v)\exp\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_1(v)\}d\Lambda_0(v)\Bigg]\end{aligned}$$

ψ can be consistently estimated by replacing the unknown quantities in (2.9) by their respective sample estimators. However, the authors suggested that this may not be the best choice if $H_{0i}(X_i) = 0$ for most of the nonzero Δ_i's. Thus, for the original case-cohort design, the authors recommended using the formula (2.8) proposed by Self and Prentice (1988) instead.

Chen and Lo (1999) improved the pseudolikelihood estimators by using a class of estimating equations based on the partial likelihood score function. Chen (2001) further improved the estimators by using a local type of average as weight in the estimating equations. Borgan et al. (2000) considered stratified case-cohort sampling designs and proposed several methods to analyze such study designs. Kulich and Lin (2004) developed a class of weighted estimating equations with time-dependent weights under the stratified case-cohort designs.

The nested case-control and case-cohort study designs have their own advantages. Either can lead to major cost savings relative to full-cohort approaches. A key advantage of the case-cohort design over the nested case-control study is its ability to use the same subcohort for several diseases or for subtypes of disease (e.g., Prentice, 1986; Wacholder et al., 1989; Langholz and Thomas, 1990; Wacholder et al., 1991). The availability of the case-cohort subcohort may be useful for study monitoring and can provide a natural comparison group at all disease occurrence times for each of the multiple study diseases. However, the choice between the two designs is less clear when only a single disease endpoint is to be studied. In the case-cohort design, a subcohort member serves in the comparison group for all cases in that individual's risk period, whereas, in the nested case-control design, a matched control does so only at the failure time(s) where the control is specifically selected. This fact can lead to a modest efficiency advantage for the case-cohort estimator in some circumstances (Kalbfleisch and Prentice, 2002). The nested case-control estimator has also some advantages over case-cohort designs. The method of analysis is simple and more readily understood. Langholz and Thomas (1991) showed that the nested case-control design may have greater efficiency than the case-cohort design when there is moderate random censoring or staggered entry into the cohort.

2.2 Correlated failure time data

The approaches discussed thus far assume the independence between failure times. In many biomedical studies, however, the independence between failure times might be violated. Such data may arise because study subjects having may be grouped in a manner that leads to dependencies within groups, or because individuals may experience multiple events. Such correlated failure time data can be mainly classified into two types : parallel and longitudinal. The failure times are unordered for parallel data, whereas the longitudinal data are sequentially ordered. We focus on the methods for parallel data in this section.

In the following subsections, we will summarize the marginal model approach which leaves the nature of dependence among related failure times completely unspecified and the frailty model approach which formulate the nature of dependence explicitly.

2.2.1 Marginal Models

Wei et al. (1989) proposed to model the marginal distribution of each failure time variable with a Cox-type proportional hazard model. In this approach, no particular structure of dependence among distinct failure times on each subject is imposed. The form of the hazard function for the kth type of failure time of the ith subject is given by

$$\lambda_k(t|\boldsymbol{Z}_{ki}) = \lambda_{k0}(t)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ki}(t)\},\ k = 1, \ldots, K,\ i = 1, \ldots, n$$

The kth failure-specific partial likelihood (Cox, 1975) is

$$L_k(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ki}(X_{ki})\}}{\sum_{l \in \mathcal{R}_k(X_{ki})} \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{li}(X_{ki})\}} \right]^{\Delta_{ki}}$$

where $\mathcal{R}_k(t) = \{i : X_{ki} \geq t\}$ is the set of subjects at risk just prior to time t with respect to the kth type of failure. The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}_k$ for $\boldsymbol{\beta}_k$ is defined as the solution to the likelihood equation $\partial \log L_k(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \boldsymbol{0}$. Under certain regularity conditions, $\hat{\boldsymbol{\beta}}$ is shown to be consistent for $\boldsymbol{\beta}_k$ and $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_1^T - \boldsymbol{\beta}_1^T, \ldots, \hat{\boldsymbol{\beta}}_K^T - \boldsymbol{\beta}_K^T)^T$ is shown to converge asymptotically to a zero mean multivariate normal random variable with covariance matrix

$\boldsymbol{Q}$ where

$$\hat{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{D}_{11}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_1) & \dots & \boldsymbol{D}_{11}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_K) \\ \vdots & & \vdots \\ \boldsymbol{D}_{K1}(\hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\beta}}_1) & \dots & \boldsymbol{D}_{KK}(\hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\beta}}_K) \end{bmatrix}$$

and where $\boldsymbol{D}_{kl}(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l)$ can be consistently estimated by

$$\begin{aligned}
\hat{\boldsymbol{D}}_{kl}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l) &= \hat{\boldsymbol{A}}_k^{-1}(\hat{\boldsymbol{\beta}}_k)\hat{\boldsymbol{B}}_{kl}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l)\hat{\boldsymbol{A}}_l^{-1}(\hat{\boldsymbol{\beta}}_l), \\
\hat{\boldsymbol{A}}_k(\boldsymbol{\beta}_k) &= n^{-1}\sum_{j=1}^{n} \Delta_{kj}\left[\frac{\boldsymbol{S}_k^{(2)}(\boldsymbol{\beta}_k, X_{kj})}{S_k^{(0)}(\boldsymbol{\beta}_k, X_{kj})} - \left(\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_k, X_{kj})}{S_k^{(0)}(\boldsymbol{\beta}_k, X_{kj})}\right)^{\otimes 2}\right], \\
\hat{\boldsymbol{B}}_{kl}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l) &= n^{-1}\sum_{j=1}^{n} \boldsymbol{W}_{kj}(\hat{\boldsymbol{\beta}}_k)\boldsymbol{W}_{lj}(\hat{\boldsymbol{\beta}}_l)^T, \\
\boldsymbol{W}_{kj}(\boldsymbol{\beta}_k) &= \left\{\Delta_{kj} - \sum_{m=1}^{n} \frac{\Delta_{km} Y_{kj}(X_{km}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{kj}(X_{km})\}}{n S_k^{(0)}(\boldsymbol{\beta}_k, X_{km})}\right\} \\
&\times \left\{\boldsymbol{Z}_{kj}(X_{km}) - \frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_k, X_{km})}{S_k^{(0)}(\boldsymbol{\beta}_k, X_{km})}\right\} \\
\text{and } \boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}_k, t) &= n^{-1}\sum_{i=1}^{n} Y_{ki}(t)\boldsymbol{Z}_{ki}^{\otimes d}(t)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ki}(t)\},\ d = 0, 1, 2
\end{aligned}$$

Lee et al. (1992) proposed to use similar approach for data that consist of large numbers of small groups of correlated failure time observations. The marginal hazard function $\lambda_{ik}(t)$ for the kth member in the ith stratum conditional on $\boldsymbol{Z}_{ik} = \boldsymbol{z}_{ik}$ has the usual proportional hazards form :

$$\lambda_0(t)\exp\{\boldsymbol{\beta}_0' \boldsymbol{z}_{ik}(t)\}$$

where $\lambda_0(t)$ is the common baseline hazard function. Under the independence working correlation assumption analogous to GEE for longitudinal data, the pseudo partial likelihood for the estimation of $\boldsymbol{\beta}_0$ is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left[\frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(X_{ik})\}}{\sum_{j=1}^{n}\sum_{m=1}^{K} Y_{jm}(X_{ik})\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{jm}(X_{ik})\}}\right]^{\Delta_{ik}}$$

The corresponding score function is given by

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left\{\boldsymbol{Z}_{ik}(t) - \frac{\sum_{m=1}^{K}\boldsymbol{S}_m^{(1)}(\boldsymbol{\beta}, X_{ik})}{\sum_{m=1}^{K}S_m^{(0)}(\boldsymbol{\beta}, X_{ik})}\right\} \tag{2.10}$$

The estimator $\hat{\boldsymbol{\beta}}$ is the solution to $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$. Under certain regularity conditions, $\hat{\boldsymbol{\beta}}$ is shown to be consistent for $\boldsymbol{\beta}_0$ and the distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is shown to converge asymptotically to a normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}(\boldsymbol{\beta}_0)$ which can be consistently estimated by a sandwich type covariance estimator, $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{B}}(\hat{\boldsymbol{\beta}})\boldsymbol{I}^{-1}(\hat{\boldsymbol{\beta}})$ where

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left[\frac{\sum_{m=1}^{K}\boldsymbol{S}_m^{(2)}(\boldsymbol{\beta}, X_{ik})}{S_m^{(0)}(\boldsymbol{\beta}, X_{ik})} - \left(\frac{\boldsymbol{S}_m^{(1)}(\boldsymbol{\beta}, X_{ik})}{S_m^{(0)}(\boldsymbol{\beta}, X_{ik})}\right)^{\otimes 2}\right], \\
\hat{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{m=1}^{K}\hat{\zeta}_{ik}\hat{\zeta}_{im}^{T},\ \hat{\zeta}_{ik} = \int_0^{\tau}\left\{\boldsymbol{Z}_{ik}(s) - \frac{\sum_{k=1}^{K}\boldsymbol{S}_k^{(1)}(\hat{\boldsymbol{\beta}}, s)}{\sum_{k=1}^{K}S_k^{(0)}(\hat{\boldsymbol{\beta}}, s)}\right\}d\hat{M}_{ik}(s) \\
\hat{M}_{ik}(s) &= N_{ik}(s) - \int_0^{s}Y_{ik}(u)\exp\{\boldsymbol{\beta}_0'\boldsymbol{Z}_{ik}(u)\}d\hat{\Lambda}(u) \text{ and} \\
\hat{\Lambda}(s) &= \int_0^{s}\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}dN_{ik}(u)}{\sum_{i=1}^{n}\sum_{k=1}^{K}Y_{ik}(u)\exp\{\hat{\boldsymbol{\beta}}^{T}\boldsymbol{Z}_{ik}(u)\}}
\end{aligned}$$

Liang et al. (1986) proposed another class of estimating functions for $\boldsymbol{\beta}_0$ under the proportional hazard model considered by Lee et al. (1992). Their estimating function is similar to (2.10), but they replaced $\sum_{k=1}^{K}\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}, t)/\sum_{k=1}^{K}S_k^{(0)}(\boldsymbol{\beta}, t)$, the conditional expected value of the covariate vector for an individual observed to fail at time t, with the average of all possible pairs collected from different clusters. Specifically, the score functions is given by

$$\sum_{i=1}^{n}\sum_{k=1}^{K}I(n_i(X_{ik}) > 0)\Delta_{ik}\left\{\boldsymbol{Z}_{ik}(X_{ik}) - n^{-1}(X_{ik})\sum_{j\neq i}\sum_{l}\boldsymbol{e}_{ik,jl}(\boldsymbol{\beta}, X_{ik})\right\},$$

where

$$\begin{aligned}
\boldsymbol{e}_{ik,jl}(\boldsymbol{\beta}, t) &= \frac{Y_{ik}(t)\boldsymbol{Z}_{ik}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\} + Y_{jl}(t)\boldsymbol{Z}_{jl}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_{jl}(t)\}}{Y_{ik}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\} + Y_{jl}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_{jl}(t)\}} \text{ and} \\
n_i(t) &= \sum_{j\neq i}\sum_{l}Y_{jl}(t)
\end{aligned}$$

Under some regularity conditions, the resulting estimator, $\tilde{\boldsymbol{\beta}}$, is shown to be consistent for $\boldsymbol{\beta}_0$ and the distribution of $n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is shown to converge asymptotically to a normal random variable with mean zero.

Spiekerman and Lin (1998) and Clegg et al. (1999) independently extended the models considered by Wei et al. (1989) and Lee et al. (1992) to a general model which includes these two as special cases. They also developed the large-sample theory for the resulting estimator of the regression parameter $\boldsymbol{\beta}_0$. Spiekerman and Lin (1998) also established the uniform consistency and joint weak convergence of the Aalen-Breslow type estimators for the cumulative baseline hazard functions.

All procedures discussed in this section thus far were based on the independence working model which weighs all observations equally. (Cai and Prentice, 1995, 1997) proposed weighted estimating equations to enhance the efficiency of the estimators for $\boldsymbol{\beta}_0$. They suggested the use of the inverse matrix of the correlation functions between counting process martingales. Their results indicated that the efficiency improvements for the resulting estimators are good if the correlations among failure times are high and censoring is not very heavy.

2.2.2 Frailty Models

The marginal model approach previously discussed in this section does not model the intra-subject correlation explicitly. When the interest resides in estimating the effect of risk factors and the correlation among the failure times are considered as a nuisance, the marginal model approach suits this purpose very well. However, in some settings, one might be interested in the strength and nature of dependencies among the failure time components. For such cases, the so-called frailty models have been proposed and studied by many authors.

The frailty model explicitly formulates the nature of the underlying dependence structure through an unobservable random variable. This unknown factor is usually called individual heterogeneity or frailty. The key assumption is that the failure times are conditionally independent given the value of the frailty. To illustrate this idea, consider a Cox proportional hazards model for subject i with respect to the kth event :

$$\lambda_{ik}(t|W_i) = w_i \lambda_0(t) \exp\{\boldsymbol{\beta}_0^T \boldsymbol{z}_{ik}(t)\} \tag{2.11}$$

where the frailty terms $\{W_i\}, i = 1, \ldots, n$ are assumed to be independent and to arise from a common parametric density. The commonly used one is the gamma distribution, mostly for mathematical convenience. Various choices are possible for this density, which include the positive stable distributions, the inverse Gaussian distributions and the log-normal distributions.

The parameter estimates are obtained through the EM algorithm, making use of the partial likelihood expression in the maximization step as shown in Klein (1992). An alternative approach is to use a penalized partial likelihood for the estimation of the shared frailty (Therneau and Grambsch, 2001).

Note that $\boldsymbol{\beta}_0$ in (2.11) generally needs to be interpreted conditionally on the unobserved frailty. There has been extensive debate over whether the unconditional specification of the marginal hazard approach or the conditional specification of the frailty model approach is more naturally related to the underlying mechanisms. The marginal model approach is model-free regarding dependence assumptions. Since we do not impose a specific model on the correlation structure, it is robust to the misspecification of correlation structure. This approach may be advantageous over the frailty model approach when the main purpose of the analysis is in finding the effect of covariates when the dependence is a nuisance. On the other hand, the frailty model approach is particularly sensible, when the purpose is to assess the dependence. Thus, the choice of the model depends on the goal of the specific study.

2.3 Case-control family studies

The literature reviewed in section 2.2 assume that the multivariate failure times data are from prospective studies assuming random samples. However, for the multivariate failure times data from the so-called case-control family studies, this assumption does not hold. Case-control family studies have been used to assess the familial aggregation of a disease and the relationship between the disease and genetic or environmental risk factors. Such a case-control family study identifies a sample of cases who develop the disease of interest and an independent sample of matched controls who are free of disease at the time of ascertainment. From each identified individual(proband), information collected includes disease outcomes(age of onset or age of censoring) and risk factors of the proband and the relatives. Due to this retrospective sampling of probands, methods for analyzing the correlated failure times data discussed thus far cannot be directly applied. The retrospective likelihood for the case-control family study can be constructed as follows. Consider a matched case-control family study where one case proband is matched in age with one control proband. Each matched set contains one case family and one control family, and there are a total of n matched set. Let i denote the family ($i = 1, \ldots, 2n$) and k denote the member in the family ($j = 0, \ldots, m_i$). Let the first n families be case families and the remaining families be control families. Δ_{ik} is an indicator variable for whether the individual developed the disease($\Delta_{ik} = 1$) or not($\Delta_{ik} = 0$), and t_{ik} denotes the age of onset if $\Delta_{ik} = 1$ and censoring time if $\Delta_{ik} = 0$. Let $(\boldsymbol{T}_i, \boldsymbol{\Delta}_i) = \{(T_{i0}, \ldots, T_{im_i}), (\Delta_{i0}, \ldots, \Delta_{im_i})\}$. The superscript -1 is used to denote a vector with its first component removed. Let $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i0}, \ldots, \boldsymbol{Z}_{im_i})$ denote the associated covariates for the ith family of size $m_i + 1$, with the first component in the vectors corresponding to the proband. Then, the retrospective likelihood for the case-control family study is given by

$$L = \prod_{i=1}^{2n} \Pr\{(\boldsymbol{T}_i^{-1}, \boldsymbol{\Delta}_i^{-1}, \boldsymbol{Z}_i | (T_{i1}, \Delta_{i1})\},$$

The likelihood can be factored as

$$L = \prod_{i=1}^{2n} \Pr\{\boldsymbol{Z}_{i0} | (T_{i0}, \Delta_{i0})\} \times \Pr\{\boldsymbol{Z}_i^{-1} | \boldsymbol{Z}_{i0}, (T_{i0}, \Delta_{i0})\} \times \Pr\{(\boldsymbol{T}_i^{-1}, \boldsymbol{\Delta}_i^{-1} | \boldsymbol{Z}_i, (T_{i0}, \Delta_{i0})\}$$

The second factor of the likelihood , $\Pr\{\boldsymbol{Z}_i^{-1} | \boldsymbol{Z}_{i0}, (T_{i0}, \Delta_{i0})\}$, can be ignored due to the reproducibility assumption for marginal models (Whittemore, 1995), i.e., $\Pr\{(T_{ik}, \Delta_{ik}) | \boldsymbol{Z}_i\} =$

$\Pr\{(T_{ik}, \Delta_{ik})|\boldsymbol{Z}_{ik}\}$. Thus, the likelihood can be reduced to

$$L = \prod_{i=1}^{2n} \Pr\{\boldsymbol{Z}_{i0}|(T_{i0}, \Delta_{i0})\} \times \Pr\{(\boldsymbol{T}_i^{-1}, \boldsymbol{\Delta}_i^{-1}|\boldsymbol{Z}_i, (T_{i0}, \Delta_{i0})\} \tag{2.12}$$

Li et al. (1998) proposed a parametric likelihood approach. They assumed the marginal distribution of ages of onset for each individual follows a proportional hazards model, which is given by

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\boldsymbol{\beta}_0'\boldsymbol{Z}\},$$

where some parametric model with a finite number of unknown parameters η for the baseline hazard function $\lambda_0(t)$ was assumed. Li *et al.*(1998) replaced the first part of the likelihood in (2.12) by the conditional likelihood of Prentice and Breslow (1978) to account for the matching. The Clayton model (Clayton, 1978) was used to specify the multivariate distribution of age of onset for the second part of the likelihood. Shih and Chatterjee (2002) extended this parametric model by allowing for a semiparametric modeling of $\Lambda_0(\cdot)$. They proposed a Nelson-Aalen type of estimator for the cumulative baseline hazard function. However, their approach required iterative procedure for the estimation of the parameters and the asymptotic properties of the resulting estimators are not yet developed. An alternative approach was proposed by Hsu et al. (2004). They studied the random effect or frailty model, where the term frailty represents the common unobserved risks shared by the family members. The objective of this approach is to make inference about individual families, while the marginal model coefficients (Li et al., 1998; Shih and Chatterjee, 2002) describe the effect of explanatory variables on the population average.

2.4 Additive hazards models

All the work discussed thus far was about the proportional hazards regression model, which assumes multiplicative risk models. The risk difference is another commonly used measure of association in epidemiology. The risk difference is more relevant to public health because it translates directly into the number of disease cases that would be avoided by eliminating a particular exposure (Kulich and Lin, 2000). When the risk difference is the measure of interest, the additive hazards model provides a useful alternative to the proportional hazards model. For example, in studies of excess risk, where the background risk and excess risk typically can have very different temporal forms, additive risk models seem to be biologically more plausible than proportional hazards models (Huffer and McKeague, 1999). The hazard function under the additive risk model for the failure time T associated with $\boldsymbol{Z}(\cdot)$ takes the form

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{\beta}_0^T \boldsymbol{Z}(t) \tag{2.13}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\beta}_0$ is a p-vector of regression parameter. Lin and Ying (1994) proposed an estimator for model (2.13) and derived the asymptotic properties. They proposed to estimate $\boldsymbol{\beta}_0$ from the following estimating function which mimics the partial likelihood score function for the proportional hazards model

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}\{dN_i(t) - Y_i(t)\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)dt\}, \qquad (2.14) \\ \text{where } \bar{\boldsymbol{Z}}(t) &= \sum_{j=1}^{n} Y_j(t)\boldsymbol{Z}_j(t) / \sum_{j=1}^{n} Y_j(t) \end{aligned}$$

We obtain $\hat{\boldsymbol{\beta}}$ by solving $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$ for $\boldsymbol{\beta}$, which has the explicit form

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} \int_0^{\tau} Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\} dN_i(t)\right]$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Under some regularity conditions, $\hat{\boldsymbol{\beta}}$ is shown to be consistent for $\boldsymbol{\beta}_0$ and the distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is shown to converge asymptotically to a p-variate normal with 0 and with a covariance matrix which can be consistently estimated by a sandwich type covariance estimator $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, where

$$\boldsymbol{A} = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dt, \;\; \boldsymbol{B} = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t)\}^{\otimes 2} dN_i(t)$$

The estimators for the cumulative baseline hazard $\Lambda_0(t)$ and the survival function $\mathrm{S}(t;\boldsymbol{z})$ were proposed and their asymptotic properties were provided. To ensure the monotonicity, modified estimators, $\hat{\Lambda}_0^*(t) = \max_{s\le t} \hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, s)$, $\hat{S}^*(t;\boldsymbol{z}) = \min_{s\le t} \hat{S}(s;\boldsymbol{z})$, have been proposed. These estimators were shown to be asymptotically equivalent to and preserve the asymptotic properties of the original estimators.

Kulich and Lin (2000) applied additive hazards model to case-cohort study. They proposed a weighted estimating equation which modified (2.14) as

$$\boldsymbol{U}_H(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho_i \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}_H(t)\}\{dN_i(t) - Y_i(t)\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)dt\}, \tag{2.15}$$

$$\text{where } \bar{\boldsymbol{Z}}_H(t) = \sum_{j=1}^{n} \rho_j Y_j(t)\boldsymbol{Z}_j(t) / \sum_{j=1}^{n} \rho_j Y_j(t),\ \rho_i = \Delta_i + (1-\Delta_i)\xi_i/p_i \text{ and } p_i = \Pr(\xi_i = 1)$$

The resulting estimator also has a closed form :

$$\hat{\boldsymbol{\beta}}_H = \left[\sum_{i=1}^{n} \rho_i \int_0^{\tau} Y_i(t)\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}_H(t)\}^{\otimes 2}dt\right]^{-1} \left[\sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}_H(t)\}dN_i(t)\right]$$

They considered two subcohort sampling settings : (i) independent Bernoulli sampling with arbitrary selection probabilities and (ii) stratified simple random sampling with fixed sample size. Under some regularity conditions, $n^{1/2}(\hat{\boldsymbol{\beta}}_H - \boldsymbol{\beta}_0)$ was shown to be asymptotically zero-mean normal and consistent estimators for the covariance matrices were proposed under both settings. Also, an estimator for the cumulative baseline hazard $\Lambda_0(t)$ is proposed and is given as

$$\hat{\Lambda}_{0H}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n \rho_j Y_j(s)} - \int_0^t \hat{\boldsymbol{\beta}}_H \bar{\boldsymbol{Z}}_H(s)ds$$

and it is also shown that $n^{1/2}(\hat{\Lambda}_{0H}(t) - \Lambda_0(t))$ converges weakly to a zero-mean Gaussian process on $[0,\tau]$ and the consistent estimator for the covariance matrix is proposed.

The additive hazards model has been applied to interval censored data by Lin et al. (1998) and Martinussen and Scheike (2002), to frailty models by Lin and Ying (1997), to cumulative incidence rates by Shen and Cheng (1999), and to competing risks analysis of the case-cohort studies by Sun et al. (2004). All this work has assumed mutual independence of the survival times.

Yin and Cai (2004) proposed a marginal additive hazards model approach for the correlated or clustered survival data. They proposed the following additive hazards model

$$\lambda_{ikl}(t; \boldsymbol{Z}_{ikl}) = \lambda_{0k}(t) + \boldsymbol{\beta}_{0k}^T \boldsymbol{Z}_{ikl}(t) \tag{2.16}$$

where k denotes the failure type, l denotes the subject and i denotes the cluster. An estimating function for $\boldsymbol{\beta}_{0k}$ assuming working independence has been proposed and is given by

$$\begin{aligned} \boldsymbol{U}_k(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{l=1}^{L}\int_0^{\tau}\{\boldsymbol{Z}_{ikl}(t) - \bar{\boldsymbol{Z}}_k(t)\}\{dN_{ikl}(t) - Y_{ikl}(t)\boldsymbol{\beta}^T\boldsymbol{Z}_{ikl}(t)dt\}, \quad (2.17)\\ \text{where } \bar{\boldsymbol{Z}}_k(t) &= \sum_{j=1}^{n}\sum_{l=1}^{L} Y_{jkl}(t)\boldsymbol{Z}_{jkl}(t) / \sum_{j=1}^{n}\sum_{l=1}^{L} Y_{jkl}(t) \end{aligned}$$

The resulting estimator, $\hat{\boldsymbol{\beta}}_k$, which is the solution to $\boldsymbol{U}_k(\boldsymbol{\beta}) = \boldsymbol{0}$ is given by

$$\hat{\boldsymbol{\beta}}_k = \left[\sum_{i=1}^{n}\sum_{l=1}^{L}\int_0^{\tau} Y_{ikl}(t)\{\boldsymbol{Z}_{ikl}(t) - \bar{\boldsymbol{Z}}_k(t)\}^{\otimes 2}dt\right]^{-1}\left[\sum_{i=1}^{n}\sum_{l=1}^{L}\int_0^{\tau}\{\boldsymbol{Z}_{ikl}(t) - \bar{\boldsymbol{Z}}_k(t)\}dN_{ikl}(t)\right]$$

Under some regularity conditions, as $n \to \infty$, $n^{1/2}\{(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})^T, \ldots, (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})^T\}^T$ is shown to converge in distribution to a zero-mean $(p \times K)$-dimensional normal random vector. A consistent estimator of the covariance matrix is proposed.

A natural estimator for the baseline cumulative hazard function for the kth failure type is proposed and is given by

$$\hat{\Lambda}_{0k}(t; \hat{\boldsymbol{\beta}}_k) = \int_0^t \frac{\sum_{i=1}^{n}\sum_{l=1}^{L} dN_{ikl}(s) - Y_{ikl}(s)\hat{\boldsymbol{\beta}}_k^T\boldsymbol{Z}_{ikl}(s)ds}{\sum_{i=1}^{n}\sum_{l=1}^{L} Y_{ikl}(s)}$$

The estimators of the cumulative hazard function and the survival function for a specific subject with the covariate vector $\boldsymbol{Z}_0(t)$ are proposed and are given by $\hat{\Lambda}_k(t; \hat{\boldsymbol{\beta}}_k, \boldsymbol{Z}_0) = \hat{\Lambda}_{0k}(t; \hat{\boldsymbol{\beta}}_k) + \int_0^t \hat{\boldsymbol{\beta}}_k^T \boldsymbol{Z}_0(u)du$ and $\hat{\mathrm{S}}_k(t; \boldsymbol{Z}_0) = \exp\{-\hat{\Lambda}_k(t; \hat{\boldsymbol{\beta}}_k^T, \boldsymbol{Z}_0)\}$. The asymptotic properties of these estimators are well established and the modified estimators, which ensure monotonicity, are also proposed and are given by $\hat{\Lambda}_k^*(t) = \max_{s \le t}\hat{\Lambda}_{0k}(s)$ and $\hat{\mathrm{S}}_k^*(t; \boldsymbol{Z}_0) =$

$\min_{s \le t} \hat{\mathrm{S}}^*_k(s; \boldsymbol{Z}_0)$ for $k = 1, \ldots, K$. It was shown that these modified estimators still preserve the asymptotic properties of the original estimators by similar arguments as in Lin and Ying (1994).

Pipper and Martinussen (2004) also proposed a marginal additive hazards model approach for clustered survival data. They also studied parametric shared frailty models to estimate measures of dependence between failure times in a cluster, as well as marginal parameters. To estimate both regression parameters and the association parameter in parametric shared frailty models with marginal additive hazards, new estimating equations which are derived by using the intensities in the observed filtration and the working independence estimators and mimicking the way in which Lin and Ying (1994) obtained their estimator of the regression parameter are proposed.

The additive and multiplicative risk models provide two major frameworks for studying the association between risk factors and disease occurrence or death. Most modern analyses of survival data focus on multiplicative models for relative risk using proportional hazards models, mostly due to desirable theoretical properties along with the simple interpretation of the results and the wide availability of computer programs. However, in many biomedical studies, proportional hazards assumption might not be valid or the investigators are more interested in risk differences than relative risks. In such cases, the additive risk model could be a practical alternative to the proportional hazards model. O'Neill (1986) has also shown that use of the proportional hazards model can result in serious bias when the additive hazards model is correct.

CHAPTER 3

MARGINAL HAZARDS REGRESSION FOR CASE-CONTROL WITHIN COHORT STUDIES WITH POSSIBLY CORRELATED FAILURE TIME DATA

3.1 Introduction

Case-control study design is an efficient and economic method to ascertain a large number of cases in a relatively short period of time. Often, the case-control study is conducted within a well-defined cohort. For example, in occupational epidemiology, a commonly used approach is to conduct a case-control study nested within a cohort that has already been enumerated. The reason for conducting a case-control study even when a cohort can be enumerated is usually that more information is needed than is readily available from records and it would be too expensive to seek this information for everyone in the cohort Rothman (2002). Thus, such a case-control study could greatly reduce the cost while achieving the same goals as a cohort study. Failure time models from such retrospective case-control studies have been studied in the literature (Thomas, 1977; Prentice and Breslow, 1978; Borgan et al., 1995; Binder, 1992; Samuelsen, 1997; Chen and Lo, 1999; Lin, 2000; Chen, 2001). An important assumption for these conventional case-control studies is the statistical independence among subjects. However, in many biomedical studies, this assumption might not hold. For example, in a retrospective cohort dental study (Caplan and Weintraub, 1997; Caplan et al., 2005), it was of interest to evaluate the degree to which pulpal involvement affects tooth survival. Root canal filled (RCF) teeth were used as an indicator of pulpal involvement. In this study, cases were defined as those who lost the RCF tooth, while controls were defined as those who did not lose the RCF tooth during the study period. After cases and controls were sampled, a non-RCF tooth was matched to the RCF tooth within each subject. The survival times of the two teeth within the same subject could be correlated and thus the independence assumption might not be valid. The primary goal of the study is to evaluate the effect of pulpal involvement on tooth survival. The fact that the survival times of the teeth from the same individual are correlated is considered as a nuisance. In such case, a marginal model approach is appealing. Examples like this one are very common in biomedical studies. For example, case-control family studies have been frequently used to assess familial aggregation of a disease and the relationship between the disease and genetic or environmental risk factors. In such studies, independent cases and controls are identified and information are collected for both cases and controls and their relatives. Since related individuals share common genetic or environmental factors, their failure times could be correlated.

There is an extensive literature of statistical methods for correlated failure time data from prospective cohort studies (Wei et al., 1989; Lee et al., 1992; Lin, 1994;Cai and Prentice, 1995,1997; Spiekerman and Lin, 1998; Clegg et al., 1999). However, these methods cannot be directly applied to correlated failure time data from case-control within cohort studies. Work for correlated failure time data from case-control within cohort studies has been limited. Some efforts have been made to analyze failure time data from case-control family studies (e.g.,Li et al., 1998; Shih and Chatterjee, 2002; Hsu et al., 1999; Hsu et al., 2004). For the case-control family studies, investigators are usually interested in estimating the strength of dependence of failure times within family. Consequently, most of the methods concentrated on frailty models or parametric approach, with the exception of Shih and Chatterjee (2002). In Shih and Chatterjee (2002), the authors considered a quasi-partial-likelihood approach for simultaneously estimating the parameters in the marginal proportional hazard model and the association among family members. However, the asymptotic properties of the proposed estimator were not clear and estimation of the variance of their estimator relied on a bootstrap method. When the correlation of the failure times is not of interest, as in the aforementioned dental study, statistical inference procedure that is easy to conduct and has nice asymptotic properties remains to be developed. It is desirable to develop hazard regression models for the correlated failure time data from case-control within cohort studies which account for the possible correlation within subject while avoiding the specification of the correlation structure.

For univariate failure time data from complex sampling designs, Binder (1992) proposed an estimating equation approach for fitting Cox's proportional hazards models for complex survey data and Lin (2000) studied the theoretical aspects of estimating procedures by Binder (1992) and extended it to the super-population approach. The sampling weights used in Binder (1992) are proportional to the inverse of the sampling probability. Samuelsen (1997) considered the same type of weights when nested case-control design is involved and Chen (2001) proposed a more efficient estimator by using different forms of weights. All these methods assume independent failure times and cannot be directly applied to multivariate failure time data.

In this chapter, we propose a weighted estimating equation approach for estimating the parameters in the marginal hazards regression models for the correlated failure time data from case-control studies within cohort. The rest of this chapter is organized as follows. The proposed model and method of estimation are presented in Section 3.2, followed by the study of the asymptotics in Section 3.3. In Section 3.4, we report some simulation results. The methodology is illustrated in Section 3.5 using the aforementioned dental study. In Section 3.6, we give a few concluding remarks.

3.2 Modeling and Estimation

3.2.1 Marginal hazards Model

Let i indicate cluster and k indicate member within cluster. Let T_{ik} denote the failure time for member k of cluster i. In the aforementioned retrospective dental study example,

i would indicate the patient, k would indicate the tooth within the patient, and T_{i1}, T_{i2} would represent the failure time of the index tooth and the matching tooth, respectively, for patient i. Let C_{ik} denote the censoring time. We assume that C_{ik} is independent of the failure process conditional on covariates. Without loss of generality, we assume that there are K members in each cluster. Varying cluster sizes can be accommodated by setting the corresponding C_{ik} to be equal to zero. In many practical cases, $C_{ik} = C_i$ for $k = 1, \ldots, K$. The observed time is $X_{ik} = \min(T_{ik}, C_{ik})$ and $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ is an indicator for failure. Note that the 'at risk' indicator process is given by $Y_{ik}(t) = I(X_{ik} \geq t)$ for member k of cluster i and let $N_{ik} = I(X_{ik} \leq t,\ \Delta_{ik} = 1)$ denote the counting process corresponding to T_{ik}. Let $\lambda_{ik}(t)$ and $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) \exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\}\lambda_0(u)du$ denote the corresponding marginal hazards function and a martingale with respect to the marginal filtration $\mathcal{F}_{ik}(t) = \sigma\{N_{ik}(s), Y_{ik}(s), \boldsymbol{Z}_{ik}(s) :\ 0 \leq s \leq t\}$. Note that $M_{ik}(t)$ are not martingales with respect to the joint filtration generated by all of the failure, censoring, and covariate history up to time t, $\mathcal{F}(t) = \vee_{i=1}^{n} \vee_{k=1}^{K} \mathcal{F}_{ik}(t)$, due to the intraclass dependence. Let τ denote the study end time.

Suppose that T_{ik} arises from a marginal intensity process model of the form (Lee et al., 1992)

$$\lambda_{ik}(t) = Y_{ik}(t)\lambda_0(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)\}, \tag{3.1}$$

where $\boldsymbol{Z}_{ik}(t) = (Z_{1ik}(t), \ldots, Z_{pik}(t))'$ is a p-dimensional vector of covariates for member k of cluster i, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed and unknown parameters. We assume that all the time-dependent covariates in $Z_{ik}(t)$ are "external", i.e., they are not affected by the disease processes, as described by (Kalbfleisch and Prentice (2002)).

3.2.2 Estimation of Regression Parameters and Cumulative Baseline Hazard Function

Under the case-control within cohort study design, suppose we select $\widetilde{n}_1$ cases and $\widetilde{n}_0$ controls from the n_1 cases and n_0 controls, respectively, in the population. Let $n = n_1 + n_0$ and $\widetilde{n} = \widetilde{n}_1 + \widetilde{n}_0$. Each case (control) has the same probability $\widetilde{n}_1/n_1$ ($\widetilde{n}_0/n_0$) to be selected.

Let π_i denote this inclusion probability for the ith cluster and ξ_i denote the indicator for being selected. The inclusion probability π_i is allowed to depend on $\mathcal{F}(\tau)$. We will refer to these cases and controls as the index member and use $k = 1$ to indicate them. Note that by the study design described in the previous section, K members in the ith stratum have the same inclusion statuses, *i.e.*, $\pi_{ik} = \pi_i$ and $\xi_{ik} = \xi_i$ for $k = 1, \ldots, K$.

We propose the following weighted estimating equations for estimating $\boldsymbol{\beta}_0$:

$$\widehat{\boldsymbol{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_{ik}(t) = \mathbf{0}, \tag{3.2}$$

where

$$w_i = \frac{\xi_i}{\pi_i},\ \widehat{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}\ (d = 0, 1),$$

and $\quad \boldsymbol{a}^{\otimes 0} = 1, \boldsymbol{a}^{\otimes 1} = \boldsymbol{a},$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$ for a vector $\boldsymbol{a}$.

It is assumed that $\pi_i > 0$ for all i.

Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. To predict the t-year survival probability for future patients with specific covariates, a Breslow-Aalen type estimator of the baseline cumulative hazard function is proposed and is given by:

$$\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, t) = \int_0^t \frac{\sum_{i=1}^{n}\sum_{k=1}^{K} w_i dN_{ik}(s)}{\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(s) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{ik}(s)}} \text{ where } w_i = \frac{\xi_i}{\pi_i}. \tag{3.3}$$

Note that when $K = 1$, *i.e.* when failure time data are from traditional case-control studies without correlated components from the same cluster, the proposed estimators reduce to the ones studied by Binder (1992) and Lin (2000) for complex survey data for univariate failure time. When all the subjects are sampled, i.e. $\xi_i = 1$, $\pi_i = 1$, $i = 1, \ldots, n$, the proposed estimators reduce to the one studied by Lee et al. (1992) for random samples.

Suppose the information on the observed failure times of all the cohort members are available. Under such situation, using only the inclusion probability π_i might not be efficient since it does not fully use the available information. Note that calculation of π_i only requires the cohort size n_0, n_1 and the sample size $\widetilde{n}_0, \widetilde{n}_1$. Thus, in an attempt to increase the efficiency of the estimator, a different type of weight which uses all the available information is desired. To this end, we consider a local average estimator. The idea of the local average estimator is to replace each missing covariate term by an appropriate local average. This estimator was considered by Chen (2001) for independent data. We propose the following weighted estimating equations for estimating $\boldsymbol{\beta}_0$. The form of the weighted estimating equations is the same as the one with inclusion probabilities except that we replace π_i with a local average. Specifically,

$$\widehat{\boldsymbol{U}}_c(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}_c^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}_c^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_{ik}(t) = \mathbf{0}, \tag{3.4}$$

where

$$w_i = \frac{\xi_i}{r_n(X_{i1}, \Delta_{i1})}, \ \widehat{\boldsymbol{S}}_c^{(d)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}, \ (d = 0, 1)$$

and

$$\begin{aligned} r_n(t, d) &= \frac{\sum_{j=1}^{n} \Delta_{j1}\xi_j I\{X_{j1} \in [t_{l-1}, t_l)\}}{\sum_{j=1}^{n} \Delta_{j1} I\{X_{j1} \in [t_{l-1}, t_l)\}} \qquad \text{if } d = 1 \text{ and } t \in [t_{l-1}, t_l)\,, \\ &= \frac{\sum_{j=1}^{n} (1 - \Delta_{j1})\,\xi_j I\{X_{j1} \in [s_{m-1}, s_m)\}}{\sum_{j=1}^{n} (1 - \Delta_{j1})\, I\{X_{j1} \in [s_{m-1}, s_m)\}} \quad \text{if } d = 0 \text{ and } t \in [s_{m-1}, s_m)\,. \end{aligned}$$

for some $1 \leq l \leq a_n$ and $1 \leq m \leq b_n$ where $0 = t_0 \leq t_1 \leq \ldots \leq t_{a_n} = \tau$ and $0 = s_0 \leq s_1 \leq \ldots \leq s_{b_n} = \tau$ are two partitions of $[0, \tau)$. With additional assumptions, the estimator using this weight function is expected to be more efficient than the previous one using the inclusion probability in the sense that the former results in a parameter estimator with smaller asymptotic variance. Note that as pointed out by Samuelsen et al., 2005, this local average method can be described by a procedure called "Post-stratification" in survey sampling literature. Specifically, after cases and controls are sampled, we divide the cohort as well as the sampled data into strata constructed from using the additional information (in this case, the failure times and censoring times for all the cohort members). Then, we construct the weighted estimating functions as if the data were collected originally by stratified case-control sampling. The Breslow-Aalen type estimator of the baseline cumulative hazard function $\widehat{\Lambda}_0^c(\widehat{\boldsymbol{\beta}}_c, t)$ will be in the form of (3.3) with $w_i = \xi_i / r_n(X_{i1}, \Delta_{i1})$.

3.3 Asymptotic Properties

In this section, we describe the asymptotic properties of the proposed estimates. We introduce the following notation for convenience:

$$
\begin{aligned}
&\boldsymbol{S}^{(d)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} Y_{ik}(t) \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}, \\
&\boldsymbol{s}^{(d)}(\boldsymbol{\beta}, t) = \mathrm{E}\{\boldsymbol{S}^{(d)}(\boldsymbol{\beta}, t)\} \ (d = 0, 1, 2), \quad \boldsymbol{e}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}, \\
&\boldsymbol{v}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t) s^{(0)}(\boldsymbol{\beta}, t) - \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}, t)^2}, \ \widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) = \boldsymbol{Z}_{ik}(t) - \boldsymbol{e}(\boldsymbol{\beta}, t), \text{ and} \\
&\mathbf{M}_{\widetilde{\boldsymbol{z}}, ik}(\boldsymbol{\beta}) = \int_0^\tau \widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) dM_{ik}(t)
\end{aligned}
$$

We assume the following set of conditions hold :

(A) $(\boldsymbol{T}_i, \boldsymbol{C}_i, \boldsymbol{Z}_i), i = 1, \ldots, n$ are independent and identically distributed.

(B) $\Pr(Y(\tau) > 0) > 0$.

(C) $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(u)| < C_z < \infty$ almost surely for some constant C_z.

(D) The matrix $\boldsymbol{A}(\boldsymbol{\beta}_0) = \int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, t) \lambda_0(t) dt$ is positive definite.

Note that the conditions $(A) - (D)$ entail the following conditions (E) - (H):

(E) (Finite interval) $\int_0^\tau \lambda_0(t) dt < \infty$.

(F) (Asymptotic stability) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ that satisfies the following conditions, as $n \to \infty$,

(i) there exists scalar, vector and matrix functions $s^{(0)}, \boldsymbol{s}^{(1)}$ and $\boldsymbol{s}^{(2)}$ defined on $\mathcal{B} \times [0, \tau]$ such that for $d = 0, 1, 2$, $\sup_{\substack{t \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} ||\boldsymbol{S}^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{s}^{(d)}(\boldsymbol{\beta}, t)|| \xrightarrow{p} 0$;

(ii) there exists a matrix $\boldsymbol{Q}(\boldsymbol{\beta})$ such that $n^{-1}\sum_{i=1}^{n}\operatorname{Var}(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{Z}},ik}(\boldsymbol{\beta}_0)) \longrightarrow \boldsymbol{Q}(\boldsymbol{\beta}_0)$.

(G) (Asymptotic regularity) For all $\boldsymbol{\beta} \in \mathcal{B},\ t \in [0,\tau]:\ \boldsymbol{s}^{(1)}(\boldsymbol{\beta},t) = \frac{\partial}{\partial \boldsymbol{\beta}} s^{(0)}(\boldsymbol{\beta},t), \boldsymbol{s}^{(2)}(\boldsymbol{\beta},t) = \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} s^{(0)}(\boldsymbol{\beta},t)$ where $\boldsymbol{s}^{(d)}(\cdot,t)$ $(d=0,1,2)$ are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $t \in [0,\tau]$ and are bounded on $\mathcal{B} \times [0,\tau]$, $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0,\tau]$.

(H) (Lindeberg condition) There exists a $\delta > 0$ $s.t.$ as $n \to \infty$

$$n^{-1/2} \sup_{i,k,t} \|\boldsymbol{Z}_{ik}(t)\| Y_{ik}(t) I\left\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t) > -\delta \|\boldsymbol{Z}_{ik}(t)\|\right\} \xrightarrow{p} 0.$$

The following additional conditions are also needed to ensure the desired asymptotic convergence of case-control samples:

(I) (Nontrivial samples) For $s = 0,1$ as $n \to \infty$,

(i) $\frac{\widetilde{n}_s}{n_s}$ converges to a constant $\alpha_s \in (0,1)$ where α_s is the realization of a function $\alpha(W)$ of a random variable W evaluated at $W = s$, i.e. $\alpha(W)|_{W=s} = \alpha_s$;

(ii) $\frac{\widetilde{n}_{sh}}{n_s}$ converges to a constant $w_{sh} \in (0,1)$ for all $h = 1,\ldots,H_s$ where H_s is the number of post-stratified groups in sth stratum.

(J) (Nontrivial cases) $\frac{n_s}{n}$ converges to a constant $p_s \in [0,1]$ for $s = 0,1$ as $n \to \infty$ where $p_1 + p_0 = 1$.

(K) (Asymptotic normality of samples) For all $k = 1,\ldots,K$, as $n \to \infty$,

$$n^{-1} \sup_{i,t} \exp\left\{2\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)\right\} \xrightarrow{p} 0,\ n^{-1} \sup_{i,t} \|\boldsymbol{Z}_{ik}(t)\|^2 \exp\left\{2\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)\right\} \xrightarrow{p} 0$$

(L) (Asymptotic stability) For $d = 0, 1, 2$, as $n \to \infty$,

(i) $\sup_{t,\boldsymbol{\beta}} \left\| \widehat{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta},\ t) - \boldsymbol{s}^{(d)}(\boldsymbol{\beta},\ t) \right\| \xrightarrow{p} 0,$ and there exists a positive-definite matrix $\boldsymbol{V}^*(\boldsymbol{\beta})$ such that $\mathrm{Var}\left(n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0) \middle| \Delta_{i1} \right) \longrightarrow \boldsymbol{V}^*(\boldsymbol{\beta}_0)$

(ii) $\sup_{t,\boldsymbol{\beta}} \left\| \widehat{\boldsymbol{S}}_c^{(d)}(\boldsymbol{\beta},\ t) - \boldsymbol{s}^{(d)}(\boldsymbol{\beta},\ t) \right\| \xrightarrow{p} 0,$ and there exists a positive-definite matrix $\boldsymbol{V}_c^*(\boldsymbol{\beta})$ such that $\mathrm{Var}\left(n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0) \middle| X_{i1}, \Delta_{i1} \right) \longrightarrow \boldsymbol{V}_c^*(\boldsymbol{\beta}_0)$

Here and in what follows $\|\cdot\|$ is the Euclidean norm for vectors or matrices.

3.3.1 Asymptotic Properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}(\widehat{\boldsymbol{\beta}}, t)$

We summarize the asymptotic behavior of the regression parameter estimator under the inclusion probability approach in the following theorem :

Theorem 3.1 *Under the regularity conditions (A) - (L), $\widehat{\boldsymbol{\beta}}$ solving (3.2) is a consistent estimator of $\boldsymbol{\beta}_0$. Also $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \boldsymbol{V}(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ where*

$$\boldsymbol{Q}(\boldsymbol{\beta}) = \mathrm{E}\left[\left(\sum_{k=1}^{K} \boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\boldsymbol{\beta}) \right)^{\otimes 2} \right],\ \boldsymbol{V}(\boldsymbol{\beta}) = \mathrm{E}\left[\frac{1 - \alpha(\Delta_{11})}{\alpha(\Delta_{11})} \mathrm{Var}\left(\sum_{k=1}^{K} \boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\boldsymbol{\beta}) \middle| \Delta_{11} \right) \right]$$

Note that $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ has two sources of variations: $\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{Q}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the cohort and $\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{V}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the case-control sample from the cohort.

$\boldsymbol{A}(\boldsymbol{\beta}_0)$, $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\boldsymbol{V}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}), \widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\beta}})$ and $\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\beta}})$ where

$$
\begin{aligned}
\widehat{\boldsymbol{A}}(\boldsymbol{\beta}) &= -n^{-1}\frac{\partial \widehat{\boldsymbol{U}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} w_i \left(\sum_{k=1}^{K}\widehat{\mathbf{M}}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2}, \\
\widehat{\boldsymbol{V}}(\boldsymbol{\beta}) &= \sum_{s=0}^{1}\widehat{p}_s\frac{1-\widehat{\alpha}_s}{\widehat{\alpha}_s}\widehat{\mathrm{Var}}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\middle|\,\Delta_{11}=s\right), \\
\widehat{\mathbf{M}}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) &= \Delta_{ik}\left\{\boldsymbol{Z}_{ik}(X_{ik}) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta};X_{ik})}{\widehat{S}^{(0)}(\boldsymbol{\beta};X_{ik})}\right\} \\
&- n^{-1}\sum_{j=1}^{n} w_j \sum_{l=1}^{K}\frac{\Delta_{jl}Y_{ik}(X_{jl})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jl})}}{\widehat{S}^{(0)}(\boldsymbol{\beta};X_{jl})}\left\{\boldsymbol{Z}_{ik}(X_{jl}) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta};X_{jl})}{\widehat{S}^{(0)}(\boldsymbol{\beta};X_{jl})}\right\}, \\
\widehat{p}_s &= \frac{n_s}{n},\ \widehat{\alpha}_s = \frac{\widetilde{n}_s}{n_s},\ w_i = \frac{\xi_i}{\pi_i} \quad \text{and}
\end{aligned}
$$

$\widehat{\mathrm{Var}}\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\middle|\,\Delta_{11}=s\right)$ is a sample variance of $\left\{\sum_{k=1}^{K}\widehat{\mathbf{M}}_{\widetilde{\boldsymbol{z}},1k}(\widehat{\boldsymbol{\beta}})\middle|\,\Delta_{11}=s\right\}$ for $s=0,1$.

We summarize the asymptotic properties of $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)$ in the following theorem :

Theorem 3.2 *Under the regularity conditions (A) - (L), $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)$ is a consistent estimator of $\Lambda_0(t)$. Also, $n^{1/2}\left(\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)-\Lambda_0(t)\right)$ converges weakly to a zero-mean Gaussian process with covariance function $\phi(t_1,t_2)(\boldsymbol{\beta}_0)+\sigma(t_1,t_2)(\boldsymbol{\beta}_0)$ at (t_1,t_2) where*

$$
\begin{aligned}
\phi(t_1,t_2)(\boldsymbol{\beta}) &= \mathrm{E}\left[\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1)\right)\left(\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\right)\right], \\
\sigma(t_1,t_2)(\boldsymbol{\beta}) &= \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Cov}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\middle|\,\Delta_{11}\right)\right], \\
\phi_{ik}(\boldsymbol{\beta},t) &= \int_0^t \frac{dM_{ik}(u)}{s^{(0)}(\boldsymbol{\beta},u)} + \boldsymbol{r}(\boldsymbol{\beta},t)^T\boldsymbol{A}^{-1}(\boldsymbol{\beta})\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}), \quad \text{and} \\
\boldsymbol{r}(\boldsymbol{\beta},t) &= -\int_0^t \boldsymbol{e}(\boldsymbol{\beta},u)d\Lambda_0(u).
\end{aligned}
$$

$\phi(t_1,t_2)(\boldsymbol{\beta}_0)$ and $\sigma(t_1,t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}(t_1,t_2)(\widehat{\boldsymbol{\beta}})$ and $\widehat{\sigma}(t_1,t_2)(\widehat{\boldsymbol{\beta}})$ where

$$
\begin{aligned}
\widehat{\phi}(t_1,t_2)(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n} w_i \left(\sum_{k=1}^{K}\widehat{\phi}_{ik}(\boldsymbol{\beta},t_1)\right)\left(\sum_{m=1}^{K}\widehat{\phi}_{im}(\boldsymbol{\beta},t_2)\right),\\
\widehat{\sigma}(t_1,t_2)(\boldsymbol{\beta}) &= \sum_{s=0}^{1}\widehat{p}_s\frac{1-\widehat{\alpha}_s}{\widehat{\alpha}_s}\widehat{\mathrm{Cov}}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\middle|\,\Delta_{11}=s\right),\\
\widehat{\phi}_{ik}(\boldsymbol{\beta},t) &= \int_0^t \frac{d\widehat{M}_{ik}(u)}{\widehat{S}^{(0)}(\boldsymbol{\beta},u)} + \boldsymbol{R}(\boldsymbol{\beta},t)^T\widehat{\boldsymbol{A}}^{-1}(\boldsymbol{\beta})\widehat{\mathbf{M}}_{\tilde{\boldsymbol{Z}},ik}(\boldsymbol{\beta}),\\
\int_0^t \frac{d\widehat{M}_{ik}(u)}{\widehat{S}^{(0)}(\boldsymbol{\beta},u)} &= \frac{\Delta_{ik}I(X_{ik}\le t)}{\widehat{S}^{(0)}(\boldsymbol{\beta},X_{ik})}\\
&- n^{-1}\sum_{j=1}^{n}w_j\sum_{l=1}^{K}\frac{\Delta_{jl}I(X_{jl}\le t)Y_{ik}(X_{jl})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jl})}}{\widehat{S}^{(0)}(\boldsymbol{\beta},X_{jl})^2},\\
\boldsymbol{R}(\boldsymbol{\beta},t) &= -n^{-1}\sum_{i=1}^{n}w_i\sum_{l=1}^{K}\frac{\Delta_{il}I(X_{il}\le t)\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta},X_{il})}{\widehat{S}^{(0)}(\boldsymbol{\beta},X_{il})^2} \quad \text{and}
\end{aligned}
$$

$\widehat{\mathrm{Cov}}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\middle|\,\Delta_{11}=s\right)$ is a sample covariance for $\left\{\left(\sum_{k=1}^{K}\widehat{\phi}_{1k}(\widehat{\boldsymbol{\beta}},t_1),\sum_{m=1}^{K}\widehat{\phi}_{1m}(\widehat{\boldsymbol{\beta}},t_2)\right)\middle|\,\Delta_{11}=s\right\}$ for $s=0,1$.

3.3.2 Asymptotic Properties of $\widehat{\boldsymbol{\beta}}_c$ and $\widehat{\Lambda}^c(\widehat{\boldsymbol{\beta}}_c,t)$

Now, we describe the asymptotic properties of the model parameter estimators under the local average approach. The asymptotic variance-covariance matrix is shown to have the form of a proportionally allocated stratified sample. This is due to the post-stratification argument when the original sampling is either simple random sampling or stratified simple random sampling (Cochran, 1977). Our original sampling scheme is a stratified simple random sampling where the strata are defined by case-status. Thus, we do not need the additional assumptions imposed in Chen (2001) for local average method while those assumptions are needed for other sampling schemes such as nest case-control sampling (Samuelsen et al., 2005). We summarize the asymptotic behavior of the regression parameter estimator and $\widehat{\Lambda}_0^c(\widehat{\boldsymbol{\beta}}_c,t)$ under the local average approach in the following two theorems:

Theorem 3.3 *Under the regularity conditions (A) - (L), $\widehat{\boldsymbol{\beta}}_c$ solving (3.4) is a consistent estimator of $\boldsymbol{\beta}_0$. Also, $n^{1/2}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}_c(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \boldsymbol{V}_c(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ where*

$$\boldsymbol{V}_c(\boldsymbol{\beta}) = \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Var}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\middle| X_{11}, \Delta_{11}\right)\right]$$

Note that $\boldsymbol{V}_c(\boldsymbol{\beta}_0)$ is not larger than $\boldsymbol{V}(\boldsymbol{\beta}_0)$ since

$$\begin{aligned}\boldsymbol{V}_c(\boldsymbol{\beta}_0) &= \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{E}\left\{\mathrm{Var}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\middle| X_{11}, \Delta_{11}\right)\middle|\Delta_{11}\right\}\right]\\ &\le \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Var}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\middle|\Delta_{11}\right)\right]\end{aligned}$$

Hence, $\boldsymbol{\Sigma}_c(\boldsymbol{\beta}_0)$ is not larger than $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$. $\boldsymbol{A}(\boldsymbol{\beta}_0), \boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\boldsymbol{V}_c(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}_c(\widehat{\boldsymbol{\beta}}_c)$, $\widehat{\boldsymbol{Q}}_c(\widehat{\boldsymbol{\beta}}_c)$ and $\widehat{\boldsymbol{V}}_c(\widehat{\boldsymbol{\beta}}_c)$ where

$$\begin{aligned}\widehat{\boldsymbol{A}}_c(\boldsymbol{\beta}) &= -n^{-1}\frac{\partial\widehat{\boldsymbol{U}}_c(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}_c(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}w_i\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}^c_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2},\\ \widehat{\boldsymbol{V}}_c(\boldsymbol{\beta}) &= \sum_{s=0}^{1}\sum_{h=1}^{H_s}\widehat{w}_{sh}\frac{1-\widehat{\alpha}_s}{\widehat{\alpha}_s}\widehat{\mathrm{Var}}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})|X_{11}=h, \Delta_{11}=s\right),\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{Var}}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})|X_{11}=h,\Delta_{11}=s\right) &= n^{-1}\sum_{i=1}^{n}w_i\\ \times\left[\sum_{k=1}^{K}\widehat{\mathbf{M}}^c_{\widetilde{\boldsymbol{z}},ik}(\widehat{\boldsymbol{\beta}}_c)\right. &- \left.\widehat{\mathrm{E}}\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\middle| X_{i1}=h, \Delta_{i1}=s\right)\right]^{\otimes 2},\end{aligned}$$

$$\widehat{\mathbf{M}}^c_{\widetilde{\boldsymbol{z}},il}(\boldsymbol{\beta}) = \Delta_{il}\left\{\boldsymbol{Z}_{il}(X_{il}) - \frac{\widehat{\boldsymbol{S}}^{(1)}_c(\boldsymbol{\beta}, X_{il})}{\widehat{S}^{(0)}_c(\boldsymbol{\beta}, X_{il})}\right\} - n^{-1}\sum_{j=1}^{n}w_j$$

$$\times\ \sum_{k=1}^{K}\frac{\Delta_{jk}Y_{il}(X_{jk})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{il}(X_{jk})}}{\widehat{S}^{(0)}_c(\boldsymbol{\beta}, X_{jk})}\left\{\boldsymbol{Z}_{il}(X_{jk}) - \frac{\widehat{\boldsymbol{S}}^{(1)}_c(\boldsymbol{\beta}, X_{jk})}{\widehat{S}^{(0)}_c(\boldsymbol{\beta}, X_{jk})}\right\},\ w_i = \frac{\xi_i}{r_n(X_{i1}, \Delta_{i1})}\ \text{and}$$

$$\widehat{\mathrm{E}}\left[\left(\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{Z}},ik}(\boldsymbol{\beta}_0)\middle| X_{i1}=h,\Delta_{i1}=s\right)\right] \text{ is a local average of}$$

$$\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}^c_{\widetilde{\boldsymbol{Z}},ik}(\widehat{\boldsymbol{\beta}}_c)\middle| X_{i1}=h,\Delta_{i1}=s\right), i=1,\ldots,n \text{ using the partitions.}$$

Theorem 3.4 *Under some regularity conditions (A) - (L), $\widehat{\Lambda}_0^c(\widehat{\boldsymbol{\beta}}_c, t)$ is a consistent estimator of $\Lambda_0(t)$. Also, $n^{1/2}\left(\widehat{\Lambda}_0^c(\widehat{\boldsymbol{\beta}}_c, t)-\Lambda_0(t)\right)$ converges weakly to a zero-mean Gaussian process with covariance function $\phi(t_1,t_2)(\boldsymbol{\beta}_0)+\sigma_c(t_1,t_2)(\boldsymbol{\beta}_0)$ at (t_1,t_2) where*

$$\sigma_c(t_1,t_2)(\boldsymbol{\beta}) = \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Cov}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\middle| X_{11},\Delta_{11}\right)\right]$$

$\phi(t_1,t_2)(\boldsymbol{\beta}_0)$ and $\sigma_c(t_1,t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}_c(t_1,t_2)(\widehat{\boldsymbol{\beta}}_c)$ and $\widehat{\sigma}_c(t_1,t_2)(\widehat{\boldsymbol{\beta}}_c)$ where

$$\begin{aligned}
\widehat{\phi}_c(t_1,t_2)(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n} w_i\left(\sum_{k=1}^{K}\widehat{\phi}^c_{ik}(\boldsymbol{\beta},t_1)\right)\left(\sum_{l=1}^{K}\widehat{\phi}^c_{il}(\boldsymbol{\beta},t_2)\right),\\
\widehat{\sigma}_c(t_1,t_2)(\boldsymbol{\beta}) &= \sum_{s=0}^{1}\sum_{h=1}^{H_s}\widehat{w}_{sh}\frac{1-\widehat{\alpha}_s}{\widehat{\alpha}_s}\widehat{\mathrm{Cov}}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)|X_{11}=h,\Delta_{11}=s\right),\\
\widehat{\phi}^c_{ik}(\boldsymbol{\beta},t) &= \int_0^t \frac{d\widehat{M}^c_{ik}(u)}{\widehat{S}^{(0)}_c(\boldsymbol{\beta},u)} + \boldsymbol{R}_c(\boldsymbol{\beta},t)^T\widehat{\boldsymbol{A}}_c^{-1}(\boldsymbol{\beta})\widehat{\mathbf{M}}^c_{\widetilde{\boldsymbol{Z}},ik}(\boldsymbol{\beta}),\\
\int_0^t \frac{d\widehat{M}^c_{ik}(u)}{\widehat{S}^{(0)}_c(\boldsymbol{\beta},u)} &= \frac{\Delta_{ik}I(X_{ik}\le t)}{\widehat{S}^{(0)}_c(\boldsymbol{\beta},X_{ik})}\\
&- n^{-1}\sum_{j=1}^{n}w_j\sum_{l=1}^{K}\frac{\Delta_{jl}I(X_{jl}\le t)Y_{ik}(X_{jl})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jl})}}{\widehat{S}^{(0)}_c(\boldsymbol{\beta},X_{jl})^2},\\
\boldsymbol{R}_c(\boldsymbol{\beta},t) &= -n^{-1}\sum_{i=1}^{n}w_i\sum_{l=1}^{K}\frac{\Delta_{il}I(X_{il}\le t)\widehat{\boldsymbol{S}}^{(1)}_c(\boldsymbol{\beta},X_{il})}{\widehat{S}^{(0)}_c(\boldsymbol{\beta},X_{il})^2} \quad \text{and}\\
&\widehat{\mathrm{Cov}}\left(\sum_{k=1}^{K}\phi_{1k}(\boldsymbol{\beta},t_1),\sum_{m=1}^{K}\phi_{1m}(\boldsymbol{\beta},t_2)\middle| X_{11}=h,\Delta_{11}=s\right) = n^{-1}\sum_{i=1}^{n}w_i\\
&\times \left[\sum_{k=1}^{K}\widehat{\phi}^c_{ik}(\widehat{\boldsymbol{\beta}},t_1)-\widehat{\mathrm{E}}\left(\sum_{k=1}^{K}\phi_{ik}(\boldsymbol{\beta},t_1)\middle| X_{i1}=h,\Delta_{i1}=s\right)\right]\\
&\times \left[\sum_{m=1}^{K}\widehat{\phi}^c_{im}(\widehat{\boldsymbol{\beta}},t_2)-\widehat{\mathrm{E}}\left(\sum_{m=1}^{K}\phi_{im}(\boldsymbol{\beta},t_2)\middle| X_{i1}=h,\Delta_{i1}=s\right)\right].
\end{aligned}$$

The proofs of the theorems 3.1 and 3.4 are outlined in the last section of this chapter. The consistency of the estimators for the hazards regression parameters were shown via an extension of Foutz (1977)'s theorem. The key steps to show the asymptotic normality involved the decomposition of the weighted estimating functions to the pseudo partial likelihood score function for the full cohort data plus a term involving the sampling of the case-control samples from the full cohort. This was based on a modified version of lemma 1 in Lin et al. (2000), the strong embedding theorem (Shorack and Wellner, 1986), and the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988). The martingale convergence results (Andersen and Gill, 1982) and the theory of modern empirical processes (van der Vaart and Wellner, 1996) were used to show the asymptotic normality of the pseudo partial likelihood score function for the full cohort data (Spiekerman and Lin, 1998). However, for the second part, the martingale convergence results can no longer be applied since the weights are not predictable. Thus, the theory of modern empirical processes (van der Vaart and Wellner, 1996) which does not require the predictability condition was employed for the second part. The asymptotic theory for sampling from finite population (Hájek, 1960) is also needed since it involves the sampling without replacement from the cohort. The asymptotic independence of the two terms and the Taylor expansion ensure the desired asymptotic normality of the estimators for the hazards regression parameters. The uniform consistency of the cumulative baseline hazards estimators and the weak convergence to a tight Gaussian processes were shown via similar arguments mentioned above.

3.4 Simulations

Extensive Monte Carlo simulations have been conducted to examine the finite sample properties of the proposed procedures. For each cluster, mimicking the setup for our motivating dental study example, failure times for two members (K=2) were generated via a multivariate extension of the Clayton and Cuzick (1985), in which the joint survival function for (T_1, T_2) given (Z_1, Z_2) is $\mathrm{S}(t_1, t_2; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = \{\mathrm{S}_1(t_1; \boldsymbol{Z}_1)^{-1/\theta} + \mathrm{S}_2(t_2; \boldsymbol{Z}_2)^{-1/\theta} - 1\}^{-\theta}$ where

$S_k(t; \boldsymbol{Z}) = \Pr(T_k > t|\boldsymbol{Z}_k)$, $k = 1, 2$, is the marginal survival function for T_i given covariate $\boldsymbol{Z}_i$. We considered a binary covariate with all the first members having 1 and second members having 0 which is the case for the dental study example. We also considered a continuous covariate where the continuous covariate was generated from the standard normal distribution. Exponential and Weibull distributions were considered for the marginal distribution of the failure times. The parameter θ represents the degree of dependence of T_1 and T_2. The smaller the value of θ, the stronger the dependence between T_1 and T_2. Values of 4, 1.25, 0.67, and 0.1 were considered for θ. This corresponds to a correlation of 0.237, 0.573, 0.762, and 0.987 when $\beta = 0$ and no censoring. We used values of 0 and log(2) for the regression parameter β. λ_0 was set to 1 for exponential failure times. For Weibull failure times, the scale parameter and the shape parameter were set to 1 and 0.5, respectively. Cohort sizes of n = 1000, 2000 were considered. We conducted simple random sampling without replacement within cases and controls independently. Approximately 80% and 90% of censoring proportions were considered for each setup and 90% of the cases and the same number of controls were sampled. The censoring times were generated from uniform(0, c) independently from the failure times where c was determined to achieve the desired censoring proportions. For each of the configuration studied, 2000 simulations were carried out.

Table 3.1 presents simulation summary statistics with marginal distribution with $\lambda_0 = 1$ and for the binary covariate where the value of the first member is equal to one and the value of the second member is equal to zero($Z_{i1} = 1$ and $Z_{i2} = 0$). "mean $\widehat{\boldsymbol{\beta}}_0$" denotes the average of the estimates of $\boldsymbol{\beta}_0$, "indep. s.e." denotes the average of the estimates of standard errors based on independence assumption, "proposed s.e." denotes the average of the estimates of standard errors based on the proposed method, "true S.D." denotes the sample standard deviation of the 2,000 estimates, and "95% coverage" denotes the coverage rate of the nominal 95% confidence interval. Note that the sample size for the case-control sample increases with increasing event proportion in our setup since we sample 90% of the cases and the same number of controls. The simulation results suggest that the coefficient estimates are approximately unbiased for the samples considered when $\beta = 0$, while the coefficient estimates are relatively biased(4 - 12 %) when $\beta = log(2)$ with small cohort and sample sizes($n = 1000$, event proportion = 10%). However, as the cohort size or sample size increases, the coefficient estimates improve and are approximately unbiased. The proposed estimated standard errors provide a very good estimate of the true variability of $\widehat{\boldsymbol{\beta}}$ while standard errors based on independence assumption do not. As expected, the variance of $\widehat{\boldsymbol{\beta}}$ decreases as cohort size or sample size increases. The coverage rate of the nominal 95% confidence intervals using the proposed method are in the 93% - 96% range in most of the cases considered. Table 3.2 provides simulation summary statistics for the standard normal covariate and Table 3 shows the results for the same setup as in Table 3.1 except that the marginal distribution follows Weibull distribution with the scale parameter and the shape parameter being set to 1 and 0.5, respectively. The findings are similar to those of Table 3.1.

We have also conducted simulations to compare the estimates with inclusion probability and the local average method. Exponential failure times were generated with $\lambda_0 = 0.4$ and 0.25 for $\beta = 0$ and $\log(2)$, respectively. The covariate Z was uniformly distributed on five points $m/5, 1 \leq m \leq 5$. We considered the situation when the censoring times were dependent on covariates. For each cluster i, the censoring time was generated from uniform distributions on the interval with length 0.4 and centered at $m^*/5$ where m^* was chosen such that it satisfies $(m^* - 1)/5 < \sum_{k=1}^{K} Z_{ik}/K \leq m^*/5(m^* = 1, \ldots, 5)$. Cohort sizes of n = 1000 and 2000 were considered. Under these setups, the proportion of failures is about 0.230 when $\beta = 0$ and is about 0.228 when $\beta = \log(2)$. For the local average approach, the partitions of the time interval $[0, \tau)$ were defined as $[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, $[0.4, 0.5)$, $[0.5, 0.6)$, $[0.6, 0.7)$ and $[0.7, \tau)$ for cases and $[0, 0.4)$, $[0.4, 0.5)$, $[0.5, 0.6)$, $[0.6, 0.7)$, $[0.7, 0.8)$, $[0.8, 0.9)$, $[0.9, 1.0)$ and $[1.0, \tau)$ for controls where τ was set to a value bigger than the maximum value of the failure and censoring times of the first members, say $\max_i(\boldsymbol{T}_{i1}, \boldsymbol{C}_{i1}, i = 1, \ldots, n) + 0.1$. Eighty percent of the cases were sampled and the same number of controls were sampled. Table 3.4 displays a comparison between the estimators using inclusion probabilities and local average method. Both methods perform reasonably well under the settings considered. The results indicate that the local average method is more efficient than the inclusion probabilities method when the censoring time depends on the covariate, especially when the correlations of the failure times within a cluster is very high ($\theta = 0.1$).

3.5 Application to the Retrospective Dental Study

We applied our proposed method to data from the retrospective dental study of pulpal involvement and tooth survival described in section 1. The sample was drawn from the population of enrollees in the Kaiser Permanente Dental Care Program (KPDCP), a dental HMO located in Portland, OR (Caplan and Weintraub, 1997). Enrollees are current or retired employees(or their dependents) of companies with dental insurance through KPDCP. As an indicator of pulpal involvement, root canal filled (RCF) teeth were used. Cases were defined as those who lost the RCF tooth during 1987-1994 period, while controls were defined as those who did not lose the RCF tooth during that period. After cases and controls were sampled, a non-RCF tooth was matched to the RCF tooth within each subject. For a matched non-RCF tooth, the contralateral tooth was selected if it was present. If that tooth was missing or already had RCF on the RCF tooth's access date(index date), the tooth of the same type(anterior, premolar or molar) adjacent to the contralateral tooth was selected. A total of 406 charts was requested, including 232 randomly selected from among 272 cases, and 174 randomly selected from among 1523 controls. Two-hundred-and-two (202) subjects were identified following the study eligibility criteria. Each of them has one RCT tooth and a matched non-RCT tooth. Subject- and tooth- level covariates were then ascertained for the RCF tooth and the matching non-RCF tooth from the electronic databases and from radiographs (bitewing, periapical, panoramic) and clinical periodontal recordings taken most recently before the RCF tooth's access date. Databases and charts were examined to determine all treatment received by the study teeth between the index date and 12/31/94, and the most recent radiograph was examined to validate extraction status. For both RCF and non-RCF teeth, follow-up started on the index date and continued through the date of extraction or 12/31/94, whichever came first. If an initially non-RCF tooth was accessed endodontically during that interval, the tooth was censored on its endodontic access date.

We applied the proposed method to this data set to investigate the effect of RCF on tooth survival. We also analyze the data using the unweighted method where the sampling scheme was not taken into consideration. For the analyses, we included RCF status, Tooth type, Interaction between RCF status and Tooth type, Proximal contacts and Number of pockets $\geq$ 5mm as covariates and studied the effect of RCF on tooth survival. Tooth type is molar and nonmolar. There were 176 molars and 228 nonmolars among 202 subjects. Proximal contacts (PCs) are where teeth contact adjacent teeth in the same arch. RCF and non-RCF teeth were classified as into one of four mutually exclusive groups: PC2 (teeth contacting adjacent teeth on both front and back sides); PC1 (teeth contacting adjacent teeth only on one side); PC0 (teeth with no adjacent contacting teeth); or PCABUT (teeth that were abuments for bridges). Ninety percent of the sampled teeth falls either in PC1 or PC2. Periodontal pockets are the spaces between the teeth and gums. Pocket depths had been recorded at six sites per tooth. 5mm was chosen as a binary threshold representing "deep pockets", and the number of pockets of this depth (out of six possible sites per tooth) were counted. Two hundred and seventy nine teeth(70%) do not have any periodontal pockets $\geq$ 5mm.

Table 3.6 provides hazard ratio(HR) estimates, the estimated standard errors, and the associated p-values for the proposed method and naive(unweighted) method. The results show strong evidence of significant RCF effect among molars. It indicates that for molars, the hazard rate with RCF is approximately seven times as higher as that without RCF. However, no statistically significant effect was seen among non-molars. The HR estimates for the molars and nonmolars using naive method are biased and are 1.5 to 3 times higher than those using the proposed method. For other variables, the teeth with 2 PCs and the number of pockets show statistically significant effect. The hazard rate with the teeth with 2 PCs is approximately 10 times lower than those with 0 PCs. Having one more pocket $\geq$ 5mm increases the hazard rate by approximately 30%, however, this effect is marginal (p-value = 0.09).

3.6 Concluding Remarks

Motivated by the aforementioned dental study (Caplan and Weintraub, 1997; Caplan et al., 2005), we have proposed methods of fitting marginal hazard regression models for the multivariate failure time data from case-control within cohort studies. The primary interest of the study was to evaluate the effect of pulpal involvement on tooth survival. The correlation between two teeth within the same subject is considered as a nuisance. This naturally led us to consider marginal hazard regression models. Weighted estimating equations are proposed for the estimation of the regression parameter. A Breslow-Aalen type estimator is proposed for the cumulative baseline hazard functions. The proposed estimators are shown to be consistent and asymptotically normally distributed. Two types of weights were considered in estimation: the inverse of the inclusion probabilities and the local average. The latter requires the additional information on the observed failure times of all the cohort members. It is more efficient than the inclusion probability estimator when the censoring time is dependent on some covariates which the failure time is also dependent on.

3.7 Proofs of the theorems

The following lemmas will be frequently used in proving the theorems.

Lemma 1 *Let $\boldsymbol{f}_n(t)$ and $g_n(t)$ be two sequences of bounded functions. For some constant τ, assume that the following conditions (a) - (c) hold where*

(a) $\sup_{0\le t\le\tau}\|\boldsymbol{f}_n(t)-\boldsymbol{f}(t)\|\longrightarrow 0$*, for some bounded function $\boldsymbol{f}(t)$,*

(b) $\{\boldsymbol{f}_n(t)\}$ *are monotone on* $[0,\tau]$ *and*

(c) $\sup_{0\le t\le\tau}|g_n(t)-g(t)|\longrightarrow 0$ *where $g(t)$ is continuous on $[0,\tau]$. Then*

$$\sup_{0\le t\le\tau}\left\|\int_0^t \boldsymbol{f}_n(s)dg_n(s)-\int_0^t \boldsymbol{f}(s)dg(s)\right\|\longrightarrow 0,$$
$$\sup_{0\le t\le\tau}\left\|\int_0^t g_n(s)d\boldsymbol{f}_n(s)-\int_0^t g(s)d\boldsymbol{f}(s)\right\|\longrightarrow 0.$$

This lemma is a simple extension of lemma 1 of Lin et al., 2000. The proof follows that of Lin et al. (2000) by replacing $|.|$ with $\|.\|$.

Lemma 2 *Let $\boldsymbol{W}_n(t)$ and $G_n(t)$ be two sequences of bounded processes. For some constant τ, assume that the following conditions (a) - (c) hold where* *(a)* $\sup_{0\le t\le\tau}\|\boldsymbol{W}_n(t)-\boldsymbol{W}(t)\| \xrightarrow{p} 0$ *for some bounded process* $\boldsymbol{W}(t)$*,*

(b) $\boldsymbol{W}_n(t)$ *is monotone on* $[0,\tau]$ *and*

(c) $G_n(t)$ *converges to a zero-mean process with continuous sample paths. Then*

$$\sup_{0\le t\le\tau}\left\|\int_0^t \{\boldsymbol{W}_n(s)-\boldsymbol{W}(t)\}dG_n(s)\right\| \xrightarrow{p} 0,\ \sup_{0\le t\le\tau}\left\|\int_0^t G_n(s)d\{\boldsymbol{W}_n(s)-\boldsymbol{W}(s)\}\right\| \xrightarrow{p} 0.$$

Proof of lemma 2 Let $G(t)$ be a limiting process of $G_n(t)$. Then, by the strong embedding theorem (Shorack and Wellner, 1986, p47-48), we can construct a new probability space wherein $(\boldsymbol{W}_n, G_n)$ converges almost surely to $(\boldsymbol{W}, G)$. Since $\boldsymbol{W}_n(t)$ is monotone and $G(t)$ is continuous, by applying lemma 1, we have

$$\sup_{0\le t\le\tau}\left\|\int_0^t \boldsymbol{W}_n(s)dG_n(s)-\int_0^t \boldsymbol{W}(s)dG(s)\right\| \xrightarrow{a.s.} 0,$$

and

$$\sup_{0\le t\le\tau}\left\|\int_0^t \boldsymbol{W}_n(s)dG_n(s)-\int_0^t \boldsymbol{W}(s)dG(s)\right\| \xrightarrow{p} 0 \tag{3.5}$$

in the original probability space. By (3.5), the following also holds

$$\sup_{0\le t\le\tau}\left\|\int_0^t \boldsymbol{W}(s)d\{G_n(s)-G(s)\}\right\| \xrightarrow{p} 0 \tag{3.6}$$

by replacing $\boldsymbol{W}_n(s)$ with $\boldsymbol{W}(s)$. Now, one can write

$$\begin{aligned}\int_0^t \{\boldsymbol{W}_n(s)-\boldsymbol{W}(s)\}dG_n(s) &= \left\{\int_0^t \boldsymbol{W}_n(s)dG_n(s)-\int_0^t \boldsymbol{W}(s)dG(s)\right\}\\ &- \int_0^t \boldsymbol{W}(s)d\{G_n(s)-G(s)\}.\end{aligned}$$

Each of the two terms on the right-hand side of this equation converges to zero uniformly in t in probability by (3.5) and (3.6). Thus, $\int_0^t\{\boldsymbol{W}_n(s)-\boldsymbol{W}(s)\}dG_n(s)$ converges to zero uniformly in t in probability as $n\to\infty$. The other expression follows from the integration by parts formula.

Lemma 3 *Suppose a cohort of size n can be divided into S mutually exclusive strata and this stratification is based on a discrete random variable W whose information is available for all the cohort members. Let n_s denote the size of the sth stratum ($s = 1, \ldots, S$). Let X_{sj}'s be independent and identically distributed random variables and $\boldsymbol{\xi}_{sj} = (\xi_{s1}, \ldots, \xi_{sn_s})$ be a random vector of $\widetilde{n}_s$ ones and $n_s - \widetilde{n}_s$ zeros with each permutation equally likely. Let $\widetilde{n}_s = \sum_{j=1}^{n_s} \xi_{sj}$ denote the sample size drawn from the sth stratum. Then*

$$U_n = n^{-1/2} \sum_{s=1}^{S} \sum_{j=1}^{n_s} \left(\frac{\xi_{sj}}{\widetilde{n}_s / n_s} - 1 \right) X_{sj}$$

converges to a zero-mean normal random variable with the following covariance function

$$\mathrm{E}\left\{ \left(\frac{1}{\alpha(W)} - 1 \right) \mathrm{Var}(X_{W1} | W) \right\}$$

provided that

(a) $\dfrac{n_s}{n} \xrightarrow{p} p_s \equiv P(W = s) \in (0, 1)$ *and* $\dfrac{\widetilde{n}_s}{n_s} \xrightarrow{p} \alpha_s \in (0, 1)$ *as* $n \to \infty$, *where* α_s *is the realization of a function* $\alpha(W)$ *of a random variable* W *evaluated at* $W = s$, *i.e.* $\alpha(W)|_{W=s} = \alpha_s$,

(b) $S_s^2 = \dfrac{1}{n_s - 1} \displaystyle\sum_{j=1}^{n_s} (X_{sj} - \bar{X}_s)^2 \xrightarrow{p} \sigma_s^2 = \mathrm{Var}(X_{W1} | W = s) \neq 0$ *where* $\bar{X}_s = \dfrac{1}{n_s} \displaystyle\sum_{j=1}^{n_s} X_{sj}$, *and*

(c) $\dfrac{\max(X_{sj} - \bar{X}_s)^2}{\sum_{j=1}^{n_s} (X_{sj} - \bar{X}_s)^2} \longrightarrow 0$ *as* $n \to \infty$ *for* $s = 1, \ldots, S$.

This is simply applying the Hájek (1960)'s asymptotic theory of random sampling without replacement from a finite population within each strata. Specifically, write

$$
\begin{aligned}
U_n &= \sum_{s=1}^{S} \frac{1}{\sqrt{n}} n_s \sum_{j=1}^{n_s} \left(\frac{\xi_{sj}}{\widetilde{n}_s} - \frac{1}{n_s} \right) X_{sj} \\
&= \sum_{s=1}^{S} \sqrt{\frac{n_s}{n}} \cdot \sqrt{\left(\frac{n_s}{\widetilde{n}_s} - 1 \right) S_s^2} \cdot \frac{\sum_{j=1}^{n_s} \left(\frac{\xi_{sj}}{\widetilde{n}_s} - \frac{1}{n_s} \right) X_{sj}}{\sqrt{\frac{1}{n_s} \left(\frac{n_s}{\widetilde{n}_s} - 1 \right) S_s^2}} \\
&= \sum_{s=1}^{S} \sqrt{\frac{n_s}{n} \left(\frac{n_s}{\widetilde{n}_s} - 1 \right) S_s^2} \cdot U_n^{(s)}.
\end{aligned}
$$

On the basis of the conditions (a), (b) and (c), conditional on $\mathcal{F}(\tau)$, $U_n^{(s)}$ converges to a standard normal random variable by the Hájek (1960)'s asymptotic theory of random sampling without replacement from a finite population. Since the resulting standard normal random variable does not depend on $\mathcal{F}(\tau)$, this is true unconditionally. Note that the sampling was conducted independently across the strata. Then, together with the conditions (a) and (b), we can use the Slutsky's theorem and conclude that U_n converges to a normal random variable with mean zero and with the following covariance function

$$
\sum_{s=1}^{S} p_s \left(\frac{1}{\alpha_s} - 1 \right) \mathrm{Var}(X_{W1} | W = s) = \mathrm{E}\left[\left(\frac{1}{\alpha(W)} - 1 \right) \mathrm{Var}(X_{W1} | W) \right].
$$

Lemma 4 *Suppose that within each stratum S we have further classified the sample into H_s mutually exclusive groups based on a discrete random variable V and the sizes of strata, n_{sh}, for $h = 1, \ldots, H_s$, $s = 1, \ldots, S$, are known. Let $\widetilde{n}_{sh} = \sum_{j=1}^{n_{sh}} \xi_{shj}$ denote the sample size drawn from the hth stratum in the sth stratum. Then*

$$
U_n = n^{-1/2} \sum_{s=1}^{S} \sum_{h=1}^{H_s} \sum_{j=1}^{n_{sh}} \left(\frac{\xi_{shj}}{\widetilde{n}_{sh}/n_{sh}} - 1 \right) X_{shj}
$$

converges to a zero-mean normal random variable with the following covariance function

$$
\mathrm{E}\left\{ \left(\frac{1}{\alpha(W)} - 1 \right) \mathrm{Var}(X_{WV1} | W, V) \right\}
$$

provided that

(a) $\frac{n_s}{n} \xrightarrow{p} p_s \equiv P(W=s) \in (0,1), \frac{n_{sh}}{n_s} \xrightarrow{p} q_{sh} \equiv P(V=h|W=S) \in (0,1),$ *and* $\frac{\widetilde{n}_s}{n_s} \xrightarrow{p} \alpha_s \in (0,1)$ *as* $n \to \infty$, *where* α_s *is the realization of a function* $\alpha(W)$ *of a random variable* W *evaluated at* $W=s$, *i.e.* $\alpha(W)|_{W=s} = \alpha_s$,

(b) $S_{sh}^2 = \frac{1}{n_{sh}-1}\sum_{j=1}^{n_{sh}}(X_{shj}-\bar{X}_{sh})^2 \xrightarrow{p} \sigma_{sh}^2 = \mathrm{Var}(X_{WV1}|W=s, V=h) \neq 0$

where $\bar{X}_{sh} = \frac{1}{n_{sh}}\sum_{j=1}^{n_{sh}} X_{shj}$, *and*

(c) $\frac{\max(X_{shj}-\bar{X}_{sh})^2}{\sum_{j=1}^{n_{sh}}(X_{shj}-\bar{X}_{sh})^2} \longrightarrow 0$ *as* $n \to \infty$ *for* $s = 1, \ldots, S$ *and* $h = 1, \ldots, H_s$.

Proof of lemma 4 The number of samples in each stratum after post-stratification, $\widetilde{n}_{sh}$, are not fixed, but random. Note that conditioning on $\widetilde{\boldsymbol{n}} = (\widetilde{\boldsymbol{n}}_1, \ldots, \widetilde{\boldsymbol{n}}_s)$ where $\widetilde{\boldsymbol{n}}_s = (\widetilde{n}_{s1}, \ldots, \widetilde{n}_{sH_s})$, $s = 1, \ldots, S$, we can apply standard results from independent simple random sampling within strata. Thus,

$$\mathrm{E}(U_n|\widetilde{\boldsymbol{n}}, \mathcal{F}(\tau)) = \sqrt{n}^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\sum_{j=1}^{n_{sh}} \mathrm{E}\left\{\left(\frac{\xi_{shj}}{\widetilde{n}_{sh}/n_{sh}} - 1\right) X_{shj}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right\} = 0, \text{ since}$$

$$\mathrm{E}\left(\xi_{shj}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right) = \frac{\widetilde{n}_{sh}}{n_{sh}}, \text{ and } \mathrm{Var}(U_n|\widetilde{\boldsymbol{n}}, \mathcal{F}(\tau))$$

$$= n^{-1}\sum_{s=1}^{S} \mathrm{Var}\left\{\sum_{h=1}^{H_s}\sum_{j=1}^{n_{sh}}\left(\frac{\xi_{shj}}{\widetilde{n}_{sh}/n_{sh}} - 1\right) X_{shj}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right\}$$

$$= n^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\left(\frac{n_{sh}}{\widetilde{n}_{sh}}\right)^2 \mathrm{Var}\left(\sum_{j=1}^{n_{sh}} \xi_{shj}X_{shj}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right)$$

$$= n^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\left(\frac{n_{sh}}{\widetilde{n}_{sh}}\right)^2 \left\{\sum_{j=1}^{n_{sh}} \mathrm{Var}\left(\xi_{shj}X_{shj}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right)\right.$$

$$\left. + \sum_{j,k=1, j\neq k}^{n_{sh}} \mathrm{Cov}\left(\xi_{shj}X_{shj}, \xi_{shk}X_{shk}|\widetilde{n}_{sh}, \mathcal{F}(\tau)\right)\right\}$$

$$= n^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\left(\frac{n_{sh}}{\widetilde{n}_{sh}}\right)^2 \left\{\frac{\widetilde{n}_{sh}}{n_{sh}}\left(1 - \frac{\widetilde{n}_{sh}}{n_{sh}}\right)\sum_{j=1}^{n_{sh}} X_{shj}^2 + \frac{\widetilde{n}_{sh}(\widetilde{n}_{sh}-n_{sh})}{n_{sh}^2(n_{sh}-1)}\right.$$

$$\left. \times \sum_{j,k=1, j\neq k}^{n_{sh}} X_{shj}X_{shk}\right\} \left(\text{since } \mathrm{Cov}(\xi_{shj}, \xi_{shk}) = \frac{\widetilde{n}_{sh}(\widetilde{n}_{sh}-n_{sh})}{n_{sh}^2(n_{sh}-1)}\right)$$

$$
\begin{aligned}
&= n^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\left(\frac{1}{\widetilde{n}_{sh}}\right)(n_{sh}-\widetilde{n}_{sh})\left\{\sum_{j=1}^{n_{sh}}X_{shj}^2-\frac{1}{n_{sh}-1}\sum_{j,k=1,j\neq k}^{n_{sh}}X_{shj}X_{shk}\right\}\\
&= n^{-1}\sum_{s=1}^{S}\sum_{h=1}^{H_s}\left(\frac{1}{\widetilde{n}_{sh}}\right)(n_{sh}-\widetilde{n}_{sh})\frac{n_{sh}}{n_{sh}-1}\sum_{j=1}^{n_{sh}}(X_{shj}-\bar{X}_{sh})^2\\
&= n^{-1}\sum_{s=1}^{S}n_s^2\sum_{h=1}^{H_s}\left(\frac{n_{sh}}{n_s}\right)^2\left(\frac{1}{\widetilde{n}_{sh}}-\frac{1}{n_{sh}}\right)S_{sh}^2
\end{aligned}
$$

Now the results from conditioning only on $\mathcal{F}(\tau)$ can be obtained as follows.

$$
\begin{aligned}
\mathrm{E}(U_n|\mathcal{F}(\tau)) &= \mathrm{E}_{\widetilde{\boldsymbol{n}}}(\mathrm{E}(U_n|\ \widetilde{\boldsymbol{n}},\mathcal{F}(\tau))) = 0 \text{ and}\\
\mathrm{Var}(U_n|\ \mathcal{F}(\tau)) &= \mathrm{E}_{\widetilde{\boldsymbol{n}}}(\mathrm{Var}(U_n|\ \widetilde{\boldsymbol{n}},\mathcal{F}(\tau))) + \mathrm{Var}_{\widetilde{\boldsymbol{n}}}(\mathrm{E}(U_n|\ \widetilde{\boldsymbol{n}},\mathcal{F}(\tau)))\\
&= \mathrm{E}_{\widetilde{\boldsymbol{n}}}\left(\sum_{s=1}^{S}\left(\frac{n_s^2}{n}\right)\sum_{h=1}^{H_s}\left(\frac{n_{sh}}{n_s}\right)^2\left(\frac{1}{\widetilde{n}_{sh}}-\frac{1}{n_{sh}}\right)S_{sh}^2\middle|\mathcal{F}(\tau)\right)
\end{aligned}
$$

Note that given $\mathcal{F}(\tau)$, $\widetilde{n}_{sh}$ is random, but $\widetilde{n}_s, n_{sh}$, and n_s are not random; $\mathrm{E}_{\widetilde{\boldsymbol{n}}}(Z(\widetilde{\boldsymbol{n}}))$ is the average of $Z(\widetilde{\boldsymbol{n}})$ over all possible values of $\widetilde{\boldsymbol{n}}$, and $\mathrm{E}_{\widetilde{\boldsymbol{n}}}(\widetilde{\boldsymbol{n}}_{sh}|\mathcal{F}(\tau)) = \widetilde{n}_s W_{sh}$, $\mathrm{Var}_{\widetilde{\boldsymbol{n}}}(\widetilde{\boldsymbol{n}}_{sh}|\mathcal{F}(\tau)) = \widetilde{n}_s W_{sh}(1-W_{sh})$ where $W_{sh} = n_{sh}/n_s$. Since

$$
\begin{aligned}
\frac{1}{\widetilde{n}_{sh}} &= \frac{1}{\widetilde{n}_s W_{sh}}\cdot\frac{1}{1-\left(1-\frac{\widetilde{n}_{sh}}{\widetilde{n}_s W_{sh}}\right)}\\
&= \frac{1}{\widetilde{n}_s W_{sh}}\left(1+\frac{\widetilde{n}_s W_{sh}-\widetilde{n}_{sh}}{\widetilde{n}_s W_{sh}}+\frac{1}{(1-a^*)^3}\frac{(\widetilde{n}_s W_{sh}-\widetilde{n}_{sh})^2}{(\widetilde{n}_s W_{sh})^2}\right)
\end{aligned}
$$

by the Taylor series expansion where a^* lies on the line segment between 0 and $1-\widetilde{n}_{sh}/\widetilde{n}_s W_{sh}$,

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{\widetilde{n}_{sh}}\middle|\mathcal{F}(\tau)\right) &= \mathrm{E}\left(\frac{1}{\widetilde{n}_s W_{sh}}+\frac{\widetilde{n}_s W_{sh}-\widetilde{n}_{sh}}{(\widetilde{n}_s W_{sh})^2}+\frac{1}{(1-a^*)^3}\frac{(\widetilde{n}_s W_{sh}-\widetilde{n}_{sh})^2}{(\widetilde{n}_s W_{sh})^3}\right)\\
&= \frac{1}{\widetilde{n}_s W_{sh}}+\frac{\mathrm{E}(\widetilde{n}_s W_{sh}-\widetilde{n}_{sh})}{(\widetilde{n}_s W_{sh})^2}+\frac{1}{(1-a^*)^3}\frac{\mathrm{E}(\widetilde{n}_s W_{sh}-\widetilde{n}_{sh})^2}{(\widetilde{n}_s W_{sh})^3}\\
&= \frac{1}{\widetilde{n}_s W_{sh}}+0+\frac{1}{(1-a^*)^3}\frac{1-W_{sh}}{(\widetilde{n}_s W_{sh})^2}\\
&= \frac{1}{\widetilde{n}_s W_{sh}}+O(\widetilde{n}_s^{-2})
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(U_n|\ \mathcal{F}(\tau)) &= \sum_{s=1}^{S}\frac{n_s^2}{n}\left\{\sum_{h=1}^{H_s} W_{sh}^2\left(\frac{1}{\widetilde{n}_s W_{sh}} + O(\widetilde{n}_s^{-2}) - \frac{1}{n_{sh}}\right) S_{sh}^2\right\} \\
&= \sum_{s=1}^{S}\frac{n_s^2}{n}\left\{\sum_{h=1}^{H_s}\left(\frac{W_{sh}}{\widetilde{n}_s} + O(\widetilde{n}_s^{-2}) - \frac{W_{sh}^2}{n_{sh}}\right) S_{sh}^2\right\} \\
&= \sum_{s=1}^{S}\frac{n_s}{n}\left\{\sum_{h=1}^{H_s}\left(\frac{n_s}{\widetilde{n}_s} - 1\right) W_{sh}S_{sh}^2 + O(\widetilde{n}_s^{-1})\right\} \\
&= \sum_{s=1}^{S}\frac{n_s}{n}\sum_{h=1}^{H_s}\left(\frac{n_s}{\widetilde{n}_s} - 1\right) W_{sh}S_{sh}^2 + O(\widetilde{n}_s^{-1})
\end{aligned}$$

The first term is the variance of a proportionally allocated stratified sample while the second term is $O(\widetilde{n}_s^{-1}) = O(n^{-1}) = o(1)$ by condition (a). Thus, the asymptotic behavior of U_n conditioning on $\mathcal{F}(\tau)$ under post-stratification can be asserted to be the same as that under a proportionally allocated stratified sample. Therefore, by conditions (a), (b) and (c), U_n conditioning on $\mathcal{F}(\tau)$ converges to a zero-mean normal random variable with the following covariance function

$$\lim_{n\to\infty}\sum_{s=1}^{S}\frac{n_s}{n}\sum_{h=1}^{H_s}\left(\frac{n_s}{\widetilde{n}_s} - 1\right) W_{sh}S_{sh}^2 + o(1) = \mathrm{E}\left\{\left(\frac{1}{\alpha(W)} - 1\right)\mathrm{Var}(X_{WV1}|W, V)\right\}$$

Since the covariance function does not depend on $\mathcal{F}(\tau)$, this argument is true unconditionally.

Proof of theorem 3.1 We first consider the proof of the consistency of $\widehat{\boldsymbol{U}}(\boldsymbol{\beta}_0)$. Denote n^{-1} times $\widehat{\boldsymbol{U}}(\boldsymbol{\beta})$ by $\boldsymbol{U}_n(\boldsymbol{\beta})$. Based on a straightforward extension of Foutz (1977), one can show $\widehat{\boldsymbol{\beta}}$ to be consistent for $\boldsymbol{\beta}_0$ provided: (i) $\partial\boldsymbol{U}_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$, (ii)$\partial\boldsymbol{U}_n(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0^T$ is negative definite with probability going to one as $n\to\infty$, (iii) $\partial\boldsymbol{U}_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ converges to $\boldsymbol{A}(\boldsymbol{\beta}_0)$ in probability uniformly for $\boldsymbol{\beta}$ in an open neighborhood about $\boldsymbol{\beta}_0$, and (iv) $\boldsymbol{U}_n(\boldsymbol{\beta})\to 0$ in probability.

One can write

$$\frac{\partial \boldsymbol{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = -n^{-1}\int_0^\tau \widehat{\boldsymbol{V}}(\boldsymbol{\beta},t)d\widehat{N}(t) \text{ where } \widehat{N}(t)=\sum_{i=1}^n\sum_{k=1}^K w_i N_{ik}(t),$$

$$w_i = \frac{\xi_i}{\pi_i} \text{ and } \widehat{\boldsymbol{V}}(\boldsymbol{\beta},t) = \left\{\frac{\widehat{\boldsymbol{S}}^{(2)}(\boldsymbol{\beta},t)\widehat{S}^{(0)}(\boldsymbol{\beta},t)-\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta},t)^{\otimes 2}}{\widehat{S}^{(0)}(\boldsymbol{\beta},t)^2}\right\}. \tag{3.7}$$

Then, (i) is clearly satisfied on the basis of (4.15) and by the continuity of each component. Now, following Andersen and Gill (1982),

$$\begin{aligned}\left\|\left(-\frac{\partial \boldsymbol{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right)-\boldsymbol{A}(\boldsymbol{\beta})\right\| &\le \left\|\sum_{k=1}^K\int_0^\tau\{\widehat{\boldsymbol{V}}(\boldsymbol{\beta},t)-\boldsymbol{v}(\boldsymbol{\beta},t)\}n^{-1}d\bar{N}_k(t)\right\| \\ &+ \left\|\sum_{k=1}^K\int_0^\tau\{\widehat{\boldsymbol{V}}(\boldsymbol{\beta},t)-\boldsymbol{v}(\boldsymbol{\beta},t)\}dn^{-1}\sum_{i=1}^n(w_i-1)N_{ik}(t)\right\| \\ &+ \left\|\sum_{k=1}^K\int_0^\tau\boldsymbol{v}(\boldsymbol{\beta},t)n^{-1}d\bar{M}_k(t)\right\| + \left\|\sum_{k=1}^K n^{-1}\sum_{i=1}^n(w_i-1)\int_0^\tau\boldsymbol{v}(\boldsymbol{\beta},t)dM_{ik}(t)\right\| \\ &+ \left\|\int_0^\tau\boldsymbol{v}(\boldsymbol{\beta},t)\{\widehat{S}^{(0)}(\boldsymbol{\beta},t)-s^{(0)}(\boldsymbol{\beta},t)\}\lambda_0(t)dt\right\|\end{aligned} \tag{3.8}$$

where $\bar{N}_k(t)=\sum_{i=1}^n N_{ik}(t)$ and $\bar{M}_k(t)=\sum_{i=1}^n M_{ik}(t)$.

Each of the terms on the right side of the above inequality will be shown to converge to zero, uniformly in $\boldsymbol{\beta}\in\mathcal{B}$ in the following.

The Lenglart inequality (Andersen and Gill, 1982, p1115) implies that, for any $\delta,\ \rho > 0$, there exists n_0 such that for $n \ge n_0$,

$$P[n^{-1}\bar{N}_k(\tau)>c] \le \frac{\delta}{c} + P[\int_0^\tau S^{(0)}(\boldsymbol{\beta}_0;t)\lambda_0(t)dt>\delta]$$

By Condition (F), for $\delta > \int_0^\tau s^{(0)}(\boldsymbol{\beta}_0,t)\lambda_0(t)dt$, $P[\int_0^\tau S^{(0)}(\boldsymbol{\beta}_0;t)\lambda_0(t)dt>\delta]\to 0$ as $n\to\infty$. Then, $\lim_{c\uparrow\infty}\lim_{n\to\infty}P[n^{-1}\bar{N}_k(\tau)>c]=0$. Conditions (G) and (L) imply $\sup_{t\in[0,\tau],\ \boldsymbol{\beta}\in\mathcal{B}} \|\widehat{\boldsymbol{V}}(\boldsymbol{\beta};t)-\boldsymbol{v}(\boldsymbol{\beta},t)\| \xrightarrow{p} 0$ as $n\to\infty$. Thus, it follows that the first term on the right side of (4.16) converges to zero in probability, uniformly in $\boldsymbol{\beta}\in\mathcal{B}$, as $n\to\infty$.

By applying lemma 3, $n^{-1/2}\sum_{i=1}^n(w_i-1)N_{ik}(t)$ can be shown to be asymptotically normally distributed with mean zero. Here we have two strata (Cases and Controls) and the conditions (I), (J), (L) and the fact that $\max_i |N_{ik} - n_1^{-1}\bar{N}_k| \leq \max_i |N_{ik}| = 1$ ensure the conditions (a), (b) and (c) in lemma 3 are satisfied. This implies $n^{-1}\sum_{i=1}^{n}(w_i - 1)N_{ik}(t)$ converges to zero in probability. Thus, together with $\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|\widehat{\boldsymbol{V}}(\boldsymbol{\beta}; t) - \boldsymbol{v}(\boldsymbol{\beta}, t)\| \xrightarrow{p} 0$ as $n \to \infty$, it follows that the second term on the right side of (4.16) also converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$, as $n \to \infty$.

$n^{-1}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}, t)dM_{ik}(t)$ is a local square integrable martingale. Hence, the Lenglart inequality (Andersen and Gill, 1982, p1115) implies that, for any $\delta,\ \rho\ >\ 0$, there exists n_0 such that for $n \geq n_0$,

$$\mathrm{P}\left[\left\|n^{-1}\int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta}, t)\}_{ll'}\, d\bar{M}_k(t)\right\| > \rho\right] \;\leq\; \frac{\delta}{\rho^2} \;+\; P\left[n^{-1}\int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta}, t)\}_{ll'}^2\, S^{(0)}(\boldsymbol{\beta}, t)\lambda_0(t)dt > \delta\right]$$

where the subscript ll' denotes the (l, l') element of the indicated matrix. The boundedness conditions (E), (F) and (G) ensure that the second term on the right side of the above inequality converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \to \infty$ for any δ. Since δ can be arbitrarily small, it follows that the left side of the above inequality also converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \to \infty$. Therefore, the third term on the right side of (4.16) also converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$, as $n \to \infty$.

The fourth term on the right side of (4.16) can be shown to converge to zero by applying lemma 3. Without loss of generality, we assume for $s = 1$, $i = 1, \ldots, n_1$ denote cases and for $s = 0$, $i = 1, \ldots, n_0$ denote controls. Then, for all $k = 1, \ldots, K$ and $s = 0, 1$, one can write

$$
\begin{aligned}
& \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dM_{ik}(t) - n_s^{-1} \sum_{i=1}^{n_s} \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dM_{ik}(t) \right| \\
= & \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} \{dN_{ik}(t) - Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t)\} \right. \\
- & \left. n_s^{-1} \sum_{i=1}^{n_s} \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} \{dN_{ik}(t) - Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t)\} \right| \\
\leq & \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dN_{ik}(t) \right| + \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} dN_{ik}(t) \right| \\
+ & \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| + \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right|
\end{aligned}
$$

where the subscript ll' denotes the (l, l') element of the indicated matrix. Thus

$$
\begin{aligned}
& \max_i \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dM_{ik}(t) - n_s^{-1} \sum_{i=1}^{n_s} \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dM_{ik}(t) \right|^2 \\
& \leq \max_i \left(\left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dN_{ik}(t) \right| + \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} dN_{ik}(t) \right| \right. \\
& \left. + \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| + \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| \right)^2
\end{aligned}
$$

Note that

$$
\begin{aligned}
\max_i \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} dN_{ik}(t) \right| & \leq \sup_{\boldsymbol{\beta},t} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}|, \\
\max_i \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} dN_{ik}(t) \right| & \leq \sup_{\boldsymbol{\beta},t} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}|, \\
\max_i \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau) \text{ and} \\
\max_i \left| \int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'} n_s^{-1} \sum_{i=1}^{n_s} Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau),
\end{aligned}
$$

To verify the condition (c) in lemma 3, it suffices to show that $n_s^{-1} \sup_{\boldsymbol{\beta},t} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}|$ and $n_s^{-1} \sup_{\boldsymbol{\beta},t,i} |\{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau)$ converge to zero in probability as $n \to \infty$. This holds by our conditions (E), (F), (G) and (K). Conditions (a) and (b) in lemma 3 are satisfied on the basis of conditions (I), (J) and (L). This implies $\sum_{k=1}^{K} n^{-1} \sum_{i=1}^{n} (w_i - 1) \int_0^{\tau} \boldsymbol{v}(\boldsymbol{\beta}, t) dM_{ik}(t)$ converges to zero in probability.

On the basis of conditions (D), (E) and (F), the last term on the right side of (4.16) can be shown to converge to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \to \infty$. Therefore,

$$-\frac{\partial \boldsymbol{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\beta}) \text{ as } n \to \infty \text{ uniformly in } \boldsymbol{\beta} \in \mathcal{B}$$

and, thus, (ii) and (iii) are satisfied.

For (iv), we can show that $n^{-1/2}\widehat{\boldsymbol{U}}(\boldsymbol{\beta})$ is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \boldsymbol{M}_{\widetilde{\boldsymbol{z}},ik}$. Specifically, write

$$\begin{aligned}
n^{1/2}\boldsymbol{U}_n(\boldsymbol{\beta}) &= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_{ik}(t) \\
&= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_{ik}(t) \\
&= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} w_i \widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) dM_{ik}(t) \\
&+ n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} w_i \left\{ \boldsymbol{e}(\boldsymbol{\beta}, t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_{ik}(t) \\
&= U_1 + U_2
\end{aligned}$$

Now, we will show that U_2 converges to zero in probability as $n \to \infty$. Write $U_2 = U_{21} + U_{22}$ where

$$\begin{aligned}
U_{21} &= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \left\{ \boldsymbol{e}(\boldsymbol{\beta}, t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} dM_{ik}(t) \text{ and} \\
U_{22} &= \sum_{k=1}^{K} \int_0^{\tau} \left\{ \boldsymbol{e}(\boldsymbol{\beta}, t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} d \left\{ n^{-1/2} \sum_{i=1}^{n} (w_i - 1) M_{ik}(t) \right\}
\end{aligned}$$

Note that, for fixed t, $n^{-1/2} \sum_{i=1}^{n} M_{ik}(t)$ is a sum of i.i.d. zero-mean random variables. Based on conditions (C) and (E), $M_{ik}(t)$ is of bounded variation and therefore can be written as a difference of two monotone functions in t. It then follows from the example of 2.11.16

of van der Vaart and Wellner (1996, p215) that $n^{-1/2}\sum_{i=1}^n M_{ik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{W}_M(t)$. It can be shown that $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^4 \le C\{\Lambda_0(t)-\Lambda_0(s)\}^2$ for some constant $C>0$. Specifically, $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^4 = 3(\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^2)^2$ since $\mathcal{W}_M(t)$ is a zero-mean normal random variable for a fixed t. Then $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^2 = \mathrm{E}\,\mathcal{W}_M(t)^2 + \mathrm{E}\,\mathcal{W}_M(s)^2 - 2\,\mathrm{E}\,\mathcal{W}_M(t)\mathcal{W}_M(s) = \mathrm{E}\,\mathcal{W}_M(t)^2 - \mathrm{E}\,\mathcal{W}_M(s)^2$ for $s \le t$. Since $\mathrm{E}\,\mathcal{W}_M(t)^2 = \mathrm{E}\,M_{ik}(t)^2 = \mathrm{E}\left[\int_0^t Y_{ik}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik(u)}}\lambda_0(u)du\right]$, $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^2 = \mathrm{E}\left[\int_s^t Y_{ik}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik(u)}}\lambda_0(u)du\right] \le e^{C_z}\,\mathrm{E}\left[\int_s^t \lambda_0(u)du\right] = \widetilde{C}_z(\Lambda_0(t)-\Lambda_0(s))$ by the boundedness condition (C). Since $\Lambda_0(\cdot)$ is differentiable and $\lambda_0(\cdot)$ is bounded on $[0,\tau]$, $\exists$ a constant M, such that $\Lambda_0(t)-\Lambda_0(s) \le M(t-s)$ for $s \le t$. Therefore, $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^2 \le C_z^*(t-s)$ and $\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^4 \le 3(\mathrm{E}\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\}^2)^2 \le \widetilde{C}_z^*(t-s)^2$ for some constant C_z^*.

Then, by the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988, p53), $\mathcal{W}_M(t)$ has continuous sample paths. In addition, since $\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)$ and $\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)$ are of bounded variations and $\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)$ is bounded away from 0, based on conditions (C), (G) and (L), $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)}$ is of bounded variation and can be written as a sum of two monotone functions in t, respectively. Specifically, $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)} = \boldsymbol{Z}_{k1}^*(t) - \boldsymbol{Z}_{k2}^*(t)$ where both $\boldsymbol{Z}_{k1}^*(t)$ and $\boldsymbol{Z}_{k2}^*(t)$ are nonnegative, monotone in t and bounded. Hence, it follows from lemma 2 that

$$U_{21} = \sum_{k=1}^K \int_0^\tau \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)} - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}^{(0)}(\boldsymbol{\beta},t)} \right\} n^{-1/2}\sum_{i=1}^n dM_{ik}(t) \xrightarrow{p} 0 \text{ as } n \to \infty.$$

In similar manners, U_{22} will be shown to converge to zero in probability as $n \to \infty$. The weak convergence of $n^{-1/2}\sum_{i=1}^n (w_i-1)M_{ik}(t)$ to a zero-mean Gaussian process follows from lemma 3 and the example 3.6.14 of van der Vaart and Wellner (1996, p356). By employing similar argument for $n^{-1/2}\sum_{i=1}^n (w_i-1)\int_0^\tau \{\boldsymbol{v}(\boldsymbol{\beta},t)\}_{ll'}dM_{ik}(t)$, conditions (E), (I), (J), (K) and (L) ensure that the conditions (a), (b) and (c) in lemma 3 are satisfied. The limiting process can be shown to have continuous sample paths via the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988, p53). Specifically, let $\mathcal{W}_M^*(t)$ be the limiting process of $n^{-1/2}\sum_{i=1}^n (w_i-1)M_{ik}(t)$. Then $\mathrm{E}\{\mathcal{W}_M^*(t)-\mathcal{W}_M^*(s)\}^4 = 3(\mathrm{E}\{\mathcal{W}_M^*(t)-\mathcal{W}_M^*(s)\}^2)^2$ since

$\mathcal{W}_M^*(t)$ is a zero-mean normal random variable for a fixed t. Thus, for $s \le t$,

$$\begin{aligned} & \mathrm{E}\{\mathcal{W}_M^*(t) - \mathcal{W}_M^*(s)\}^2 = \mathrm{E}\left\{\left(\frac{1}{\alpha(\Delta_{11})} - 1\right) \mathrm{Var}(\mathcal{W}_M(t) - \mathcal{W}_M(s)|\Delta_{11})\right\} \\ \le \quad & C_\alpha \, \mathrm{E}\{\mathrm{Var}(\mathcal{W}_M(t) - \mathcal{W}_M(s)|\Delta_{11})\} \text{ for some constant } C_\alpha \ge \max\left(\frac{1}{\alpha_0} - 1, \frac{1}{\alpha_1} - 1\right) \\ \le \quad & C_\alpha \, \mathrm{Var}(\mathcal{W}_M(t) - \mathcal{W}_M(s)) \end{aligned}$$

Therefore, by the same argument for $\mathcal{W}_M(\cdot)$, $\mathrm{E}\{\mathcal{W}_M^*(t) - \mathcal{W}_M^*(s)\}^4 \le C_\alpha^*(t-s)^2$ for some constant C_α^* and $\mathcal{W}_M^*(t)$ has continuous sample paths by the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988, p53). It follows from lemma 2 that U_{22} converges to zero in probability as $n \to \infty$. Hence, U_2 converges to zero in probability as $n \to \infty$.

Now, one can write $U_1 = U_{11} + U_{12}$ where

$$U_{11} = n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) \quad \text{and} \quad U_{12} = n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} (w_i - 1) \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})$$

Then, under the regularity conditions, the first term is asymptotically zero-mean normal with covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ by Spiekerman and Lin (1998).

The second term can be shown to be asymptotically zero-mean normal with covariance matrix

$\boldsymbol{V}(\boldsymbol{\beta}_0)$ by lemma 3. Specifically, write

$$
\begin{aligned}
&\left|\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) - n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right| \\
&= \left|\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l\{dN_{ik}(t) - Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\}\right. \\
&\left. - n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l\{dN_{ik}(t) - Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\}\right| \\
&\leq \left|\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l dN_{ik}(t)\right| + \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l dN_{ik}(t)\right| \\
&+ \left|\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right| + \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\widetilde{\boldsymbol{Z}}_{ik}(t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right| \\
&\leq \left|\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t)\right| + \left|\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{e}(\boldsymbol{\beta},t)\}_l dN_{ik}(t)\right| \\
&+ \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t)\right| + \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{e}(\boldsymbol{\beta},t)\}_l dN_{ik}(t)\right| \\
&+ \left|\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right| + \left|\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{e}(\boldsymbol{\beta},t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right| \\
&+ \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right| \\
&+ \left|n_s^{-1}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\int_0^{\tau}\{\boldsymbol{e}(\boldsymbol{\beta},t)\}_l Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}d\Lambda_0(t)\right|
\end{aligned}
$$

where the subscript l denote the l element of the indicated vector. Thus, for $s = 0, 1$,

$$
\begin{aligned}
& \max_i \left| \sum_{k=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) - n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) \right|^2 \\
\leq \; & \max_i \left(\left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t) \right| + \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l dN_{ik}(t) \right| \right. \\
+ \; & \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t) \right| + \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l dN_{ik}(t) \right| \\
+ \; & \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| + \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| \\
+ \; & \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| \\
+ \; & \left. \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| \right)^2
\end{aligned}
$$

Note that

$$
\begin{aligned}
\max_i \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t) \right| & \leq \sup_{t,i} \sum_{k=1}^{K} |\boldsymbol{Z}_{ik}(t)| \\
\max_i \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l dN_{ik}(t) \right| & \leq \sup_{t,i} \sum_{k=1}^{K} |\boldsymbol{Z}_{ik}(t)| \\
\max_i \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l dN_{ik}(t) \right| & \leq \sup_{\boldsymbol{\beta},t} \sum_{k=1}^{K} |\boldsymbol{e}(\boldsymbol{\beta}, t)_l| \\
\max_i \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l dN_{ik}(t) \right| & \leq \sup_{\boldsymbol{\beta},t} \sum_{k=1}^{K} |\boldsymbol{e}(\boldsymbol{\beta}, t)_l| \\
\max_i \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} \sum_{k=1}^{K} |\boldsymbol{Z}_{ik}(t)| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau) \\
\max_i \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{Z}_{ik}(t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} \sum_{k=1}^{K} |\boldsymbol{Z}_{ik}(t)| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau) \\
\max_i \left| \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} \sum_{k=1}^{K} |\boldsymbol{e}(\boldsymbol{\beta}, t)_l| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau) \\
\max_i \left| n_s^{-1} \sum_{i=1}^{n_s} \sum_{k=1}^{K} \int_0^\tau \{\boldsymbol{e}(\boldsymbol{\beta}, t)\}_l Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} d\Lambda_0(t) \right| & \leq \sup_{\boldsymbol{\beta},t,i} \sum_{k=1}^{K} |\boldsymbol{e}(\boldsymbol{\beta}, t)_l| \, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} \Lambda_0(\tau)
\end{aligned}
$$

To verify the condition (c) in lemma 3, it suffices to show that $n_s^{-1}\sum_{k=1}^K \sup_{t,i} |\boldsymbol{Z}_{ik}(t)|$, $n_s^{-1}\sum_{k=1}^K \sup_{\boldsymbol{\beta},t} |\boldsymbol{e}(\boldsymbol{\beta},t)_l|$, $n_s^{-1}\sum_{k=1}^K \sup_{\boldsymbol{\beta},t,i} |\boldsymbol{Z}_{ik}(t)|\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}\Lambda_0(\tau)$, and $n_s^{-1}\sum_{k=1}^K \sup_{\boldsymbol{\beta},t,i} |\boldsymbol{e}(\boldsymbol{\beta},t)_l|\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}\Lambda_0(\tau)$ converge to zero in probability as $n \to \infty$ for $s = 0, 1$. This holds by our conditions (C), (E), (G) and (K). Thus, condition (c) in lemma 3 is satisfied. Conditions (a) and (b) in lemma 3 are satisfied on the basis of conditions (I), (J) and (L). This implies $n^{-1/2}\sum_{i=1}^n (w_i - 1)\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})$ converges to a mean-zero normal random variable.

Note that U_{11} and U_{12} are independent since

$$\mathrm{Cov}(U_{11}, U_{12}) = \mathrm{E}(U_{11}U_{12}) = \mathrm{E}(\mathrm{E}(U_{11}U_{12}|\mathcal{F}(\tau))) = \mathrm{E}(U_{11}\,\mathrm{E}(U_{12}|\mathcal{F}(\tau))) = 0.$$

Therefore, $n^{1/2}\boldsymbol{U}_n(\boldsymbol{\beta})$ is asymptotically normally distributed with mean zero and with finite variance $\boldsymbol{Q}(\boldsymbol{\beta}_0)+\boldsymbol{V}(\boldsymbol{\beta}_0)$. Hence $\boldsymbol{U}_n(\boldsymbol{\beta})$ converges to zero in probability. Thus, (iv) is satisfied. By (i),(ii),(iii) and (iv), it follows that there is a unique sequence $\widehat{\boldsymbol{\beta}}$ $s.t.$ $\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) = 0$ with probability converging to one as $n \to 0$ and with $\widehat{\boldsymbol{\beta}}$ converging in probability to $\boldsymbol{\beta}_0$ by extension of Foutz (1977, Thm.2,).

The asymptotic normality of $\widehat{\boldsymbol{\beta}}$ follows from the consistency of $\widehat{\boldsymbol{\beta}}$ and a Taylor series expansion of $\widehat{\boldsymbol{U}}(\boldsymbol{\beta})$.

Proof of theorem 3.2 One can make decomposition

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}_0(\widehat{\boldsymbol\beta},t)-\Lambda_0(t)\} \\
= \;& n^{1/2}\left\{\widehat{\Lambda}_0(\widehat{\boldsymbol\beta},t)-\int_0^t \frac{d\widehat{N}(u)}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right\} + n^{1/2}\left\{\int_0^t \frac{d\widehat{N}(u)}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}-\Lambda_0(t)\right\} \\
= \;& n^{1/2}\int_0^t\left(\frac{1}{n\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)\left(d\widehat{N}(u)-\sum_{k=1}^K\sum_{i=1}^n w_iY_{ik}(u)e^{\boldsymbol\beta_0^T\boldsymbol Z_{ik}(t)}d\Lambda_0(u)\right) \\
+ \;& n^{1/2}\int_0^t\left(\frac{1}{n\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)\sum_{k=1}^K\sum_{i=1}^n w_iY_{ik}(u)e^{\boldsymbol\beta_0^T\boldsymbol Z_{ik}(t)}d\Lambda_0(u) \\
+ \;& n^{1/2}\int_0^t\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\left(d\widehat{N}(u)-n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)d\Lambda_0(u)\right) \\
= \;& n^{1/2}\int_0^t\left(\frac{1}{n\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)d\widehat{M}(u) \\
+ \;& n^{1/2}\int_0^t\left(\frac{1}{n\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)d\Lambda_0(u) \\
+ \;& n^{1/2}\int_0^t\frac{1}{n\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}d\widehat{M}(u) \quad \text{where} \qquad (3.9) \\
& \widehat{M}(t)=\sum_{k=1}^K\sum_{i=1}^n\left(w_iN_{ik}(t)-\int_0^t w_iY_{ik}(t)e^{\boldsymbol\beta^T\boldsymbol Z_{ik}(u)}d\Lambda_0(u)\right)
\end{aligned}
$$

One can write the first term of (4.25) as

$$
\begin{aligned}
& \sum_{k=1}^K\int_0^t\left(\frac{1}{\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)dn^{-1/2}\bar{M}_k(u) \\
+ \;& \sum_{k=1}^K\int_0^t\left(\frac{1}{\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)}-\frac{1}{\widehat{S}^{(0)}(\boldsymbol\beta_0,u)}\right)d\left\{n^{-1/2}\sum_{i=1}^n(w_i-1)M_{ik}(u)\right\} \qquad (3.10)
\end{aligned}
$$

By the Taylor expansion of $\widehat{S}^{(0)}(\widehat{\boldsymbol\beta},u)^{-1}$ around $\boldsymbol\beta_0$, the first term of (3.10), can be shown to be equivalent to

$$\sum_{k=1}^K\int_0^t -\frac{\widehat{\boldsymbol S}^{(1)}(\boldsymbol\beta^*,u)^T}{\widehat{S}^{(0)}(\boldsymbol\beta^*,u)^2}(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)dn^{-1/2}\sum_{i=1}^n M_{ik}(u) \qquad (3.11)$$

where $\boldsymbol\beta^*$ is on the line segment between $\widehat{\boldsymbol\beta}$ and $\boldsymbol\beta_0$. Again, $\widehat{S}^{(0)}$ and $\widehat{\boldsymbol S}^{(1)}$ are bounded and sums of monotone functions. Then, together with the consistency of $\widehat{\boldsymbol\beta}$, $\widehat{S}^{(0)}(\boldsymbol\beta^*,u)$, $\widehat{\boldsymbol S}^{(1)}(\boldsymbol\beta^*,u)$ and the weak convergence of $n^{-1/2}\sum_{i=1}^n M_{ik}(t)$ with continuous sample paths, (4.26) converges to 0 uniformly in t in probability by applying lemma 2. By the same argument, together with the weak convergence of $n^{-1/2}\sum_{i=1}^{n}(w_i-1)M_{ik}(t)$ with continuous sample paths, it follows from lemma 2 that the second term of (3.10) converges to 0 uniformly in t in probability. Combining these results, the first term of (4.25) converges to 0 uniformly in t in probability. Again, by the Taylor expansion of $\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},u)^{-1}$ around $\boldsymbol{\beta}_0$, it can be shown that the second term of (4.25) is equal to

$$n^{1/2}\int_0^t\left(-\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}_0,u)\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,u)^T}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,u)^2}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)d\Lambda_0(u) \tag{3.12}$$

Since $n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{A}^{-1}(\boldsymbol{\beta}^*)n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}w_iM_{\tilde{\boldsymbol{z}},ik}+o_p(1)$, it follows from the consistency of $\widehat{S}^{(0)}(\boldsymbol{\beta}_0,u)$, $\widehat{S}^{(0)}(\boldsymbol{\beta}^*,u)$, $\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,u)$, $\widehat{\boldsymbol{\beta}}$ and the boundedness condition on $\Lambda_0(\cdot)$ that

$$(4.27)=\left(-\int_0^t\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0,u)^T}{s^{(0)}(\boldsymbol{\beta}_0,u)}d\Lambda_0(u)\right)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}w_i\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)+o_p(1)$$

One can write the third term of (4.25) as

$$\sum_{k=1}^{K}\int_0^t\frac{1}{\widehat{S}^{(0)}(\boldsymbol{\beta}_0,u)}dn^{-1/2}\sum_{i=1}^{n}M_{ik}(u)+\sum_{k=1}^{K}\int_0^t\frac{1}{\widehat{S}^{(0)}(\boldsymbol{\beta}_0,u)}d\left\{n^{-1/2}\sum_{i=1}^{n}(w_i-1)M_{ik}(u)\right\} \tag{3.13}$$

Since $\widehat{S}^{(0)}(\boldsymbol{\beta}_0,u)^{-1}$ is a sum of two monotone functions in t and converges uniformly to $s^{(0)}(\boldsymbol{\beta}_0,u)$, and $n^{-1/2}\sum_{i=1}^{n}M_{ik}(t)$ converges to a zero-mean Gaussian process with continuous sample path, it follows from lemma 2 that the first term of (3.13) is asymptotically equivalent to

$$\sum_{k=1}^{K}\int_0^t\frac{1}{s^{(0)}(\boldsymbol{\beta}_0,u)}d\left\{n^{-1/2}\sum_{i=1}^{n}M_{ik}(u)\right\}$$

by applying lemma 2. By the same argument, the second term of (3.13) is asymptotically equivalent to

$$\sum_{k=1}^{K}\int_0^t\frac{1}{s^{(0)}(\boldsymbol{\beta}_0,u)}d\left\{n^{-1/2}\sum_{i=1}^{n}(w_i-1)M_{ik}(u)\right\}$$

By combining the results, we have

$$\begin{aligned} n^{1/2}(\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t) \;\; - \;\; \Lambda_0(t)) \;&=\; n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\left\{\int_0^t \frac{dM_{ik}(u)}{s^{(0)}(\boldsymbol{\beta}_0,u)} + \boldsymbol{r}(\boldsymbol{\beta}_0,t)^T\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\right\} \\ &+\; n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}(w_i-1)\left\{\int_0^t \frac{dM_{ik}(u)}{s^{(0)}(\boldsymbol{\beta}_0,u)} + \boldsymbol{r}(\boldsymbol{\beta}_0,t)^T\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\mathbf{M}_{\widetilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\right\} \\ &=\; U_1^{\Lambda}(\boldsymbol{\beta}_0,t) \;+\; U_2^{\Lambda}(\boldsymbol{\beta}_0,t) \end{aligned}$$

The first term converges weakly to a zero-mean Gaussian process with covariance function $\phi(t_1,t_2)$ at (t_1,t_2) (Spiekerman and Lin, 1998). The weak convergence of the second term to a zero-mean Gaussian process with covariance function $\sigma(t_1,t_2)$ at (t_1,t_2) follows from lemma 3 and the example 3.6.14 of van der Vaart and Wellner (1996, p356). Note that these two terms are independent since $\mathrm{Cov}(U_1^{\Lambda},U_2^{\Lambda}) = \mathrm{E}(U_1^{\Lambda}U_2^{\Lambda}) = \mathrm{E}(\mathrm{E}(U_1^{\Lambda}U_2^{\Lambda}|\mathcal{F}(\tau))) = \mathrm{E}(U_1^{\Lambda}\,\mathrm{E}(U_2^{\Lambda}|\mathcal{F}(\tau))) = 0$. This completes the proof.

Proof of theorem 3.3 and 3.4 The consistency of $\widehat{\boldsymbol{\beta}}_c$ and $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}_c,t)$, and the asymptotic normality of $n^{1/2}(\widehat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0)$ and the weak convergence of $n^{1/2}(\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}_c,t)-\Lambda_0(t))$ can be shown by similar arguments used for proving theorem 3.1 and 3.2 replacing lemma 3 with lemma 4. Some conditions used in proving lemma 3 also need to be replaced. Specifically, $\sup_{t,\boldsymbol{\beta}}\left\|\widehat{\boldsymbol{S}}^{(d)}(\boldsymbol{\beta},\ t)-\boldsymbol{s}^{(d)}(\boldsymbol{\beta},\ t)\right\| \xrightarrow{p} 0$, as $n\to\infty$ for $d=0,1,2$ in the asymptotic stability condition for case-control samples, (L), needs to be replaced by
$\sup_{t,\boldsymbol{\beta}}\left\|\widehat{\boldsymbol{S}}_c^{(d)}(\boldsymbol{\beta},\ t)-\boldsymbol{s}^{(d)}(\boldsymbol{\beta},\ t)\right\| \xrightarrow{p} 0$, as $n\to\infty$ for $d=0,1,2$.

TABLE 3.1: Summary of simulation results. $Z_{i1} = 1$ and $Z_{i2} = 0$.

β_0	n	event proportion	θ	mean $\hat{\beta}_0$	indep. s.e.	proposed s.e.	true S.D.	95% Coverage
0	1000	10%	0.1	0.019	0.1434	0.2094	0.2177	0.913
			0.67	0.035	0.1446	0.2976	0.3096	0.944
			1.25	0.048	0.1450	0.3110	0.3268	0.948
			4	0.038	0.1451	0.3234	0.3337	0.951
		20%	0.1	0.001	0.1003	0.0822	0.0839	0.931
			0.67	-0.003	0.1003	0.1348	0.1339	0.954
			1.25	0.007	0.1004	0.1435	0.1437	0.947
			4	0.001	0.1005	0.1522	0.1507	0.948
	2000	10%	0.1	0.010	0.1005	0.1491	0.1589	0.919
			0.67	0.019	0.1009	0.2090	0.2101	0.942
			1.25	0.030	0.1013	0.2181	0.2139	0.955
			4	0.024	0.1011	0.2248	0.2240	0.949
		20%	0.1	0.003	0.0708	0.0579	0.0577	0.943
			0.67	0.001	0.0708	0.0952	0.0960	0.945
			1.25	0.005	0.0708	0.1013	0.1024	0.948
			4	0.003	0.0709	0.1071	0.1061	0.948
0.693	1000	10%	0.1	0.734	0.1741	0.2483	0.2659	0.826
			0.67	0.768	0.1795	0.3886	0.4203	0.927
			1.25	0.789	0.1810	0.4136	0.4423	0.946
			4	0.781	0.1818	0.4347	0.4564	0.950
		20%	0.1	0.699	0.1195	0.1028	0.1027	0.949
			0.67	0.706	0.1201	0.1708	0.1730	0.947
			1.25	0.709	0.1201	0.1821	0.1852	0.942
			4	0.699	0.1200	0.1933	0.1940	0.946
	2000	10%	0.1	0.708	0.1209	0.1800	0.1891	0.913
			0.67	0.734	0.1229	0.2747	0.2859	0.939
			1.25	0.737	0.1232	0.2884	0.2964	0.945
			4	0.732	0.1232	0.2990	0.3164	0.939
		20%	0.1	0.693	0.0842	0.0725	0.7235	0.948
			0.67	0.694	0.0843	0.1203	0.1172	0.960
			1.25	0.702	0.0845	0.1284	0.1286	0.950
			4	0.698	0.0845	0.1359	0.1363	0.945

TABLE 3.2: Summary of simulation results. $Z_{ik} \sim N(0,1)$.

β_0	n	event proportion	θ	mean $\hat{\beta}_0$	indep. s.e.	proposed s.e.	true S.D.	95% Coverage
0	1000	10%	0.1	0.003	0.0722	0.1359	0.1382	0.946
			0.67	0.001	0.0723	0.1558	0.1727	0.916
			1.25	-0.005	0.0724	0.1589	0.1679	0.934
			4	0.002	0.0721	0.1616	0.1653	0.933
		20%	0.1	0.002	0.0504	0.0742	0.0744	0.941
			0.67	0.000	0.0503	0.0805	0.0816	0.949
			1.25	-0.001	0.0503	0.0814	0.0839	0.933
			4	-0.002	0.0505	0.0833	0.0846	0.948
	2000	10%	0.1	-0.001	0.0504	0.0956	0.0952	0.949
			0.67	-0.003	0.0505	0.1114	0.1116	0.945
			1.25	0.001	0.0507	0.1137	0.1193	0.930
			4	-0.003	0.0506	0.1157	0.1133	0.951
		20%	0.1	0.001	0.0355	0.0524	0.0525	0.950
			0.67	-0.001	0.0355	0.0567	0.0566	0.950
			1.25	0.001	0.0355	0.0576	0.0583	0.946
			4	-0.003	0.0355	0.0585	0.0587	0.952
0.693	1000	10%	0.1	0.708	0.0763	0.1573	0.1678	0.922
			0.67	0.709	0.0767	0.1717	0.1863	0.914
			1.25	0.704	0.0763	0.1735	0.1873	0.919
			4	0.706	0.0761	0.1763	0.1895	0.917
		20%	0.1	0.698	0.0526	0.0836	0.0843	0.939
			0.67	0.700	0.0527	0.0865	0.0891	0.937
			1.25	0.695	0.0525	0.0866	0.0864	0.950
			4	0.700	0.0527	0.0873	0.0890	0.938
	2000	10%	0.1	0.705	0.0529	0.1133	0.1166	0.937
			0.67	0.702	0.0531	0.1239	0.1293	0.932
			1.25	0.704	0.0529	0.1244	0.1279	0.931
			4	0.695	0.0530	0.1265	0.1298	0.934
		20%	0.1	0.692	0.0370	0.0595	0.0609	0.940
			0.67	0.694	0.0370	0.0610	0.0617	0.944
			1.25	0.696	0.0370	0.0616	0.0594	0.952
			4	0.695	0.0370	0.0617	0.0635	0.940

TABLE 3.3: Summary of simulation results. $Z_{i1} = 1$, $Z_{i2} = 0$. $T_{ik} \sim Weibull(1, 0.5)$.

β_0	n	event proportion	θ	mean $\hat{\beta}_0$	indep. s.e.	proposed s.e.	true S.D.	95% Coverage
0	1000	10%	0.1	0.013	0.1410	0.2106	0.2181	0.923
			0.67	0.030	0.1423	0.2918	0.3052	0.934
			1.25	0.041	0.1425	0.3030	0.3187	0.939
			4	0.026	0.1424	0.3130	0.3183	0.951
		20%	0.1	0.002	0.1000	0.0853	0.0850	0.945
			0.67	0.005	0.1002	0.1362	0.1387	0.943
			1.25	0.005	0.1000	0.1435	0.1435	0.952
			4	0.003	0.1003	0.1517	0.1553	0.944
	2000	10%	0.1	0.002	0.0992	0.1490	0.1536	0.934
			0.67	0.022	0.0996	0.2048	0.2090	0.947
			1.25	0.022	0.0996	0.2122	0.2143	0.953
			4	0.022	0.0996	0.2182	0.2249	0.944
		20%	0.1	0.002	0.0706	0.0602	0.0608	0.942
			0.67	0.000	0.0706	0.0960	0.0953	0.952
			1.25	0.003	0.0706	0.1015	0.1041	0.946
			4	0.001	0.0706	0.1065	0.1051	0.953
0.693	1000	10%	0.1	0.729	0.1742	0.2617	0.2856	0.845
			0.67	0.766	0.1792	0.3990	0.4212	0.935
			1.25	0.773	0.1809	0.4169	0.4536	0.943
			4	0.770	0.1831	0.4373	0.4919	0.944
		20%	0.1	0.698	0.1215	0.1106	0.1098	0.945
			0.67	0.704	0.1221	0.1808	0.1811	0.946
			1.25	0.710	0.1221	0.1912	0.1926	0.946
			4	0.702	0.1222	0.2019	0.2043	0.950
	2000	10%	0.1	0.713	0.1212	0.1892	0.1976	0.920
			0.67	0.735	0.1230	0.2803	0.2860	0.949
			1.25	0.739	0.1235	0.2922	0.3075	0.941
			4	0.730	0.1230	0.3003	0.3042	0.961
		20%	0.1	0.697	0.0857	0.0778	0.0815	0.937
			0.67	0.698	0.0858	0.1275	0.1281	0.952
			1.25	0.700	0.0859	0.1349	0.1311	0.961
			4	0.700	0.0859	0.1415	0.1399	0.953

TABLE 3.4: Summary of simulation results. Inclusion probabilities vs. Local averages. The covariate is uniformly distributed on five points, $m/5$, $1 \leq m \leq 5$.

n	β_0	θ	Approach	mean $\hat{\beta}_0$	proposed s.e.	true S.D.	95% Coverage
1000	0	0.1	Inclusion Probabilities	0.004	0.2407	0.2399	0.949
			Local Average	-0.004	0.2214	0.2251	0.945
		0.67	Inclusion Probabilities	0.001	0.2728	0.2703	0.949
			Local Average	0.001	0.2624	0.2674	0.944
		1.25	Inclusion Probabilities	0.002	0.2779	0.2803	0.953
			Local Average	0.004	0.2682	0.2779	0.948
		4	Inclusion Probabilities	-0.006	0.2846	0.2845	0.954
			Local Average	-0.009	0.2759	0.2816	0.943
	log(2)	0.1	Inclusion Probabilities	0.702	0.2478	0.2503	0.950
			Local Average	0.699	0.2288	0.2343	0.944
		0.67	Inclusion Probabilities	0.702	0.2821	0.2864	0.946
			Local Average	0.693	0.2708	0.2800	0.940
		1.25	Inclusion Probabilities	0.706	0.2866	0.2899	0.948
			Local Average	0.692	0.2760	0.2839	0.943
		4	Inclusion Probabilities	0.710	0.2928	0.2981	0.952
			Local Average	0.697	0.2835	0.2875	0.952
2000	0	0.1	Inclusion Probabilities	-0.001	0.1697	0.1712	0.949
			Local Average	-0.000	0.1568	0.1571	0.949
		0.67	Inclusion Probabilities	0.004	0.1930	0.1883	0.956
			Local Average	0.003	0.1866	0.1839	0.954
		1.25	Inclusion Probabilities	-0.000	0.1961	0.1940	0.956
			Local Average	-0.002	0.1903	0.1912	0.951
		4	Inclusion Probabilities	-0.001	0.2002	0.1962	0.957
			Local Average	-0.005	0.1955	0.1946	0.952
	log(2)	0.1	Inclusion Probabilities	0.693	0.1747	0.1761	0.948
			Local Average	0.696	0.1621	0.1641	0.946
		0.67	Inclusion Probabilities	0.691	0.1992	0.2027	0.943
			Local Average	0.695	0.1928	0.1943	0.945
		1.25	Inclusion Probabilities	0.690	0.2023	0.2052	0.953
			Local Average	0.694	0.1966	0.1990	0.945
		4	Inclusion Probabilities	0.692	0.2070	0.2104	0.951
			Local Average	0.692	0.2020	0.2062	0.946

TABLE 3.5: Baseline characteristics of KPDCP data

		RCF Frequency	%	Non-RCF Frequency	%
Tooth type	Molar	88	43.6	88	43.6
	Non-molar	114	56.4	114	56.4
Proximal contacts	BA*	16	7.9	9	4.5
	NBA**/0	7	3.5	7	3.5
	NBA/1	58	28.7	61	30.2
	NBA/2	121	59.9	125	61.8
Pockets	0	138	68.3	141	69.8
	1	30	14.9	27	13.4
	2	17	8.4	21	10.4
	3	9	4.4	9	4.4
	4	7	3.5	1	0.5
	5	0	0.0	3	1.5
	6	1	0.5	0	0.0
DF coronal surfaces	0	7	3.5	29	14.4
	1	9	4.4	21	10.4
	2	50	24.8	49	24.3
	3	136	67.3	103	50.9
DF roots surfaces	0	143	70.8	178	88.1
	1	52	25.7	19	9.4
	2	7	3.5	5	2.5

* : Bridge abutment, ** :Non-Bridge abutment

TABLE 3.6: Data analysis for KPDCP data

		Proposed Method			Unweighted Method		
Variable	Level	HR	s.e	p-value	HR	s.e	p-value
RCF(Molar)		6.9	0.44	< 0.01	9.1	0.40	< 0.01
RCF(Non-molar)		1.8	0.57	0.30	4.7	0.30	< 0.01
Proximal Contacts							
	PC1	0.3	0.81	0.17	0.5	0.47	0.10
	PC2	0.1	0.81	0.02	0.2	0.46	< 0.01
	PCABUT	0.5	0.97	0.44	0.6	0.54	0.33
Number of Pockets$\geq 5mm$		1.3	0.15	0.09	1.2	0.09	0.08

CHAPTER 4

MARGINAL HAZARDS MODEL FOR CASE-COHORT STUDIES WITH MULTIPLE DISEASE OUTCOMES

4.1 Introduction

In large cohort studies, the major effort and cost typically arise from the assembling of covariate measurements. To reduce the cost in such studies and achieve the same goals as a cohort study, several study designs have been proposed. Case-cohort study design is one of the most widely used ones, especially when the disease rate is low. Under the case-cohort design, a random sample called subcohort is selected from the entire cohort. The covariate measurements are only assembled for the subjects in the subcohort and all the cases (failures) who experience the disease of interest regardless of whether or not they are in the subcohort. A key advantage of the case-cohort design is its ability to use the same subcohort for several diseases or for subtypes of disease (e.g., Prentice, 1986; Wacholder et al., 1989; Langholz and Thomas, 1990; Wacholder et al., 1991). For example, the case-cohort design was implemented in the Busselton Health Study (Cullen, 1972). The Busselton Population Health Surveys are a series of cross-sectional health surveys conducted in the town of Busselton in Western Australia. Every 3 years from 1966 to 1981, general health information for adult participants were collected by means of questionnaire and clinical visit. It was of interest to study the relationship between serum ferritin and coronary heart disease and stroke events. To reduce costs and preserve stored serum, case-cohort sampling was used. In order to compare the effect of serum ferritin on coronary heart disease and stroke, times to coronary heart disease and stroke events need to be modeled simultaneously. Since times to coronary heart disease and stroke events observed from the same subject could be correlated, valid statistical methods which take it into consideration need to be developed.

For data from case-cohort study for a single disease outcome, various estimating procedures have been proposed in the literature. Prentice (1986) and Self and Prentice (1988) first considered Cox model and proposed a pseudolikelihood approach based on the partial likelihood function, where the risk set was appropriately estimated to incorporate the case-cohort design. Lin and Ying (1993), and Barlow (1994) further discussed pseudolikelihood method and provided different ways to obtain an easily computed variance for the estimators of regression parameters. Chen and Lo (1999) improved the pseudolikelihood estimators by using a class of estimating equations based on the partial likelihood score function. Chen (2001) further improved the estimators by using a local type of average as weight in the estimating equations. Borgan et al. (2000) considered stratified case-cohort sampling designs and proposed several methods to analyze such study designs. Kulich and Lin (2004) developed a class of weighted estimating equations with time-dependent weights under the stratified case-cohort designs.

Despite the progress in the methods for analyzing case-cohort data for a single disease outcome, methodologies to address analysis of case-cohort data with multiple diseases outcomes have been limited. A commonly used method for dealing with multiple diseases is to analyze each disease separately. This approach does not allow comparison of the risk factors for different diseases, because it does not account for the induced correlation between outcomes (Langholz and Thomas, 1990). Statistical methods which account for the correlation between outcomes are needed.

In this chapter, we propose a weighted estimating equation approach for estimating the parameters in the marginal hazards regression models for the multivariate failure time data from case-cohort studies with multiple disease outcomes. The rest of this chapter is organized as follows. We present the proposed model and method of estimation in Section 4.2. In Section 4.3, the asymptotic properties of the proposed estimators are studied. The finite sample properties are investigated by simulations in Section 4.4. The methodology is illustrated in Section 4.5 using the aforementioned Busselton Health Study.

4.2 Model and Estimation

Suppose that there are n independent subjects in a cohort study and there are K disease outcomes of interest. Consider independent failure time response vectors $\boldsymbol{T}_i = (T_{i1}, \cdots, T_{iK})^T$, $i = 1, \cdots, n$. For example, (T_{i1}, T_{i2}) may denote time for CHD and time for stroke for subject i. Let C_{ik} denote the potential censoring time for outcome k of subject i. We assume that C_{ik} is independent of the disease processes. In most practical cases, $C_{ik} = C_i$ for $k = 1, \ldots, K$. The observed time is $X_{ik} = \min(T_{ik}, C_{ik})$. Let $N_{ik}(t)$ denote the counting process for outcome k of subject i, $Y_{ik}(t) = I(X_{ik} \geq t)$ denote an 'at risk' indicator process and $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ denote an indicator for failure, where $I(\cdot)$ is an indicator function. Let $Z_{ik}(t)$ be a $p \times 1$ covariate vector corresponding to the kth disease outcome for subject i at time t. We assume that all the time-dependent covariates in $Z_{ik}(t)$ are "external", i.e., they are not affected by the disease processes, as described by Kalbfleisch and Prentice (2002). Let $\lambda_{ik}(t)$ denote the corresponding marginal hazards function and $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) \exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\} \lambda_{0k}(u) du$ denote a martingale with respect to the marginal filtration $\mathcal{F}_{ik}(t) = \sigma\{N_{ik}(s), Y_{ik}(s), \boldsymbol{Z}_{ik}(s) : 0 \leq s \leq t\}$. Let $\boldsymbol{X}_i = (X_{i1}, \cdots, X_{iK})^T, i = 1, \cdots, n$, denote the observed failure time vector and $\boldsymbol{Z}_i(\cdot) = (\boldsymbol{Z}_{i1}, \cdots, \boldsymbol{Z}_{iK})^T$ denote the covariate vector. Let τ denote the study end time.

Under the case-cohort design, suppose we select a subcohort of fixed size $\tilde{n}$ from the cohort by simple random sampling without replacement. This sampling may be done prospectively or retrospectively. Let ξ_i denote the indicator for the ith subject being selected into the subcohort and $\pi_i = \Pr(\xi_i = 1) = \tilde{\alpha} = \tilde{n}/n$ denote the selection probability of the ith subject. Here $\xi_1, \ldots, \xi_n$ are correlated due to the sampling scheme. We assume that complete covariate histories $\boldsymbol{Z}_{ik}(t)(0 \leq t \leq \tau)$ are available for all the subcohort members and for the cases outside the subcohort. For all the others, we assume that their censoring time information are available. Thus, the observable information for the kth disease outcome of the ith subject when $\xi_i = 1$ or $\Delta_{ik} = 1$ is $\{X_{ik}, \Delta_{ik}, \xi_i, \boldsymbol{Z}_{ik}(t), 0 \le t \le X_{ik}\}$ and when $\xi_i = 0$ and $\Delta_{ik} = 0$ is $\{X_{ik}, \Delta_{ik}, \xi_i\}$.

4.2.1 Multiplicative intensity models

Suppose that T_{ik} arises from a marginal intensity process model of the form (Cox, 1972)

$$\lambda_{ik}(t|\boldsymbol{Z}_{ik}(t)) = Y_{ik}(t)\ \lambda_{0k}(t)\ e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)}, \tag{4.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease outcome k and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of fixed and unknown parameters. Note that a subject may experience all K diseases, may also experience only some, or even none of the events of interest due to right censoring. Model (4.1) can incorporate failure type specific effects and includes the Wei et al. (1989) model, $\lambda_{ik}(t; \boldsymbol{Z}^*_{ik}(t)) = \lambda_{0k}(t)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{Z}^*_{ik}(t)\}$, as a special case, i.e., disease specific effects can be obtained by defining $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_k^T, ..., \boldsymbol{\beta}_K^T)^T$ and $\boldsymbol{Z}_{ik}(t) = [\mathbf{0}_{i1}^T, \mathbf{0}_{i2}^T, \cdots, \mathbf{0}_{i(k-1)}^T, \{\boldsymbol{Z}^*_{ik}(t)\}^T, \mathbf{0}_{i(k+1)}^T, \cdots, \mathbf{0}_{iK}^T]^T$, where $\mathbf{0}$ are zero vectors. Notice the equivalence of the risk scores under both notations: $\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t) = \boldsymbol{\beta}_k^T \boldsymbol{Z}^*_{ik}(t)$. The baseline hazard function is explicitly disease-specific.

4.2.2 Estimation

If the elements of each $\boldsymbol{T}_i$ were statistically independent and the data were complete, the relative risk parameter $\boldsymbol{\beta}$ in (4.1) could be estimated by solving the partial likelihood (Cox, 1972) score equation $\boldsymbol{U}(\boldsymbol{\beta}) = \mathbf{0}_{p\times 1}$, where

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \left\{ \boldsymbol{Z}_{ik}(u) - \frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}, u)}{S_k^{(0)}(\boldsymbol{\beta}, u)} \right\} dN_{ik}(u), \tag{4.2}$$

and $\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}; t) = n^{-1} \sum_{i=1}^{n} Y_{ik}(t) \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}$ for $d = 0, 1$. Here, for a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = aa^T$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 0} = 1$. This estimating equation can be solved iteratively, for example, by Newton-Raphson or Fisher Scoring method (Thistead, 1988).

Since the elements of each $\boldsymbol{T}_i$ are not statistically independent and the data are not complete, (4.2) cannot be calculated. Thus, we consider the following pseudo-likelihood score equations $\boldsymbol{U}^I(\boldsymbol{\beta}) = \mathbf{0}_{p\times 1}$, where

$$\boldsymbol{U}^I(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{ik}(u) - \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},u)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},u)}\right\} dN_{ik}(u), \tag{4.3}$$

$\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t) = n^{-1}\sum_{i=1}^{n}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}$ for $d = 0, 1$ and $\rho_{ik}(t)$ is a possibly time-dependent weight function which has the following form:

$$\rho_{ik}(t) \quad = \xi_i/\widehat{\alpha}_k(t) \quad \text{where} \quad \widehat{\alpha}_k(t) = \frac{\sum_{i=1}^{n}\xi_i Y_{ik}(t)}{\sum_{i=1}^{n} Y_{ik}(t)}$$

The estimator of the hazards regression parameter $\boldsymbol{\beta}_0$ is defined as the solution to this equation and is denoted by $\widehat{\boldsymbol{\beta}}_I$. We will call this type of estimator as Estimator I.
Here $\widehat{\alpha}_k(t)$ is the estimator of the true sampling probability $\tilde{\alpha}$ and denotes the number of sampled subjects divided by the number of subjects remaining in the risk set at time t. This type of the weight function has been considered in the univariate failure time context. It was first considered by Barlow (1994). Borgan et al. (2000) considered the same type of the weight functions for stratified case-cohort studies(Estimator I). The estimator considered by Self and Prentice (1988) is a special case and can be obtained by replacing $\widehat{\alpha}_k(t)$ by $\tilde{\alpha}$.

Let $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$. A Breslow-Aalen type estimator of the baseline cumulative hazard function is given by $\widehat{\Lambda}_{0k}^I(\widehat{\boldsymbol{\beta}}_I, t)$, where

$$\widehat{\Lambda}_{0k}^I(\boldsymbol{\beta}, t) = \int_0^t \frac{\sum_{i=1}^{n} dN_{ik}(u)}{n\widehat{S}_k^{(0)}(\boldsymbol{\beta},u)}. \tag{4.4}$$

Note that $\widehat{\alpha}_k(t)$ does not include the cases outside the subcohort and Estimator I needs the covariate measurement of the cases outside the subcohort only at their failure times. However, when the complete covariate measurement history is available for the cases outside the subcohort, Estimator I might not be very efficient since it discards some of the available information. To make better use of the available information, we consider the following pseudo-likelihood equations $\boldsymbol{U}^{II}(\boldsymbol{\beta}) = \mathbf{0}_{p\times 1}$, where

$$\boldsymbol{U}^{II}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{ik}(u) - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)}\right\} dN_{ik}(u), \tag{4.5}$$

$\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta};t) = n^{-1}\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}$ for $d = 0, 1$ and $\omega_{ik}(t)$ is a possibly time-dependent weight function which has the following form:

$$\omega_{ik}(t) = \Delta_{ik} + (1-\Delta_{ik})\xi_i/\widehat{\alpha}_k^{II}(t) \quad \text{where} \quad \widehat{\alpha}_k^{II}(t) = \frac{\sum_{i=1}^{n}\xi_i(1-\Delta_{ik})Y_{ik}(t)}{\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t)}$$

The estimator of the hazards regression parameter $\boldsymbol{\beta}_0$ is defined as the solution to this equation and is denoted by $\widehat{\boldsymbol{\beta}}_{II}$. We will call this type of estimator as Estimator II.

This weight function is defined to be equal to one for the cases regardless of their subcohort membership and to $\widehat{\alpha}_k^{II}(t)^{-1}$ for the sampled censored individuals. Thus, $\widehat{\alpha}_k^{II}(t)$ is constructed using only censored individuals. Unlike the weight function for Estimator I, it uses the information from all the individuals sampled. Consequently, it is anticipated that this results in a more efficient estimator. This approach also has been considered in the univariate failure time data. It was first proposed by Kalbfleisch and Lawless (1988) and they considered a time-invariant version of the weight functions, i.e., they used $\widetilde{\alpha}$ instead of $\widehat{\alpha}_k^{II}(t)$. Borgan et al. (2000) considered the same type of the weight functions in the univariate failure time data from stratified case-cohort studies(Estimator II). To be able to use this approach, one is required to assess complete covariate histories for the cases throughout their at-risk periods, which might not be always available for prospective studies. In case of having complete covariate histories for the cases, using this type of weights is expected to improve efficiencies.

The Breslow-Aalen type estimator of the cumulative baseline hazard function will have the following form:

$$\widehat{\Lambda}_{0k}^{II}(\boldsymbol{\beta},t) = \int_0^t \frac{\sum_{i=1}^{n} dN_{ik}(u)}{n\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)}. \tag{4.6}$$

4.3 Asymptotic properties

In this section, we describe the asymptotic properties of the proposed estimates. We define the following notation for convenience: For $k = 1, \ldots, K$,

$$\begin{aligned}
&\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta}, t) = \mathrm{E}\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t)\} \ (d = 0, 1, 2), \quad \boldsymbol{e}_k(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}, t)}{s_k^{(0)}(\boldsymbol{\beta}, t)}, \\
&\boldsymbol{v}_k(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}_k^{(2)}(\boldsymbol{\beta}, t) s_k^{(0)}(\boldsymbol{\beta}, t) - \boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}, t)^{\otimes 2}}{s_k^{(0)}(\boldsymbol{\beta}, t)^2}, \\
&\boldsymbol{A}_k(\boldsymbol{\beta}) = \int_0^\tau \boldsymbol{v}_k(\boldsymbol{\beta}, t) s_k^{(0)}(\boldsymbol{\beta}, t) \lambda_{0k}(t) dt, \\
&\tilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) = \boldsymbol{Z}_{ik}(t) - \boldsymbol{e}_k(\boldsymbol{\beta}, t), \quad \text{and} \quad \mathbf{M}_{\tilde{\boldsymbol{z}}, ik}(\boldsymbol{\beta}) = \int_0^\tau \tilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) dM_{ik}(t).
\end{aligned}$$

Here and hereafter the norms for the vector $\boldsymbol{a}$, matrix $\boldsymbol{A}$, and function f are defined as the following:

$$\|\boldsymbol{a}\| = \max_i |a_i|, \ \ \|\boldsymbol{A}\| = \max_{i,j} |A_{ij}|, \|f\| = \sup_t |f(t)|$$

4.3.1 Asymptotic properties of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\Lambda}_{0k}^I(t)$

We summarize the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_I$ in the following theorem :

Theorem 4.1 *Under the conditions in the Appendix, $\widehat{\boldsymbol{\beta}}_I$ solving (4.3) is a consistent estimator of $\boldsymbol{\beta}_0$. Also $n^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}_I(\boldsymbol{\beta}_0) = \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)\}\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ where*

$$\begin{aligned}
\boldsymbol{A}(\boldsymbol{\beta}) &= \sum_{k=1}^K \boldsymbol{A}_k(\boldsymbol{\beta}), \ \boldsymbol{Q}(\boldsymbol{\beta}) = \mathrm{E}\left(\sum_{k=1}^K \boldsymbol{M}_{\tilde{\boldsymbol{z}}, 1k}(\boldsymbol{\beta})\right)^{\otimes 2}, \\
\boldsymbol{V}(\boldsymbol{\beta}) &= \mathrm{E}\left(\sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{1k}(\boldsymbol{\beta}, t) d\Lambda_{0k}(t)\right)^{\otimes 2} \ \textit{and}\ \boldsymbol{R}_{ik}(\boldsymbol{\beta}, t) = Y_{ik}(t)\left(\boldsymbol{Z}_{ik} - \boldsymbol{e}_k(\boldsymbol{\beta}, t)\right) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}
\end{aligned}$$

$\boldsymbol{A}(\boldsymbol{\beta}_0)$, $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}_I), \widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\beta}}_I)$ and $\frac{1-\widetilde{\alpha}}{\widetilde{\alpha}}\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\beta}}_I)$ where

$$\begin{aligned}
\widehat{\boldsymbol{A}}(\boldsymbol{\beta}) &= -n^{-1}\frac{\partial \boldsymbol{U}^I(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n \frac{\xi_i}{\widetilde{\alpha}}\left(\sum_{k=1}^K \widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2}, \\
\widehat{\boldsymbol{V}}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^n \frac{\xi_i}{\widetilde{\alpha}}\left(\sum_{k=1}^K \int_0^\tau \widehat{\boldsymbol{R}}_{ik}(\boldsymbol{\beta},t)d\widehat{\Lambda}_{0k}^I(\boldsymbol{\beta},t)\right)^{\otimes 2}, \\
\int_0^\tau \widehat{\boldsymbol{R}}_{ik}(\boldsymbol{\beta},t)d\widehat{\Lambda}_{0k}^I(\boldsymbol{\beta},t) &= n^{-1}\sum_{j=1}^n \frac{\Delta_{jk}Y_{ik}(X_{jk})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jk})}}{\widehat{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\left(\boldsymbol{Z}_{ik}(X_{jk}) - \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{jk})}{\widehat{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\right), \\
\text{and}\quad \widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) &= \Delta_{ik}\left(\boldsymbol{Z}_{ik}(X_{ik}) - \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{ik})}{\widehat{S}_k^{(0)}(\boldsymbol{\beta};X_{ik})}\right) \\
&\quad - n^{-1}\sum_{j=1}^n \frac{\Delta_{jk}Y_{ik}(X_{jk})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jk})}}{\widehat{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\left(\boldsymbol{Z}_{ik}(X_{jk}) - \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{jk})}{\widehat{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\right).
\end{aligned}$$

To study the asymptotic properties of $\widehat{\Lambda}_{0k}^I(\widehat{\boldsymbol{\beta}}_I,t)(k=1,\ldots,K)$, we define the following metric space. Let $D[0,\tau]^K$ be a metric space consisting of right-continuous functions $\boldsymbol{f}(t)$ with left-hand limits where $\boldsymbol{f}(t) = \{f_1(t),\ldots,f_K(t)\}^T$ and $f_k(t) : [0,\tau] \to \mathcal{R}$. The metric for this space is defined as $d_k(\boldsymbol{f},\boldsymbol{g}) = \sup_{k,t\in[0,\tau]}\{|f_k(t)-g_k(t)| : 1 \le k \le K\}$ for $\boldsymbol{f},\boldsymbol{g} \in D[0,\tau]^K$. We summarize the asymptotic properties of $\widehat{\Lambda}_{0k}(\widehat{\boldsymbol{\beta}},t)(k=1,\ldots,K)$ in the following theorem.

Theorem 4.2 *Under the conditions in the Appendix, for each $k=1,\ldots,K$, $\widehat{\Lambda}_{0k}^I(\widehat{\boldsymbol{\beta}}_I,t)$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0,\tau]$. Also, $\boldsymbol{W}(t) = n^{1/2}[\{\widehat{\Lambda}_{01}^I(\widehat{\boldsymbol{\beta}}_I,t) - \Lambda_{0k}(t)\},\ldots,\{\widehat{\Lambda}_{0K}^I(\widehat{\boldsymbol{\beta}}_I,t) - \Lambda_{0K}(t)\}]^T$ converges weakly to a zero-mean Gaussian process $\mathcal{W}(t)$ in $D[0,\tau]^K$ where $\mathcal{W}(t) = (\mathcal{W}_1(t),\ldots,\mathcal{W}_K(t))^T$. The covariance function between $\mathcal{W}_j(t_1)$ and $\mathcal{W}_k(t_2)$ is*

$$\phi_{jk}(t_1,t_2)(\boldsymbol{\beta}_0) = \mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0,t_1)\nu_{1k}(\boldsymbol{\beta}_0,t_2)\} + \frac{1-\alpha}{\alpha}\mathrm{E}\{\psi_{1j}(\boldsymbol{\beta}_0,t_1)\psi_{1k}(\boldsymbol{\beta}_0,t_2)\}$$

where

$$
\begin{aligned}
\nu_{ik}(\boldsymbol{\beta}, t) &= \boldsymbol{r}_k(\boldsymbol{\beta}, t)^T \boldsymbol{A}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}, t) + \int_0^t \{s_k^{(0)}(\boldsymbol{\beta}, u)\}^{-1} dM_{ik}(u), \\
\psi_{ik}(\boldsymbol{\beta}, t) &= \Bigg\{ \boldsymbol{r}_k(\boldsymbol{\beta}, t)^T \boldsymbol{A}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \int_0^{\tau} \boldsymbol{R}_{im}(\boldsymbol{\beta}, u) d\Lambda_{0m}(u) \\
&+ \int_0^t Y_{ik}(u) \left(e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}, u)}{\mathrm{E}\, Y_{1k}(u)} \right) \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}, u)} \Bigg\}, \quad \text{and} \\
\boldsymbol{r}_k(\boldsymbol{\beta}, t) &= -\int_0^t \boldsymbol{e}_k(\boldsymbol{\beta}, u) d\Lambda_{0k}(u).
\end{aligned}
$$

$\phi_{jk}(t_1, t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}_{jk}(t_1, t_2)(\widehat{\boldsymbol{\beta}}_I)$ where

$$
\begin{aligned}
\widehat{\phi}_{jk}(t_1, t_2)(\boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^{n} \frac{\xi_i}{\widetilde{\alpha}} \widehat{\nu}_{ij}(\boldsymbol{\beta}, t_1) \widehat{\nu}_{ik}(\boldsymbol{\beta}, t_2) + \frac{1-\widetilde{\alpha}}{\widetilde{\alpha}} n^{-1} \sum_{i=1}^{n} \frac{\xi_i}{\widetilde{\alpha}} \widehat{\psi}_{ij}(\boldsymbol{\beta}, t_1) \widehat{\psi}_{ik}(\boldsymbol{\beta}, t_2), \\
\widehat{\nu}_{ik}(\boldsymbol{\beta}, t) &= \widehat{\boldsymbol{r}}_k(\boldsymbol{\beta}, t)^T \widehat{\boldsymbol{A}}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}) + \int_0^t \{\widehat{S}_k^{(0)}(\boldsymbol{\beta}, u)\}^{-1} d\widehat{M}_{ik}(u), \\
\widehat{\psi}_{ik}(\boldsymbol{\beta}, t) &= \Bigg\{ \widehat{\boldsymbol{r}}_k(\boldsymbol{\beta}, t)^T \widehat{\boldsymbol{A}}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \int_0^{\tau} \widehat{\boldsymbol{R}}_{im}(\boldsymbol{\beta}) d\widehat{\Lambda}_{0m}^{I}(\boldsymbol{\beta}, t) \\
&+ n^{-1} \sum_{j=1}^{n} \frac{\Delta_{jk} I(X_{jk} \le t) Y_{ik}(X_{jk})}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}, X_{jk})} \left(e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(X_{jk})} - \frac{\widehat{S}_k^{(0)}(\boldsymbol{\beta}; X_{jk})}{\widehat{\mathrm{E}} Y_{1k}(X_{jk})} \right) \\
\widehat{\boldsymbol{r}}_k(\boldsymbol{\beta}, t) &= -n^{-1} \sum_{i=1}^{n} \frac{\Delta_{ik} I(X_{ik} \le t) \widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}; X_{ik})}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}; X_{ik})^2}, \\
\int_0^t \frac{d\widehat{M}_{ik}(u)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}, u)} &= \frac{\Delta_{ik} I(X_{ik} \le t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}; X_{ik})} \\
&- n^{-1} \sum_{j=1}^{n} \frac{\Delta_{jk} I(X_{jk} \le t) Y_{ik}(X_{jk}) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(X_{jk})}}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}; X_{jk})^2} \\
\text{and } \widehat{\mathrm{E}} Y_{1k}(t) &= n^{-1} \sum_{i=1}^{n} Y_{ik}(t)
\end{aligned}
$$

4.3.2 Asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$

In this subsection, we will study the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$. As described in the Appendix, the techniques used in proving the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$ are very similar to those used for Estimator I. We summarize the asymptotic behavior of the regression parameter estimator $\widehat{\boldsymbol{\beta}}_{II}$ in the following theorem :

Theorem 4.3 *Under the conditions in the Appendix, $\widehat{\boldsymbol{\beta}}_{II}$ solving (4.5) is a consistent estimator of $\boldsymbol{\beta}_0$. Also, $n^{1/2}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}_{II}(\boldsymbol{\beta}_0) = \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)\}\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ where*

$$\boldsymbol{V}_{II}(\boldsymbol{\beta}) = \mathrm{E}\left[\sum_{k=1}^{K}(1-\Delta_{1k})\int_0^{\tau}\left\{\boldsymbol{R}_{1k}(\boldsymbol{\beta},u) - \frac{Y_{1k}(u)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\}d\Lambda_{0k}(u)\right]^{\otimes 2}.$$

$\boldsymbol{A}(\boldsymbol{\beta}_0)$, $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}^{II}(\widehat{\boldsymbol{\beta}}_{II}), \widehat{\boldsymbol{Q}}^{II}(\widehat{\boldsymbol{\beta}}_{II})$ and $\frac{1-\widetilde{\alpha}}{\widetilde{\alpha}}\widehat{\boldsymbol{V}}_{II}(\widehat{\boldsymbol{\beta}}_{II})$ where

$$\begin{aligned}
\widehat{\boldsymbol{A}}^{II}(\boldsymbol{\beta}) &= -n^{-1}\frac{\partial \boldsymbol{U}^{II}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}^{II}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}^{II}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2}, \\
\widehat{\boldsymbol{V}}_{II}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\sum_{k=1}^{K}\left[n^{-1}(1-\Delta_{ik})\sum_{j=1}^{n}\frac{\Delta_{jk}}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\right. \\
&\times \left.\left\{\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta};X_{jk}) - \frac{Y_{ik}(X_{jk})\widehat{\mathrm{E}}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta};X_{jk})\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(X_{jk})\right)}\right\}\right]^{\otimes 2}, \\
\widehat{\mathbf{M}}^{II}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) &= \Delta_{ik}\left\{\boldsymbol{Z}_{ik}(X_{ik}) - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{ik})}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{ik})}\right\} \\
&- n^{-1}\sum_{j=1}^{n}\frac{\Delta_{jk}Y_{ik}(X_{jk})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jk})}}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\left\{\boldsymbol{Z}_{ik}(X_{jk}) - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{jk})}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})}\right\}, \\
\widehat{\mathrm{E}}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)\right) &= n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})\frac{\xi_i}{\widetilde{\alpha}}\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta},t), \\
\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta},t) &= \left\{\boldsymbol{Z}_{ik}(t) - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},t)}\right\}Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}, \quad \text{and} \\
\widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(t)\right) &= n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t).
\end{aligned}$$

The asymptotic properties of $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II},t)(k=1,\ldots,K)$ are summarized in the following theorem.

Theorem 4.4 *Under the conditions in the Appendix, for each $k = 1,\ldots,K$, $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II},t)$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0,\tau]$. Also, $\boldsymbol{W}^{II}(t) = n^{1/2}[\{\widehat{\Lambda}^{II}_{01}(\widehat{\boldsymbol{\beta}}_{II},t) - \Lambda_{01}(t)\},\ldots,\{\widehat{\Lambda}^{II}_{0K}(\widehat{\boldsymbol{\beta}}_{II},t) - \Lambda_{0K}(t)\}]^T$ converges weakly to a zero-mean Gaussian process $\mathcal{W}^{II}(t)$*

in $D[0,\tau]^K$ *where* $\mathcal{W}^{II}(t) = (\mathcal{W}_1^{II}(t), \ldots, \mathcal{W}_K^{II}(t))^T$. *The covariance function between* $\mathcal{W}_j^{II}(t_1)$ *and* $\mathcal{W}_k^{II}(t_2)$ *is*

$$\phi_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}_0) = \mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0,t_1)\nu_{1k}(\boldsymbol{\beta}_0,t_2)\} + \frac{1-\alpha}{\alpha}\,\mathrm{E}\left\{\psi_{1j}^{II}(\boldsymbol{\beta}_0,t_1)\psi_{1k}^{II}(\boldsymbol{\beta}_0,t_2)\right\}$$

where

$$\begin{aligned}
\nu_{ik}(\boldsymbol{\beta},t) =&\ \boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta},u) + \int_0^t \{s_k^{(0)}(\boldsymbol{\beta},u)\}^{-1}dM_{ik}(u), \text{ and}\\
\psi_{ik}^{II}(\boldsymbol{\beta},t) =&\ \Bigg[\boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^K(1-\Delta_{im})\\
&\times\int_0^\tau\left\{\boldsymbol{R}_{im}(\boldsymbol{\beta},u) - \frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right\}d\Lambda_{0m}(u)\\
&+(1-\Delta_{ik})\int_0^t Y_{ik}(u)\left(e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)} - \frac{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)}\right)}{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right)\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta},u)}\Bigg].
\end{aligned}$$

$\phi_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}_{jk}^{II}(t_1,t_2)(\widehat{\boldsymbol{\beta}}_{II})$ where

$$\begin{aligned}
\widehat{\phi}_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}) =&\ n^{-1}\sum_{i=1}^n\frac{\xi_i}{\widetilde{\alpha}}\widehat{\nu}_{ij}^{II}(\boldsymbol{\beta},t_1)\widehat{\nu}_{ik}^{II}(\boldsymbol{\beta},t_2) + \frac{1-\widetilde{\alpha}^{-1}}{n}\sum_{i=1}^n\frac{\xi_i}{\widetilde{\alpha}}\widehat{\psi}_{ij}^{II}(\boldsymbol{\beta},t_1)\widehat{\psi}_{ik}^{II}(\boldsymbol{\beta},t_2),\\
\widehat{\nu}_{ik}^{II}(\boldsymbol{\beta},t) =&\ \widehat{\boldsymbol{r}}_k^{II}(\boldsymbol{\beta},t)^T\widehat{\boldsymbol{A}}^{II}(\boldsymbol{\beta})^{-1}\sum_{k=1}^K\widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},ik}^{II}(\boldsymbol{\beta}) + \int_0^t\{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)\}^{-1}d\widehat{M}_{ik}^{II}(u),\\
\widehat{\psi}_{ik}^{II}(\boldsymbol{\beta},t) =&\ \Bigg[\widehat{\boldsymbol{r}}_k^{II}(\boldsymbol{\beta},t)^T\widehat{\boldsymbol{A}}^{II}(\boldsymbol{\beta})^{-1}\sum_{m=1}^K(1-\Delta_{im})\int_0^\tau\Bigg\{\widehat{\boldsymbol{R}}_{im}^{II}(\boldsymbol{\beta},u)\\
&-\frac{Y_{im}(u)\widehat{\mathrm{E}}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta},u)\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\Bigg\}d\widehat{\Lambda}_{0m}^{II}(\boldsymbol{\beta},u) + (1-\Delta_{ik})\int_0^t Y_{ik}(u)\\
&\times\left(e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)} - \frac{\widehat{\mathrm{E}}\left((1-\Delta_{1k})\,Y_{1k}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)}\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right)\frac{d\widehat{\Lambda}_{0k}^{II}(\boldsymbol{\beta},u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)}\Bigg],\\
\widehat{\boldsymbol{r}}_k^{II}(\boldsymbol{\beta},t) =&\ -n^{-1}\sum_{i=1}^n\frac{\Delta_{ik}I(X_{ik}\le t)\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta};X_{ik})}{\widetilde{S}^{(0)}(\boldsymbol{\beta};X_{ik})^2},\\
\int_0^t\frac{d\widehat{M}_{ik}^{II}(u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)} =&\ \frac{\Delta_{ik}I(X_{ik}\le t)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{ik})} - n^{-1}\sum_{j=1}^n\frac{\Delta_{jk}I(X_{jk}\le t)Y_{ik}(X_{jk})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(X_{jk})}}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta};X_{jk})^2}, \quad \text{and}\\
\widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)}\right) =&\ n^{-1}\sum_{i=1}^n(1-\Delta_{ik})\frac{\xi_i}{\widetilde{\alpha}}Y_{ik}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}
\end{aligned}$$

The proofs of the theorems are outlined in the appendix.

4.3.3 Stratified case-cohort sampling

The main purpose of case-cohort study designs is to reduce the cost of assembling expensive covariate measurement. Thus, we assume that these expensive covariate measurements are available only for the subcohort members or cases outside the subcohort. However, some covariate information such as gender or race might be available for all the cohort members. Several authors including Borgan et al. (2000), and Kulich and Lin (2004) considered a stratified case-cohort study designs for univariate failure time based on this extra information. Our proposed estimating procedures for multiple disease outcomes can be easily extended to this stratified case-cohort study designs.

Suppose the cohort is divided into Q mutually exclusive strata based on a discrete variable S which are available for all the cohort members. This S may involve $\boldsymbol{X}$, $\boldsymbol{Z}(.)$ and some other variables related to $\boldsymbol{X}$ and $\boldsymbol{Z}(.)$ at the time of sampling. As in Kulich and Lin (2004), we require that S affects the failure time only through the covariates. Then, we assume we select the subcohort by simple random sampling without replacement within each strata and this selection of the subcohort is independent across the strata. Let n_q denote the number of subjects in stratum q, $\tilde{n}_q$ denote the size of subcohort in stratum q, and ξ_{qi} be the indicator for subject i being sampled in the subcohort in stratum q. Then, for each $q = 1, \ldots, Q$, the selection probability of the ith subject being sampled in the subcohort in stratum q is $\Pr(\xi_{qi} = 1) = \frac{\tilde{n}_q}{n_q} = \tilde{\alpha}_q$. We will assume that $\widetilde{\alpha}_q \to \alpha_q \in (0, 1)$ as $n \to \infty$ for $q = 1, \ldots, Q$. Now, Model (4.1) can be extended to

$$\lambda_{qik}(t|\boldsymbol{Z}_{qik}(t)) = Y_{qik}(t)\ \lambda_{0k}(t)\ e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{qik}(t)}, \tag{4.7}$$

for $q = 1, \ldots, Q$, $i = 1, \ldots, n_q$, and $k = 1, \ldots, K$ where the subscript qik indexes the quantity for outcome k of subject i in stratum q.

The estimating function (4.3) and a Breslow-Aalen type estimator of the baseline cumulative hazard function (4.4) can be extended to

$$\boldsymbol{U}^{I}_{st}(\boldsymbol{\beta}) = \sum_{q=1}^{Q}\sum_{i=1}^{n_q}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{qik}(u) - \frac{\widehat{\widetilde{\boldsymbol{S}}}_k^{(1)}(\boldsymbol{\beta},u)}{\widehat{\widetilde{S}}_k^{(0)}(\boldsymbol{\beta},u)}\right\} dN_{qik}(u), \tag{4.8}$$

and

$$\widehat{\widetilde{\Lambda}}^{I}_{0k}(\boldsymbol{\beta},t) = \int_0^t \frac{\sum_{q=1}^{Q}\sum_{i=1}^{n_q} dN_{qik}(u)}{n\widehat{\widetilde{S}}_k^{(0)}(\boldsymbol{\beta},u)} \tag{4.9}$$

where $\widehat{\widetilde{\boldsymbol{S}}}_k^{(d)}(\boldsymbol{\beta},t) = n^{-1}\sum_{q=1}^{Q}\sum_{i=1}^{n_q}\rho_{qik}(t)Y_{qik}(t)\boldsymbol{Z}_{qik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{qik}(t)}$ for $d = 0, 1$ and $\rho_{qik}(t)$ is an extension of $\rho_{ik}(t)$ which has the following form:

$$\rho_{qik}(t) = \xi_{qi}/\widehat{\alpha}_{qk}(t) \quad \text{where} \quad \widehat{\alpha}_{qk}(t) = \frac{\sum_{i=1}^{n_q}\xi_{qi}Y_{qik}(t)}{\sum_{i=1}^{n_q}Y_{qik}(t)}.$$

Similarly, the estimating function (4.5) and a Breslow-Aalen type estimator of the baseline cumulative hazard function (4.6) can be extended to

$$\boldsymbol{U}^{II}_{st}(\boldsymbol{\beta}) = \sum_{q=1}^{Q}\sum_{i=1}^{n_q}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{qik}(u) - \frac{\widetilde{\widetilde{\boldsymbol{S}}}_k^{(1)}(\boldsymbol{\beta},u)}{\widetilde{\widetilde{S}}_k^{(0)}(\boldsymbol{\beta},u)}\right\} dN_{qik}(u), \tag{4.10}$$

and

$$\widetilde{\widetilde{\Lambda}}^{II}_{0k}(\boldsymbol{\beta},t) = \int_0^t \frac{\sum_{q=1}^{Q}\sum_{i=1}^{n_q} dN_{qik}(u)}{n\widetilde{\widetilde{S}}_k^{(0)}(\boldsymbol{\beta},u)} \tag{4.11}$$

where $\widetilde{\widetilde{\boldsymbol{S}}}_k^{(d)}(\boldsymbol{\beta},t) = n^{-1}\sum_{q=1}^{Q}\sum_{i=1}^{n_q}\omega_{qik}(t)Y_{qik}(t)\boldsymbol{Z}_{qik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{qik}(t)}$ for $d = 0, 1$ and $\omega_{qik}(t)$ is an extension of $\omega_{ik}(t)$ which has the following form:

$$\omega_{qik}(t) = \Delta_{qik} + (1-\Delta_{qik})\xi_{qi}/\widehat{\alpha}^{II}_{qk}(t) \quad \text{where} \quad \widehat{\alpha}^{II}_{qk}(t) = \frac{\sum_{i=1}^{n_q}\xi_{qi}(1-\Delta_{qik})Y_{qik}(t)}{\sum_{i=1}^{n_q}(1-\Delta_{qik})Y_{qik}(t)}$$

The asymptotic properties of the proposed estimators under stratified case-cohort study designs follow from the similar arguments for proving theorems 4.1 - 4.4. This is because the sampling of the subcohort are independent across the strata and we can use the same arguments used for proving theorems 4.1 - 4.4 within each strata. Thus, consistency and asymptotic normality of the proposed estimators remain unchanged, but some parts of the asymptotic variances need to be replaced by the following quantities:

In theorem 4.1, $\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)$ should be replaced by $\sum_{q=1}^{Q} v_q \frac{1-\alpha_q}{\alpha_q}\boldsymbol{V}^{st}(\boldsymbol{\beta}_0)$ where

$$\begin{aligned}\boldsymbol{V}^{st}(\boldsymbol{\beta}) &= \mathrm{E}\left(\sum_{k=1}^{K}\int_0^{\tau}\boldsymbol{R}_{q1k}(\boldsymbol{\beta},t)d\Lambda_{0k}(t)\right)^{\otimes 2},\\ \boldsymbol{R}_{qik}(\boldsymbol{\beta},t) &= Y_{qik}(t)\left(\boldsymbol{Z}_{qik}(t)-\boldsymbol{e}_k(\boldsymbol{\beta},t)\right)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{qik}(t)} \quad \text{and} \quad v_q = \Pr(S=q).\end{aligned}$$

In theorem 4.2, $\frac{1-\alpha}{\alpha}\mathrm{E}\{\psi_{1j}(\boldsymbol{\beta}_0,t_1)\psi_{1k}(\boldsymbol{\beta}_0,t_2)\}$ should be replaced by $\sum_{q=1}^{Q} v_q \frac{1-\alpha_q}{\alpha_q}\mathrm{E}\{\psi_{q1j}(\boldsymbol{\beta}_0,t_1)\psi_{q1k}(\boldsymbol{\beta}_0,t_2)\}$ where

$$\begin{aligned}\psi_{qik}(\boldsymbol{\beta},t) &= \Bigg\{\boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^{K}\int_0^{\tau}\boldsymbol{R}_{qim}(\boldsymbol{\beta},u)d\Lambda_{0m}(u)\\ &+ \int_0^t Y_{qik}(u)\left(e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{qik}(u)}-\frac{s_k^{(0)}(\boldsymbol{\beta},u)}{\mathrm{E}\,Y_{q1k}(u)}\right)\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta},u)}\Bigg\}.\end{aligned}$$

In theorem 4.3, $\frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ should be replaced by $\sum_{q=1}^{Q} v_q \frac{1-\alpha_q}{\alpha_q}\boldsymbol{V}_{II}^{st}(\boldsymbol{\beta}_0)$ where

$$\begin{aligned}\boldsymbol{V}_{II}^{st}(\boldsymbol{\beta}) &= \mathrm{E}\Bigg[\sum_{k=1}^{K}(1-\Delta_{q1k})\int_0^{\tau}\{\boldsymbol{R}_{q1k}(\boldsymbol{\beta},u)\\ &- \frac{Y_{q1k}(u)\,\mathrm{E}\left((1-\Delta_{q1k})\boldsymbol{R}_{q1k}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{q1k})Y_{q1k}(u)\right)}\Bigg\}d\Lambda_{0k}(u)\Bigg]^{\otimes 2}.\end{aligned}$$

Finally, in theorem 4.4, $\frac{1-\alpha}{\alpha}\mathrm{E}\{\psi_{1j}^{II}(\boldsymbol{\beta}_0,t_1)\psi_{1k}^{II}(\boldsymbol{\beta}_0,t_2)\}$ should be replaced by $\sum_{q=1}^{Q} v_q \frac{1-\alpha_q}{\alpha_q}\mathrm{E}\{\psi_{q1j}^{II}(\boldsymbol{\beta}_0,t_1)\psi_{q1k}^{II}(\boldsymbol{\beta}_0,t_2)\}$ where

$$\begin{aligned}\psi_{qik}^{II}(\boldsymbol{\beta},t) &= \Bigg[\boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^{K}(1-\Delta_{qim})\int_0^{\tau}\{\boldsymbol{R}_{qim}(\boldsymbol{\beta},u)\\ &- \frac{Y_{qim}(u)\,\mathrm{E}\left((1-\Delta_{q1m})\boldsymbol{R}_{q1m}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{q1m})Y_{q1m}(u)\right)}\Bigg\}d\Lambda_{0m}(u)+(1-\Delta_{qik})\\ &\times \int_0^t Y_{qik}(u)\left(e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{qik}(u)}-\frac{\mathrm{E}\left((1-\Delta_{q1k})\,Y_{q1k}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{q1k}(u)}\right)}{\mathrm{E}\left((1-\Delta_{q1k})\,Y_{q1k}(u)\right)}\right)\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta},u)}\Bigg].\end{aligned}$$

4.4 Simulations

We conducted extensive simulation studies to investigate the finite sample properties of the proposed methods. Multivariate failure times were generated from the multivariate Clayton-Cuzick model (Clayton and Cuzick, 1985; Oakes, 1989) in which the joint survival function for $(T_1, \cdots, T_K)$ given $(Z_1, \cdots, Z_K)$ is:

$$S(t_1, \cdots, t_K | Z_1, \cdots, Z_K) = \left[\sum_{k=1}^{K} \exp\left\{ \frac{\int_0^{t_k} \lambda_{0k}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_k} dt}{\theta} \right\} - (K-1) \right]^{-\theta},$$

where K takes integer values. We took the marginal distribution of T_k to be exponential with failure rate $\lambda_{0k} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_k}$. Note that $\theta(> 0)$ is a parameter which represents the degree of dependence of T_k and $T_{k'}(k, k' = 1, \cdots, K)$. Smaller θ induces larger correlation. We considered two types of events (K=2). λ_{0k} was set to be equal to 2 for $k = 1$ and 4 for $k = 2$. Covariates were simulated from Bernoulli distribution with probability 0.5 and standard normal distributions. Values of 0 and log(2) were used for β; and values of 4, 1.25, 0.8 or 0.1 were considered for θ. The censoring time distribution were generated from uniform distribution $[0, u]$ with u chosen to depend on the desired percentage of censoring. We considered 97%, 90%, and 75% censoring. For each configuration, we simulated full cohort samples of size $n = 1000$ and then selected two case-cohort samples from each full cohort data. For a heavy censoring (97%), we also considered $n = 3000$ to have adequate sample size. The size of the random subcohort $\tilde{n}$ was set to have either the same expected number of controls and cases or twice as many controls as cases. The sampling was conducted via simple random sampling with fixed sample size. For each parameter combination, we ran $R = 2,000$ simulations.

Tables 4.1 and 4.2 present simulation summary statistics with Bernoulli covariate Z_{ik} with $\Pr(Z_{ik} = 1) = 0.5$ for $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$, respectively. "mean $\widehat{\boldsymbol{\beta}}_I$" or "mean $\widehat{\boldsymbol{\beta}}_{II}$" denotes the average of the estimates of $\boldsymbol{\beta}_0$, "proposed S.E." denotes the average of the estimates of standard errors based on the proposed method, "true S.D." denotes the sample standard deviation of the 2,000 estimates, and "95% C.I." denotes the coverage rate of the nominal 95% confidence interval. The simulation results suggest that the coefficient estimates are approximately unbiased for the samples considered when $\beta = 0$, while the coefficient estimates are relatively biased(4 - 10 %) when $\beta = log(2)$ with small event proportion (3%). The proposed estimated standard errors provide a very good estimate of the true variability of $\hat{\boldsymbol{\beta}}$ in most of the cases. As expected, the variance of $\hat{\boldsymbol{\beta}}$ decreases as the subcohort sample size increases. The coverage rate of the nominal 95% confidence intervals using the proposed method are in the 93% - 96% range in most of the cases considered. However, when the event proportion is very small (3%) and the expected number of cases and controls are the same in the sample ($\tilde{n} = 31$), the proposed estimated standard errors were not very accurate and the coverage rate of the nominal 95% confidence intervals using the proposed method tended to be underestimated (90.6 % - 93.4 %). The magnitude of biases, inaccuracy and underestimation were bigger for nonzero true regression parameter($\beta = \log(2)$) than $\beta = 0$. However, as the subcohort sample size increases to $\tilde{n} = 62$, the results improve. Overall, $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ showed similar results, however, as expected, $\widehat{\beta}_{II}$ was more efficient than $\widehat{\beta}_I$ in the sense that the variabilities of the regression parameter estimates were smaller for $\widehat{\beta}_{II}$.

Tables 4.3 and 4.4 provide simulation summary statistics for $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ with the standard normal covariate, respectively. The findings are similar to those of tables 4.1 and 4.2. However, the overall performance of the proposed estimators is better for the Bernoulli covariate than the standard normal one. The proposed estimated coefficients and standard errors are more accurate for the Bernoulli covariate. For a small event proportion (3%) and $\beta = \log(2)$, with the normal covariate, the magnitude of biases get bigger (up to 25 %) and the coverage rate of the nominal 95% confidence intervals tend to be more underestimated (83.8 % - 91.2 %). As subcohort size increases to $\tilde{n} = 62$, the results improve. However, unlike the Bernoulli covariate case, the improved results are still not satisfactory. Thus, we increased the cohort size to $n = 3000$ and ran the simulation under the event proportion being equal to 3% and $\beta = \log(2)$. Table 4.5 shows the results. Both methods perform reasonably well under the settings considered. This indicates that at least 100 cases are needed for valid estimates.

4.5 Analysis of Busselton Health Study

We applied the proposed methods to analyze data from Busselton Health Study (Cullen, 1972; Knuiman et al., 2003). The Busselton Health Surveys are a series of cross-sectional health surveys conducted in the town of Busselton in Western Australia. Every 3 years from 1966 to 1981, general health information for adult participants were collected by means of questionnaire and clinical visit. The population of this study is based on the 1,612 men and women aged 40-89 years who participated in the 1981 Busselton Health Survey and had no history of diagnosed CHD or stroke at that time. For both CHD or stroke, follow-up started on the 1981 survey and continued through the date of first CHD event and the date of first stroke event or December 31, 1998, whichever comes first. The subjects were treated as censored if they left Western Australia in the middle.

It was proposed that body iron stores are positively related to coronary heart disease risk (Sullivan, 1996). However, the accumulated epidemiologic evidence has been inconsistent and it is of interest to examine this hypothesis in this population. There are several measures of stored body iron and serum ferritin is regarded as the best biochemical measure of body iron store (Cook et al., 1974). To reduce costs and preserve stored serum, a case-cohort sampling was used.

We used a subset of the data for the analysis. We consider the case-cohort study to be based on this subset. There were 1,212 cohort members with 217 CHD cases and 118 stroke cases. The subcohort size was 360. Ferritin assays were conducted for all the cases and subcohort members in the total cohort. Because of overlap between CHD/stroke cases and the random subcohort, the total number of assayed sera samples was 536.

We applied our proposed methodology to this data set to study and compare the effect of serum ferritin level on the risk of CHD and of stroke. For the analyses, we included several variables as covariates to control for confounding factors. These variables were age (years), blood pressure treatment, systolic blood pressure (mmHg), BMI, cholesterol (mmol/liter), triglycerides (mmol/liter), diabetes treatment, hemoglobin (g/100 ml), and smoking (never, former, current). The total number of cohort members we considered for the analyses was

1,212 and the case-cohort analysis was conducted on the 536 subjects. The log of the serum ferritin level was used in the model as the main risk factor and we allowed for serum ferritin level to have different effects on the risk of CHD and stroke. We also considered whether the effect of serum ferritin level on the risk of CHD and stroke was different by gender.

Table 4.6 shows the baseline characteristics of the subcohort sample by gender. About 44 % (n=159) of the subcohort members were men and 56% (n=201) were women. The average age was around 59. The average of the Ferritin levels for men was about two times higher (214.2 μg/L) than that for women (95.8 μg/L). The average of hemoglobin levels for men was slightly higher (149.2 g/100ml) than that for women (137.2 g/100ml). More women seemed to receive blood pressure treatment (21.9 %) than men did (15.7 %). There were more current or former smokers for men. Other characteristics were similar for both genders. These patterns and the average values were also similar to those from full cohort members, which means the subcohort was a well representative of the full cohort.

Table 4.7 provides the results from the full model. Here full model means all the cardiovascular risk factors were included in the model as covariates. As shown in the table, the hazard ratio estimates for log of ferritin levels on CHD and stroke were similar (1.1 and 1.2) but 95 % C.I. indicated that neither of them were statistically significant at the level of alpha = 0.05 since both included 1 in the intervals. We performed a Wald-type of test to see whether the common ferritin effect on CHD and stroke can be assumed. The test statistic was 0.3296 with corresponding p-value being equal to 0.57. Thus, there was weak evidence for a different ferritin effect on CHD and stroke. We refit the model assuming the common ferritin effect on CHD and stroke. 4.8 provides the results from the model with common ferritin effect. The results showed weak evidence of the effect of ferritin level on the risk of CHD and stroke. The hazard ratio estimates for log of ferritin level on the risk of CHD and stroke was 1.2 with standard error of 0.11. Ninety-five percent C.I. indicated that this effect was not statistically significant at the level of alpha = 0.05.

As mentioned above, we also fit the model which allowed different ferritin effect on CHD and stroke by gender. Table 4.9 provides the results from this model. The results showed that, both for men and women, there is no significant effect of ferritin level on the risk of

CHD and stroke. This was also true after we assumed common effect of ferritin level on the risk of CHD and stroke for both men and women, and refit the model.

4.6 Concluding remarks

We have proposed methods of fitting marginal hazard regression models for case-cohort studies with multiple disease outcomes. Weighted estimating equations were proposed for the estimation of the regression parameter. A Breslow-Aalen type estimator was proposed for the cumulative baseline hazard functions. Two different types of weights were considered in estimation: Estimator I and Estimator II. The former was a multivariate extension of Self and Prentice (1988)'s estimator for univariate failure time data while the latter was a multivariate extension of Kalbfleisch and Lawless (1988)'s estimator for univariate failure time data. The proposed estimators were shown to be consistent and asymptotically normally distributed. The latter was shown to be more efficient by the simulations results since the former does not use the covariate information on cases outside the subcohort. This was shown to be easily extended to a stratified case-cohort studies.

In this work, we have proved the properties based on simple random sampling without replacement for the subcohort. Other types of sampling schemes such as Bernoulli sampling of the subcohort can be considered as well. Under Bernoulli sampling scheme, the main asymptotic results can be easily shown to remain unchanged.

4.7 Proofs of the theorems

Outline of the Proofs of Theorem 4.1 - 4.4

We assume the following set of conditions hold :

(A) $(\boldsymbol{T}_i, \boldsymbol{C}_i, \boldsymbol{Z}_i), i = 1, \ldots, n$ are independent and identically distributed.

(B) $\Pr(Y(\tau) > 0) > 0$.

(C) $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(u)| < C_z < \infty$ almost surely for some constant C_z.

(D) The matrix $\boldsymbol{A}_k(\boldsymbol{\beta}_0) = \int_0^\tau \boldsymbol{v}_k(\boldsymbol{\beta}_0, t) s_k^{(0)}(\boldsymbol{\beta}_0, t)\lambda_{0k}(t)dt$ is positive definite.

Note that the conditions $(A) - (D)$ entail the following conditions (E) - (H):

(E) (Finite interval) $\int_0^\tau \lambda_{0k}(t)dt < \infty$, for all $k = 1, \ldots, K$.

(F) (Asymptotic stability) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ that satisfies the following conditions, as $n \to \infty$,

(i) For all $k = 1, \ldots, K$, there exists scalar, vector and matrix functions $s^{(0)}, \boldsymbol{s}^{(1)}$ and $\boldsymbol{s}^{(2)}$ defined on $\mathcal{B} \times [0, \tau]$ such that for $d = 0, 1, 2$,

$$\sup_{\substack{t \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} ||\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{s}_k^{(d)}(\boldsymbol{\beta}, t)|| \xrightarrow{p} 0;$$

(ii) there exists a matrix $\boldsymbol{Q}(\boldsymbol{\beta})$ such that $n^{-1}\sum_{i=1}^n \mathrm{Var}(\sum_{k=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)) \longrightarrow \boldsymbol{Q}(\boldsymbol{\beta}_0)$.

(G) (Asymptotic regularity) For all $\boldsymbol{\beta} \in \mathcal{B}$, $t \in [0, \tau]$ and $k = 1, \ldots, K$: $\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}, t) = \frac{\partial}{\partial \boldsymbol{\beta}} s_k^{(0)}(\boldsymbol{\beta}, t), \boldsymbol{s}_k^{(2)}(\boldsymbol{\beta}, t) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} s^{(0)}(\boldsymbol{\beta}, t)$, where $\boldsymbol{s}^{(d)}(\cdot, t)(d = 0, 1, 2)$ are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$, $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$.

(H) (Lindeberg condition) There exists a $\delta > 0$ $s.t.$ as $n \to \infty$

$$n^{-1/2} \sup_{i,k,t} \|\boldsymbol{Z}_{ik}(t)\| Y_{ik}(t) I\left\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t) > -\delta \|\boldsymbol{Z}_{ik}(t)\|\right\} \xrightarrow{p} 0.$$

The following additional conditions are also needed to ensure the desired asymptotic convergence of case-cohort samples:

(I) (Nontrivial subcohort) As $n \to \infty$, $\tilde{\alpha} = \frac{\tilde{n}}{n}$ converges to a constant $\alpha \in (0, 1)$.

(J) (Nontrivial cases) $\frac{n_s}{n}$ converges to a constant $p_s \in [0, 1]$ for $s = 0, 1$ as $n \to \infty$ where $p_1 + p_0 = 1$.

(K) (Asymptotic normality of samples) For all $k = 1, \ldots, K$, as $n \to \infty$,

$$n^{-1} \sup_{i,t} \exp\left\{2\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)\right\} \xrightarrow{p} 0,\ n^{-1} \sup_{i,t} \|\boldsymbol{Z}_{ik}(t)\|^2 \exp\left\{2\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)\right\} \xrightarrow{p} 0$$

(L) (Asymptotic stability) As $n \to \infty$,

(i) There exists a positive definite matrix $\boldsymbol{V}(\boldsymbol{\beta}_0)$ such that

$$\text{Var}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t)d\Lambda_{0k}(t)\right\} \longrightarrow \boldsymbol{V}(\boldsymbol{\beta}_0) \text{ in probability}$$

where $\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t)$ is defined in theorem 1;

(ii) There exists a positive definite matrix $\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ such that

$$\text{Var}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\Delta_{ik})\int_0^\tau \left(\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) - \frac{Y_{ik}(t)\,\text{E}(1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0, t)}{\text{E}(1-\Delta_{1k})Y_{1k}(t)}\right)d\Lambda_{0k}(t)\right\}$$

$$\longrightarrow \boldsymbol{V}_{II}(\boldsymbol{\beta}_0) \text{ in probability.}$$

The following lemma together with lemmas 1 - 3 in Chapter 3 will be frequently used in proving the theorems.

Lemma 5 *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ be a random vector containing $\tilde{n}$ ones and $n - \tilde{n}$ zeros, with each permutation equally likely. Let $B_i(t)$, $i = 1, \ldots, n$, be i.i.d. real-valued random processes on $[0, \tau]$ with $\text{E}\{B_i(t)\} = \mu_B(t)$, $\text{Var}\{B_i(0)\} < \infty$ and $\text{Var}\{B_i(\tau)\} < \infty$. Let $\boldsymbol{B}(t) = (B_1(t), \ldots, B_n(t))$ and $\boldsymbol{\xi}$ be independent. Suppose that almost all paths of $B_i(t)$ have finite variation. Then,*

$$n^{-1/2}\sum_{i=1}^{n}\xi_i\{B_i(t) - \mu_B(t)\} \tag{4.12}$$

converges weakly in $\ell^\infty[0, \tau]$ to a zero-mean Gaussian process and therefore

$$n^{-1}\sum_{i=1}^{n}\xi_i\{B_i(t) - \mu_B(t)\} \tag{4.13}$$

converges in probability to 0 uniformly in t.

This lemma is an extension of the proposition in Kulich and Lin (2000).

Proof. Suppose first that the $B_i(t)$'s have nondecreasing sample paths then the finite-dimensional convergence follows from Hájek (1960)'s central limit theorem for finite population sampling while the tightness follows from Example 3.6.14 of van der Vaart and Wellner (1996). In the general case, since almost every path $b(t)$ of $B(t)$ have finite variation, $b(t)$ can be written as $b_1^*(t) - b_2^*(t)$, where $b_1^*(t)$ and $b_2^*(t)$ are nonnegative, nondecreasing in t. Hence $B_i(t) = B_{i1}^*(t) - B_{i2}^*(t)$, where $B_{i1}^*(t)$ and $B_{i2}^*(t)$ are marginally tight since they meet the condition of Example 3.6.14 of van der Vaart and Wellner (1996). This implies that they are jointly tight. The joint finite-dimensional convergence of the normalized $n^{-1/2}\sum_{i=1}^n \xi_i\{B_{i1}^*(t) - \mu_{B_{i1}^*}(t)\}$ and $n^{-1/2}\sum_{i=1}^n \xi_i\{B_{i2}^*(t) - \mu_{B_{i1}^*}(t)\}$ follows again from Hájek (1960)'s central limit theorem for finite population sampling. Therefore, $n^{-1/2}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_{B_i}(t)\}$ converges weakly in $\ell^\infty[0,\tau]$ to zero mean Gaussian processes. It then follows that $n^{-1}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_{B_i}(t)\}$ converges to 0 in probability uniformly in t. This completes the proof of lemma 5.

Note that for our case, ξ_i is the subcohort membership indicator where the sampling of the subcohort was conducted by simple random sampling without replacement. Thus, it is clear that our ξ_i's satisfy the conditions in lemma 5. Also note that for the sampling from finite population, $\mu_B(t) = n^{-1}\sum_{i=1}^n B_i(t)$ and thus $n^{-1/2}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_{B_i}(t)\} = n^{-1/2}\sum_{i=1}^n \left(\xi_i - \frac{\tilde{n}}{n}\right) B_i(t) = n^{-1/2}\tilde{\alpha}\sum_{i=1}^n \left(\frac{\xi_i}{\tilde{\alpha}} - 1\right) B_i(t)$.

Before we move onto the proofs of the theorems, we investigate the asymptotic properties of the time-varying sampling probability estimator $\hat{\alpha}_k(t) = \frac{\sum_{i=1}^n \xi_i Y_{ik}(t)}{\sum_{i=1}^n Y_{ik}(t)}$. These asymptotic properties will be frequently used in proving the theorems.

For each k, it follows from the Taylor expansion of $\widehat{\alpha}_k(t)^{-1}$ around $\widetilde{\alpha}$,

$$\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1} = -\frac{1}{\alpha_*^2(t)}\left(\widehat{\alpha}_k(t) - \widetilde{\alpha}\right) = \frac{\widetilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{1}{\sum_{i=1}^n Y_{ik}(t)} \left\{\sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(t)\right\}$$

where $\alpha_*(t)$ is on the line segment between $\widehat{\alpha}_k(t)$ and $\widetilde{\alpha}$. Then,

$$n^{1/2}\left(\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}\right) = \frac{\widetilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{n}{\sum_{i=1}^n Y_{ik}(t)} n^{-1/2} \left\{\sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(t)\right\}$$

By Glivenko-Cantelli lemma, $n^{-1}\sum_{i=1}^n Y_{ik}(t)$ converges to $\mathrm{E}\, Y_{1k}(t)$ in probability uniformly in t. In view of lemma 5, $n^{-1/2}\sum_{i=1}^n (\frac{\xi_i}{\widetilde{\alpha}} - 1) Y_{ik}(t)$ converges to a zero-mean Gaussian process since $Y_{ik}(t)$ is bounded and monotone function in t. This implies $n^{-1}\sum_{i=1}^n (\frac{\xi_i}{\widetilde{\alpha}} - 1) Y_{ik}(t)$ converges to 0 in probability uniformly in t and consequently, $\widehat{\alpha}_k(t)$ and $\widetilde{\alpha}$ converges to the same limit uniformly in t. This ensures $\alpha_*(t)$ also converges to the same limit as $\widetilde{\alpha}$. Combining these results, it follows from Slutsky's theorem that

$$\sqrt{n}\left(\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}\right) = \frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\right\}$$
$$+\left(\frac{\widetilde{\alpha}}{\alpha_*(t)^2}\cdot\frac{n}{\sum_{i=1}^{n}Y_{ik}(t)} - \frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)$$
$$= \frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\right\} + o_p(1) \tag{4.14}$$

Now we prove theorem 4.1.

Proof of Theorem 4.1 We first consider the proof for the consistency of $\widehat{\boldsymbol{\beta}}_I$. Denote n^{-1} times $\boldsymbol{U}^I(\boldsymbol{\beta})$ by $\boldsymbol{U}_n^I(\boldsymbol{\beta})$. Based on a straightforward extension of Foutz (1977), one can show $\widehat{\boldsymbol{\beta}}_I$ to be consistent for $\boldsymbol{\beta}_0$ provided: (i) $\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$, (ii)$\partial\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0^T$ is negative definite with probability going to one as $n\to\infty$, (iii) $\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ converges to $\boldsymbol{A}(\boldsymbol{\beta}_0)$ in probability uniformly for $\boldsymbol{\beta}$ in an open neighborhood about $\boldsymbol{\beta}_0$, and (iv) $\boldsymbol{U}_n^I(\boldsymbol{\beta})\to 0$ in probability.

One can write

$$\begin{aligned}\frac{\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T} &= -n^{-1}\sum_{k=1}^{K}\int_0^\tau \widehat{\boldsymbol{V}}_k(\boldsymbol{\beta},t)d\sum_{i=1}^{n}N_{ik}(t) \quad \text{where}\\ \widehat{\boldsymbol{V}}_k(\boldsymbol{\beta},t) &= \frac{\widehat{\boldsymbol{S}}_k^{(2)}(\boldsymbol{\beta},t)\widehat{S}_k^{(0)}(\boldsymbol{\beta},t) - \widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)^{\otimes 2}}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)^2}\end{aligned} \tag{4.15}$$

Then, (i) is clearly satisfied on the basis of (4.15) and by the continuity of each component. Now, following Andersen and Gill (1982),

$$\left\|\left(-\frac{\partial\boldsymbol{U}_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\right) - \boldsymbol{A}(\boldsymbol{\beta})\right\| \le \left\|\sum_{k=1}^{K}\int_0^\tau\{\widehat{\boldsymbol{V}}_k(\boldsymbol{\beta},t) - \boldsymbol{v}_k(\boldsymbol{\beta},t)\}n^{-1}d\sum_{i=1}^{n}N_{ik}(t)\right\|$$
$$+\left\|\sum_{k=1}^{K}\int_0^\tau \boldsymbol{v}_k(\boldsymbol{\beta},t)n^{-1}d\sum_{i=1}^{n}M_{ik}(t)\right\| + \left\|\int_0^\tau \boldsymbol{v}_k(\boldsymbol{\beta},t)\{S_k^{(0)}(\boldsymbol{\beta},t) - s_k^{(0)}(\boldsymbol{\beta},t)\}\lambda_{0k}(t)dt\right\| \tag{4.16}$$

Each of the terms on the right side of the above inequality will be shown to converge to zero, uniformly in $\boldsymbol{\beta}\in\mathcal{B}$ in the following.

To show the first term on the right side of (4.16), we will first show that

$$\sup_{\substack{t \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} \left\| \widehat{\boldsymbol{V}}_k(\boldsymbol{\beta}, t) - \boldsymbol{v}_k(\boldsymbol{\beta}, t) \right\| \xrightarrow{p} 0 \text{ as } n \to \infty \text{ for } k = 1, \ldots, K.$$

It suffices to show that $\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t)\| \xrightarrow{p} 0$ as $n \to \infty$ for $d = 0, 1, 2$. One can write

$$\begin{aligned} \widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^{n} \left(\frac{\xi_i}{\widetilde{\alpha}} - 1 \right) \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t) \\ &- n^{-1} \sum_{i=1}^{n} (\widetilde{\alpha}^{-1} - \widehat{\alpha}_k(t)^{-1}) \xi_i \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t) \end{aligned}$$

Then

$$\begin{aligned} \left\| \widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t) \right\| \leq & \left\| n^{-1} \sum_{i=1}^{n} \left(\frac{\xi_i}{\widetilde{\alpha}} - 1 \right) \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t) \right\| \\ & + \left| (\widetilde{\alpha}^{-1} - \widehat{\alpha}_k(t)^{-1}) \right| n^{-1} \sum_{i=1}^{n} \xi_i \left| \boldsymbol{Z}_{ik}(t)^{\otimes d} \right| e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t) \end{aligned} \tag{4.17}$$

For each $j (j = 1, \ldots, p)$, by the condition on $Z_{ikj}(t)$, the total variation of $Z_{ikj}(t)^{\otimes d} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t)$ is also finite on $[0, \tau]$. Thus, by lemma 5, the first term on the right-hand side of (4.17) converges to 0 in probability uniformly in t. The second term on the right-hand side of (4.17) also converges to 0 in probability uniformly in t since $\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}$ was shown to converge to 0 in probability uniformly in t and $n^{-1} \sum_{i=1}^{n} \xi_i |\boldsymbol{Z}_{ik}(t)^{\otimes d}| e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t)$ converges to a finite quantity $\tilde{\alpha}\, \mathrm{E}(|\boldsymbol{Z}_{ik}^{\otimes d}| e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)} Y_{ik}(t))$ in probability uniformly in t and $\boldsymbol{\beta}$ by lemma 5. Combining these results, $\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t)$ and $\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t)$ were shown to converge to the same limit uniformly and consequently, we have

$$\sup_{\substack{t \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} \left\| \widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{s}_k^{(d)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{p} 0 \text{ as } n \to \infty \text{ for } d = 0, 1 \tag{4.18}$$

Since $s_k^{(0)}(\boldsymbol{\beta}, t)$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$ by condition (G), it follows from the above convergence results that for $k = 1, \ldots, K$, $\widehat{\boldsymbol{V}}_k(\boldsymbol{\beta}, t)$ converges to $\boldsymbol{v}_k(\boldsymbol{\beta}, t)$ in probability uniformly in t and $\boldsymbol{\beta}$.

The Lenglart inequality (Andersen and Gill, 1982, p1115) implies that, for any $\delta,\ \rho\ >\ 0$, there exists n_0 such that for $n \geq n_0$,

$$P[n^{-1}\bar{N}_k(\tau) > c] \leq \frac{\delta}{c} + P[\int_0^\tau S_k^{(0)}(\boldsymbol{\beta}_0; t)\lambda_{0k}(t)dt > \delta]$$

By Condition (F), for $\delta > \int_0^\tau s_k^{(0)}(\boldsymbol{\beta}_0, t)\lambda_{0k}(t)dt$, $P[\int_0^\tau S_k^{(0)}(\boldsymbol{\beta}_0; t)\lambda_{0k}(t)dt > \delta] \rightarrow 0$ as $n \rightarrow \infty$. Then, $\lim_{c\uparrow\infty} \lim_{n\rightarrow\infty} P[n^{-1}\bar{N}_k(\tau) > c] = 0$. Thus, it follows that the first term on the right side of (4.16) converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$, as $n \rightarrow \infty$.

For the second term on the right side of (4.16), $n^{-1}\sum_{i=1}^n \int_0^\tau \boldsymbol{v}_k(\boldsymbol{\beta}, t)dM_{ik}(t)$ is a local square integrable martingale. Hence, the Lenglart inequality (Andersen and Gill, 1982, p1115) implies that, for any $\delta,\ \rho\ >\ 0$, there exists n_0 such that for $n \geq n_0$,

$$\mathrm{P}\left[\left\|n^{-1}\int_0^\tau \{\boldsymbol{v}_k(\boldsymbol{\beta}, t)\}_{ll'}\, d\bar{M}_k(t)\right\| > \rho\right]\ \leq\ \frac{\delta}{\rho^2} + P\left[n^{-1}\int_0^\tau \{\boldsymbol{v}_k(\boldsymbol{\beta}, t)\}_{ll'}^2\, S_k^{(0)}(\boldsymbol{\beta}, t)\lambda_{0k}(t)dt > \delta\right]$$

where the subscript ll' denotes the (l, l') element of the indicated matrix. The boundedness conditions (E), (F) and (G) ensure that the second term on the right side of the above inequality converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \rightarrow \infty$ for any δ. Since δ can be arbitrarily small, it follows that the left side of the above inequality also converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \rightarrow \infty$. Therefore, the second term on the right side of (4.16) also converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$, as $n \rightarrow \infty$.

Again, the conditions (D), (E) and (F) ensure the boundedness of $\sup_{t,\boldsymbol{\beta}}\{\boldsymbol{v}_k(\boldsymbol{\beta}, t)\}_{ll'}$ and $\Lambda_{0k}(\tau)$ for $k = 1, \ldots, K$ and $l, l' = 1, \ldots, p$. Thus, together with the uniform convergence of $\widehat{S}_k^{(0)}(\boldsymbol{\beta}, t)$ to $s_k^{(0)}(\boldsymbol{\beta}, t)$ in probability, the last term on the right side of (4.16) converges to zero in probability, uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ as $n \rightarrow \infty$. Hence,

$$-\frac{\partial \boldsymbol{U}_n^I(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\beta}) \text{ as } n \rightarrow \infty \text{ uniformly in } \boldsymbol{\beta} \in \mathcal{B}$$

and, thus, (ii) and (iii) are satisfied.

For (iv), we will show that $n^{-1/2}\boldsymbol{U}^I(\boldsymbol{\beta}_0)$ is asymptotically equivalent to
$n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)+n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K(1-\frac{\xi_i}{\tilde{\alpha}})\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)$. Specifically, one can decompose $n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)$ into two parts:

$$\begin{aligned} n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0) &= n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K\int_0^\tau\left\{\boldsymbol{Z}_{ik}(t)-\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}dN_{ik}(t) \\ &+ n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K\int_0^\tau\left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}-\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}dN_{ik}(t) \end{aligned} \tag{4.19}$$

The first term on the right-hand side of (4.19) is the pseudo partial likelihood score function for the full cohort data. This was shown to be asymptotically equivalent to
$n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)$ (Spiekerman and Lin, 1998). The second term on the right-hand side of (4.19) can be further decomposed as

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K\int_0^\tau\left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}-\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}dN_{ik}(t) \\ =\ & \sum_{k=1}^K\int_0^\tau\left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}-\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(t)\right\} \\ +\ & n^{-1/2}\sum_{k=1}^K\int_0^\tau\left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}-\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}\sum_{i=1}^n Y_{ik}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}d\Lambda_{0k}(t) \end{aligned} \tag{4.20}$$

Note that, for fixed t, $n^{-1/2}\sum_{i=1}^n M_{ik}(t)$ is a sum of i.i.d. zero-mean random variables. Based on conditions (C) and (E), $M_{ik}(t)$ is of bounded variation and therefore can be written as a difference of two monotone functions in t. It then follows from the example of 2.11.16 of van der Vaart and Wellner (1996, p215) that $n^{-1/2}\sum_{i=1}^n M_{ik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{W}_{Mk}(t)$. It can be shown that $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^4 \le C\{\Lambda_{0k}(t)-\Lambda_{0k}(s)\}^2$ for some constant $C>0$. Specifically, $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^4 = 3(\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^2)^2$ since $\mathcal{W}_{Mk}(t)$ is a zero-mean normal random variable for a fixed t. Then $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^2 = \mathrm{E}\,\mathcal{W}_{Mk}(t)^2+\mathrm{E}\,\mathcal{W}_{Mk}(s)^2-2\,\mathrm{E}\,\mathcal{W}_{Mk}(t)\mathcal{W}_{Mk}(s) = \mathrm{E}\,\mathcal{W}_{Mk}(t)^2-\mathrm{E}\,\mathcal{W}_{Mk}(s)^2$ for $s\le t$. Since $\mathrm{E}\,\mathcal{W}_{Mk}(t)^2 = \mathrm{E}\,M_{ik}(t)^2 = \mathrm{E}\left[\int_0^t Y_{ik}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)}\lambda_{0k}(u)du\right]$, $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^2 = \mathrm{E}\left[\int_s^t Y_{ik}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)}\lambda_{0k}(u)du\right]$

$\leq e^{C_z}\,\mathrm{E}\left[\int_s^t \lambda_{0k}(u)du\right] = \tilde{C}_z(\Lambda_{0k}(t) - \Lambda_{0k}(s))$ by the boundedness condition (C). Since $\Lambda_{0k}(\cdot)$ is differentiable and $\lambda_0(\cdot)$ is bounded on $[0,\tau]$, $\exists$ a constant M, such that $\Lambda_{0k}(t) - \Lambda_{0k}(s) \leq M(t-s)$ for $s \leq t$. Therefore, $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^2 \leq C_z^*(t-s)$ and $\mathrm{E}\{\mathcal{W}_{Mk}(t)-\mathcal{W}_{Mk}(s)\}^4 \leq 3(\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2)^2 \leq \tilde{C}_z^*(t-s)^2$ for some constant C_z^*. Then, by the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988, p53), $\mathcal{W}_{Mk}(t)$ has continuous sample paths. In addition, since $\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)$ and $S_k^{(0)}(\boldsymbol{\beta},t)$ are of bounded variations and $S_k^{(0)}(\boldsymbol{\beta},t)$ is bounded away from 0, based on conditions (C), (F) and (G), $\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)}{S_k^{(0)}(\boldsymbol{\beta},t)}$ is of bounded variation and can be written as a sum of two monotone functions in t, respectively. Specifically, $\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)}{S_k^{(0)}(\boldsymbol{\beta},t)} = \boldsymbol{Z}_{k1}^*(t) - \boldsymbol{Z}_{k2}^*(t)$ where both $\boldsymbol{Z}_{k1}^*(t)$ and $\boldsymbol{Z}_{k2}^*(t)$ are nonnegative, monotone in t and bounded. Since $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)}$ is also of bounded variation based on (4.18) and conditions (C) and (G), by the same argument, we can write $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)} = \boldsymbol{Z}_{k1}^{**}(t) - \boldsymbol{Z}_{k2}^{**}(t)$ where both $\boldsymbol{Z}_{k1}^{**}(t)$ and $\boldsymbol{Z}_{k2}^{**}(t)$ are nonnegative, monotone in t and bounded. Therefore, $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)}$ is also a sum of two monotone functions. Based on condition (F) and result in (4.18), it can be shown that both $\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)}{S_k^{(0)}(\boldsymbol{\beta},t)}$ and $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)}$ converge to the same limit uniformly. Hence, it follows from lemma 2 that

$$\begin{aligned}
& \sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)}{S_k^{(0)}(\boldsymbol{\beta},t)} - \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)}\right\} n^{-1/2}\sum_{i=1}^{n} dM_{ik}(t) \\
= & \sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta},t)}{S_k^{(0)}(\boldsymbol{\beta},t)} - \frac{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta},t)}{s_k^{(0)}(\boldsymbol{\beta},t)}\right\} n^{-1/2}\sum_{i=1}^{n} dM_{ik}(t) \\
- & \sum_{k=1}^{K}\int_0^\tau \left\{\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta},t)} - \frac{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta},t)}{s_k^{(0)}(\boldsymbol{\beta},t)}\right\} n^{-1/2}\sum_{i=1}^{n} dM_{ik}(t) \xrightarrow{p} 0 \text{ as } n \to \infty
\end{aligned}$$

Thus, the first term on the right-hand side of (4.20) converges to 0 in probability uniformly in t.

To investigate the asymptotic properties of the second term on the right-hand side of (4.20), we first study the asymptotic expansion of $n^{1/2}\left\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t) - \widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\right\}$ $(d = 0,1)$.

One can write

$$
\begin{aligned}
n^{1/2}\left\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)-\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\right\} &= n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k(t)^{-1}\right)\xi_i\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)\\
&+ n^{-1}\sum_{i=1}^{n}\left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}n^{-1/2}\sum_{j=1}^{n}\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right)Y_{jk}(t)\right\}\xi_i\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)+o_p(1)\ \text{(by (5.6))}\\
&= n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\left\{n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\boldsymbol{Z}_{jk}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{jk}(t)}Y_{jk}(t)\right\}+o_p(1) \qquad (4.21)
\end{aligned}
$$

It follows from lemma 5 that $n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\boldsymbol{Z}_{jk}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{jk}(t)}Y_{jk}(t)$ converges to $\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta},t)$ in probability uniformly in t. Thus, from (4.21)

$$
\begin{aligned}
& n^{1/2}\left\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)-\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\right\}\\
= \; & n^{-1/2}\sum_{i=1}^{n}\left\{\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}Y_{ik}(t)+\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta},t)\right\}+o_p(1)\\
= \; & n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\left\{\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}-\frac{\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta},t)}{\mathrm{E}\,Y_{1k}(t)}\right\}+o_p(1) \qquad (4.22)
\end{aligned}
$$

Now, the second term on the right-hand side of (4.20) can be further decomposed as

$$
\begin{aligned}
& n^{-1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}-\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\}\sum_{i=1}^{n}Y_{ik}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}d\Lambda_{0k}(t)\\
= \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)-\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)\right\}d\Lambda_{0k}(t)\\
- \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{S_k^{(0)}(\boldsymbol{\beta}_0,t)-\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)\right\}\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}d\Lambda_{0k}(t)\\
= \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)-\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)\right\}d\Lambda_{0k}(t)\\
- \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{S_k^{(0)}(\boldsymbol{\beta}_0,t)-\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)\right\}\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\\
- \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{S_k^{(0)}(\boldsymbol{\beta}_0,t)-\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)\right\}\left\{\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}-\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)\right\}d\Lambda_{0k}(t)\\
= \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)-\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)\right\}d\Lambda_{0k}(t)\\
- \;& n^{1/2}\sum_{k=1}^{K}\int_0^{\tau}\left\{S_k^{(0)}(\boldsymbol{\beta}_0,t)-\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)\right\}\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)+o_p(1) \qquad (4.23)
\end{aligned}
$$

The last equality holds since $\frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}$ converges to $\boldsymbol{e}_k(\boldsymbol{\beta},t)$ in probability uniformly in t, $n^{1/2}\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)-\widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\}(d=0,1)$ converges weakly to a zero-mean Gaussian process and $\Lambda_{0k}(t)$ is bounded on $t\in[0,\tau]$. Then, based on (4.22),

$$
\begin{aligned}
(4.23) = \;& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\left\{\boldsymbol{Z}_{ik}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}-\frac{\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\mathrm{E}\,Y_{1k}(t)}\right\}d\Lambda_{0k}(t)\\
- \;& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}-\frac{s_k^{(0)}(\boldsymbol{\beta}_0,t)}{\mathrm{E}\,Y_{1k}(t)}\right\}\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)+o_p(1)\\
= \;& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\left(\boldsymbol{Z}_{ik}-\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)\right)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}d\Lambda_{0k}(t)\\
- \;& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\left(\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}_0,t)-s_k^{(0)}(\boldsymbol{\beta}_0,t)\boldsymbol{e}_k(\boldsymbol{\beta}_0,t)\right)d\Lambda_{0k}(t)+o_p(1)\\
= \;& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^{\tau}\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)+o_p(1)
\end{aligned}
$$

where $\boldsymbol{R}_{ik}(\boldsymbol{\beta}, t) = Y_{ik}(t)\widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}$, where $\widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) = \boldsymbol{Z}_{ik}(t) - \boldsymbol{e}_k(\boldsymbol{\beta}, t)$. The last equality holds since $\boldsymbol{s}_k^{(1)}(\boldsymbol{\beta}_0, t) - s_k^{(0)}(\boldsymbol{\beta}_0, t)\boldsymbol{e}_k(\boldsymbol{\beta}, t) = 0$. Therefore, the second term on the right-hand side of (4.20) is asymptotically equivalent to
$n^{-1/2}\sum_{i=1}^n \sum_{k=1}^K \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t)$. Combining the above results, we have shown that $n^{1/2}\boldsymbol{U}^I(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^n \sum_{k=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0) + n^{-1/2}\sum_{i=1}^n \sum_{k=1}^K \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) \tag{4.24}$$

Under the regularity conditions, the first term on the right-hand side of (4.24) is asymptotically zero-mean normal with covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0) = \mathrm{E}\left(\sum_{k=1}^K \mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\right)^{\otimes 2}$ by Spiekerman and Lin (1998).

The second term on the right-hand side of (4.24) can be shown to be asymptotically zero-mean normal with covariance matrix $\boldsymbol{V}(\boldsymbol{\beta}_0)$ by Hájek (1960)'s central limit theorem for finite population sampling. Specifically, let $\boldsymbol{a} = (a_1, \ldots, a_p)^T$ be a $p \times 1$ real valued vector. Then, one can write

$$\begin{aligned}
&\left|\boldsymbol{a}^T \left(\sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t)\right) - n^{-1}\sum_{i=1}^n \boldsymbol{a}^T \left(\sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t)\right)\right| \\
= &\left|\sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \widetilde{\boldsymbol{Z}}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) - n^{-1}\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \widetilde{\boldsymbol{Z}}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| \\
\leq &\left|\sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \widetilde{\boldsymbol{Z}}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| + \left|n^{-1}\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \widetilde{\boldsymbol{Z}}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| \\
\leq &\left|\sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| + \left|\sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| \\
+ &\left|n^{-1}\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right| \\
+ &\left|n^{-1}\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t)\right|
\end{aligned}$$

Thus,

$$\max_i \left| \boldsymbol{a}^T \left(\sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) \right) - n^{-1} \sum_{i=1}^n \boldsymbol{a}^T \left(\sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) \right) \right|^2$$

$$\leq \max_i \left(\left| \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| + \left| \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \right.$$

$$+ \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right|$$

$$\left. + \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \right)^2$$

Note that

$$\max_i \left| \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \leq \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$$

$$\max_i \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \leq \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$$

$$\max_i \left| \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \leq \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$$

$$\max_i \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) Y_{ik}(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} d\Lambda_{0k}(t) \right| \leq \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$$

To use Hájek (1960)'s theorem, the following conditions need to be verified. As $n \to \infty$,

(a) $\widetilde{\alpha}$ converges to a constant $\alpha \in (0, 1)$;

(b) $n^{-1} \max_i \left| \boldsymbol{a}^T \sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) - n^{-1} \sum_{i=1}^n \boldsymbol{a}^T \sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) \right|^2$ converges to zero in probability, and

(c) $(n-1)^{-1} \sum_{i=1}^n \left(\boldsymbol{a}^T \sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) - n^{-1} \sum_{i=1}^n \boldsymbol{a}^T \sum_{k=1}^K \int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) \right)^2$ converges to $\sigma^2 \neq 0$.

To verify (b), it suffices to show that $n^{-1} \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{Z}_{ik}(t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$ and $n^{-1} \sup_{t,i} \sum_{k=1}^K \left| \boldsymbol{a}^T \boldsymbol{e}_k(\boldsymbol{\beta}, t) \right| e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)} \Lambda_{0k}(\tau)$ converge to zero in probability as $n \to \infty$. This holds by our conditions (C), (E), (G) and (K). (a) and (c) are satisfied on the basis of condi-

tions (I), (J) and (L)(i). This implies $n^{-1/2}\sum_{i=1}^{n}\boldsymbol{a}^T\left\{\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\right\}$ converges to a mean-zero normal random variable. Therefore, by Cramer-Wold device, $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)$ converges to a p-variate mean-zero normal random variable with variance $\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)=\frac{1-\alpha}{\alpha}\,\mathrm{E}\left[\sum_{k=1}^{K}\int_0^\tau \boldsymbol{R}_{1k}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\right]^{\otimes 2}$.
Note that $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)$ and $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\frac{\xi_i}{\widetilde{\alpha}})\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)$ are independent since

$$\begin{aligned}
& \mathrm{Cov}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0), n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\right) \\
= \;& \mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\right\} \\
= \;& \mathrm{E}\left\{\mathrm{E}\left(n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\middle|\mathcal{F}(\tau)\right)\right\} \\
= \;& \mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\mathrm{E}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\middle|\mathcal{F}(\tau)\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t)\right\}=0
\end{aligned}$$

Therefore, $n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with finite variance $\boldsymbol{Q}(\boldsymbol{\beta}_0)+\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)$. Hence $\boldsymbol{U}_n(\boldsymbol{\beta})$ converges to zero in probability. Thus, (iv) is satisfied.
By (i),(ii),(iii) and (iv), it follows that there is a unique sequence $\widehat{\boldsymbol{\beta}}_I$ $s.t.$ $\boldsymbol{U}_I(\widehat{\boldsymbol{\beta}})=0$ with probability converging to one as $n\to 0$ and with $\widehat{\boldsymbol{\beta}}_I$ converging in probability to $\boldsymbol{\beta}_0$ by extension of (Foutz, 1977, Thm.2).

The asymptotic normality of $\widehat{\boldsymbol{\beta}}_I$ follows from the consistency of $\widehat{\boldsymbol{\beta}}_I$ and a Taylor series expansion of $\boldsymbol{U}^I(\boldsymbol{\beta})$.

Proof of Theorem 4.2 One can make decomposition

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}^I_{0k}(\widehat{\boldsymbol{\beta}}_I,t)-\Lambda_{0k}(t)\} \\
= \;& n^{1/2}\left\{\widehat{\Lambda}^I_{0k}(\widehat{\boldsymbol{\beta}}_I,t)-\int_0^t \frac{d\left(\sum_{i=1}^n N_{ik}(u)\right)}{n\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}\right\} + n^{1/2}\left\{\int_0^t \frac{d\left(\sum_{i=1}^n N_{ik}(u)\right)}{n\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}-\Lambda_{0k}(t)\right\} \\
= \;& n^{1/2}\int_0^t \left(\frac{1}{n\widehat{S}^{(0)}_k(\widehat{\boldsymbol{\beta}}_I,u)}-\frac{1}{n\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}\right) d\sum_{i=1}^n M_{ik}(u) \\
+ \;& n^{1/2}\int_0^t \left(\frac{1}{\widehat{S}^{(0)}_k(\widehat{\boldsymbol{\beta}}_I,u)}-\frac{1}{\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}\right) S^{(0)}_k(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(u) \\
+ \;& \int_0^t \frac{1}{\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)} dn^{-1/2}\sum_{i=1}^n M_{ik}(u) \\
+ \;& n^{1/2}\int_0^t \left(\frac{S^{(0)}_k(\boldsymbol{\beta}_0,u)-\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}{\widehat{S}^{(0)}_k(\boldsymbol{\beta}_0,u)}\right) d\Lambda_{0k}(u) \qquad (4.25)
\end{aligned}
$$

By the Taylor expansion of $\widehat{S}^{(0)}_k(\widehat{\boldsymbol{\beta}}_I,u)^{-1}$ around $\boldsymbol{\beta}_0$, the first term of (4.25), can be shown to be equivalent to

$$\int_0^t \left(-\frac{\widehat{\boldsymbol{S}}^{(1)}_k(\boldsymbol{\beta}^*,u)^T}{\widehat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)^2}\right)(\widehat{\boldsymbol{\beta}}_I-\boldsymbol{\beta}_0)d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\} \qquad (4.26)$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\hat{\boldsymbol{\beta}}_I$ and $\boldsymbol{\beta}_0$. Again, $\hat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)$ and $\hat{\boldsymbol{S}}^{(1)}_k(\boldsymbol{\beta}^*,u)$ are of bounded variations and $\hat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)$ is bounded away from 0, therefore, $\frac{\hat{\boldsymbol{S}}^{(1)}_k(\boldsymbol{\beta}^*,u)}{\hat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)^2}$ can be expressed as a sum of two monotone functions in t. Then, together with the consistency of $\hat{\boldsymbol{\beta}}_I$, the uniform convergence of $\hat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)$ and $\hat{\boldsymbol{S}}^{(1)}_k(\boldsymbol{\beta}^*,u)$, and the weak convergence of $n^{-1/2}\sum_{i=1}^n M_{ik}(t)$ with continuous sample paths, (4.26) converges to 0 uniformly in t in probability by applying lemma 2.

Again, by the Taylor expansion of $\widehat{S}^{(0)}_k(\widehat{\boldsymbol{\beta}}_I,u)^{-1}$ around $\boldsymbol{\beta}_0$, the second term on the right-hand side of (4.25) is equivalent to

$$-n^{1/2}\int_0^t \frac{\widehat{\boldsymbol{S}}^{(1)}_k(\boldsymbol{\beta}^*,u)^T}{\widehat{S}^{(0)}_k(\boldsymbol{\beta}^*,u)^2}(\hat{\boldsymbol{\beta}}_I-\boldsymbol{\beta}_0)S^{(0)}_k(\boldsymbol{\beta}_0,u)d\Lambda_{0k}(u) \qquad (4.27)$$

By the consistency of $\widehat{\boldsymbol{\beta}}_I$, the uniform consistency of $\widehat{S}_k^{(0)}(\boldsymbol{\beta}^*, u)$, $S_k^{(0)}(\boldsymbol{\beta}_0, u)$, and $\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}^*, u)$, and the boundedness of $\Lambda_{0k}(u)$ on $[0, \tau]$, we have

$$-n^{1/2}\int_0^t \frac{\widehat{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}^*, u)^T}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}^*, u)^2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)S_k^{(0)}(\boldsymbol{\beta}_0, u)d\Lambda_{0k}(u) = n^{1/2}\boldsymbol{r}_k(\boldsymbol{\beta}_0, t)^T\left(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\right) + o_p(1),$$

where $\boldsymbol{r}_k(\boldsymbol{\beta}, t) = -\int_0^t \boldsymbol{e}_k(\boldsymbol{\beta}, u)d\Lambda_{0k}(u)$. Since $\hat{S}_k^{(0)}(\boldsymbol{\beta}_0, u)^{-1}$ can be written a sum of two monotone functions in t and converges uniformly to $s_k^{(0)}(\boldsymbol{\beta}_0, u)^{-1}$, where $s_k^{(0)}(\boldsymbol{\beta}_0, u)$ is bounded away from 0, and $n^{-1/2}\sum_{i=1}^n M_{ik}(u)$ converges to a zero-mean Gaussian process with continuous sample path, it follows from lemma 2 that the third term on the right-hand side of (4.25) is asymptotically equivalent to

$$\int_0^t \frac{1}{s_k^{(0)}(\boldsymbol{\beta}_0, u)}d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\}$$

For the last term on the right-hand side of (4.25), it follows from (4.21) and the uniform convergence of $\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0, t)^{-1}$ to $s_k^{(0)}(\boldsymbol{\beta}_0, t)^{-1}$, where $s_k^{(0)}(\boldsymbol{\beta}_0, t)$ is bounded away from 0 that

$$\begin{aligned} & n^{1/2}\int_0^t\left(\frac{S_k^{(0)}(\boldsymbol{\beta}_0, u) - \widehat{S}_k^{(0)}(\boldsymbol{\beta}_0, u)}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0, u)}\right)d\Lambda_{0k}(u) \\ = & \int_0^t \frac{1}{\widehat{S}_k^{(0)}(\boldsymbol{\beta}_0, u)}n^{-1/2}\sum_{i=1}^n\left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(u)\left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}_0, u)}{\operatorname{E}Y_{1k}(u)}\right\}d\Lambda_{0k}(u) + o_p(1) \\ = & n^{-1/2}\sum_{i=1}^n\left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^t Y_{ik}(u)\left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}_0, u)}{\operatorname{E}Y_{1k}(u)}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} + o_p(1) \end{aligned}$$

Now, by combining the above results and using the asymptotic expansion of $n^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ where

$$\begin{aligned} n^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0) = & \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\Big\{n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0) \\ & + n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K\left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau \boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t)d\Lambda_{0k}(t)\Big\} + o_p(1) \end{aligned}$$

we have

$$
\begin{aligned}
& n^{1/2}(\widehat{\Lambda}_{0k}^{I}(\widehat{\boldsymbol{\beta}}_I, t) - \Lambda_{0k}(t)) \\
= \;& \boldsymbol{r}_k(\boldsymbol{\beta}_0, t)^T \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1} \left\{ n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0) + n^{-1/2} \sum_{i=1}^{n} \sum_{m=1}^{K} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^\tau \boldsymbol{R}_{im}(\boldsymbol{\beta}_0, u) d\Lambda_{0m}(u) \right\} \\
+ \;& \int_0^t \frac{1}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} d\left\{ n^{-1/2} \sum_{i=1}^{n} M_{ik}(u) \right\} + n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \\
\times \;& \int_0^t Y_{ik}(u) \left(e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}_0, u)}{\mathrm{E}\, Y_{1k}(u)} \right) \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} \\
= \;& n^{-1/2} \sum_{i=1}^{n} \left[\left\{ \boldsymbol{r}_k(\boldsymbol{\beta}_0, t)^T \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}}, im}(\boldsymbol{\beta}_0) + \int_0^t \frac{1}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} dM_{ik}(u) \right\} \right. \\
+ \;& \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \left\{ \boldsymbol{r}_k(\boldsymbol{\beta}_0, t)^T \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1} \sum_{m=1}^{K} \int_0^\tau \boldsymbol{R}_{im}(\boldsymbol{\beta}_0, u) d\Lambda_{0m}(u) \right. \\
+ \;& \left. \left. \int_0^t Y_{ik}(u) \left(e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}_0, u)}{\mathrm{E}\, Y_{1k}(u)} \right) \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} \right] + o_p(1) \qquad (4.28) \\
= \;& n^{-1/2} \sum_{i=1}^{n} \nu_{ik}(\boldsymbol{\beta}_0, t) + n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \psi_{ik}(\boldsymbol{\beta}_0, t) + o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_{ik}(\boldsymbol{\beta}, t) &= \boldsymbol{r}_k(\boldsymbol{\beta}, t)^T \boldsymbol{A}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}}, im}(\boldsymbol{\beta}) + \int_0^t \frac{1}{s_k^{(0)}(\boldsymbol{\beta}, u)} dM_{ik}(u) \text{ and} \\
\psi_{ik}(\boldsymbol{\beta}, t) &= \boldsymbol{r}_k(\boldsymbol{\beta}, t)^T \boldsymbol{A}(\boldsymbol{\beta})^{-1} \sum_{m=1}^{K} \int_0^\tau \boldsymbol{R}_{im}(\boldsymbol{\beta}, u) d\Lambda_{0m}(u) \\
&+ \int_0^t Y_{ik}(u) \left(e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(u)} - \frac{s_k^{(0)}(\boldsymbol{\beta}, u)}{\mathrm{E}\, Y_{1k}(u)} \right) \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}, u)}.
\end{aligned}
$$

Now, let $\boldsymbol{W}^{(1)}(t) = (W_1^{(1)}(t), \ldots, W_K^{(1)}(t))^T$ where $W_k^{(1)}(t) = n^{-1/2} \sum_{i=1}^{n} \nu_{ik}(\boldsymbol{\beta}_0, t)$ and $\boldsymbol{W}^{(2)}(t) = (W_1^{(2)}(t), \ldots, W_K^{(2)}(t))^T$ where $W_k^{(2)}(t) = n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \psi_{ik}(\boldsymbol{\beta}_0, t)$ for $k = 1, \ldots, K$. Then, $\boldsymbol{W}^{(1)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}^{(1)}(t) = (\mathcal{W}_1^{(1)}(t), \ldots, \mathcal{W}_K^{(1)}(t))^T$ in $D[0, \tau]^K$ where the covariance function between $\mathcal{W}_j^{(1)}(t_1)$ and $\mathcal{W}_k^{(1)}(t_2)$ is $\mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0, t_1)\nu_{1k}(\boldsymbol{\beta}_0, t_2)\}$ by Spiekerman and Lin (1998, Thm.2,). $\boldsymbol{W}^{(2)}(t)$ also can be shown to converge weakly to a zero-mean Gaussian process $\mathcal{W}^{(2)}(t) = (\mathcal{W}_1^{(2)}(t), \ldots, \mathcal{W}_K^{(2)}(t))^T$.

Specifically, $\psi_{ik}(\boldsymbol{\beta}_0, t)$ is of bounded variation since $\boldsymbol{r}_k(\boldsymbol{\beta}_0, t)$, $Y_{ik}(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t)}$ and $\operatorname{E} Y_{1k}(t)$ are of bounded variations, $\operatorname{E} Y_{ik}(t)$ and $\boldsymbol{s}_k^{(0)}(\boldsymbol{\beta}, t)$ are bounded away from zero, and $\boldsymbol{A}(\boldsymbol{\beta}_0)$ is positive definite based on conditions (B), (C), (D), (F), and (G). Thus, for any finite number of time points $(t_1, \ldots, t_L)$, the finite dimensional distribution of $\boldsymbol{W}^{(2)}(t)$ is asymptotically the same as those of $\mathcal{W}^{(2)}(t)$ by lemma 5 and Cramer-Wold device. Now, if we show the tightness of $\boldsymbol{W}^{(2)}(t)$, the proof for the weak convergence is completed. Since the space $D[0, \tau]^K$ is equipped with the uniform metric, it suffices to show the marginal tightness of $W_k^{(2)}(t)$ for each k. The marginal tightness follows directly by applying lemma 5 to $W_k^{(2)}(t)$. Thus, $\boldsymbol{W}^{(2)}(t)$ converges weakly to a zero-mean Gaussian process where the covariance function between $\mathcal{W}_j^{(2)}(t_1)$ and $\mathcal{W}_k^{(2)}(t_2)$ is $\frac{1-\alpha}{\alpha} \operatorname{E}\{\psi_{1j}(\boldsymbol{\beta}_0, t_1)\psi_{1k}(\boldsymbol{\beta}_0, t_2)\}$. Note that $\mathcal{W}^{(1)}(t)$ and $\mathcal{W}^{(2)}(t)$ are independent since

$$
\begin{aligned}
& \operatorname{Cov}\left(n^{-1/2}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0, t_1), n^{-1/2}\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0, t_2)\right) \\
= \; & \operatorname{E}\left\{n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0, t_1)\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0, t_2)\right\} \\
= \; & \operatorname{E}\left\{\operatorname{E}\left(n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0, t_1)\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0, t_2)\middle| \mathcal{F}(\tau)\right)\right\} \\
= \; & \operatorname{E}\left\{n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0, t_1)\sum_{j=1}^{n}\operatorname{E}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\middle| \mathcal{F}(\tau)\right)\psi_{jm}(\boldsymbol{\beta}_0, t_2)\right\} = 0.
\end{aligned}
$$

Therefore, $\boldsymbol{W}(t) = \boldsymbol{W}^{(1)}(t) + \boldsymbol{W}^{(2)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}(t) = \mathcal{W}^{(1)}(t) + \mathcal{W}^{(2)}(t)$ where the covariance function between $\mathcal{W}_j(t_1)$ and $\mathcal{W}_k(t_2)$ is $\operatorname{E}\{\nu_{1j}(\boldsymbol{\beta}_0, t_1)\nu_{1k}(\boldsymbol{\beta}_0, t_2)\} + \frac{1-\alpha}{\alpha} \operatorname{E}\{\psi_{1j}(\boldsymbol{\beta}_0, t_1)\psi_{1k}(\boldsymbol{\beta}_0, t_2)\}$. This completes the proof of theorem 4.2.

Proofs of Theorems 4.3 and 4.4 The asymptotic properties of Estimator II can be shown by the similar arguments used for Estimator I. However, the resulting asymptotic properties need some modifications and will involves $(1-\Delta_{ik})$. This is because the asymptotic expansion of $n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1} - \widetilde{\alpha}^{-1}\right)$ includes the terms involving $(1-\Delta_{ik})$. In addition, the asymptotic expansion of $n^{1/2}\left(\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}, t)\right)$ $(d = 0, 1)$ is different from the one using

Estimator I and includes the terms involving $(1-\Delta_{ik})$ as well. Specifically, for each k, it follows from the Taylor expansion of $\widehat{\alpha}_k^{II}(t)^{-1}$ around $\widetilde{\alpha}$,

$$\begin{aligned}\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1} &= -\frac{1}{\alpha_{**}(t)^2}\left(\widehat{\alpha}_k^{II}(t)-\widetilde{\alpha}\right)\\ &= \frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{1}{\sum_{i=1}^n(1-\Delta_{ik})Y_{ik}(t)}\left\{\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}\end{aligned}$$

where $\alpha_{**}(t)$ is on the line segment between $\widehat{\alpha}_k^{II}(t)$ and $\widetilde{\alpha}$. Then,

$$n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)=\frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{n}{\sum_{i=1}^n(1-\Delta_{ik})Y_{ik}(t)}n^{-1/2}\left\{\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}$$

By Glivenko-Cantelli lemma, $n^{-1}\sum_{i=1}^n(1-\Delta_{ik})Y_{ik}(t)$ converges to $\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)$ in probability uniformly in t. In view of lemma 5, $n^{-1/2}\sum_{i=1}^n\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})Y_{ik}(t)$ converges to a zero-mean Gaussian process since $(1-\Delta_{ik})Y_{ik}(t)$ is bounded and monotone function in t. This implies $n^{-1}\sum_{i=1}^n(\frac{\xi_i}{\widetilde{\alpha}}-1)(1-\Delta_{ik})Y_{ik}(t)$ converges to 0 in probability uniformly in t and consequently, $\widehat{\alpha}_k^{II}(t)$ and $\widetilde{\alpha}$ converges to the same limit uniformly in t. This ensures $\alpha_{**}(t)$ also converges to the same limit as $\widetilde{\alpha}$. Combining these results, it follows from Slutsky's theorem that

$$\begin{aligned}&n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)=\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\left\{\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}\\ &+\left(\frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{n}{\sum_{i=1}^n(1-\Delta_{1k})Y_{ik}(t)}-\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\\ &=\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\left\{\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}\;+\;o_p(1). \qquad (4.29)\end{aligned}$$

Likewise, for each k,

$$
\begin{aligned}
& n^{1/2}\left\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)-\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\right\} = n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)} \\
+ \quad & n^{-1/2}\sum_{i=1}^{n}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{II}(t)^{-1}\right)(1-\Delta_{ik})\xi_i Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)} \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)} \\
+ \quad & n^{-1}\sum_{i=1}^{n}\left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\sum_{j=1}^{n}\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right)(1-\Delta_{jk})Y_{jk}(t)\right\} \\
\times \quad & (1-\Delta_{ik})\xi_i Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)} \; + \; o_p(1) \qquad\qquad (\text{ by } (5.24)\ ) \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)} + n^{-1/2}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik}) \\
\times \quad & \frac{Y_{ik}(t)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\left\{n^{-1}\sum_{j=1}^{n}(1-\Delta_{jk})\frac{\xi_j}{\widetilde{\alpha}}Y_{jk}(t)\boldsymbol{Z}_{jk}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{jk}(t)}\right\} \; + \; o_p(1) \qquad (4.30)
\end{aligned}
$$

It follows from lemma 5 that $n^{-1}\sum_{j=1}^{n}(1-\Delta_{jk})\frac{\xi_j}{\widetilde{\alpha}}Y_{jk}(t)\boldsymbol{Z}_{jk}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{jk}(t)}$ converges to $\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)}\right)$ for $d=0,1$, in probability uniformly in t. Thus, from (4.30)

$$
\begin{aligned}
& n^{1/2}\left\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)-\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)\right\} = n^{-1/2}\sum_{i=1}^{n}\left\{\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}\right. \\
+ \; & \left.\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})\frac{Y_{ik}(t)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)}\right)\right\} + o_p(1) \\
= \; & n^{-1/2}\sum_{i=1}^{n}(1-\Delta_{ik})\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t) \\
\times \; & \left\{\boldsymbol{Z}_{ik}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}-\frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)}\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\} + o_p(1) \qquad (4.31)
\end{aligned}
$$

It then follows from lemma 5 that both $n^{1/2}\left\{\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right\}$ and $n^{1/2}\left\{\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)-\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta},t)\right\}$ converge weakly to zero-mean Gaussian processes, respectively. Consequently, both $\left\{\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right\}$ and $\left\{\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta},t)-\boldsymbol{s}_k^{(d)}(\boldsymbol{\beta},t)\right\}$ converge to 0 in probability uniformly in t, respectively.

Now, one can decompose $n^{-1/2}U^{II}(\boldsymbol{\beta}_0)$ into two parts:

$$n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\boldsymbol{Z}_{ik}(u) - \frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,u)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\} dN_{ik}(u) + n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,u)}{S_k^{(0)}(\boldsymbol{\beta}_0,u)} - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\} dN_{ik}(u) \quad (4.32)$$

While the first term on the right-hand side of (4.32) remains the same as that of Estimator I, the second term needs modifications since it involves the weight functions. Specifically, the second term on the right-hand side of (4.32) can be further decomposed as

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,u)}{S_k^{(0)}(\boldsymbol{\beta}_0,u)} - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\} dN_{ik}(u) \\ = & \sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,u)}{S_k^{(0)}(\boldsymbol{\beta}_0,u)} - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\} d\left\{n^{-1/2}\sum_{i=1}^{n} M_{ik}(u)\right\} \\ + & n^{-1/2}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,u)}{S_k^{(0)}(\boldsymbol{\beta}_0,u)} - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\} \sum_{i=1}^{n} Y_{ik}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)}d\Lambda_{0k}(u) \quad (4.33) \end{aligned}$$

Based on conditions (C) and (G), $\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}_0,u)(d=0,1)$ are of bounded variations and $\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)$ is bounded away from 0. Therefore, $\frac{\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}$ is of bounded variations. Along with the uniform convergence of $\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}_0,u)$ to $\boldsymbol{s}_k^{\otimes d}(\boldsymbol{\beta}_0,u)(d=0,1)$, by the same arguments used for proving theorem 1, the first term on the right-hand side of (4.33) converges to 0 in probability uniformly in t as $n\to\infty$.

By the same argument used for theorem 1, the second term on the right-hand side of (4.33) can be shown to be equivalent to

$$\begin{aligned} & n^{-1/2}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t)}{S_k^{(0)}(\boldsymbol{\beta}_0,t)} - \frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}\right\} \sum_{i=1}^{n} Y_{ik}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)}d\Lambda_{0k}(t) \\ = & n^{1/2}\sum_{k=1}^{K}\int_0^\tau \left\{\boldsymbol{S}_k^{(1)}(\boldsymbol{\beta}_0,t) - \widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)\right\} d\Lambda_{0k}(t) \\ - & n^{1/2}\sum_{k=1}^{K}\int_0^\tau \left\{S_k^{(0)}(\boldsymbol{\beta}_0,t) - \widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,t)\right\} \boldsymbol{e}_k(\boldsymbol{\beta}_0,t)d\Lambda_{0k}(t) + o_p(1) \quad (4.34) \end{aligned}$$

The above equality holds since $\frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}_0,t)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,t)}$ converges to $\boldsymbol{e}_k(\boldsymbol{\beta}_0, t)$ in probability uniformly in t, $n^{1/2}\{\boldsymbol{S}_k^{(d)}(\boldsymbol{\beta}_0, t) - \widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}_0, t)\}(d = 0, 1)$ converges weakly to a zero-mean Gaussian process and $\Lambda_{0k}(t)$ is bounded on $t \in [0, \tau]$. Then, based on (4.31), the right-hand side of (4.34) is equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t) \\
& \times \left\{\boldsymbol{Z}_{ik}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)} - \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)}\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\} d\Lambda_{0k}(t) \\
& - n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t) \\
& \times \left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)} - \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)}\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\} \boldsymbol{e}_k(\boldsymbol{\beta}_0, t) d\Lambda_{0k}(t) + o_p(1) \\
& = n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \qquad (4.35) \\
& \times \int_0^\tau \left(\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) - \frac{Y_{ik}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0, t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right) d\Lambda_{0k}(t) + o_p(1).
\end{aligned}
$$

where $\boldsymbol{R}_{ik}(\boldsymbol{\beta}, t) = Y_{ik}(t)\widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)}$. Combining the above results, we have shown that $n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0) + n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \\
& \times \int_0^\tau \left(\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0, t) - \frac{Y_{ik}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0, t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right) d\Lambda_{0k}(t). \qquad (4.36)
\end{aligned}
$$

The first term on the right-hand side of (4.36) was again asymptotically zero-mean normal with covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0) = \mathrm{E}\left(\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\right)^{\otimes 2}$ by Spiekerman and Lin (1998). The second term on the right-hand side of (4.36) can be shown to be asymptotically zero-mean normal with covariance matrix $\frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ where

$$
\boldsymbol{V}_{II}(\boldsymbol{\beta}) = \mathrm{E}\left[\sum_{k=1}^{K}(1-\Delta_{1k})\int_0^\tau \left\{\boldsymbol{R}_{1k}(\boldsymbol{\beta}, u) - \frac{Y_{1k}(u)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}, u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\} d\Lambda_{0k}(u)\right]^{\otimes 2}.
$$

by Hájek (1960)'s central limit theorem for finite population sampling. Then, together with the independence of the first term and the second term of (4.36), it follows that $n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0)$ converges to zero mean normal random variable with finite covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$. Now, the consistency of $\widehat{\boldsymbol{\beta}}_{II}$ and the asymptotic normality of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ will follow from the similar arguments used for proving theorem 1 if we replace $\widehat{\alpha}_k(t), \widehat{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t)(d = 0, 1), \boldsymbol{U}^I(\boldsymbol{\beta})$ by $\widehat{\alpha}_k^{II}(t)$, $\widetilde{\boldsymbol{S}}_k^{(d)}(\boldsymbol{\beta}, t)(d = 0, 1)$, and $\boldsymbol{U}^{II}(\boldsymbol{\beta})$, and their corresponding asymptotic properties which we have derived.

The asymptotic properties of $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$ can also be shown by the similar arguments used for proving theorem 2 with some modifications. Specifically,

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t) - \Lambda_{0k}(t)\} \\
= \;& n^{1/2}\int_0^t \left\{ \frac{1}{n\widetilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, u)} - \frac{1}{n\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} d\sum_{i=1}^n M_{ik}(u) \\
+ \;& n^{1/2}\int_0^t \left\{ \frac{1}{\widetilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, u)} - \frac{1}{\widetilde{S}^{(0)}(\boldsymbol{\beta}_0, u)} \right\} S_k^{(0)}(\boldsymbol{\beta}_0, u) d\Lambda_{0k}(u) \\
+ \;& \int_0^t \frac{1}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0, u)} d\left\{ n^{-1/2}\sum_{i=1}^n M_{ik}(u) \right\} \\
+ \;& n^{1/2}\int_0^t \left\{ \frac{S_k^{(0)}(\boldsymbol{\beta}_0, u) - \widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0, u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} d\Lambda_{0k}(u)
\end{aligned}
\tag{4.37}
$$

By the Taylor expansion of $\widetilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, u)^{-1}$ around $\boldsymbol{\beta}_0$, the first term on the right-hand side of (4.37) is equivalent to

$$
\int_0^t \left(-\frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}^{**}, u)^T}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}^{**}, u)^2} \right) (\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0) d\left\{ n^{-1/2}\sum_{i=1}^n M_{ik}(u) \right\} \tag{4.38}
$$

where $\boldsymbol{\beta}^{**}$ is on the line segment between $\widehat{\boldsymbol{\beta}}_{II}$ and $\boldsymbol{\beta}_0$. Then, as $n \to \infty$, (4.38) converges to 0 uniformly in t in probability by lemma 2 since $\frac{\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta},u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta},u)}$ is of bounded variation, $\widehat{\boldsymbol{\beta}}_{II}$ is consistent for $\boldsymbol{\beta}_0$, and $n^{-1/2}\sum_{i=1}^n M_{ik}(u)$ converges weakly to a zero-mean Gaussian process with continuous sample path.

Again, it follows from the Taylor expansion of $\widetilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, u)^{-1}$ around $\boldsymbol{\beta}_0$, the uniform convergence of $\widetilde{\boldsymbol{S}}_k^{(1)}(\boldsymbol{\beta}, u)$ and $\widetilde{\boldsymbol{S}}_k^{(0)}(\boldsymbol{\beta}, u)$, the consistency of $\widehat{\boldsymbol{\beta}}_{II}$ for $\boldsymbol{\beta}_0$ and the boundedness of

$\Lambda_{0k}(t)$ on $[0,\tau)$ that the second term is asymptotically equivalent to

$$n^{1/2}\boldsymbol{r}_k(\boldsymbol{\beta}_0,t)^T\left(\widehat{\boldsymbol{\beta}}_{II}-\boldsymbol{\beta}_0\right).$$

The third term can be shown to be asymptotically equivalent to

$$\int_0^t \frac{1}{s_k^{(0)}(\boldsymbol{\beta}_0,u)} d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\}$$

by lemma 2 since $\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)$ is of bounded variation, converges uniformly to $s_k(\boldsymbol{\beta}_0,u)$ where $s_k(\boldsymbol{\beta}_0,u)$ is bounded away from 0, and $n^{-1/2}\sum_{i=1}^n M_{ik}(u)$ converges weakly to a zero-mean Gaussian process with continuous sample path.

For the last term on the right-hand side of (4.37), it follows from (4.31) and the uniform convergence of $\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,t)^{-1}$ to $s_k^{(0)}(\boldsymbol{\beta}_0,t)^{-1}$, where $s_k^{(0)}(\boldsymbol{\beta}_0,t)$ is bounded away from 0 that

$$n^{1/2}\int_0^t\left\{\frac{S_k^{(0)}(\boldsymbol{\beta}_0,u)-\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}{\widetilde{S}_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\}d\Lambda_{0k}(u)=n^{-1/2}\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})$$
$$\times\int_0^t Y_{ik}(u)\left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)}-\frac{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(u)}\right)}{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0,u)}$$

Now by combining the above results and using the asymptotic expansion of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{II}-\boldsymbol{\beta}_0)$ where

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{II}-\boldsymbol{\beta}_0)=\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\left\{n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0)+n^{-1/2}\sum_{i=1}^n\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\right.$$
$$\left.\times\quad\sum_{m=1}^K(1-\Delta_{im})\int_0^\tau\left\{\boldsymbol{R}_{im}(\boldsymbol{\beta}_0,t)-\frac{Y_{im}(t)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta}_0,t)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(t)\right)}\right\}d\Lambda_{0m}(t)\right\}+o_p(1),$$

we have

$$
\begin{aligned}
& n^{1/2}\left\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II},t)-\Lambda_{0k}(t)\right\} \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left[\left\{\boldsymbol{r}_k(\boldsymbol{\beta}_0,t)^T\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0)+\int_0^t\frac{1}{s_k^{(0)}(\boldsymbol{\beta}_0,u)}dM_{ik}(u)\right\}\right. \\
+ \;& \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\left\{\boldsymbol{r}_k(\boldsymbol{\beta}_0,t)^T\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\sum_{m=1}^{K}(1-\Delta_{im})\int_0^{\tau}\left(\boldsymbol{R}_{im}(\boldsymbol{\beta}_0,u)\right.\right. \\
- \;& \left.\frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta}_0,u)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right)d\Lambda_{0m}(u)+(1-\Delta_{ik})\int_0^t Y_{ik}(u)\left(e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)}\right. \\
- \;& \left.\left.\left.\frac{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(u)}\right)}{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right)\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\right]+o_p(1) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t)+n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\psi_{ik}^{II}(\boldsymbol{\beta}_0,t)+o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_{ik}(\boldsymbol{\beta},t) \;=\;& \boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta},t)+\int_0^t\{s_k^{(0)}(\boldsymbol{\beta},u)\}^{-1}dM_{ik}(u)\quad\text{and} \\
\psi_{ik}^{II}(\boldsymbol{\beta},t) \;=\;& \boldsymbol{r}_k(\boldsymbol{\beta},t)^T\boldsymbol{A}(\boldsymbol{\beta})^{-1}\sum_{m=1}^{K}(1-\Delta_{im})\int_0^{\tau}\left(\boldsymbol{R}_{im}(\boldsymbol{\beta},u)\right. \\
-\;& \left.\frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right)d\Lambda_{0m}(u)+(1-\Delta_{ik})\int_0^t Y_{ik}(u)\left(e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)}\right. \\
-\;& \left.\frac{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)}\right)}{\mathrm{E}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right)\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta},u)}.
\end{aligned}
$$

The asymptotic properties of $n^{1/2}\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II},t)-\Lambda_{0k}(t)\}$ follow from the similar arguments used for proving theorem 4.2. This complete the proofs of theorems 4.3 and 4.4.

TABLE 4.1: Summary of simulation results for $\widehat{\beta}_I$: $Z_{ik} \sim Bin(0.5)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.014	0.4189	0.4546	0.914
			0.67	0.013	0.4184	0.4353	0.932
			1.25	0.000	0.4199	0.4494	0.922
			4	0.005	0.4233	0.4608	0.920
		62	0.1	0.007	0.3469	0.3596	0.925
			0.67	0.006	0.3440	0.3463	0.932
			1.25	0.005	0.3470	0.3564	0.927
			4	-0.006	0.3485	0.3741	0.924
	10%	111	0.1	0.005	0.2145	0.2224	0.942
			0.67	0.007	0.2138	0.2170	0.951
			1.25	-0.002	0.2139	0.2188	0.941
			4	0.008	0.2139	0.2255	0.935
		222	0.1	-0.002	0.1781	0.1811	0.944
			0.67	0.008	0.1776	0.1765	0.945
			1.25	0.003	0.1776	0.1792	0.945
			4	0.003	0.1776	0.1824	0.939
	25%	333	0.1	-0.006	0.1210	0.1224	0.950
			0.67	0.003	0.1211	0.1241	0.946
			1.25	-0.002	0.1211	0.1239	0.947
			4	0.000	0.1211	0.1248	0.944
		666	0.1	-0.005	0.0991	0.1002	0.949
			0.67	0.002	0.0992	0.1013	0.947
			1.25	-0.002	0.0991	0.0983	0.957
			4	0.003	0.0992	0.0989	0.950
log(2)	3%	31	0.1	0.763	0.4279	0.4700	0.906
			0.67	0.749	0.4297	0.4408	0.928
			1.25	0.749	0.4333	0.4610	0.925
			4	0.738	0.4326	0.4680	0.917
		62	0.1	0.733	0.3562	0.3673	0.920
			0.67	0.719	0.3527	0.3514	0.934
			1.25	0.729	0.3567	0.3605	0.928
			4	0.716	0.3562	0.3755	0.917
	10%	111	0.1	0.710	0.2183	0.2259	0.938
			0.67	0.709	0.2169	0.2214	0.950
			1.25	0.702	0.2172	0.2204	0.942
			4	0.697	0.2169	0.2298	0.937
		222	0.1	0.697	0.1817	0.1828	0.950
			0.67	0.705	0.1809	0.1819	0.950
			1.25	0.702	0.1810	0.1823	0.952
			4	0.695	0.1809	0.1858	0.948
	25%	333	0.1	0.693	0.1243	0.1232	0.953
			0.67	0.701	0.1238	0.1272	0.945
			1.25	0.695	0.1236	0.1248	0.946
			4	0.695	0.1236	0.1269	0.940
		666	0.1	0.692	0.1026	0.1005	0.958
			0.67	0.697	0.1020	0.1039	0.943
			1.25	0.692	0.1018	0.0999	0.954
			4	0.696	0.1019	0.1015	0.961

TABLE 4.2: Summary of simulation results for $\widehat{\beta}_{II}$: $Z_{ik} \sim Bin(0.5)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.009	0.4243	0.4508	0.917
			0.67	0.001	0.4245	0.4338	0.931
			1.25	-0.005	0.4250	0.4444	0.932
			4	-0.009	0.4295	0.4570	0.923
		62	0.1	0.006	0.3499	0.3570	0.930
			0.67	0.005	0.3472	0.3445	0.937
			1.25	0.003	0.3501	0.3529	0.931
			4	-0.007	0.3518	0.3719	0.927
	10%	111	0.1	0.004	0.2114	0.2178	0.941
			0.67	0.007	0.2109	0.2130	0.948
			1.25	-0.001	0.2110	0.2151	0.940
			4	0.008	0.2109	0.2212	0.936
		222	0.1	-0.003	0.1766	0.1791	0.945
			0.67	0.007	0.1761	0.1751	0.953
			1.25	0.003	0.1762	0.1773	0.944
			4	0.003	0.1763	0.1807	0.943
	25%	333	0.1	-0.007	0.1165	0.1175	0.951
			0.67	0.003	0.1166	0.1194	0.947
			1.25	-0.001	0.1166	0.1182	0.951
			4	0.001	0.1166	0.1196	0.944
		666	0.1	-0.006	0.0978	0.0985	0.948
			0.67	0.001	0.0979	0.0998	0.946
			1.25	-0.002	0.0978	0.0964	0.958
			4	0.003	0.0979	0.0976	0.952
log(2)	3%	31	0.1	0.756	0.4324	0.4654	0.909
			0.67	0.740	0.4334	0.4364	0.934
			1.25	0.738	0.4365	0.4499	0.932
			4	0.730	0.4374	0.4608	0.924
		62	0.1	0.730	0.3588	0.3647	0.920
			0.67	0.717	0.3552	0.3489	0.935
			1.25	0.726	0.3595	0.3579	0.930
			4	0.714	0.3589	0.3728	0.919
	10%	111	0.1	0.706	0.2149	0.2206	0.940
			0.67	0.710	0.2135	0.2175	0.947
			1.25	0.701	0.2138	0.2156	0.943
			4	0.696	0.2135	0.2249	0.940
		222	0.1	0.696	0.1801	0.1814	0.946
			0.67	0.705	0.1793	0.1798	0.952
			1.25	0.701	0.1793	0.1797	0.949
			4	0.696	0.1793	0.1844	0.944
	25%	333	0.1	0.692	0.1196	0.1179	0.952
			0.67	0.699	0.1190	0.1231	0.940
			1.25	0.694	0.1188	0.1178	0.956
			4	0.695	0.1188	0.1211	0.944
		666	0.1	0.692	0.1012	0.0991	0.957
			0.67	0.697	0.1006	0.1029	0.947
			1.25	0.692	0.1003	0.0979	0.959
			4	0.695	0.1004	0.0999	0.955

TABLE 4.3: Summary of simulation results for $\widehat{\beta}_I$: $Z_{ik} \sim N(0,1)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	-0.003	0.2224	0.2321	0.924
			0.67	0.001	0.2206	0.2388	0.911
			1.25	0.002	0.2207	0.2367	0.916
			4	0.010	0.2202	0.2521	0.904
		62	0.1	-0.001	0.1774	0.1828	0.918
			0.67	0.000	0.1755	0.1810	0.914
			1.25	0.003	0.1769	0.1855	0.913
			4	0.008	0.1756	0.1907	0.901
	10%	111	0.1	0.004	0.1085	0.1144	0.940
			0.67	-0.002	0.1086	0.1100	0.949
			1.25	-0.002	0.1083	0.1138	0.941
			4	-0.005	0.1082	0.1093	0.948
		222	0.1	0.005	0.0892	0.0930	0.943
			0.67	-0.002	0.0895	0.0901	0.944
			1.25	-0.002	0.0894	0.0917	0.943
			4	-0.002	0.0896	0.0885	0.949
	25%	333	0.1	0.002	0.0608	0.0625	0.944
			0.67	-0.001	0.0608	0.0610	0.950
			1.25	-0.001	0.0608	0.0606	0.949
			4	-0.001	0.0608	0.0608	0.951
		666	0.1	0.000	0.0497	0.0505	0.947
			0.67	0.001	0.0496	0.0499	0.947
			1.25	0.000	0.0497	0.0497	0.951
			4	0.000	0.0497	0.0501	0.948
log(2)	3%	31	0.1	0.845	0.3263	0.4055	0.895
			0.67	0.857	0.3322	0.4113	0.875
			1.25	0.851	0.3144	0.4047	0.867
			4	0.877	0.3463	0.4606	0.860
		62	0.1	0.764	0.2184	0.2504	0.911
			0.67	0.758	0.2158	0.2522	0.907
			1.25	0.759	0.2151	0.2501	0.904
			4	0.779	0.2195	0.2617	0.894
	10%	111	0.1	0.729	0.1336	0.1484	0.924
			0.67	0.725	0.1328	0.1429	0.933
			1.25	0.721	0.1327	0.1486	0.919
			4	0.719	0.1319	0.1404	0.933
		222	0.1	0.712	0.1044	0.1121	0.932
			0.67	0.706	0.1037	0.1063	0.939
			1.25	0.705	0.1037	0.1086	0.939
			4	0.706	0.1035	0.1051	0.950
	25%	333	0.1	0.703	0.0724	0.0761	0.945
			0.67	0.701	0.0712	0.0719	0.950
			1.25	0.700	0.0709	0.0741	0.940
			4	0.700	0.0709	0.0748	0.929
		666	0.1	0.695	0.0560	0.0565	0.954
			0.67	0.694	0.0545	0.0557	0.951
			1.25	0.695	0.0544	0.0561	0.944
			4	0.696	0.0543	0.0553	0.948

TABLE 4.4: Summary of simulation results for $\widehat{\beta}_{II}$: $Z_{ik} \sim N(0,1)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_{II}$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.004	0.2284	0.2320	0.919
			0.67	0.004	0.2235	0.2383	0.916
			1.25	0.005	0.2239	0.2312	0.923
			4	0.003	0.2209	0.2346	0.910
		62	0.1	0.005	0.1822	0.1789	0.937
			0.67	0.002	0.1805	0.1785	0.924
			1.25	0.004	0.1793	0.1782	0.917
			4	-0.001	0.1783	0.1840	0.913
	10%	111	0.1	0.001	0.1064	0.1079	0.950
			0.67	0.003	0.1065	0.1096	0.942
			1.25	0.002	0.1067	0.1087	0.948
			4	0.001	0.1073	0.1110	0.940
		222	0.1	0.003	0.0883	0.0878	0.942
			0.67	0.001	0.0885	0.0903	0.944
			1.25	0.002	0.0885	0.0883	0.953
			4	0.001	0.0890	0.0899	0.951
	25%	333	0.1	-0.004	0.0585	0.0595	0.946
			0.67	0.000	0.0584	0.0595	0.942
			1.25	0.002	0.0584	0.0586	0.950
			4	0.001	0.0584	0.0566	0.954
		666	0.1	-0.002	0.0490	0.0493	0.949
			0.67	0.001	0.0490	0.0494	0.950
			1.25	0.003	0.0490	0.0492	0.948
			4	0.001	0.0490	0.0476	0.957
log(2)	3%	31	0.1	0.808	0.2554	0.3029	0.843
			0.67	0.814	0.2547	0.3114	0.836
			1.25	0.799	0.2533	0.3036	0.838
			4	0.797	0.2524	0.2994	0.853
		62	0.1	0.748	0.2044	0.2236	0.897
			0.67	0.751	0.2054	0.2306	0.886
			1.25	0.742	0.2048	0.2217	0.912
			4	0.749	0.2034	0.2251	0.892
	10%	111	0.1	0.715	0.1182	0.1283	0.908
			0.67	0.718	0.1170	0.1255	0.914
			1.25	0.716	0.1167	0.1276	0.913
			4	0.720	0.1170	0.1255	0.919
		222	0.1	0.706	0.0969	0.1016	0.937
			0.67	0.705	0.0959	0.1000	0.936
			1.25	0.702	0.0959	0.0989	0.937
			4	0.707	0.0962	0.0990	0.939
	25%	333	0.1	0.698	0.0624	0.0651	0.935
			0.67	0.696	0.0614	0.0628	0.941
			1.25	0.696	0.0611	0.0619	0.940
			4	0.698	0.0610	0.0625	0.937
		666	0.1	0.694	0.0529	0.0549	0.938
			0.67	0.695	0.0516	0.0532	0.945
			1.25	0.694	0.0514	0.0517	0.950
			4	0.695	0.0513	0.0514	0.942

TABLE 4.5: Summary of simulation results: cohort size = 3,000, event proportion = 3 %, $\beta_0 = \log(2)$

Z	n	θ	mean $\widehat{\beta}_I$	proposed S.E.$_I$	true S.D.$_I$	95% C.I.$_I$	mean $\widehat{\beta}_{II}$	proposed S.E.$_{II}$	true S.D.$_{II}$	95% C.I.$_{II}$
Bin(0.5)	93	0.1	0.710	0.2655	0.2759	0.936	0.707	0.2652	0.2724	0.943
		0.67	0.713	0.2662	0.2823	0.930	0.710	0.2971	0.2785	0.934
		1.25	0.717	0.2675	0.2811	0.934	0.713	0.2672	0.2766	0.937
		4	0.724	0.2663	0.2741	0.931	0.723	0.2662	0.2700	0.937
	186	0.1	0.705	0.2173	0.2223	0.940	0.704	0.2172	0.2211	0.940
		0.67	0.703	0.2189	0.2260	0.941	0.702	0.2209	0.2233	0.945
		1.25	0.706	0.2188	0.2266	0.930	0.705	0.2188	0.2246	0.932
		4	0.706	0.2177	0.2177	0.955	0.706	0.2177	0.2161	0.956
$N(0,1)$	93	0.1	0.732	0.1517	0.1721	0.915	0.723	0.1435	0.1615	0.910
		0.67	0.735	0.1513	0.1670	0.924	0.731	0.1445	0.1580	0.914
		1.25	0.735	0.1517	0.1648	0.927	0.726	0.1436	0.1554	0.931
		4	0.732	0.1513	0.1724	0.918	0.730	0.1442	0.1573	0.908
	186	0.1	0.713	0.1197	0.1261	0.927	0.708	0.1160	0.1231	0.924
		0.67	0.715	0.1194	0.1238	0.940	0.712	0.1160	0.1196	0.933
		1.25	0.715	0.1201	0.1243	0.942	0.709	0.1159	0.1197	0.936
		4	0.711	0.1195	0.1283	0.924	0.709	0.1157	0.1191	0.934

TABLE 4.6: Baseline characteristics of Busselton Health Study (subcohort sample)

Variables	Male (n=159) mean (sd) or %	Female (n=201) mean (sd) or %
Ferritin (μg/L)	214.2 (177.04)	95.8 (80.95)
Log(Ferritin)	5.0 (0.90)	4.2 (0.94)
Age (years)	59.5 (10.92)	59.4 (11.35)
BMI (kg/m^2)	26.1 (3.69)	25.5 (4.21)
Cholesterol (mmol/L)	6.2 (1.10)	6.4 (1.27)
Haemoglobin (g/100mL)	149.2 (9.87)	137.2 (9.07)
Systolic Blood Pressure (mmHg)	133.1 (20.01)	132.8 (20.25)
Triglycerides (mmol/L)	1.7 (1.16)	1.4 (0.86)
Diabetes Treatment (%)	1.9	2.5
Blood Pressure Treatment (%)	15.7	21.9
Smoke (Never)	32.7	68.2
Smoke (Former)	46.5	18.9
Smoke (Current)	20.8	12.9

TABLE 4.7: Analysis of Busselton Health Study

Variables	$\widehat{\beta}_I$	S.E.$_I$	HR$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	HR$_{II}$	95% C.I.$_{II}$
Ferritin on CHD	0.16	0.113	1.17	(0.94, 1.46)	0.11	0.112	1.12	(0.90, 1.40)
Ferritin on Stroke	0.20	0.156	1.22	(0.90, 1.66)	0.20	0.159	1.22	(0.90, 1.67)
Age	0.06	0.009	1.06	(1.04, 1.08)	0.06	0.008	1.06	(1.05, 1.08)
BMI	0.03	0.021	1.03	(0.99, 1.08)	0.04	0.021	1.04	(1.00, 1.08)
Cholesterol	-0.03	0.065	0.97	(0.86, 1.10)	0.01	0.058	1.00	(0.90, 1.13)
Triglycerides	0.23	0.076	1.26	(1.08, 1.46)	0.20	0.059	1.22	(1.09, 1.37)
Diabetes Treatment	0.19	0.457	1.21	(0.49, 2.98)	0.12	0.413	1.13	(0.50, 2.54)
Haemoglobin	-0.01	0.008	0.99	(0.97, 1.01)	0.02	0.007	1.00	(0.99, 1.02)
BPT†	0.26	0.207	1.30	(0.87, 1.95)	0.29	0.196	1.33	(0.91, 1.95)
SBP‡	0.01	0.005	1.01	(1.00, 1.02)	0.01	0.005	1.01	(1.00, 1.02)
Smoke (Former)	0.32	0.206	1.37	(0.92, 2.06)	0.34	0.198	1.40	(0.95, 2.06)
Smoke (Current)	0.49	0.241	1.62	(1.01, 2.60)	0.42	0.225	1.52	(0.98, 2.36)

†: Blood Pressure Treament, ‡: Systolic Blood Pressure

TABLE 4.8: Analysis of Busselton Health Study with Common Ferritin Effect

Variables	$\widehat{\beta}_I$	S.E.$_I$	HR$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	HR$_{II}$	95% C.I.$_{II}$
Ferritin	0.17	0.110	1.19	(0.96, 1.47)	0.14	0.107	1.16	(0.94, 1.43)
Age	0.06	0.009	1.06	(1.04, 1.08)	0.06	0.008	1.06	(1.05, 1.08)
BMI	0.03	0.021	1.03	(0.99, 1.08)	0.04	0.021	1.04	(1.00, 1.08)
Cholesterol	-0.03	0.064	0.97	(0.86, 1.10)	0.01	0.058	1.00	(0.90, 1.12)
Triglycerides	0.23	0.076	1.26	(1.08, 1.46)	0.20	0.059	1.22	(1.09, 1.37)
Diabetes Treatment	0.19	0.459	1.21	(0.50, 2.97)	0.13	0.412	1.13	(0.51, 2.54)
Haemoglobin	-0.01	0.008	0.99	(0.97, 1.01)	-0.01	0.007	0.99	(0.98, 1.01)
BPT	0.26	0.207	1.30	(0.87, 1.95)	0.28	0.195	1.33	(0.91, 1.95)
SBP	0.01	0.005	1.01	(1.00, 1.02)	0.01	0.005	1.01	(1.00, 1.02)
Smoke (Former)	0.32	0.205	1.38	(0.92, 2.06)	0.34	0.196	1.41	(0.96, 2.07)
Smoke (Current)	0.49	0.241	1.62	(1.01, 2.60)	0.42	0.225	1.52	(0.98, 2.37)

TABLE 4.9: Analysis of Busselton Health Study Considering Gender Effect

Variables	$\widehat{\beta}_I$	S.E.$_I$	HR$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	HR$_{II}$	95% C.I.$_{II}$
Ferritin on CHD (M)†	-0.01	0.154	1.00	(0.74, 1.35)	0.06	0.164	1.06	(0.77, 1.46)
Ferritin on CHD (W)‡	-0.15	0.135	0.86	(0.66, 1.12)	-0.06	0.150	0.96	(0.73, 1.27)
Ferritin on Stroke (M)†	-0.01	0.189	0.99	(0.68, 1.43)	0.03	0.198	1.03	(0.70, 1.52)
Ferritin on Stroke (W)‡	0.16	0.267	1.18	(0.70, 2.00)	0.35	0.299	1.42	(0.78, 2.56)
Age	0.07	0.009	1.07	(1.05, 1.09)	0.06	0.008	1.06	(1.04, 1.08)
BMI	0.02	0.023	1.02	(0.98, 1.07)	0.04	0.022	1.04	(0.99, 1.08)
Cholesterol	0.11	0.070	1.12	(0.98, 1.29)	0.06	0.067	1.07	(0.94, 1.22)
Triglycerides	0.17	0.069	1.18	(1.03, 1.35)	0.18	0.058	1.20	(1.07, 1.34)
Diabetes Treatment	-0.16	0.417	0.85	(0.38, 1.93)	0.16	0.408	1.17	(0.53, 2.60)
Haemoglobin	-0.01	0.007	0.99	(0.98, 1.01)	-0.01	0.007	0.99	(0.97, 1.00)
BPT	0.45	0.199	1.56	(1.06, 2.32)	0.31	0.199	1.36	(0.92, 2.00)
SBP	0.01	0.005	1.01	(1.00, 1.02)	0.01	0.005	1.01	(1.00, 1.02)
Smoke (Former)	-0.01	0.220	0.99	(0.64, 1.52)	0.12	0.221	1.12	(0.73, 1.73)
Smoke (Current)	0.10	0.232	1.10	(0.70, 1.73)	0.31	0.229	1.18	(0.87, 2.13)

†: For men, ‡: For women

CHAPTER 5

ADDITIVE HAZARDS MODEL FOR CASE-COHORT STUDIES WITH MULTIPLE DISEASE OUTCOMES

5.1 Introduction

The additive and multiplicative risk models provide the two principal frameworks for studying the association between risk factors and disease occurrence or death. Most modern analyses of survival data focus on multiplicative models for relative risk using proportional hazards models, mostly due to desirable theoretical properties along with the simple interpretation of the results and the wide availability of computer programs. However, epidemiologists often are interested in the risk difference attributed to the exposure, and the risk difference is known to be more relevant to public health because it translates directly into the number of disease cases that would be avoided by eliminating a particular exposure (Kulich & Lin, 2000). Consequently, the additive hazards model, which model the risk differences, has often been suggested as an alternative to the proportional hazards model.

For univariate failure time data, Lin and Ying (1994) proposed a semiparametric estimating procedure using estimating equation approach and derived the large-sample theory of the proposed estimators. Yin and Cai (2004) extended this procedure to the multivariate failure time data using marginal model approach. Pipper and Martinussen (2004) also extended this procedure to the clustered failure time data and studied parametric shared frailty models as well.

All the aforementioned work assume the data are obtained fully for all the members in the entire cohort. However, conducting epidemiologic cohort studies could be prohibitively expensive and thus it might not be always feasible to obtain data for the full cohort. The case-cohort study design (Prentice, 1986) is one of several study designs which have been proposed in an attempt to reduce costs in expensive epidemiological cohort studies. The key idea of this study design is to obtain the covariate measurements only on a random sample (subcohort) from the entire cohort and all the subjects in the cohort who experience the disease of interest (cases). The major cost typically arise from the assembling of covariate measurements and much of the covariate information on disease-free subjects (controls) is largely redundant. Thus, the case-cohort study designs are particularly useful for large-scale cohort studies with low disease rate or for cohort studies where the measurements of covariates are expensive. A key advantage of the case-cohort design is its ability to use the same subcohort for several diseases or for subtypes of disease (e.g. Prentice, 1986; Wacholder et al., 1989; Langholz and Thomas, 1990; Wacholder et al., 1991). For example, the case-cohort design was implemented in the Busselton Health Study (Cullen, 1972). In this study, it was of interest to study the relationship between serum ferritin and coronary heart disease and stroke events. To reduce costs and preserve stored serum, case-cohort sampling was used. In order to compare the effect of serum ferritin on coronary heart disease and stroke, times to coronary heart disease and stroke events need to be modeled simultaneously. Since times to coronary heart disease and stroke events observed from the same subject could be correlated, valid statistical method needs to take it into consideration.

For a single disease outcome, Kulich and Lin (2000) developed the semiparametric inference procedure for the case-cohort data. Sun, Sun and Flournoy (2004) extended this approach to competing risks analysis. Despite the progress in the methods for analyzing case-cohort data, methodologies to address analysis of case-cohort data with multiple diseases outcomes have been limited. A valid statistical methods which account for the correlation between outcomes is needed.

In this chapter, we propose a weighted estimating equation approach for estimating the parameters in the marginal additive hazards regression models for the multivariate failure time data from case-cohort studies with multiple disease outcomes.

The rest of this chapter is organized as follows. The proposed model and method of estimation are presented in Section 5.2. In Section 5.3, the asymptotic properties of the proposed estimators are studied. The outlines of the proofs for the asymptotic results are provided in the appendix. The finite sample properties are investigated by simulations in Section 5.4. The methodology is illustrated in Section 5.5 using the aforementioned Busselton Health Study. In Section 5.6, we give a few concluding remarks.

5.2 Model and Estimation

Suppose that there are n independent subjects in a cohort study and there are K different disease outcomes of interest. Consider independent failure time response vectors $\boldsymbol{T}_i = (T_{i1}, \cdots, T_{iK})^T, i = 1, \cdots, n$. For example, (T_{i1}, T_{i2}) may denote time for CHD and time for stroke for subject i. Let C_{ik} denote the potential censoring time for outcome k of subject i. We assume that C_{ik} is independent of the disease processes. In most practical cases, $C_{ik} = C_i$ for $k = 1, \ldots, K$. The observed time is $X_{ik} = \min(T_{ik}, C_{ik})$. Let $N_{ik}(t)$ denote the counting process for outcome k of subject i, $Y_{ik}(t) = I(X_{ik} \geq t)$ denote an 'at risk' indicator process and $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ denote an indicator for failure, where $I(\cdot)$ is an indicator function. Let $Z_{ik}(t)$ be a $p \times 1$ covariate vector corresponding to the kth disease outcome for subject i at time t. We assume that all the time-dependent covariates in $Z_{ik}(t)$ are "external", i.e., they are not affected by the disease processes, as described by Kalbfleisch and Prentice (2002). Let $\lambda_{ik}(t)$ denote the corresponding marginal hazards function and $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) \left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du$ denote a martingale with respect to the marginal filtration $\mathcal{F}_{ik}(t) = \sigma\{N_{ik}(s), Y_{ik}(s), \boldsymbol{Z}_{ik}(s) : \ 0 \leq s \leq t\}$. Let $\boldsymbol{X}_i = (X_{i1}, \cdots, X_{iK})^T, i = 1, \cdots, n$, denote the observed failure time vector and $\boldsymbol{Z}_i(\cdot) = (\boldsymbol{Z}_{i1}(\cdot), \cdots, \boldsymbol{Z}_{iK}(\cdot))^T$ denote the covariate vector. Let τ denote the study end time.

Under the case-cohort design, suppose we select a subcohort of fixed size $\tilde{n}$ from the cohort by simple random sampling without replacement. This sampling may be done prospectively or retrospectively. Let ξ_i denote the indicator for the ith subject being selected into the subcohort and $\pi_i = \Pr(\xi_i = 1) = \tilde{\alpha} = \tilde{n}/n$ denote the selection probability of the ith subject. Here $\xi_1, \ldots, \xi_n$ are correlated due to the sampling scheme. We assume that complete covariate histories $\boldsymbol{Z}_{ik}(t)(0 \leq t \leq \tau)$ are available for all the subcohort members and for the cases outside the subcohort. For all the others, we assume that their censoring time information are available. Thus, the observable information for the kth disease outcome of the ith subject when $\xi_i = 1$ or $\Delta_{ik} = 1$ is $\{X_{ik}, \Delta_{ik}, \xi_i, \boldsymbol{Z}_{ik}(t), 0 \leq t \leq X_{ik}\}$ and when $\xi_i = 0$ and $\Delta_{ik} = 0$ is $\{X_{ik}, \Delta_{ik}, \xi_i\}$.

5.2.1 Additive hazards models

In this subsection, we study marginal additive hazards regression model for multiple disease outcomes data from case-cohort studies.

We consider the following additive hazards model for T_{ik}

$$\lambda_{ik}(t|\boldsymbol{Z}_{ik}) = \lambda_{0k}(t) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(t), \tag{5.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease outcome k and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of fixed unknown parameters. Note that a subject may experience all K diseases, may also experience only some, or even none of the events of interest due to right censoring. The baseline hazard function is explicitly disease-specific.

5.2.2 Estimation

If the data were complete, the true regression parameter $\boldsymbol{\beta}_0$ in (5.1) could be estimated by solving the estimating function (Yin and Cai, 2004)

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k(t)\}\{dN_{ik}(t) - Y_{ik}(t)\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)dt\}, \tag{5.2}$$

where

$$\overline{\boldsymbol{Z}}_k(t) = \frac{\sum_{i=1}^n Y_{ik}(t)\boldsymbol{Z}_{ik}(t)}{\sum_{i=1}^n Y_{ik}(t)}$$

There exists an explicit solution $\widehat{\boldsymbol{\beta}}$ to the estimating equations $\boldsymbol{U}(\boldsymbol{\beta}) = \mathbf{0}_{p\times 1}$ and has the following form:

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau Y_{ik}(t)\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k(t)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k(t)\} dN_{ik}(t)\right]$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$.

For data from case-cohort studies, (5.2) cannot be calculated since the data are not complete. Thus, we consider the following weighted estimating function

$$\boldsymbol{U}^I(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^\rho(t)\}\{dN_{ik}(t) - \rho_{ik}(t)Y_{ik}(t)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)dt\}, \tag{5.3}$$

where

$$\overline{\boldsymbol{Z}}_k^\rho(t) = \sum_{i=1}^n \rho_{ik}(t)\boldsymbol{Z}_{ik}(t)Y_{ik}(t) / \sum_{i=1}^n \rho_{ik}(t)Y_{ik}(t)$$

and $\rho_{ik}(t)$ is a possibly time-dependent weight function which has the following form:

$$\rho_{ik}(t) \quad = \xi_i/\widehat{\alpha}_k(t) \quad \text{where} \quad \widehat{\alpha}_k(t) = \frac{\sum_{i=1}^n \xi_i Y_{ik}(t)}{\sum_{i=1}^n Y_{ik}(t)}$$

The estimator of the hazards regression parameter $\boldsymbol{\beta}_0$ is defined as the solution to this equation and is denoted by $\widehat{\boldsymbol{\beta}}_I$. $\widehat{\boldsymbol{\beta}}_I$ has the following explicit form:

$$\widehat{\boldsymbol{\beta}}_I = \left[\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \rho_{ik}(t)Y_{ik}(t)\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^\rho(t)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^\rho(t)\} dN_{ik}(t)\right]$$

We will call this type of estimator as Estimator I. Here $\widehat{\alpha}_k(t)$ is the estimator of the true sampling probability $\tilde{\alpha}$ and denotes the proportion of sampled subjects among the number of subjects remaining in the risk set at time t. This type of weight function has been considered for multiplicative hazards models in the univariate failure time data. It was first considered by Barlow (1994). Borgan *et al.*(2000) considered the same type of the weight functions for stratified case-cohort studies. The estimator considered by Self and Prentice (1988) is a special case and can be obtained by replacing $\widehat{\alpha}_k(t)$ by $\tilde{\alpha}$.

Let $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$. A Breslow-Aalen type estimator of the baseline cumulative hazard function is given by $\widehat{\Lambda}_{0k}^I(\widehat{\boldsymbol{\beta}}_I, t)$, where

$$\widehat{\Lambda}_{0k}^I(\boldsymbol{\beta}, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_{ik}(u) - \rho_{ik}(u)Y_{ik}(u)\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(u)du\}}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)}$$

Note that the construction of $\widehat{\alpha}_k(t)$ does not involve the cases outside the subcohort and $\widehat{\boldsymbol{\beta}}_I$ requires the covariate measurement of the cases outside the subcohort only at their failure times. However, when the complete covariate measurement history is available for the cases outside the subcohort, $\widehat{\boldsymbol{\beta}}_I$ might not be very efficient since it discards some of the available information. To make better use of the available information, we consider the following pseudo-likelihood function

$$\boldsymbol{U}^{II}(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \omega_{ik}(t)\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^\omega(t)\}\{dN_{ik}(t) - Y_{ik}(t)\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)dt\}, \tag{5.4}$$

where

$$\overline{\boldsymbol{Z}}_k^\omega(t) = \sum_{i=1}^n \omega_{ik}(t)\boldsymbol{Z}_{ik}(t)Y_{ik}(t) / \sum_{i=1}^n \omega_{ik}(t)Y_{ik}(t)$$

and $\omega_{ik}(t)$ is a possibly time-dependent weight function which has the following form:

$$\omega_{ik}(t) = \Delta_{ik} + (1 - \Delta_{ik})\xi_i / \widehat{\alpha}_k^{II}(t) \quad \text{where} \quad \widehat{\alpha}_k^{II}(t) = \frac{\sum_{i=1}^n \xi_i(1 - \Delta_{ik})Y_{ik}(t)}{\sum_{i=1}^n (1 - \Delta_{ik})Y_{ik}(t)}$$

The estimator of the hazards regression parameter $\boldsymbol{\beta}_0$ is defined as the solution to this equation and is denoted by $\widehat{\boldsymbol{\beta}}_{II}$. We will call this type of estimator as Estimator II.

This weight function is defined to be equal to one for the cases regardless of their subcohort membership and to $\widehat{\alpha}_k^{II}(t)^{-1}$ for the sampled censored individuals. Thus, the construction of $\widehat{\alpha}_k^{II}(t)$ should involve only censored individuals. Unlike the weight function for $\widehat{\boldsymbol{\beta}}_I$, it uses the information from all the individuals sampled. Consequently, it is anticipated that this results in a more efficient estimator. This approach also has been considered for multiplicative hazards models in the univariate failure time data. It was first proposed by Kalbfleisch and Lawless (1988) and they considered a time-invariant version of the weight functions, i.e., they used $\widetilde{\alpha}$ instead of $\widehat{\alpha}_k^{II}(t)$. Borgan *et al.*(2000) considered the same type of the weight functions in the univariate failure time data from stratified case-cohort studies. For additive hazards models, Kulich and Lin (2000) considered a time-invariant version of the weight functions for the univariate failure time data. To be able to use this approach, one is required to assess complete covariate histories for the cases throughout their at-risk periods, which might not be always available for prospective studies. In case of having complete covariate histories for the cases, using this type of weights is expected to improve efficiencies. The Breslow-Aalen type estimator of the cumulative baseline hazard function is given by $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$, where

$$\widehat{\Lambda}_{0k}^{II}(\boldsymbol{\beta}, t) = \int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)\{dN_{ik}(u) - Y_{ik}(u)\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(u)du\}}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} \tag{5.5}$$

5.3 Asymptotic properties

In this section, we describe the asymptotic properties of the proposed estimates. We define the following notation for convenience: For $k = 1, \ldots, K$,

$$\boldsymbol{e}_k(t) = \frac{\mathrm{E}\left(Y_{1k}(t)\boldsymbol{Z}_{1k}(t)\right)}{\mathrm{E}(Y_{1k}(t))}, \ \boldsymbol{A}_k = \mathrm{E}\left\{\int_0^\tau Y_{1k}(t)(\boldsymbol{Z}_{1k}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2})dt\right\}$$
$$\widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t) = \boldsymbol{Z}_{ik}(t) - \boldsymbol{e}_k(\boldsymbol{\beta}, t), \quad \text{and} \quad \mathbf{M}_{\tilde{\boldsymbol{Z}}, ik}(\boldsymbol{\beta}) = \int_0^\tau \widetilde{\boldsymbol{Z}}_{ik}(\boldsymbol{\beta}, t)dM_{ik}(t).$$

Here and hereafter the norms for the vector $\boldsymbol{a}$, matrix $\boldsymbol{A}$, and function f are defined as the following:

$$\|\boldsymbol{a}\| = \max_i |a_i|, \ \|\boldsymbol{A}\| = \max_{i,j} |A_{ij}|, \|f\| = \sup_t |f(t)|$$

5.3.1 Asymptotic properties of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\Lambda}_{0k}^I(t)$

We summarize the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_I$ in the following theorem :

Theorem 5.1 *Under the conditions in the Appendix, $\widehat{\boldsymbol{\beta}}_I$ solving (5.3) is a consistent estimator of $\boldsymbol{\beta}_0$. Also $n^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and* *with variance matrix of the form* $\boldsymbol{\Sigma}_I(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}$ *where*

$$\boldsymbol{A} = \sum_{k=1}^{K}\boldsymbol{A}_k,\ \boldsymbol{Q}(\boldsymbol{\beta}) \ = \ \mathrm{E}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\boldsymbol{\beta})\right)^{\otimes 2},$$

$$\boldsymbol{V}(\boldsymbol{\beta}) = \ \mathrm{E}\left[\sum_{k=1}^{K}\int_0^{\tau}\left(\boldsymbol{R}_{1k}(\boldsymbol{\beta},t) - \frac{Y_{1k}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right)dt\right]^{\otimes 2}$$

and $\boldsymbol{R}_{ik}(\boldsymbol{\beta},t) = Y_{ik}(t)(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t))(\lambda_{0k}(t)+\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t))$.

The matrices $\boldsymbol{A}$, $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\boldsymbol{V}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{Q}}(\widehat{\boldsymbol{\beta}}_I)$ and $\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\beta}}_I)$ where

$$\widehat{\boldsymbol{A}} = -n^{-1}\frac{\partial \boldsymbol{U}^I(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2},$$

$$\widehat{\boldsymbol{V}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}\left[\sum_{k=1}^{K}\int_0^{\tau}\left(\widehat{\boldsymbol{R}}_{ik}(\boldsymbol{\beta},t) - \frac{Y_{ik}(t)\widehat{\mathrm{E}}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)\right)}{\widehat{\mathrm{E}}Y_{1k}(t)}\right)dt\right]^{\otimes 2},$$

$$\widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) = \int_0^{\tau}\left(\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\rho}(t)\right)d\widehat{M}_{ik}(\boldsymbol{\beta},t),$$

$$\widehat{M}_{ik}(\boldsymbol{\beta},t) = N_{ik}(t) - \int_0^{t}Y_{ik}(u)d\widehat{\Lambda}_{0k}^{I}(\boldsymbol{\beta},u) - \int_0^{t}Y_{ik}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)du$$

$$\widehat{\boldsymbol{R}}_{ik}(\boldsymbol{\beta},t) = Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^{\rho}(t)\right)\left(d\widehat{\Lambda}_{0k}^{I}(\boldsymbol{\beta},t)+\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\right),$$

$$\widehat{\mathrm{E}}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(t)\right) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}\left(Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^{\rho}(t)\right)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\right),$$

and $\widehat{\mathrm{E}}Y_{1k}(t) = n^{-1}\sum_{i=1}^{n}Y_{ik}(t)$.

To study the asymptotic properties of $\widehat{\Lambda}_{0k}^{I}(\widehat{\boldsymbol{\beta}}_I,t)(k=1,\ldots,K)$, we define the following metric space. Let $D[0,\tau]^K$ be a metric space consisting of right-continuous functions $\boldsymbol{f}(t)$ with left-hand limits where $\boldsymbol{f}(t) = \{f_1(t),\ldots,f_K(t)\}^T$ and $f_k(t):[0,\tau]\to\mathcal{R}$. The metric for this space is defined as $d_k(\boldsymbol{f},\boldsymbol{g}) = \sup_{k,t\in[0,\tau]}\{|f_k(t)-g_k(t)| : 1\le k\le K\}$ for $\boldsymbol{f},\boldsymbol{g}\in D[0,\tau]^K$. We summarize the asymptotic properties of $\widehat{\Lambda}_{0k}^{I}(\widehat{\boldsymbol{\beta}}_I,t)(k=1,\ldots,K)$ in the following theorem.

Theorem 5.2 *Under the conditions in the Appendix, for each* $k=1,\ldots,K$, $\widehat{\Lambda}_{0k}^{I}(\widehat{\boldsymbol{\beta}}_I,t)$ *converges in probability to* $\Lambda_{0k}(t)$ *uniformly in* $t\in[0,\tau]$. *Also,* $\boldsymbol{W}(t) = n^{1/2}[\{\widehat{\Lambda}_{01}^{I}(\widehat{\boldsymbol{\beta}}_I,t) -$

$\Lambda_{0k}(t)\}, \ldots, \{\widehat{\Lambda}^{I}_{0K}(\widehat{\boldsymbol{\beta}}_I, t) - \Lambda_{0K}(t)\}]^T$ *converges weakly to a zero-mean Gaussian process* $\mathcal{W}(t)$ *in* $D[0,\tau]^K$ *where* $\mathcal{W}(t) = (\mathcal{W}_1(t), \ldots, \mathcal{W}_K(t))^T$. *The covariance function between* $\mathcal{W}_j(t_1)$ *and* $\mathcal{W}_k(t_2)$ *is*

$$\phi_{jk}(t_1,t_2)(\boldsymbol{\beta}_0) = \mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0,t_1)\nu_{1k}(\boldsymbol{\beta}_0,t_2)\} + \frac{1-\alpha}{\alpha}\mathrm{E}\{\psi_{1j}(\boldsymbol{\beta}_0,t_1)\psi_{1k}(\boldsymbol{\beta}_0,t_2)\}$$

where

$$\begin{aligned}
\nu_{ik}(\boldsymbol{\beta},t) =&\ \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}) + \int_0^t\{\mathrm{E}\,Y_{1k}(u)\}^{-1}dM_{ik}(u),\\
\psi_{ik}(\boldsymbol{\beta},t) =&\ \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}\int_0^\tau\{\boldsymbol{R}_{im}(\boldsymbol{\beta},u)\\
&-\ \frac{Y_{im}(u)}{\mathrm{E}\,Y_{1m}(u)}\mathrm{E}\left(Y_{1m}(u)\left(\boldsymbol{Z}_{1m}(u)-\boldsymbol{e}_m(u)\right)\boldsymbol{\beta}^T\boldsymbol{Z}_{1m}(u)\right)\Big\}du\\
&+\ \int_0^t Y_{ik}(u)\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\,Y_{1k}(u)}\right\}\frac{du}{\mathrm{E}\,Y_{1k}(u)} \text{ and } \boldsymbol{r}_k(t) = -\int_0^t\boldsymbol{e}_k(u)du.
\end{aligned}$$

$\phi_{jk}(t_1,t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}_{jk}(t_1,t_2)(\widehat{\boldsymbol{\beta}}_I)$ where

$$\begin{aligned}
\widehat{\phi}_{jk}(t_1,t_2)(\boldsymbol{\beta}) =&\ n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\widehat{\nu}_{ij}(\boldsymbol{\beta},t_1)\widehat{\nu}_{ik}(\boldsymbol{\beta},t_2) + \frac{1-\widetilde{\alpha}}{\widetilde{\alpha}}n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\widehat{\psi}_{ij}(\boldsymbol{\beta},t_1)\widehat{\psi}_{ik}(\boldsymbol{\beta},t_2),\\
\widehat{\nu}_{ik}(\boldsymbol{\beta},t) =&\ \widehat{\boldsymbol{r}}_k(t)^T\widehat{\boldsymbol{A}}^{-1}\sum_{m=1}^{K}\widehat{\mathbf{M}}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}) + \int_0^t\{\widehat{\mathrm{E}}Y_{1k}(u)\}^{-1}d\widehat{M}_{ik}(\boldsymbol{\beta},u),\\
\widehat{\psi}_{ik}(\boldsymbol{\beta},t) =&\ \widehat{\boldsymbol{r}}_k(t)^T\widehat{\boldsymbol{A}}^{-1}\sum_{m=1}^{K}\int_0^\tau\Big\{\widehat{\boldsymbol{R}}_{im}(\boldsymbol{\beta},t)\\
&-\ \frac{Y_{im}(t)\widehat{\mathrm{E}}\left(Y_{1m}(t)\left(\boldsymbol{Z}_{1m}(t)-\boldsymbol{e}_m(t)\right)\boldsymbol{\beta}^T\boldsymbol{Z}_{1m}(t)\right)}{\widehat{\mathrm{E}}Y_{1m}(t)}\Bigg\}dt\\
&+\ \int_0^t Y_{ik}(u)\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u) - \frac{\widehat{\mathrm{E}}\left(Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right)}{\widehat{\mathrm{E}}Y_{1k}(u)}\right\}\frac{du}{\widehat{\mathrm{E}}Y_{1k}(u)},\\
\widehat{\mathrm{E}}\left(Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right) =&\ n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}Y_{ik}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u), \quad \text{and } \widehat{\boldsymbol{r}}_k(t) = -\int_0^\tau\overline{\boldsymbol{Z}}^{\rho}_k(t)dt.
\end{aligned}$$

5.3.2 Asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II}, t)$

In this subsection, we will study the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II}, t)$. We summarize the asymptotic behavior of the regression parameter estimator $\widehat{\boldsymbol{\beta}}_{II}$ in the following theorem :

Theorem 5.3 *Under the conditions in the Appendix, $\widehat{\boldsymbol{\beta}}_{II}$ solving (5.4) is a consistent estimator of $\boldsymbol{\beta}_0$. Also, $n^{1/2}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}_{II}(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}$ where*

$$\boldsymbol{V}_{II}(\boldsymbol{\beta}) = \mathrm{E}\left[\sum_{k=1}^{K}(1-\Delta_{1k})\int_0^\tau \left\{\boldsymbol{R}_{1k}(\boldsymbol{\beta},t) - \frac{Y_{1k}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\}dt\right]^{\otimes 2}.$$

The matrices $\boldsymbol{A}$, $\boldsymbol{Q}(\boldsymbol{\beta}_0)$ and $\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\boldsymbol{A}}^{II}$, $\widehat{\boldsymbol{Q}}^{II}(\widehat{\boldsymbol{\beta}}_{II})$ and $\widehat{\boldsymbol{V}}_{II}(\widehat{\boldsymbol{\beta}}_{II})$ where

$$\begin{aligned}
\widehat{\boldsymbol{A}}^{II} &= -n^{-1}\frac{\partial \boldsymbol{U}^{II}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ \widehat{\boldsymbol{Q}}^{II}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\left(\sum_{k=1}^{K}\widehat{\mathbf{M}}^{II}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta})\right)^{\otimes 2},\\
\widehat{\boldsymbol{V}}_{II}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}\left[\sum_{k=1}^{K}(1-\Delta_{ik})\int_0^\tau\left(\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta},t) - \frac{Y_{ik}(t)\widehat{\mathrm{E}}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right)dt\right]^{\otimes 2},\\
\widehat{\mathbf{M}}^{II}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}) &= \int_0^\tau\left(\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}^{\omega}_{k}(t)\right)d\widehat{M}^{II}_{ik}(\boldsymbol{\beta},t),\\
\widehat{M}^{II}_{ik}(\boldsymbol{\beta},t) &= N_{ik}(t) - \int_0^t Y_{ik}(u)d\widehat{\Lambda}^{II}_{0k}(\boldsymbol{\beta},u) - \int_0^t Y_{ik}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)du,\\
\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta},t) &= Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}^{\omega}_{k}(t)\right)\left(d\widehat{\Lambda}^{II}_{0k}(\boldsymbol{\beta},t) + \boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\right),\\
&\widehat{\mathrm{E}}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)\right) = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\widetilde{\alpha}}(1-\Delta_{ik})\widehat{\boldsymbol{R}}^{II}_{ik}(\boldsymbol{\beta},t)\\
&\text{and } \widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(t)\right) = n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t).
\end{aligned}$$

The asymptotic properties of $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II},t)(k=1,\ldots,K)$ are summarized in the following theorem.

Theorem 5.4 *Under the conditions in the Appendix, for each $k = 1,\ldots,K$, $\widehat{\Lambda}^{II}_{0k}(\widehat{\boldsymbol{\beta}}_{II},t)$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0,\tau]$. Also, $\boldsymbol{W}^{II}(t) = n^{1/2}[\{\widehat{\Lambda}^{II}_{01}(\widehat{\boldsymbol{\beta}}_{II},t) - \Lambda_{01}(t)\},\ldots,\{\widehat{\Lambda}^{II}_{0K}(\widehat{\boldsymbol{\beta}}_{II},t) - \Lambda_{0K}(t)\}]^T$ converges weakly to a zero-mean Gaussian process $\mathcal{W}^{II}(t)$ in $D[0,\tau]^K$ where $\mathcal{W}^{II}(t) = (\mathcal{W}^{II}_1(t),\ldots,\mathcal{W}^{II}_K(t))^T$. The covariance function between $\mathcal{W}^{II}_j(t_1)$*

and $\mathcal{W}_k^{II}(t_2)$ *is*

$$\phi_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}_0) = \mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0,t_1)\nu_{1k}(\boldsymbol{\beta}_0,t_2)\} + \frac{1-\alpha}{\alpha}\mathrm{E}\left\{\psi_{1j}^{II}(\boldsymbol{\beta}_0,t_1)\psi_{1k}^{II}(\boldsymbol{\beta}_0,t_2)\right\}$$

where

$$\begin{aligned}
\nu_{ik}(\boldsymbol{\beta},t) =&\ \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^K \mathbf{M}_{\widetilde{\boldsymbol{z}},im}(\boldsymbol{\beta}) + \int_0^t \{\mathrm{E}\,Y_{1k}(u)\}^{-1}dM_{ik}(u), \text{ and} \\
\psi_{ik}^{II}(\boldsymbol{\beta},t) =&\ \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^K(1-\Delta_{im})\int_0^\tau\left\{\boldsymbol{R}_{im}(\boldsymbol{\beta},u) - \frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta},u)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right\}du \\
&+\ (1-\Delta_{ik})\int_0^t\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\}\frac{Y_{ik}(u)du}{\mathrm{E}\,Y_{1k}(u)}.
\end{aligned}$$

$\phi_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\phi}_{jk}^{II}(t_1,t_2)(\widehat{\boldsymbol{\beta}}_{II})$ where

$$\begin{aligned}
&\widehat{\phi}_{jk}^{II}(t_1,t_2)(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n\frac{\xi_i}{\widetilde{\alpha}}\widehat{\nu}_{ij}^{II}(\boldsymbol{\beta},t_1)\widehat{\nu}_{ik}^{II}(\boldsymbol{\beta},t_2) + \frac{1-\widetilde{\alpha}}{\widetilde{\alpha}}n^{-1}\sum_{i=1}^n\frac{\xi_i}{\widetilde{\alpha}}\widehat{\psi}_{ij}^{II}(\boldsymbol{\beta},t_1)\widehat{\psi}_{ik}^{II}(\boldsymbol{\beta},t_2), \\
&\widehat{\nu}_{ik}^{II}(\boldsymbol{\beta},t) = \widehat{\boldsymbol{r}}_k^{II}(t)^T\left(\widehat{\boldsymbol{A}}^{II}\right)^{-1}\sum_{k=1}^K\widehat{\mathbf{M}}_{\widetilde{\boldsymbol{z}},ik}^{II}(\boldsymbol{\beta}) + \int_0^t\{\widehat{\mathrm{E}}Y_{1k}(u)\}^{-1}d\widehat{M}_{ik}^{II}(u), \\
&\widehat{\psi}_{ik}^{II}(\boldsymbol{\beta},t) = \left[\widehat{\boldsymbol{r}}_k^{II}(t)^T\left(\widehat{\boldsymbol{A}}^{II}\right)^{-1}\sum_{m=1}^K(1-\Delta_{im})\int_0^\tau\left\{\widehat{\boldsymbol{R}}_{im}^{II}(\boldsymbol{\beta},u)\right.\right. \\
-\ &\left.\frac{Y_{im}(u)\widehat{\mathrm{E}}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta},u)\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right\}du + (1-\Delta_{ik})\int_0^t Y_{ik}(u) \\
\times\ &\left.\left(\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u) - \frac{\widehat{\mathrm{E}}\left((1-\Delta_{1k})\,Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right)}{\widehat{\mathrm{E}}\left((1-\Delta_{1k})\,Y_{1k}(u)\right)}\right)\frac{du}{\widehat{\mathrm{E}}Y_{1k}(u)}\right], \\
&\widehat{\mathrm{E}}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right) \ = \ n^{-1}\sum_{i=1}^n\frac{\xi_i}{\widetilde{\alpha}}(1-\Delta_{ik})Y_{ik}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u), \\
&\widehat{\mathrm{E}}Y_{1k}(u) = n^{-1}\sum_{i=1}^n Y_{ik}(u) \ \text{ and } \widehat{\boldsymbol{r}}_k^{II}(t) \ = \ -\int_0^\tau\overline{\boldsymbol{Z}}_k^\omega(t)dt.
\end{aligned}$$

The proofs of the theorems are outlined in the appendix.

5.4 Simulations

Extensive simulation studies were conducted to investigate the finite sample properties of the proposed methods. Failure times were generated from a multivariate extension of the model of Clayton and Cuzick model (Clayton and Cuzick, 1985). The joint survival function for $(T_1, \cdots, T_K)$ given $(Z_1, \cdots, Z_K)$ is:

$$S(t_1, \cdots, t_K | Z_1, \cdots, Z_K) = \left[\sum_{k=1}^{K} \exp\left\{ \frac{\int_0^{t_k} \left(\lambda_{0k}(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_k(t) \right) dt}{\theta} \right\} - (K-1) \right]^{-\theta},$$

where K takes integer values and $\theta(>0)$ is a parameter which represents the degree of dependence of T_k and $T_{k'}(k, k' = 1, \cdots, K)$. Note that smaller θ induces larger correlation. We considered two types of events (K=2). λ_{0k} was set to be equal to 2 for $k = 1$ and 4 for $k = 2$. Covariates considered were Bernoulli with probability 0.5 and Uniform $(0, 3)$. We examined regression parameters at $\beta = 0$ or $\log(2)$ and considered four different values for θ (0.1, 0.8, 1.25 or 4). The censoring time distribution were generated from uniform distribution $[0, u]$ with u chosen to depend on the desired percentage of censoring. We considered 97%, 90%, and 75% censoring. For each configuration, we simulated full cohort samples of size $n = 1000$ and then selected two case-cohort samples from each full cohort data. The sampling was conducted via simple random sampling with fixed sample size. The size of the random subcohort $\tilde{n}$ was set to have either the same expected number of controls and cases or twice as many controls as cases. For each data configuration, we ran $R = 2,000$ simulations.

Tables 5.1 and 5.2 show simulation summary statistics with Bernoulli covariate Z_{ik} with $\Pr(Z_{ik} = 1) = 0.5$ for $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$, respectively. "mean $\widehat{\boldsymbol{\beta}}_I$" or "mean $\widehat{\boldsymbol{\beta}}_{II}$" denotes the average of the estimates of $\boldsymbol{\beta}_0$, "proposed S.E." denotes the average of the estimates of standard errors based on the proposed method, "true S.D." denotes the sample standard deviation of the 2,000 estimates, and "95% C.I." denotes the coverage rate of the nominal 95% confidence interval. The simulation results suggest that the coefficient estimates were approximately unbiased across the setups considered for $\beta = 0$, while a substantial overestimation of the coefficients arised (up to 15 %) for $\beta = log(2)$ with small event proportion (3%). However, as the subcohort sample size increased to $\tilde{n} = 62$, the results improved. The proposed estimated standard errors appeared to closely approximate the true variabilities of $\hat{\boldsymbol{\beta}}$s in most of the cases. The coverage rate of the nominal 95% confidence intervals using the proposed method were in the 92.4% - 95.8% range. Overall, $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ performed reasonably well and showed similar results. For all data configuration, the true variabilities of the regression parameter estimates for $\widehat{\beta}_{II}$ were smaller than those for $\widehat{\beta}_I$, however, the discrepancies were not very large. Tables 5.3 and 5.4 provide simulation summary statistics for $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ with the Uniform covariate, respectively. In general, the findings were similar to those of Tables 5.1 and 5.2.

5.5 Analysis of Busselton Health Study

We applied the proposed methods to a subset of the data from the Busselton Health Study (Cullen, 1972; Knuiman et al., 2003). The population of this study was based on the 1,612 men and women aged 40-89 years who participated in the 1981 Busselton Health Survey and had no history of diagnosed CHD or stroke at that time. This group of people was then followed for both CHD or stroke and the follow-up continued through the date of first CHD event and the date of first stroke event or December 31, 1998, whichever comes first. The follow-up also ended when the subjects left Western Australia during this period.

In this analysis, our primary interest was on the assessment and the comparison of the effect of body iron stores on the risk of CHD and stroke. Body iron stores were proposed to be positively related to coronary heart disease risk (Sullivan, 1996). However, the accumulated epidemiologic evidence has been inconsistent and it was of interest to examine this hypothesis in this population. As a measure of body iron store, serum ferritin was used where serum ferritin is regarded as the best biochemical measure of body iron store. In addition, we were also interested in whether the effect of serum ferritin on the risk of CHD and stroke differed by gender.

A case-cohort design was conducted to reduce costs and preserve stored serum in the Busselton Health Study. Our analysis was based on a subset of the Busselton Health Study.

There were 1,212 cohort members with 217 CHD cases and 118 stroke cases. The subcohort size was 360. Ferritin assays were conducted for all the cases and subcohort members. Because of overlap between CHD/stroke cases and the random subcohort, the total number of assayed sera samples was 536.

The risk difference is one common measure of risk in epidemiology. It directly quantifies the effect of serum ferritin level on the risk of CHD and stroke. Thus, in this analysis, we considered the additive hazards model, which models the risk difference. We first analysed the data by fitting the marginal additive hazards model with distinct serum ferritin effect level on the risk of CHD and of stroke. To control for confounding factors, several cardiovascular risk factors were included in the model as covariates. The risk factors included were age (years), blood pressure treatment, systolic blood pressure (mmHg), BMI, cholesterol (mmol/liter), triglycerides (mmol/liter), diabetes treatment, hemoglobin (g/100 ml), and smoking (never, former, current). For ferritin, the log of the serum ferritin level was used.

Table 5.5 shows the additive hazards estimates, standard errors of the estimates and the associated 95 % confidence intervals. Note that the values presented in this table are 10^2 times the original values for convenience. This is also the case for Tables 5.6 and 5.7 later. As shown in the table, the additive hazards estimates for log of ferritin levels on both CHD and stroke were not statistically significant at the level of $\alpha = 0.05$. A Wald-type test for a common ferritin effect on CHD and stroke showed a weak evidence of different ferritin effect on CHD and stroke(test statistics = 0.7670, p-value = 0.38). Thus, we assumed a common ferritin effect on CHD and stroke and refit the model. The results are presented in 5.6 and show weak evidence of the effect of ferritin level on the risk of CHD and stroke.

We also considered a model with gender-specific serum ferritin effect on the risk of CHD and stroke. Table 5.7 provides the results from this model. The results indicate that, both for men and women, no statistically significant effect of ferritin level on the risk of CHD and stroke could be found. This was also true after we assumed common effect of ferritin level on the risk of CHD and stroke for both men and women. Therefore, there was no obvious gender specific or overall effect of serum ferritin on the risk of CHD and stroke.

5.6 Concluding remarks

We have proposed methods of fitting marginal additive hazard regression models for case-cohort studies with multiple disease outcomes. The regression parameter estimates have a closed form and thus can be obtained from the weighted estimating equations without employing any numerical methods. A Breslow-Aalen type estimator was proposed for the cumulative baseline hazard functions. The proposed estimator were shown to have desirable asymptotic properties and to perform well under the practical sample sizes considered. The proposed procedures could be naturally extended to stratified case-cohort studies or Bernoulli sampling of the subcohort.

5.7 Proofs of the theorems

Outline of the Proofs of Theorem 5.1 - 5.4

We assume the following set of conditions hold :

(A) $(\boldsymbol{T}_i, \boldsymbol{C}_i, \boldsymbol{Z}_i), i = 1, \ldots, n$ are independent and identically distributed.

(B) $\Pr(Y(\tau) > 0) > 0$.

(C) $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(u)| < C_z < \infty$ almost surely for some constant C_z.

(D) The matrix $\boldsymbol{A}_k = \mathrm{E}\left\{\int_0^\tau Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right) dt\right\}$ is positive definite.

(E) $\int_0^\tau \lambda_{0k}(t)dt < \infty$, for all $k = 1, \ldots, K$.

(F) As $n \to \infty$, $\tilde{\alpha} = \frac{\tilde{n}}{n}$ converges to a constant $\alpha \in (0, 1)$.

(J) $\frac{n_s}{n}$ converges to a constant $p_s \in [0, 1]$ for $s = 0, 1$ as $n \to \infty$ where $p_1 + p_0 = 1$.

The following lemma along with the lemmas in the previous two chapters will be frequently used in proving the theorems.

Lemma 6 *Let $B_i(t)$, $i = 1, \ldots, n$, be i.i.d. real-valued random processes on $[0, \tau]$ with $\mathrm{E}\{B_i(t)\} = \mu_B(t)$, $\mathrm{Var}\{B_i(0)\} < \infty$ and $\mathrm{Var}\{B_i(\tau)\} < \infty$. Suppose that almost all paths of $B_i(t)$ have finite variation. Then,*

$$n^{-1/2} \sum_{i=1}^{n} \{B_i(t) - \mu_B(t)\}$$

converges weakly in $\ell^\infty[0, \tau]$ to a zero-mean Gaussian process and therefore

$$n^{-1} \sum_{i=1}^{n} \{B_i(t) - \mu_B(t)\}$$

converges in probability to 0 uniformly in t.

This lemma is given as the proposition in Kulich and Lin (2004).

Before we move onto the proofs of the theorems, we first investigate the asymptotic properties of the time-varying sampling probability estimator $\widehat{\alpha}_k(t) = \frac{\sum_{i=1}^n \xi_i Y_{ik}(t)}{\sum_{i=1}^n Y_{ik}(t)}$. These asymptotic properties will be frequently used in proving the theorems.

For each k, it follows from the Taylor expansion of $\widehat{\alpha}_k(t)^{-1}$ around $\widetilde{\alpha}$,

$$\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1} = -\frac{1}{\alpha_*(t)^2} \left(\widehat{\alpha}_k(t) - \widetilde{\alpha}\right) = \frac{\widetilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{1}{\sum_{i=1}^n Y_{ik}(t)} \left\{ \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(t) \right\}$$

where $\alpha_*(t)$ is on the line segment between $\widehat{\alpha}_k(t)$ and $\widetilde{\alpha}$. Then,

$$n^{1/2} \left(\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}\right) = \frac{\widetilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{n}{\sum_{i=1}^n Y_{ik}(t)} n^{-1/2} \left\{ \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(t) \right\}$$

$n^{-1} \sum_{i=1}^n Y_{ik}(t)$ converges to $\mathrm{E}\, Y_{1k}(t)$ in probability uniformly in t by lemma 6 since $Y_{ik}(t)$ is bounded and monotone in t, where $\mathrm{E}\, Y_{1k}(t)$ is bounded away from 0 by condition (B). In view of lemma 5, $n^{-1} \sum_{i=1}^n (\frac{\xi_i}{\widetilde{\alpha}} - 1) Y_{ik}(t)$ converges to 0 in probability uniformly in t since, again, $Y_{ik}(t)$ is bounded and monotone function in t. Consequently, $\widehat{\alpha}_k(t) - \widetilde{\alpha} = \frac{\widetilde{\alpha} n^{-1} \sum_{i=1}^n \left(\frac{\xi_i}{\widetilde{\alpha}} - 1\right) Y_{ik}(t)}{n^{-1} \sum_{i=1}^n Y_{ik}(t)}$ converges to 0 in probability uniformly in t. Hence, $\widehat{\alpha}_k(t)$ and $\widetilde{\alpha}$ converge to the same limit in probability uniformly in t. This ensures $\alpha_*(t)$ also converges to the same limit as $\widetilde{\alpha}$.

Combining these results, it follows from Slutsky's theorem that

$$
\begin{aligned}
n^{1/2}\left(\widehat{\alpha}_k(t)^{-1}-\widetilde{\alpha}^{-1}\right) &= \frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}n^{-1/2}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\right\} \\
+ & \left(\frac{\widetilde{\alpha}}{\alpha_*(t)^2}\cdot\frac{n}{\sum_{i=1}^{n}Y_{ik}(t)}-\frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}\right)n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t) \\
= & \frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}n^{-1/2}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t)\right\}+o_p(1) \qquad (5.6)
\end{aligned}
$$

Now we prove theorem 5.1.

Proof of theorem 5.1 We first consider the proof for the consistency of $\widehat{\boldsymbol{\beta}}_I$. Denote n^{-1} times $\boldsymbol{U}^I(\boldsymbol{\beta})$ by $\boldsymbol{U}_n^I(\boldsymbol{\beta})$. Based on a straightforward extension of Foutz (1977), one can show $\widehat{\boldsymbol{\beta}}_I$ to be consistent for $\boldsymbol{\beta}_0$ provided: (i) $\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$, (ii)$\partial\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0^T$ is negative definite with probability going to one as $n\to\infty$, (iii) $\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ converges to $\boldsymbol{A}$ in probability uniformly for $\boldsymbol{\beta}$ in an open neighborhood about $\boldsymbol{\beta}_0$, and (iv) $\boldsymbol{U}_n^I(\boldsymbol{\beta})\to 0$ in probability.

One can write

$$
\begin{aligned}
-\frac{\partial\boldsymbol{U}_n^I(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T} &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^T dt \\
&= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\rho_{ik}(t)Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2}-\overline{\boldsymbol{Z}}_k^\rho(t)^{\otimes 2}\right)dt \qquad (5.7)
\end{aligned}
$$

Then, (i) is clearly satisfied on the basis of (5.7).

To verify conditions (ii) and (iii), we will first show that

$$
\sup_{t\in[0,\tau]}\left\|\overline{\boldsymbol{Z}}_k^\rho(t)-\boldsymbol{e}_k(t)\right\|\xrightarrow{p}0 \text{ as } n\to\infty \text{ for } k=1,\ldots,K.
$$

It suffices to show that

$\sup_{t\in[0,\tau]}\|n^{-1}\sum_{i=1}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}-n^{-1}\sum_{i=1}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\|\xrightarrow{p}0$ as $n\to\infty$ for

$d = 0, 1$. One can write

$$\begin{aligned} & n^{-1}\sum_{i=1}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} - n^{-1}\sum_{i=1}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} \\ = \quad & n^{-1}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}} - 1\right)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} - n^{-1}\sum_{i=1}^{n}(\widetilde{\alpha}^{-1} - \widehat{\alpha}_k(t)^{-1})\xi_i Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}. \end{aligned}$$

Then,

$$\begin{aligned} & \left\| n^{-1}\sum_{i=1}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} - n^{-1}\sum_{i=1}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} \right\| \\ \leq & \left\| n^{-1}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}} - 1\right)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} \right\| + \left|(\widetilde{\alpha}^{-1} - \widehat{\alpha}_k(t)^{-1})\right| n^{-1}\sum_{i=1}^{n}\xi_i Y_{ik}(t)\left\|\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\| \end{aligned} \quad (5.8)$$

For each $j (j = 1, \ldots, p)$, the total variation of $Y_{ik}(t)Z_{ikj}(t)^{\otimes d}$ is finite on $[0, \tau]$ by condition (C). Thus, by lemma 5, the first term on the right-hand side of (5.8) converges to 0 in probability uniformly in t. The second term on the right-hand side of (5.8) also converges to 0 in probability uniformly in t since $\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}$ was shown to converge to 0 in probability uniformly in t and $n^{-1}\sum_{i=1}^{n}\xi_i Y_{ik}(t)\left\|\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\|$ converges to a finite quantity $\alpha\,\mathrm{E}(Y_{1k}(t)\|\boldsymbol{Z}_{1k}(t)^{\otimes d}\|)$ in probability uniformly in t by lemma 5. Combining these results, $n^{-1}\sum_{i=1}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ and $n^{-1}\sum_{i=1}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ were shown to converge to the same limit uniformly. Note that $n^{-1}\sum_{i=1}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ converges to $\mathrm{E}(Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d})$ for $d = 0, 1$ by lemma 6 since $Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ is of bounded variation by condition (C). Therefore, we have

$$\sup_{t\in[0,\tau]}\left\| n^{-1}\sum_{i=1}\rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d} - \mathrm{E}(Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d})\right\| \xrightarrow{p} 0 \text{ as } n \to \infty \text{ for } d = 0, 1 \quad (5.9)$$

Since $\mathrm{E}\,Y_{1k}(t)$ is bounded away from zero by condition (B), it follows from the above convergence results that for $k = 1, \ldots, K$, $\overline{\boldsymbol{Z}}_k^{\rho}(t)$ converges to $\boldsymbol{e}_k(t)$ in probability uniformly in t as $n \to \infty$.

Now, (5.7) can be written as the followings:

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \rho_{ik}(t)Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \overline{\boldsymbol{Z}}_k^{\rho}(t)^{\otimes 2}\right) dt \\
= \; & n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \frac{\xi_i}{\widetilde{\alpha}} Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \overline{\boldsymbol{Z}}_k^{\rho}(t)^{\otimes 2}\right) dt \\
+ \; & n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widehat{\alpha}_k(t)^{-1} - \widetilde{\alpha}^{-1}\right)\xi_i \int_0^{\tau} Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \overline{\boldsymbol{Z}}_k^{\rho}(t)^{\otimes 2}\right) dt \qquad (5.10)
\end{aligned}
$$

Then, by the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\rho}(t)$ to $\boldsymbol{e}_k(t)$, the first term on the right-hand side of (5.10) is asymptotically equivalent to

$$
n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \frac{\xi_i}{\widetilde{\alpha}} Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right) dt
$$

Based on the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\rho}(t)$ to $\boldsymbol{e}_k(t)$, $\widehat{\alpha}_k(t)^{-1}$ to $\widetilde{\alpha}^{-1}$ and lemma 5, the second term on the right-hand side of (5.10) converges to 0 in probability uniformly in t. Thus, by combining these results, we have

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \rho_{ik}(t)Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \overline{\boldsymbol{Z}}_k^{\rho}(t)^{\otimes 2}\right) dt \\
= \; & n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} \frac{\xi_i}{\widetilde{\alpha}} Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right) dt \; + \; o_p(1) \qquad (5.11)
\end{aligned}
$$

Since $Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right)$ is of bounded variation by condition (C) and $Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right)$'s are independent and identically distributed, it follows from lemma 5 that $n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\frac{\xi_i}{\widetilde{\alpha}}Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2}\right) dt$ converges to $\mathrm{E}\left[\sum_{k=1}^{K}\int_0^{\tau}\left\{Y_{1k}(t)(\boldsymbol{Z}_{1k}(t)^{\otimes 2} - \boldsymbol{e}_k(t)^{\otimes 2})\right\} dt\right]$ in probability as $n \to \infty$. Hence,

$$
-\frac{\partial \boldsymbol{U}_n^I(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \xrightarrow{p} \boldsymbol{A} \text{ as } n \to \infty \qquad (5.12)
$$

and, thus, (ii) and (iii) are satisfied.

For (iv), we will show that $n^{-1/2}\boldsymbol{U}^I(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0) \;+\; n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\frac{\xi_i}{\widetilde{\alpha}})$$
$$\times \quad \int_0^\tau \left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t) - \frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt.$$

Specifically, one can decompose $n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)$ into the followings :

$$\begin{aligned}
n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0) &= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)\\
&\times \left\{dM_{ik}(t)+Y_{ik}(t)\lambda_{0k}(t)dt+(1-\rho_{ik}(t))\,Y_{ik}(t)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)dt\right\}\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)dM_{ik}(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)\rho_{ik}(t)Y_{ik}(t)\lambda_{0k}(t)dt\\
&+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau(1-\rho_{ik}(t))\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)\\
&\times Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt
\end{aligned}\qquad(5.13)$$

The second term on the right-hand side of (5.13) equals to 0. The first term on the right-hand side of (5.13) can be further decomposed into the following two parts:

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left\{\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right\}dM_{ik}(t)\\
=\; & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau\left\{\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k(t)\right\}dM_{ik}(t)\\
+\; & \sum_{k=1}^{K}\int_0^\tau\left\{\overline{\boldsymbol{Z}}_k(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right\}d\left\{n^{-1/2}\sum_{i=1}^{n}M_{ik}(t)\right\}
\end{aligned}\qquad(5.14)$$

The first term on the right-hand side of (5.14) is the pseudo partial likelihood score function for the full cohort data. This was shown to be asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)$ (Yin and Cai, 2004). Note that, for fixed t, $n^{-1/2}\sum_{i=1}^{n}M_{ik}(t)$

is a sum of i.i.d. zero-mean random variables. Based on conditions (C) and (E), $M_{ik}(t)$ is of bounded variation and therefore, it follows from lemma 6 that $n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{W}_{Mk}(t)$. It can be shown that $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^4 \le C\{\Lambda_{0k}(t) - \Lambda_{0k}(s)\}^2$ for some constant $C > 0$. Specifically, $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^4 = 3(\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2)^2$ since $\mathcal{W}_{Mk}(t)$ is a zero-mean normal random variable for a fixed t. Then $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2 = \mathrm{E}\,\mathcal{W}_{Mk}(t)^2 + \mathrm{E}\,\mathcal{W}_{Mk}(s)^2 - 2\,\mathrm{E}\,\mathcal{W}_{Mk}(t)\mathcal{W}_{Mk}(s) = \mathrm{E}\,\mathcal{W}_{Mk}(t)^2 - \mathrm{E}\,\mathcal{W}_{Mk}(s)^2$ for $s \le t$. Since $\mathrm{E}\,\mathcal{W}_{Mk}(t)^2 = \mathrm{E}\,M_{ik}(t)^2 =$ $\mathrm{E}\left[\int_0^t Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du\right]$, $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2 =$ $\mathrm{E}\left[\int_s^t Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du\right]$. Note that the conditions (C) and (E) ensure the boundedness of $\lambda_{0k}(\cdot)$ and $\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(\cdot)$ on $[0,\tau]$. Thus, by mean value theorem, there exists a constant M, such that $\mathrm{E}\left[\int_s^t Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du\right] \le M(t-s)$ for $s \le t$. Therefore, $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2 \le M(t-s)$ and $\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^4 \le 3(\mathrm{E}\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2)^2 \le \widetilde{M}(t-s)^2$ for some constant $\widetilde{M}$. Then, by the Kolmogorov-Centsov Theorem (Karatzas and Shereve, 1988, p53), $\mathcal{W}_{Mk}(t)$ has continuous sample paths. In addition, since $n^{-1}\sum_{i=1}^{n} Y_{ik}(t)\boldsymbol{Z}_{ik}(t)$ and $n^{-1}\sum_{i=1}^{n} Y_{ik}(t)$ are of bounded variations and $n^{-1}\sum_{i=1}^{n} Y_{ik}(t)$ is bounded away from 0, based on conditions (B) and (C), $\overline{\boldsymbol{Z}}_k(t)$ is of bounded variation and can be written as a sum of two monotone functions in t, respectively. Specifically, $\overline{\boldsymbol{Z}}_k(t) = \boldsymbol{Z}_{k1}^*(t) - \boldsymbol{Z}_{k2}^*(t)$ where both $\boldsymbol{Z}_{k1}^*(t)$ and $\boldsymbol{Z}_{k2}^*(t)$ are nonnegative, monotone in t and bounded. Since $\overline{\boldsymbol{Z}}_k^{\rho}(t)$ is also of bounded variation based on (5.9) and conditions (B) and (C), by the same argument, we can write $\overline{\boldsymbol{Z}}_k^{\rho}(t) = \boldsymbol{Z}_{k1}^{**}(t) - \boldsymbol{Z}_{k2}^{**}(t)$ where both $\boldsymbol{Z}_{k1}^{**}(t)$ and $\boldsymbol{Z}_{k2}^{**}(t)$ are nonnegative, monotone in t and bounded. Hence, it follows from lemma 2 that

$$\sum_{k=1}^{K}\int_0^{\tau}\left\{\overline{\boldsymbol{Z}}_k(t) - \overline{\boldsymbol{Z}}_k^{\rho}(t)\right\} d\left\{n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)\right\} = \sum_{k=1}^{K}\int_0^{\tau}\left\{\overline{\boldsymbol{Z}}_k(t) - \boldsymbol{e}_k(t)\right\} d\left\{n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)\right\}$$
$$- \sum_{k=1}^{K}\int_0^{\tau}\left\{\overline{\boldsymbol{Z}}_k^{\rho}(t) - \boldsymbol{e}_k(t)\right\} d\left\{n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)\right\} \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Thus, the second term on the right-hand side of (5.14) converges to 0 in probability uniformly in t.

Now, one can write the third term on the right-hand side of (5.13) as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau (1-\rho_{ik}(t))\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt \\
&+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{-1}(t)\right)\xi_i\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt \qquad (5.15)
\end{aligned}
$$

It follows from the uniform convergence of $\overline{\boldsymbol{Z}}_k^\rho(t)$ to $\boldsymbol{e}_k(t)$ and the boundedness of $\Lambda_{0k}(\tau)$ that the first term on the right-hand side of (5.15) is asymptotically equivalent to

$$
n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt.
$$

Based on (5.6) and the uniform convergence of $\overline{\boldsymbol{Z}}_k^\rho(t)$ to $\boldsymbol{e}_k(t)$, the second term on the right-hand side of (5.15) is

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{-1}(t)\right)\xi_i\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^\rho(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt \\
= \;& n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}\,Y_{1k}(t)}n^{-1/2}\sum_{j=1}^{n}\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right)Y_{jk}(t)\right\} \\
\times \;& \xi_i\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt+o_p(1) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^\tau \frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)} \\
\times \;& \left\{n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)\right)\right\}dt+o_p(1) \qquad (5.16)
\end{aligned}
$$

It follows from lemma 5 that $n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)$ and $n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)$ converge to $\mathrm{E}\left(\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)Y_{1k}(t)\right)$ and $\mathrm{E}\left(\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)Y_{1k}(t)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)$ in probability uniformly in t, respectively. Note that

$\mathrm{E}\left((\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t))\,Y_{1k}(t)\right)=\mathrm{E}(Y_{1k}(t)\boldsymbol{Z}_{1k}(t))-\boldsymbol{e}_k(t)\,\mathrm{E}\,Y_{1k}(t)=0$. Thus, from (5.16), we have

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^{\tau}\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)} \\
& \times\left\{n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)\right)\right\}dt \\
& = n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^{\tau}\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)dt+o_p(1)
\end{aligned}$$

Therefore, the third term on the right-hand side of (5.13) is asymptotically equal to

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right) \\
& \times\int_0^{\tau}\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt \qquad (5.17)
\end{aligned}$$

where

$$\boldsymbol{R}_{ik}(\boldsymbol{\beta},t)=Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)\left(\lambda_{0k}(t)+\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\right).$$

Combining the above results, we have shown that $n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\ +n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right) \\
& \times\int_0^{\tau}\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt \qquad (5.18)
\end{aligned}$$

Under the regularity conditions, the first term on the right-hand side of (5.18) is asymptotically zero-mean normal with covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0)=\mathrm{E}\left(\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},1k}(\boldsymbol{\beta}_0)\right)^{\otimes 2}$ by Yin and Cai (2004).

On the basis of conditions (C) and (F), the second term on the right-hand side of (5.18) is asymptotically zero-mean normal random variable with covariance matrix $\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)$ where

$$\boldsymbol{V}(\boldsymbol{\beta}_0)=\mathrm{E}\left[\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0,t)-\frac{Y_{1k}(t)}{\mathrm{E}\,Y_{1k}(t)}\,\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt\right]^{\otimes 2}.$$

This result follows from applying Hájek (1960)'s central limit theorem for finite population sampling to

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{a}^T\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt$$

and then applying Cramer-Wold device where $\boldsymbol{a}=(a_1,\ldots,a_p)^T$ is a $p\times 1$ real valued vector. Note that $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)$ and $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt$ are independent since

$$\begin{aligned}
& \mathrm{Cov}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0),\ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\right.\\
\times & \left.\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt\right)\\
= & \mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\right.\\
\times & \left.\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt\right\}\\
= & \mathrm{E}\left\{\mathrm{E}\left(n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\right.\right.\\
\times & \left.\left.\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt\,\middle|\,\mathcal{F}(\tau)\right)\right\}\\
= & \mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\sum_{k=1}^{K}\mathrm{E}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\,\middle|\,\mathcal{F}(\tau)\right)\right.\\
\times & \left.\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)}{\mathrm{E}\,Y_{1k}(t)}\mathrm{E}\left(Y_{1k}(t)\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt\right\}\\
= & 0.
\end{aligned}$$

Therefore, $n^{1/2}\boldsymbol{U}_n^I(\boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with variance $\boldsymbol{Q}(\boldsymbol{\beta}_0)+\frac{1-\alpha}{\alpha}\boldsymbol{V}(\boldsymbol{\beta}_0)$. Hence $\boldsymbol{U}_n(\boldsymbol{\beta})$ converges to zero in probability. Thus, (iv) is satisfied. By (i),(ii),(iii) and (iv), it follows that there is a unique sequence $\widehat{\boldsymbol{\beta}}_I$ *s.t.* $\boldsymbol{U}^I(\widehat{\boldsymbol{\beta}}_I)=0$ with probability converging to one as $n\to 0$ and with $\widehat{\boldsymbol{\beta}}_I$ converging in probability to $\boldsymbol{\beta}_0$ by an extension of Foutz (1977, Thm.2).

The asymptotic normality of $\widehat{\boldsymbol{\beta}}_I$ follows from the consistency of $\widehat{\boldsymbol{\beta}}_I$ and a Taylor series expansion of $\boldsymbol{U}^I(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$.

Proof of Theorem 5.2 One can make decomposition

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}^I_{0k}(\widehat{\boldsymbol{\beta}}_I, t) - \Lambda_{0k}(t)\} \\
&= n^{1/2}\left\{\widehat{\Lambda}^I_{0k}(\widehat{\boldsymbol{\beta}}_I, t) - \widehat{\Lambda}^I_{0k}(\boldsymbol{\beta}_0, t)\right\} \ + \ n^{1/2}\left\{\widehat{\Lambda}^I_{0k}(\boldsymbol{\beta}_0, t) - \Lambda_{0k}(t)\right\} \\
&= n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \left(dN_{ik}(u) - \rho_{ik}(u)Y_{ik}(u)\widehat{\boldsymbol{\beta}}_I^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du \right. \\
&- \left. \int_0^t \frac{\sum_{i=1}^n \left(dN_{ik}(u) - \rho_{ik}(u)Y_{ik}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du \right\} \\
&+ n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \left(dN_{ik}(u) - \rho_{ik}(u)Y_{ik}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du \right\} \\
&- n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)\lambda_{0k}(u)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du \right\} \\
&= n^{1/2}\int_0^t \frac{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I)^T \boldsymbol{Z}_{ik}(u)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du + n^{1/2}\int_0^t \frac{\sum_{i=1}^n dM_{ik}(u)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} \\
&+ n^{1/2}\int_0^t \frac{\sum_{i=1}^n \left(1 - \rho_{ik}(u)\right) Y_{ik}(u) \left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} du
\end{aligned} \tag{5.19}
$$

By the uniform convergence of $\overline{\boldsymbol{Z}}^{\rho}_k(u)$ to $\boldsymbol{e}_k(u)$, the first term of (5.19) is asymptotically equivalent to $n^{1/2}\boldsymbol{r}_k(t)^T\left(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\right)$ where $\boldsymbol{r}_k(\boldsymbol{\beta}, t) = -\int_0^t \boldsymbol{e}_k(u)du$. Since $\left(n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{1k}(u)\right)^{-1}$ can be written a sum of two monotone functions in t and converges uniformly to $\{\mathrm{E}\left(Y_{1k}(u)\right)\}^{-1}$, where $\mathrm{E}\left(Y_{1k}(u)\right)$ is bounded away from 0, and $n^{-1/2}\sum_{i=1}^n M_{ik}(u)$ converges to a zero-mean Gaussian process with continuous sample paths, it follows from lemma 2 that the second term on the right-hand side of (5.19) is asymptotically equivalent to

$$
\int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\}
$$

The last term on the right-hand side of (5.19) can be written as

$$
\begin{aligned}
& n^{1/2}\int_0^t \frac{\sum_{i=1}^n (1-\rho_{ik}(u))\, Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \rho_{ik}(u) Y_{ik}(u)} du \\
= & \int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} n^{-1/2}\sum_{i=1}^n \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du \qquad (5.20)\\
+ & \int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} n^{-1/2}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k(t)^{-1}\right)\sum_{i=1}^n \xi_i Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du
\end{aligned}
$$

It follows from the uniform convergence of $\left\{n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{1k}(u)\right\}^{-1}$ to $\{\mathrm{E}(Y_{1k}(u))\}^{-1}$, where $\mathrm{E}(Y_{1k}(u))$ is bounded away from 0 and the boundedness of $\Lambda_{0k}(u)$ on $[0,t]$ that the first term on the right-hand side of (5.20) is asymptotically equivalent to

$$
\int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} n^{-1/2}\sum_{i=1}^n \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right) Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du. \qquad (5.21)
$$

Based on (5.6), the uniform convergence of $\left\{n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{1k}(u)\right\}^{-1}$ to $\{\mathrm{E}(Y_{1k}(u))\}^{-1}$, $n^{-1}\sum_{i=1}^n \xi_i Y_{ik}(u)$ to $\mathrm{E}\, Y_{1k}(u)$ and $n^{-1}\sum_{i=1}^n \xi_i Y_{ik}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)$ to $\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)$, the second term on the right-hand side of (5.20) is

$$
\begin{aligned}
& \int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)} n^{-1/2}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k(t)^{-1}\right)\sum_{i=1}^n \xi_i Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du \\
= & \int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)}\left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}\, Y_{1k}(u)} n^{-1/2}\sum_{j=1}^n\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right) Y_{jk}(t)\right\} \\
\times & \; n^{-1}\sum_{i=1}^n \xi_i Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du + o_p(1) \\
= & \; n^{-1/2}\sum_{i=1}^n\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^t \frac{Y_{ik}(u)}{\left(\mathrm{E}\left(Y_{1k}(u)\right)\right)^2}\left\{\mathrm{E}(Y_{1k}(u))\lambda_{0k}(u)+\mathrm{E}(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u))\right\} du + o_p(1)
\end{aligned}
\qquad (5.22)
$$

Thus, it follows from (5.21) and (5.22) that the last term on the right-hand side of (5.19) is equivalent to

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^t Y_{ik}(u) \left\{ \left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) \right. \\
- & \left. \left(\lambda_{0k}(u) + \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\, Y_{1k}(u)}\right) \right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1) \\
= & \, n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^t Y_{ik}(u) \left\{ \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\, Y_{1k}(u)} \right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1)
\end{aligned}
$$

Based on Taylor expansion of $\boldsymbol{U}^I(\widehat{\boldsymbol{\beta}}_I)$ around $\boldsymbol{\beta}_0$ and the results in (5.12) and (5.18), we have

$$
\begin{aligned}
& n^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0) \\
= \; & \boldsymbol{A}^{-1} \left\{ n^{-1/2} \sum_{i=1}^{n} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0) + n^{-1/2} \sum_{i=1}^{n} \sum_{m=1}^{K} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \right. \\
\times \; & \left. \int_0^\tau \left(\boldsymbol{R}_{im}(\boldsymbol{\beta}_0, t) - \frac{Y_{im}(t)}{\mathrm{E}\, Y_{1m}(t)} \mathrm{E}\left(Y_{1m}(t)(\boldsymbol{Z}_{1m}(t) - \boldsymbol{e}_m(t))\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1m}(t)\right) \right) \right\} du + o_p(1).
\end{aligned}
$$

Combining the above results, we have

$$
\begin{aligned}
& n^{1/2}(\widehat{\Lambda}_{0k}^I(\widehat{\boldsymbol{\beta}}_I, t) - \Lambda_{0k}(t)) \\
= & \, \boldsymbol{r}_k(t)^T \boldsymbol{A}^{-1} \left\{ n^{-1/2} \sum_{i=1}^{n} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0) + n^{-1/2} \sum_{i=1}^{n} \sum_{m=1}^{K} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^\tau (\boldsymbol{R}_{im}(\boldsymbol{\beta}_0, u) \right. \\
- & \left. \frac{Y_{im}(u)}{\mathrm{E}\, Y_{1m}(u)} \mathrm{E}\left(Y_{1m}(u)\left(\boldsymbol{Z}_{1m}(u) - \boldsymbol{e}_m(u)\right)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1m}(u)\right) \Big) du \right\} \\
+ & \int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} d\left\{ n^{-1/2} \sum_{i=1}^{n} M_{ik}(u) \right\} \\
+ & \, n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \int_0^t Y_{ik}(u) \left\{ \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\, Y_{1k}(u)} \right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} \\
= & \, n^{-1/2} \sum_{i=1}^{n} \left[\left\{ \boldsymbol{r}_k(t)^T \boldsymbol{A}^{-1} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0) + \int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} dM_{ik}(u) \right\} \right. \\
+ & \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \left\{ \boldsymbol{r}_k(t)^T \boldsymbol{A}^{-1} \sum_{m=1}^{K} \int_0^\tau (\boldsymbol{R}_{im}(\boldsymbol{\beta}_0, u) \right. \\
- & \frac{Y_{im}(u)}{\mathrm{E}\, Y_{1m}(u)} \mathrm{E}\left(Y_{1m}(u)\left(\boldsymbol{Z}_{1m}(u) - \boldsymbol{e}_m(u)\right)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1m}(u)\right) \Big) du
\end{aligned}
$$

$$+ \int_0^t Y_{ik}(u) \left\{ \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\, Y_{1k}(u)} \right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} \Bigg\}\Bigg] + o_p(1) \tag{5.23}$$

$$= n^{-1/2} \sum_{i=1}^{n} \nu_{ik}(\boldsymbol{\beta}_0, t) + n^{-1/2} \sum_{i=1}^{n} \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \psi_{ik}(\boldsymbol{\beta}_0, t) + o_p(1)$$

where

$$\begin{aligned} \nu_{ik}(\boldsymbol{\beta}, t) =& \ \boldsymbol{r}_k(t)^T \boldsymbol{A}^{-1} \sum_{m=1}^{K} \mathbf{M}_{\tilde{\boldsymbol{z}}, im}(\boldsymbol{\beta}) + \int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} dM_{ik}(u) \quad \text{and} \\ \psi_{ik}(\boldsymbol{\beta}, t) =& \ \boldsymbol{r}_k(t)^T \boldsymbol{A}^{-1} \sum_{m=1}^{K} \int_0^\tau \Big(\boldsymbol{R}_{im}(\boldsymbol{\beta}, u) \\ & - \frac{Y_{im}(u)}{\mathrm{E}\, Y_{1m}(u)} \mathrm{E}\left(Y_{1m}(u) \left(\boldsymbol{Z}_{1m}(u) - \boldsymbol{e}_m(u)\right) \boldsymbol{\beta}^T \boldsymbol{Z}_{1m}(u)\right) \Big) du \\ & + \int_0^t \left\{ \boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left(Y_{1k}(u)\boldsymbol{\beta}^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\, Y_{1k}(u)} \right\} \frac{Y_{ik}(u) du}{\mathrm{E}\, Y_{1k}(u)}. \end{aligned}$$

Now, let $\boldsymbol{W}^{(1)}(t) = (W_1^{(1)}(t), \ldots, W_K^{(1)}(t))^T$ where $W_k^{(1)}(t) = n^{-1/2} \sum_{i=1}^n \nu_{ik}(\boldsymbol{\beta}_0, t)$ and $\boldsymbol{W}^{(2)}(t) = (W_1^{(2)}(t), \ldots, W_K^{(2)}(t))^T$ where $W_k^{(2)}(t) = n^{-1/2} \sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right) \psi_{ik}(\boldsymbol{\beta}_0, t)$ for $k = 1, \ldots, K$. Then, $\boldsymbol{W}^{(1)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}^{(1)}(t) = (\mathcal{W}_1^{(1)}(t), \ldots, \mathcal{W}_K^{(1)}(t))^T$ in $D[0,\tau]^K$ where the covariance function between $\mathcal{W}_j^{(1)}(t_1)$ and $\mathcal{W}_k^{(1)}(t_2)$ is $\mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0, t_1)\nu_{1k}(\boldsymbol{\beta}_0, t_2)\}$ by Yin and Cai (2004, Thm.2). $\boldsymbol{W}^{(2)}(t)$ also can be shown to converge weakly to a zero-mean Gaussian process $\mathcal{W}^{(2)}(t) = (\mathcal{W}_1^{(2)}(t), \ldots, \mathcal{W}_K^{(2)}(t))^T$. Specifically, $\psi_{ik}(\boldsymbol{\beta}_0, t)$ is of bounded variation since $\boldsymbol{r}_k(t)$, $Y_{ik}(t)Z_{ikj}(t)$ and $\mathrm{E}\, Y_{1k}(t)$ are of bounded variations, $\mathrm{E}\, Y_{ik}(t)$ is bounded away from zero, and $\boldsymbol{A}$ is positive definite based on conditions (B), (C) and (D). Thus, for any finite number of time points $(t_1, \ldots, t_L)$, the finite dimensional distribution of $\boldsymbol{W}^{(2)}(t)$ is asymptotically the same as those of $\mathcal{W}^{(2)}(t)$ by lemma 5 and Cramer-Wold device. Now, if we show the tightness of $\boldsymbol{W}^{(2)}(t)$, the proof for the weak convergence is completed. Since the space $D[0,\tau]^K$ is equipped with the uniform metric, it suffices to show the marginal tightness of $W_k^{(2)}(t)$ for each k. The marginal tightness follows directly by applying lemma 5 to $W_k^{(2)}(t)$. Thus, $\boldsymbol{W}^{(2)}(t)$ converges weakly to a zero-mean Gaussian process where the covariance function between $\mathcal{W}_j^{(2)}(t_1)$ and $\mathcal{W}_k^{(2)}(t_2)$ is

$\frac{1-\alpha}{\alpha}\,\mathrm{E}\{\psi_{1j}(\boldsymbol{\beta}_0,t_1)\psi_{1k}(\boldsymbol{\beta}_0,t_2)\}$. Note that $\mathcal{W}^{(1)}(t)$ and $\mathcal{W}^{(2)}(t)$ are independent since

$$\begin{aligned}
&\mathrm{Cov}\left(n^{-1/2}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t_1),n^{-1/2}\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0,t_2)\right)\\
=\;&\mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t_1)\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0,t_2)\right\}\\
=\;&\mathrm{E}\left\{\mathrm{E}\left(n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t_1)\sum_{j=1}^{n}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right)\psi_{jm}(\boldsymbol{\beta}_0,t_2)\right|\mathcal{F}(\tau)\right)\right\}\\
=\;&\mathrm{E}\left\{n^{-1}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t_1)\sum_{j=1}^{n}\mathrm{E}\left(1-\frac{\xi_j}{\widetilde{\alpha}}\right|\mathcal{F}(\tau)\right)\psi_{jm}(\boldsymbol{\beta}_0,t_2)\right\}=0.
\end{aligned}$$

Therefore, $\boldsymbol{W}(t)=\boldsymbol{W}^{(1)}(t)+\boldsymbol{W}^{(2)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}(t)=\mathcal{W}^{(1)}(t)+\mathcal{W}^{(2)}(t)$ where the covariance function between $\mathcal{W}_j(t_1)$ and $\mathcal{W}_k(t_2)$ is $\mathrm{E}\{\nu_{1j}(\boldsymbol{\beta}_0,t_1)\nu_{1k}(\boldsymbol{\beta}_0,t_2)\}+\frac{1-\alpha}{\alpha}\,\mathrm{E}\{\psi_{1j}(\boldsymbol{\beta}_0,t_1)\psi_{1k}(\boldsymbol{\beta}_0,t_2)\}$. This completes the proof of theorem 5.2.

Proof of theorems 5.3 and 5.4 The asymptotic properties of $\widehat{\boldsymbol{\beta}}_{II}$ and $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II},t)$ can be shown by similar arguments used for $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\Lambda}_{0k}^{I}(\widehat{\boldsymbol{\beta}}_I,t)$. However, the resulting asymptotic properties need some modifications and will involves $(1-\Delta_{ik})$. This is because the asymptotic expansion of $n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)$ includes the terms involving $(1-\Delta_{ik})$. In addition, the asymptotic properties of $n^{-1/2}\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ and $n^{-1}\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ for $d=0,1$ need to be investigated since these include the terms involving $(1-\Delta_{ik})$ as well. Specifically, for each k, it follows from the Taylor expansion of $\widehat{\alpha}_k^{II}(t)^{-1}$ around $\widetilde{\alpha}$,

$$\begin{aligned}
\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1} &= -\frac{1}{\alpha_{**}(t)^2}\left(\widehat{\alpha}_k^{II}(t)-\widetilde{\alpha}\right)\\
&= \frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{1}{\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t)}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\},
\end{aligned}$$

where $\alpha_{**}(t)$ is on the line segment between $\widehat{\alpha}_k^{II}(t)$ and $\widetilde{\alpha}$. Then,

$$n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)=\frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{n}{\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t)}n^{-1/2}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}.$$

$n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t)$ converges to $\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)$ in probability uniformly in t by lemma 6 since $(1-\Delta_{ik})Y_{ik}(t)$ is bounded and monotone in t. In view of lemma 5, $n^{-1}\sum_{i=1}^{n}(\frac{\xi_i}{\widetilde{\alpha}}-1)(1-\Delta_{ik})Y_{ik}(t)$ converges to 0 in probability uniformly in t since $(1-\Delta_{ik})Y_{ik}(t)$ is bounded and monotone function in t. Consequently, $\widehat{\alpha}_k^{II}(t)-\widetilde{\alpha}=\frac{\widetilde{\alpha}n^{-1}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})Y_{ik}(t)}{n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})Y_{ik}(t)}$ converges to 0 in probability uniformly in t. Hence, $\widehat{\alpha}_k^{II}(t)$ and $\widetilde{\alpha}$ converges in probability to the same limit uniformly in t. This ensures $\alpha_{**}(t)$ also converges to the same limit as $\widetilde{\alpha}$. Combining these results, it follows from Slutsky's theorem that

$$
\begin{aligned}
& n^{1/2}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)=\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}\\
+ & \left(\frac{\widetilde{\alpha}}{\alpha_{**}(t)^2}\cdot\frac{n}{\sum_{i=1}^{n}(1-\Delta_{1k})Y_{ik}(t)}-\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right)\\
\times & \ n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\\
= & \frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\left\{\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\right\}\ +\ o_p(1). \qquad (5.24)
\end{aligned}
$$

Likewise, for each k,

$$
\begin{aligned}
& n^{-1/2}\left\{\sum_{i=1}^{n}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}-\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\}\\
= & \ n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\\
+ & \ n^{-1/2}\sum_{i=1}^{n}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{II}(t)^{-1}\right)(1-\Delta_{ik})\xi_iY_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\\
= & \ n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\\
+ & \ n^{-1}\sum_{i=1}^{n}\left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}n^{-1/2}\sum_{j=1}^{n}\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right)(1-\Delta_{jk})Y_{jk}(t)\right\}\\
\times & \ (1-\Delta_{ik})\xi_iY_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\ +\ o_p(1) \qquad (\text{ by } (5.24)\ )\\
= & \ n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}+n^{-1/2}\sum_{i=1}^{n}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})\\
\times & \ \frac{Y_{ik}(t)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\left\{n^{-1}\sum_{j=1}^{n}(1-\Delta_{jk})\frac{\xi_j}{\widetilde{\alpha}}Y_{jk}(t)\boldsymbol{Z}_{jk}(t)^{\otimes d}\right\}\ +\ o_p(1) \qquad (5.25)
\end{aligned}
$$

Note that $(1-\Delta_{jk})Y_{jk}(t)\boldsymbol{Z}_{jk}(t)^{\otimes d}$ is of bounded variation by condition (C). It then follows from lemma 5 that $n^{-1}\sum_{j=1}^{n}(1-\Delta_{jk})\frac{\xi_j}{\widetilde{\alpha}}Y_{jk}(t)\boldsymbol{Z}_{jk}(t)^{\otimes d}$ converges to $\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d}\right)$ for $d=0,1$, in probability uniformly in t. Thus, from (5.25)

$$\begin{aligned}
& n^{-1/2}\left\{\sum_{i=1}^{n} Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}-\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\} \\
= & \; n^{-1/2}\sum_{i=1}^{n}\left\{\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right. \\
+ & \left.\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})\frac{Y_{ik}(t)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{1k}(t)^{\otimes d}\right)\right\}+o_p(1) \\
= & \; n^{-1/2}\sum_{i=1}^{n}(1-\Delta_{ik})\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)Y_{ik}(t) \\
\times & \left\{\boldsymbol{Z}_{ik}(t)^{\otimes d}-\frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\}+o_p(1) \qquad (5.26)
\end{aligned}$$

Therefore, based on (5.24) and (5.26) and by lemma 5, both $n^{1/2}\left\{\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right\}$ and $n^{-1/2}\left\{\sum_{i=1}^{n}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}-\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\}$ converge weakly to zero-mean Gaussian processes, respectively. Consequently, both $\left\{\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right\}$ and $n^{-1}\left\{\sum_{i=1}^{n}Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}-\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}\right\}$ converge to 0 in probability uniformly in t, respectively. Note that the following uniform convergence follows directly from the above results:

$$\sup_{t\in[0,\tau]}\left\|\overline{\boldsymbol{Z}}_k^{\omega}(t)-\boldsymbol{e}_k(t)\right\|\xrightarrow{p}0 \quad \text{as } n\to\infty \qquad (5.27)$$

since $\mathrm{E}\,Y_{1k}(t)$ is bounded away from 0 by condition (B).

One can write

$$\begin{aligned}
-\frac{\partial\boldsymbol{U}_n^{II}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T} &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^{\omega}(t)\right)\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^T dt \\
&= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\omega_{ik}(t)Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)^{\otimes 2}-\overline{\boldsymbol{Z}_k^{\omega}}(t)^{\otimes 2}\right)dt \qquad (5.28)
\end{aligned}$$

Based on (5.27), (5.28), and the uniform convergence of $\left(\widehat{\alpha}_k^{II}(t) - \widetilde{\alpha}\right)$ to 0 and lemma 5, it can be shown that

$$-\frac{\partial \boldsymbol{U}_n^{II}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \xrightarrow{p} \boldsymbol{A} \ \text{ as } n \to \infty \tag{5.29}$$

by using similar arguments as in proving theorem 5.1.

One can decompose $n^{-1/2}U^{II}(\boldsymbol{\beta}_0)$ into two parts:

$$\begin{aligned} n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0) &= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\omega}(t)\right\} dM_{ik}(t) \\ &+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}(\omega_{ik}(t) - 1)\left\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\omega}(t)\right\} dM_{ik}(t) \end{aligned} \tag{5.30}$$

Based on conditions (B) and (C), $n^{-1}\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}(d = 0, 1)$ are of bounded variations and $n^{-1}\sum_{i=1}^{n}\omega_{ik}(t)Y_{ik}(t)$ is bounded away from 0. Therefore, $\overline{\boldsymbol{Z}}_k^{\omega}(t)$ is of bounded variations. Along with the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\omega}(t)$ to $\boldsymbol{e}_k(t)$, by the similar arguments used for proving theorem 5.1, the first term on the right-hand side of (5.30) is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{\boldsymbol{Z}_{ik}(t) - \boldsymbol{e}_k(t)\right\} dM_{ik}(t).$$

The second term on the right-hand side of (5.30) can be further decomposed as the following:

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}(\omega_{ik}(t) - 1)\left\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\omega}(t)\right\} dM_{ik}(t) \\ = \; & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\frac{\xi_i}{\widetilde{\alpha}} - 1\right)(1 - \Delta_{ik})\left\{\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\omega}(t)\right\} dM_{ik}(t) \\ + \; & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\widehat{\alpha}_k^{II}(t)^{-1} - \widetilde{\alpha}^{-1}\right)\xi_i(1 - \Delta_{ik})\left(\boldsymbol{Z}_{ik}(t) - \overline{\boldsymbol{Z}}_k^{\omega}(t)\right) dM_{ik}(t) \end{aligned} \tag{5.31}$$

It follows from the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\omega}(t)$ to $\boldsymbol{e}_k(t)$ that the first term on the right-hand side of (5.31) is asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)(1-\Delta_{ik})\left\{\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right\}dM_{ik}(t)\\
= \quad & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})\left\{\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right\}Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)
\end{aligned}
$$

The last equality holds since only censored observations contribute to this term.
Likewise, based on (5.24) and the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\omega}(t)$ to $\boldsymbol{e}_k(t)$, the second term on the right-hand side of (5.31) is

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\widehat{\alpha}_k^{II}(t)^{-1}-\widetilde{\alpha}^{-1}\right)\xi_i(1-\Delta_{ik})\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^{\omega}(t)\right)dM_{ik}(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{II}(t)^{-1}\right)\xi_i(1-\Delta_{ik})\left(\boldsymbol{Z}_{ik}(t)-\overline{\boldsymbol{Z}}_k^{\omega}(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt\\
&= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\left\{\frac{1}{\widetilde{\alpha}\,\mathrm{E}((1-\Delta_{1k})Y_{1k}(t))}n^{-1/2}\sum_{j=1}^{n}\left(\frac{\xi_j}{\widetilde{\alpha}}-1\right)(1-\Delta_{jk})Y_{jk}(t)\right\}\\
&\times\ \xi_i(1-\Delta_{ik})\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)Y_{ik}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(t)\right)dt+o_p(1)\\
&=n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^{\tau}(1-\Delta_{ik})\frac{Y_{ik}(t)}{\mathrm{E}((1-\Delta_{1k})Y_{1k}(t))}\\
&\times\ \left\{n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}(1-\Delta_{jk})\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)\right)\right\}dt+o_p(1) \qquad (5.32)
\end{aligned}
$$

It follows from lemma 5 that $n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}(1-\Delta_{jk})\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)$ and
$n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}(1-\Delta_{jk})\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)$ converge to
$\mathrm{E}\left\{(1-\Delta_{jk})\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)Y_{1k}(t)\right\}$ and $\mathrm{E}\left\{(1-\Delta_{jk})\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)Y_{1k}(t)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right\}$

in probability uniformly in t, respectively. Thus, from (5.32),

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^\tau (1-\Delta_{ik})\frac{Y_{ik}(t)}{\mathrm{E}((1-\Delta_{1k})Y_{1k}(t))} \\
\times & \left\{n^{-1}\sum_{j=1}^{n}\frac{\xi_j}{\widetilde{\alpha}}(1-\Delta_{jk})\left(\boldsymbol{Z}_{jk}(t)-\boldsymbol{e}_k(t)\right)Y_{jk}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{jk}(t)\right)\right\}dt \\
= & n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\xi_i}{\widetilde{\alpha}}-1\right)\int_0^\tau (1-\Delta_{1k})\frac{Y_{ik}(t)}{\mathrm{E}((1-\Delta_{1k})Y_{1k}(t))} \\
\times & \mathrm{E}\left\{(1-\Delta_{1k})\left(\boldsymbol{Z}_{1k}(t)-\boldsymbol{e}_k(t)\right)Y_{1k}(t)\left(\lambda_{0k}(t)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(t)\right)\right\}dt+o_p(1)
\end{aligned}$$

Therefore, the second term on the right-hand side of (5.30) is asymptotically equal to

$$n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})\int_0^\tau\left\{\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0,t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\}dt$$

where

$$\boldsymbol{R}_{ik}(\boldsymbol{\beta},t)=Y_{ik}(t)\left(\boldsymbol{Z}_{ik}(t)-\boldsymbol{e}_k(t)\right)\left(\lambda_{0k}(t)+\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(t)\right).$$

Combining the above results, we have shown that $n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},ik}(\boldsymbol{\beta}_0)+n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \\
\times & \int_0^\tau\left(\boldsymbol{R}_{ik}(\boldsymbol{\beta}_0,t)-\frac{Y_{ik}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta}_0,t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right)dt. \qquad (5.33)
\end{aligned}$$

The first term on the right-hand side of (5.33) is again asymptotically zero-mean normal with covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0)=\mathrm{E}\left(\sum_{k=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},1k}(\boldsymbol{\beta}_0)\right)^{\otimes 2}$ by Yin and Cai (2004).
The second term on the right-hand side of (5.33) can be shown to be asymptotically zero-mean normal with covariance matrix $\frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$ where

$$\boldsymbol{V}_{II}(\boldsymbol{\beta})=\mathrm{E}\left[\sum_{k=1}^{K}(1-\Delta_{1k})\int_0^\tau\left\{\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)-\frac{Y_{1k}(t)\,\mathrm{E}\left((1-\Delta_{1k})\boldsymbol{R}_{1k}(\boldsymbol{\beta},t)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(t)\right)}\right\}dt\right]^{\otimes 2}.$$

by Hájek (1960)'s central limit theorem for finite population sampling. Then, together with the independence of the first term and the second term of (5.33), it follows that $n^{-1/2}\boldsymbol{U}^{II}(\boldsymbol{\beta}_0)$ converges to zero mean normal random variable with finite covariance matrix $\boldsymbol{Q}(\boldsymbol{\beta}_0) + \frac{1-\alpha}{\alpha}\boldsymbol{V}_{II}(\boldsymbol{\beta}_0)$. Now, the consistency of $\widehat{\boldsymbol{\beta}}_{II}$ and the asymptotic normality of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ will follow from the similar arguments used for proving theorem 5.1 if we replace $\widehat{\alpha}_k(t)$, $n^{-1}\sum_{i=1}^n \rho_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ $(d = 0, 1)$, and $\boldsymbol{U}^I(\boldsymbol{\beta})$ by $\widehat{\alpha}_k^{II}(t)$, $n^{-1}\sum_{i=1}^n \omega_{ik}(t)Y_{ik}(t)\boldsymbol{Z}_{ik}(t)^{\otimes d}$ $(d = 0, 1)$, and $\boldsymbol{U}^{II}(\boldsymbol{\beta})$, respectively, and use their corresponding asymptotic properties we have just derived.

The asymptotic properties of $\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t)$ can also be shown by the similar arguments used for proving theorem 5.2 with some modifications. Specifically,

$$
\begin{aligned}
& n^{1/2}\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t) - \Lambda_{0k}(t)\} \\
&= n^{1/2}\left\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t) - \widehat{\Lambda}_{0k}^{II}(\boldsymbol{\beta}_0, t)\right\} \;+\; n^{1/2}\left\{\widehat{\Lambda}_{0k}^{II}(\boldsymbol{\beta}_0, t) - \Lambda_{0k}(t)\right\} \\
&= n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)\left(dN_{ik}(u) - Y_{ik}(u)\widehat{\boldsymbol{\beta}}_{II}^T\boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du\right. \\
&\quad \left. - \int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)\left(dN_{ik}(u) - Y_{ik}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du\right\} \\
&\quad + n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)\left(dN_{ik}(u) - Y_{ik}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du\right\} \\
&\quad - n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)\lambda_{0k}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du\right\} \\
&= n^{1/2}\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_{II})^T\boldsymbol{Z}_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du \;+\; n^{1/2}\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)dM_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} \\
&= n^{1/2}\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_{II})^T\boldsymbol{Z}_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du \;+\; n^{1/2}\int_0^t \frac{\sum_{i=1}^n dM_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} \\
&\quad + n^{1/2}\int_0^t \frac{\sum_{i=1}^n (\omega_{ik}(u) - 1)dM_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} \\
&= n^{1/2}\int_0^t \frac{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_{II})^T\boldsymbol{Z}_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du \;+\; n^{1/2}\int_0^t \frac{\sum_{i=1}^n dM_{ik}(u)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} \\
&\quad + n^{1/2}\int_0^t \frac{\sum_{i=1}^n \left(1 - \omega_{ik}(u)\right)Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du \qquad (5.34)
\end{aligned}
$$

By the uniform convergence of $\overline{\boldsymbol{Z}}_k^{\omega}(u)$ to $\boldsymbol{e}_k(u)$, the first term of (5.34) is asymptotically equivalent to $n^{1/2}\boldsymbol{r}_k(t)^T\left(\widehat{\boldsymbol{\beta}}_{II}-\boldsymbol{\beta}_0\right)$ where $\boldsymbol{r}_k(t)=-\int_0^t \boldsymbol{e}_k(u)du$.

The second term on the right-hand side of (5.34) is asymptotically equivalent to

$$\int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\}$$

by lemma 2, since $\left(n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{1k}(u)\right)^{-1}$ is of bounded variation, converges uniformly to $(\mathrm{E}\, Y_{1k}(u))^{-1}$ where $\mathrm{E}\, Y_{1k}(u)$ is bounded away from 0, and $n^{-1/2}\sum_{i=1}^n M_{ik}(u)$ converges to a zero-mean Gaussian process with continuous sample paths.

The last term on the right-hand side of (5.34) can be written as

$$\begin{aligned}
& n^{1/2}\int_0^t \frac{\sum_{i=1}^n (1-\omega_{ik}(u))\, Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)}du \\
= & \int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} n^{-1/2}\sum_{i=1}^n \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)du \\
+ & \int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} n^{1/2}\left(\widetilde{\alpha}^{-1}-\widehat{\alpha}_k^{II}(u)\right) \\
\times & \; n^{-1}\sum_{i=1}^n \xi_i(1-\Delta_{ik})Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)du \qquad (5.35)
\end{aligned}$$

It follows from the uniform convergence of $\left\{n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)\right\}^{-1}$ to $\{\mathrm{E}\, Y_{1k}(u)\}^{-1}$, where $\mathrm{E}\, Y_{1k}(u)$ is bounded away from 0, that the first term on the right-hand side of (5.35) is asymptotically equivalent to

$$\int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} n^{-1/2}\sum_{i=1}^n \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik})Y_{ik}(u)\left(\lambda_{0k}(u)+\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)\right)du$$

Based on the uniform convergence of $\left\{n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)\right\}^{-1}$ to $\{\mathrm{E}\, Y_{1k}(u)\}^{-1}$, $n^{-1}\sum_{i=1}^n \frac{\xi_i}{\alpha}(1-\Delta_{ik})Y_{ik}(u)$ to $\mathrm{E}((1-\Delta_{1k})Y_{1k}(u))$, $n^{-1}\sum_{i=1}^n \frac{\xi_i}{\alpha}(1-\Delta_{ik})Y_{ik}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)$ to $\mathrm{E}((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(u))$, (5.24) and lemma 5, the second term on the right-hand side of (5.35) is

$$\int_0^t \frac{1}{n^{-1}\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} n^{1/2}\left(\widetilde{\alpha}^{-1} - \widehat{\alpha}_k^{II}(u)\right)$$
$$\times n^{-1}\sum_{i=1}^n \xi_i(1-\Delta_{ik})Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right) du$$
$$= n^{-1/2}\sum_{i=1}^n \left(\frac{\xi_i}{\widetilde{\alpha}} - 1\right)(1-\Delta_{ik}) \int_0^t \frac{Y_{ik}(u)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}$$
$$\times\ \mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u))\right) \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1)$$

By combining the above results, the last term on the right-hand side of (5.34) is

$$n^{1/2}\int_0^t \frac{\sum_{i=1}^n \left(1-\omega_{ik}(u)\right) Y_{ik}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)}{\sum_{i=1}^n \omega_{ik}(u)Y_{ik}(u)} du$$
$$= n^{-1/2}\sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \int_0^t Y_{ik}(u)\left\{\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u)\right)\right.$$
$$\left. -\ \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\left(\lambda_{0k}(u) + \boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1)$$
$$= n^{-1/2}\sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \int_0^t Y_{ik}(u)$$
$$\times \left\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1)$$

Based on (5.34) and the above results, we have that

$$\begin{aligned}
& n^{1/2}(\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t) - \Lambda_{0k}(t)) \\
= \;& n^{1/2}\boldsymbol{r}_k(t)^T\left(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0\right) + \int_0^t \frac{1}{\mathrm{E}\, Y_{1k}(u)} d\left\{n^{-1/2}\sum_{i=1}^n M_{ik}(u)\right\} \\
+ \;& n^{-1/2}\sum_{i=1}^n \left(1 - \frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \\
\times \;& \int_0^t Y_{ik}(u)\left\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{ik}(u) - \frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\} \frac{du}{\mathrm{E}\, Y_{1k}(u)} + o_p(1) \quad (5.36)
\end{aligned}$$

Based on Taylor expansion of $\boldsymbol{U}^{II}(\widehat{\boldsymbol{\beta}}_{II})$ around $\boldsymbol{\beta}_0$ and the results in (5.29), (5.33) and (5.36), we have

$$
\begin{aligned}
& n^{1/2}(\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II},t)-\Lambda_{0k}(t)) \\
&= \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\left\{n^{-1/2}\sum_{i=1}^{n}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0)+n^{-1/2}\sum_{i=1}^{n}\sum_{m=1}^{K}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{im})\right. \\
&\times \left.\int_0^{\tau}\left(\boldsymbol{R}_{im}(\boldsymbol{\beta}_0,t)-\frac{Y_{im}(t)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(t)\right)}\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta}_0,t)\right)\right)dt\right\} \\
&+ \int_0^t\frac{1}{\mathrm{E}\,Y_{1k}(u)}d\left\{n^{-1/2}\sum_{i=1}^{n}M_{ik}(u)\right\}+n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)(1-\Delta_{ik}) \\
&\times \int_0^t Y_{ik}(u)\left\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)-\frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\}\frac{du}{\mathrm{E}\,Y_{1k}(u)}+o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\left[\left\{\boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta}_0)+\int_0^t\frac{1}{\mathrm{E}\,Y_{1k}(u)}dM_{ik}(u)\right\}\right. \\
&+ \left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\left\{\boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}(1-\Delta_{im})\right. \\
&\times \int_0^{\tau}\left(\boldsymbol{R}_{im}(\boldsymbol{\beta}_0,u)-\frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta}_0,u)\right)}{\mathrm{E}\left((1-\Delta_{1m})Y_{1m}(u)\right)}\right)du \\
&+(1-\Delta_{ik})\int_0^t Y_{ik}(u)\left\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{ik}(u)-\frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}_0^T\boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\} \\
&\times \left.\left.\frac{du}{\mathrm{E}\,Y_{1k}(u)}\right\}\right] + o_p(1) \qquad (5.37)\\
&= n^{-1/2}\sum_{i=1}^{n}\nu_{ik}(\boldsymbol{\beta}_0,t)+n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\widetilde{\alpha}}\right)\psi_{ik}^{II}(\boldsymbol{\beta}_0,t)+o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_{ik}(\boldsymbol{\beta},t) &= \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}\mathbf{M}_{\tilde{\boldsymbol{z}},im}(\boldsymbol{\beta})+\int_0^t\frac{1}{\mathrm{E}\,Y_{1k}(u)}dM_{ik}(u) \quad \text{and} \\
\psi_{ik}^{II}(\boldsymbol{\beta},t) &= \boldsymbol{r}_k(t)^T\boldsymbol{A}^{-1}\sum_{m=1}^{K}(1-\Delta_{im})\int_0^{\tau}\left(\boldsymbol{R}_{im}(\boldsymbol{\beta}_0,u)-\frac{Y_{im}(u)\,\mathrm{E}\left((1-\Delta_{1m})\boldsymbol{R}_{1m}(\boldsymbol{\beta}_0,u)\right)}{\mathrm{E}(1-\Delta_{1m})Y_{1m}(u)}\right)du \\
&\quad + (1-\Delta_{ik})\int_0^t\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_{ik}(u)-\frac{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\boldsymbol{\beta}^T\boldsymbol{Z}_{1k}(u)\right)}{\mathrm{E}\left((1-\Delta_{1k})Y_{1k}(u)\right)}\right\}\frac{Y_{ik}(u)du}{\mathrm{E}\,Y_{1k}(u)}.
\end{aligned}
$$

The asymptotic properties of $n^{1/2}\{\widehat{\Lambda}_{0k}^{II}(\widehat{\boldsymbol{\beta}}_{II}, t) - \Lambda_{0k}(t)\}$ follow from the similar arguments used for proving theorem 5.2. This complete the proofs of theorems 5.3 and 5.4.

TABLE 5.1: Summary of simulation results for $\widehat{\beta}_I$: $Z_{ik} \sim Bin(0.5)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.045	1.325	1.437	0.924
			0.67	0.025	1.328	1.371	0.942
			1.25	-0.001	1.327	1.420	0.937
			4	-0.020	1.331	1.449	0.927
		62	0.1	0.024	1.081	1.121	0.924
			0.67	0.008	1.075	1.083	0.933
			1.25	0.013	1.076	1.102	0.939
			4	-0.022	1.081	1.151	0.928
	10%	111	0.1	0.017	0.645	0.669	0.947
			0.67	0.016	0.643	0.651	0.956
			1.25	-0.005	0.642	0.660	0.944
			4	-0.023	0.642	0.676	0.941
		222	0.1	-0.005	0.532	0.541	0.948
			0.67	0.023	0.530	0.527	0.956
			1.25	0.011	0.530	0.534	0.945
			4	-0.008	0.530	0.542	0.943
	25%	333	0.1	-0.017	0.357	0.361	0.951
			0.67	0.010	0.357	0.366	0.948
			1.25	-0.005	0.357	0.365	0.949
			4	0.001	0.356	0.367	0.946
		666	0.1	-0.015	0.292	0.295	0.950
			0.67	0.005	0.292	0.299	0.946
			1.25	-0.005	0.291	0.289	0.957
			4	0.009	0.291	0.290	0.950
log(2)	3%	31	0.1	0.794	1.443	1.555	0.931
			0.67	0.775	1.443	1.484	0.946
			1.25	0.754	1.438	1.544	0.936
			4	0.718	1.445	1.570	0.934
		62	0.1	0.744	1.162	1.201	0.935
			0.67	0.723	1.159	1.153	0.942
			1.25	0.742	1.158	1.190	0.935
			4	0.688	1.164	1.238	0.933
	10%	111	0.1	0.730	0.725	0.751	0.944
			0.67	0.722	0.722	0.730	0.954
			1.25	0.697	0.720	0.739	0.944
			4	0.677	0.720	0.764	0.934
		222	0.1	0.697	0.602	0.607	0.946
			0.67	0.720	0.600	0.595	0.955
			1.25	0.707	0.598	0.602	0.947
			4	0.685	0.598	0.618	0.941
	25%	333	0.1	0.683	0.399	0.403	0.951
			0.67	0.711	0.400	0.407	0.948
			1.25	0.687	0.398	0.403	0.949
			4	0.697	0.398	0.410	0.944
		666	0.1	0.683	0.327	0.331	0.955
			0.67	0.703	0.326	0.331	0.951
			1.25	0.684	0.325	0.320	0.958
			4	0.703	0.326	0.325	0.956

TABLE 5.2: Summary of simulation results for $\widehat{\beta}_{II}$: $Z_{ik} \sim Bin(0.5)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.026	1.350	1.427	0.931
			0.67	0.016	1.354	1.366	0.950
			1.25	-0.019	1.350	1.404	0.942
			4	-0.030	1.355	1.426	0.935
		62	0.1	0.019	1.135	1.113	0.929
			0.67	0.007	1.090	1.081	0.937
			1.25	0.006	1.090	1.093	0.944
			4	-0.027	1.095	1.141	0.937
	10%	111	0.1	0.012	0.636	0.655	0.946
			0.67	0.016	0.634	0.638	0.956
			1.25	-0.001	0.634	0.648	0.945
			4	-0.024	0.633	0.661	0.944
		222	0.1	-0.007	0.528	0.535	0.951
			0.67	0.020	0.526	0.523	0.953
			1.25	0.011	0.526	0.529	0.946
			4	-0.001	0.525	0.536	0.946
	25%	333	0.1	-0.021	0.344	0.346	0.953
			0.67	0.008	0.344	0.352	0.948
			1.25	-0.004	0.343	0.348	0.954
			4	0.003	0.343	0.351	0.947
		666	0.1	-0.017	0.288	0.290	0.948
			0.67	0.004	0.288	0.294	0.946
			1.25	-0.004	0.287	0.283	0.957
			4	0.008	0.287	0.286	0.952
log(2)	3%	31	0.1	0.776	1.471	1.549	0.937
			0.67	0.766	1.496	1.482	0.951
			1.25	0.732	1.459	1.517	0.948
			4	0.707	1.469	1.551	0.939
		62	0.1	0.741	1.179	1.192	0.942
			0.67	0.723	1.175	1.150	0.944
			1.25	0.734	1.171	1.178	0.939
			4	0.681	1.178	1.229	0.938
	10%	111	0.1	0.720	0.717	0.738	0.945
			0.67	0.723	0.714	0.715	0.953
			1.25	0.699	0.712	0.724	0.946
			4	0.675	0.712	0.748	0.942
		222	0.1	0.693	0.598	0.602	0.948
			0.67	0.718	0.596	0.590	0.954
			1.25	0.706	0.594	0.594	0.950
			4	0.684	0.594	0.613	0.941
	25%	333	0.1	0.678	0.384	0.387	0.952
			0.67	0.707	0.384	0.392	0.948
			1.25	0.687	0.383	0.384	0.956
			4	0.698	0.383	0.391	0.941
		666	0.1	0.680	0.322	0.324	0.954
			0.67	0.702	0.322	0.326	0.951
			1.25	0.685	0.321	0.314	0.953
			4	0.702	0.322	0.319	0.958

TABLE 5.3: Summary of simulation results for $\widehat{\beta}_I$: $Z_{ik} \sim U(0,3)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	0.046	2.236	2.437	0.917
			0.67	0.029	2.233	2.366	0.927
			1.25	0.016	2.259	2.384	0.931
			4	0.007	2.261	2.365	0.934
		62	0.1	0.004	1.805	1.886	0.920
			0.67	0.012	1.796	1.864	0.929
			1.25	0.001	1.828	1.881	0.927
			4	0.022	1.814	1.828	0.929
	10%	111	0.1	-0.013	1.109	1.131	0.951
			0.67	0.003	1.119	1.139	0.955
			1.25	-0.014	1.114	1.149	0.943
			4	0.007	1.117	1.103	0.954
		222	0.1	-0.006	0.919	0.917	0.954
			0.67	-0.004	0.923	0.931	0.955
			1.25	-0.005	0.920	0.955	0.944
			4	0.011	0.921	0.928	0.944
	25%	333	0.1	-0.012	0.618	0.603	0.951
			0.67	-0.012	0.619	0.632	0.955
			1.25	-0.011	0.620	0.630	0.946
			4	-0.001	0.620	0.605	0.955
		666	0.1	-0.010	0.505	0.494	0.954
			0.67	-0.005	0.505	0.509	0.952
			1.25	0.004	0.506	0.505	0.950
			4	-0.017	0.505	0.493	0.959
log(2)	3%	31	0.1	0.788	2.479	2.695	0.911
			0.67	0.801	2.481	2.617	0.929
			1.25	0.780	2.510	2.642	0.936
			4	0.761	2.512	2.630	0.933
		62	0.1	0.708	2.005	2.088	0.923
			0.67	0.744	1.994	2.063	0.935
			1.25	0.727	2.032	2.087	0.925
			4	0.743	2.018	2.039	0.927
	10%	111	0.1	0.687	1.062	1.079	0.950
			0.67	0.715	1.068	1.089	0.949
			1.25	0.697	1.066	1.108	0.948
			4	0.718	1.067	1.054	0.956
		222	0.1	0.679	0.803	0.794	0.955
			0.67	0.689	0.805	0.821	0.949
			1.25	0.696	0.805	0.830	0.948
			4	0.710	0.805	0.800	0.952
	25%	333	0.1	0.683	0.689	0.668	0.954
			0.67	0.690	0.689	0.704	0.951
			1.25	0.688	0.691	0.704	0.945
			4	0.697	0.690	0.675	0.957
		666	0.1	0.683	0.564	0.547	0.956
			0.67	0.693	0.564	0.567	0.951
			1.25	0.702	0.565	0.566	0.948
			4	0.675	0.564	0.546	0.958

TABLE 5.4: Summary of simulation results for $\widehat{\beta}_{II}$: $Z_{ik} \sim U(0,3)$

β_0	event proportion	$\tilde{n}$	θ	mean $\widehat{\beta}_I$	proposed S.E.	true S.D.	95% C.I.
0	3%	31	0.1	-0.002	2.278	2.405	0.921
			0.67	-0.010	2.277	2.338	0.931
			1.25	-0.010	2.302	2.364	0.936
			4	-0.017	2.310	2.341	0.935
		62	0.1	-0.007	1.831	1.874	0.926
			0.67	0.005	1.820	1.854	0.933
			1.25	-0.005	1.853	1.862	0.933
			4	0.022	1.841	1.818	0.930
	10%	111	0.1	-0.016	1.095	1.108	0.953
			0.67	0.008	1.104	1.118	0.958
			1.25	-0.014	1.099	1.138	0.943
			4	-0.001	1.101	1.082	0.957
		222	0.1	-0.011	0.912	0.906	0.951
			0.67	-0.003	0.916	0.922	0.955
			1.25	-0.005	0.913	0.946	0.946
			4	0.006	0.913	0.914	0.949
	25%	333	0.1	-0.012	0.595	0.581	0.958
			0.67	-0.009	0.595	0.609	0.951
			1.25	-0.005	0.597	0.606	0.951
			4	-0.008	0.596	0.585	0.957
		666	0.1	-0.009	0.298	0.486	0.957
			0.67	-0.003	0.498	0.502	0.952
			1.25	0.004	0.499	0.498	0.954
			4	-0.016	0.498	0.487	0.954
log(2)	3%	31	0.1	0.734	2.529	2.658	0.919
			0.67	0.759	2.527	2.587	0.934
			1.25	0.737	2.558	2.624	0.938
			4	0.736	2.566	2.599	0.939
		62	0.1	0.697	2.035	2.071	0.928
			0.67	0.735	2.021	2.051	0.938
			1.25	0.719	2.061	2.068	0.932
			4	0.742	2.048	2.022	0.930
	10%	111	0.1	0.695	0.995	0.995	0.947
			0.67	0.706	0.997	1.017	0.943
			1.25	0.689	0.997	1.026	0.947
			4	0.706	0.999	1.012	0.943
		222	0.1	0.680	0.767	0.753	0.954
			0.67	0.690	0.767	0.780	0.946
			1.25	0.696	0.768	0.782	0.945
			4	0.693	0.769	0.764	0.952
	25%	333	0.1	0.682	0.664	0.642	0.956
			0.67	0.692	0.664	0.679	0.948
			1.25	0.694	0.665	0.675	0.950
			4	0.689	0.665	0.654	0.955
		666	0.1	0.683	0.556	0.537	0.957
			0.67	0.695	0.556	0.561	0.950
			1.25	0.702	0.557	0.559	0.954
			4	0.677	0.556	0.540	0.954

TABLE 5.5: Analysis of Busselton Health Study

Variables	$\widehat{\beta}_I$	S.E.$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	95% C.I.$_{II}$
Ferritin on CHD	0.15	0.106	(-0.06, 0.35)	0.11	0.106	(-0.10, 0.32)
Ferritin on Stroke	0.03	0.089	(-0.15, 0.20)	0.02	0.088	(-0.15, 0.19)
Age	0.04	0.008	(0.03, 0.06)	0.04	0.007	(0.03, 0.06)
BMI	0.01	0.021	(-0.03, 0.05)	0.02	0.021	(-0.02, 0.06)
Cholesterol	-0.07	0.065	(-0.20, 0.06)	-0.04	0.060	(-0.16, 0.07)
Triglycerides	0.26	0.118	(0.03, 0.49)	0.25	0.108	(0.04, 0.46)
Diabetes Treatment	0.71	0.939	(-1.13, 2.55)	0.74	0.911	(-1.04, 2.52)
Haemoglobin	-0.01	0.007	(-0.02, 0.01)	-0.01	0.006	(-0.02, 0.01)
BPT	0.35	0.264	(-0.17, 0.87)	0.40	0.262	(-0.12, 0.91)
SBP	0.01	0.005	(-0.01, 0.02)	0.01	0.004	(-0.01, 0.02)
Smoke (Former)	0.23	0.187	(-0.14, 0.59)	0.24	0.181	(-0.12, 0.60)
Smoke (Current)	0.29	0.200	(-0.11, 0.68)	0.22	0.186	(-0.14, 0.58)

TABLE 5.6: Analysis of Busselton Health Study with Common Ferritin Effect

Variables	$\widehat{\beta}_I$	S.E.$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	95% C.I.$_{II}$
Ferritin	0.09	0.084	(-0.08, 0.25)	0.06	0.082	(-0.10, 0.22)
Age	0.04	0.008	(0.03, 0.06)	0.04	0.007	(0.03, 0.06)
BMI	0.01	0.021	(-0.03, 0.06)	0.02	0.021	(-0.02, 0.06)
Cholesterol	-0.07	0.065	(-0.20, 0.06)	-0.04	0.060	(-0.16, 0.07)
Triglycerides	0.26	0.117	(0.02, 0.48)	0.25	0.108	(0.04, 0.46)
Diabetes Treatment	0.70	0.938	(-1.13, 2.54)	0.74	0.911	(-1.05, 2.52)
Haemoglobin	-0.01	0.007	(-0.02, 0.01)	-0.01	0.006	(-0.02, 0.01)
BPT	0.35	0.264	(-1.64, 0.87)	0.40	0.262	(-0.11, 0.92)
SBP	0.01	0.005	(-0.01, 0.02)	0.01	0.004	(-0.02, 0.02)
Smoke (Former)	0.23	0.187	(-0.14, 0.59)	0.24	0.181	(0.12, 0.59)
Smoke (Current)	0.28	0.200	(-0.11, 0.68)	0.22	0.185	(-0.15, 0.58)

TABLE 5.7: Analysis of Busselton Health Study Considering Gender Effect

Variables	$\widehat{\beta}_I$	S.E.$_I$	95% C.I.$_I$	$\widehat{\beta}_{II}$	S.E.$_{II}$	95% C.I.$_{II}$
Ferritin on CHD (M)	0.05	0.224	(-0.39, 0.49)	0.01	0.227	(-0.43, 0.46)
Ferritin on CHD (W)	-0.11	0.114	(-0.25, 0.20)	-0.01	0.113	(-0.28, 0.16)
Ferritin on Stroke (M)	-0.01	0.156	(-0.31, 0.30)	-0.01	0.155	(-0.31, 0.30)
Ferritin on Stroke (W)	0.18	0.117	(-0.23, 0.22)	-0.01	0.115	(-0.25, 0.19)
Age	0.04	0.008	(0.03, 0.06)	0.04	0.008	(0.03, 0.06)
BMI	0.01	0.021	(-0.03, 0.06)	0.02	0.021	(-0.02, 0.06)
Cholesterol	-0.03	0.066	(-0.16, 0.10)	-0.01	0.061	(-0.12, 0.12)
Triglycerides	0.23	0.117	(0.01, 0.46)	0.23	0.108	(0.01, 0.44)
Diabetes Treatment	0.76	0.926	(-1.05, 2.58)	0.81	0.878	(-0.91, 2.53)
Haemoglobin	-0.02	0.008	(-0.04, -0.01)	-0.02	0.007	(-0.03, -0.01)
BPT	0.43	0.264	(-0.09, 0.94)	0.47	0.260	(-0.04, 0.98)
SBP	0.01	0.005	(-0.01, 0.02)	0.01	0.005	(-0.01, 0.02)
Smoke (Former)	0.08	0.193	(-0.30, 0.45)	0.07	0.189	(-0.30, 0.44)
Smoke (Current)	0.23	0.202	(-0.16, 0.63)	0.17	0.187	(-0.20, 0.53)

CHAPTER 6

SUMMARY AND FUTURE RESEARCH

In this dissertation, we have studied statistical methods for multivariate failure time data arising from case-control and case-cohort studies. Specifically, the following two different scenarios were studied: 1) case-control within cohort studies with correlated failure time data, 2) case-cohort studies with multiple disease outcomes.

Case-control and case-cohort studies are often used to save costs and efforts in cohort studies. Many statistical methods have been proposed for such studies, however, most of them were limited to univariate failure time data. Multivariate failure time data are frequently encountered in many biomedical studies. Thus, the main contribution of this dissertation is to provide statistical methods which address both multivariate feature of the failure times and sampling schemes such as case-control or case-cohort study designs. Our focus was on the situation where the primary interest of the studies was on the assessment of the effect of covariate on time to main disease outcome of interest while the correlations among the failure times within each subject were considered as nuisance. This naturally led us to consider marginal hazard regression models. For the estimation of the regression parameters, we developed weighted estimating equation approach where the weights were included to appropriately account for the sampling schemes. The cumulative baseline hazard functions were also studied and Breslow-Aalen type of estimates were proposed.

In Chapter 3, we have considered the marginal proportional hazards regression models for correlated failure time data from case-control studies. Two different types of weights were considered: the inverse of the inclusion probabilities and the local average. The latter requires additional information on the observed failure times of all the cohort members but was more efficient than the former when the censoring time is dependent on some covariates which the failure time is also dependent on. In Chapter 4, we have considered the marginal proportional hazards regression models for case-cohort studies with multiple disease outcomes. Two different forms of time-varying weights were considered: one was a multivariate extension of Self and Prentice (1988)'s estimator for univariate failure time data while the other was a multivariate extension of Kalbfleisch and Lawless (1988)'s estimator for univariate failure time data. In Chapter 5, we have considered the marginal additive hazards regression models instead.

The asymptotic properties of the proposed estimators were studied and were shown to provide desirable asymptotic properties such as consistency and asymptotic normality. Most of the proofs relied on modern empirical processes theory instead of famous martingale convergence results primarily due to lack of predictability in the weights.

We investigated the finite sample properties of the proposed methods via simulation studies. Simulation results under various different setups confirmed that the proposed methods worked properly under reasonable finite sample sizes.

The proposed methods were applied to real-world data sets for illustration. We analysed the KPCDP data in Chapter 3 and the Busselton Health Study in Chapters 4 and 5.

The proposed methods in this dissertation research can be extended in several directions:

First, in this dissertation, we incorporated weights to take the sampling feature into account. Different types of weights were also considered in an effort to enhance efficiency. Robins, Rotnitzky and Zhao (1994) considered the general problem of regression models with missing covariates. They introduced a class of estimating equations which can achieve a semiparametric efficiency bound. We will explore the possibility of extending their results in multivariate survival data to obtain more efficient estimators.

Second, in some applications, as mentioned above, the proportional hazards assumption may not always be true, or one may be interested in modeling association from different aspects. Thus, in my dissertation, we considered additive hazards models for multiple disease outcome data from case-cohort studies as an alternative to multiplicative models. A natural extension would be to consider other types of models including, but not limited to the proportional odds model, the accelerated failure time model, or the semiparametric transformation model.

Third, we considered two different cohort sampling designs: case-control and case-cohort. One may be interested in applying other types of study designs. For example, a nested case-control study design is another type of cohort sampling design which has been of particular interest. Applying this study design when multiple disease end points are to be evaluated would be worth pursuing.

Last, but not least, an approach which extends to accommodate the measurement error of the covariate is highly desired. Since both case-control and case-cohort studies are mostly conducted retrospectively, the covariate measurement might be subject to errors. For example, the covariate measurements which rely on self-report or are affected by the passage of time are likely to contain errors. In such situation, it would be important to develop methods which will account for covariate measurement error.

REFERENCES

Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10**, 1100–1120.

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

Barlow, W., Ichikawa, L., Rosner, D. and Izumi, S. (1999). Analysis of case-cohort designs. *Journal of Clinical Epidemiology* **12**, 1165–1172.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for cox model with stagger entry. *The Annals of Statistics* **25**, 662–682.

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

Borgan, O., Goldstein, L. and Langholz, B. (1995). Methods for the analysis of sampled cohort data in the cox proportional hazards model. *The Annals of Statistics* **23**, 1749–1778.

Borgan, O., Langholz, B., O., S. S., Goldstein, L. and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6**, 39–58.

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

Cai, J. and Prentice, R. (1997). Regression analysis for correlated failure time data. *Lifetime Data Analysis* **3**, 197–213.

Caplan, D., Cai, J., Yin, G. and White, B. A. (2005). Root canal filled versus non-root canal filled teeth: a retrospective comparison of survival times. *Journal of Public Health Dentistry* **65**, 90–96.

Caplan, D. and Weintraub, J. (1997). Factors related to loss of root canal filled teeth. *Journal of Public Health Dentistry* **57**, 31–39.

Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society, Series B* **63**, 791–809.

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

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* **148**, 82–117.

Clegg, L. X., Cai, J. and Sen, P. K. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* **55**, 805–812.

Cochran, W. G. (1977). *Sampling Techniques.* Wiley, New York, 3rd edition.

Cook, J., Lipschitz, D., Miles, L. and Finch, C. (1974). Serum ferritin as a measure of iron stores in normal subjects. *American Journal of Clinical Nutrition* **27**, 681–687.

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

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman & Hall, London.

Cullen, K. J. (1972). Mass health examinations in the busselton population, 1966 to 1970. *Australian Journal of Medicine* **2**, 714–718.

Flemming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

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

Goldstein, L. and Langholz, B. (1992). Asymptotic theory for nested case-control sampling in the cox regression model. *The Annals of Statistics* **20**, 1903–1928.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hungar. Acad. Sci.* **5**, 361–374.

Horvitz, D. G. and Thompson, D. J. (1951). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, New York.

Hsu, L., Chen, L., Gorfine, M. and Malone, K. (2004). Semiparametric estimation of marginal hazard function from case-control family studies. *Biometrics* **60**, 936–944.

Hsu, L., Prentice, R., Zhao, L. and Fan, J. (1999). On dependence estimation using correlated failure time data from case-control family studies. *Biometrika* **86**, 743–753.

Huffer, F. W. and McKeague, I. W. (1999). Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* **86**, 114–129.

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

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, John & Sons, New York, 2nd edition.

Karatzas, I. and Shereve, S. E. (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics* **48**, 795–806.

Knuiman, M. W., D. M. L., Olynyk, J. K. and Cullen, D. J.and Bartholomew, H. C. (2003). Serum ferritin and cardiovascular disease: A 17-year follow-up study in busselton, western australia. *American Journal of Epidemiology* **158**, 144–149.

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.

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

Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association* **99**, 832–844.

Langholz, B. and Thomas, D. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *American Journal of Epidemiology* **131**, 169–176.

Langholz, B. and Thomas, D. (1991). Efficiency of cohort sampling designs: some surprising results. *Biometrics* **47**, 1653–1571.

Lee, E. W., Wei, L. J. and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In Klein, J. P. and Goel, P. K., editors, *Survival Analysis: State of the Art.*, pages 237 – 247. Kluwer Academic Publishers, Dordrecht.

Li, H., Yang, P. and Schwartz, A. G. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Li, H., Yang, P. and Schwartz, A. G. (1998). Analysis of age of onset data from case-control family studies. *Biometrics* **54**, 1030–1039.

Liang, K. Y., Self, S. G. and Chang, Y. (1986). Modeling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society, Series B* **55**, 441–453.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* **13**, 2233–2247.

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

87, 37–47.

Lin, D. Y., Oakes, D. and Ying, Z. (1998). Additive hazards regression for current status data. *Biometrika* **85**, 289–198.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62**, 711–730.

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

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Lin, D. Y. and Ying, Z. (1997). Additive hazards regression models for survival data. In Lin, D. Y. and Fleming, T. R., editors, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pages 185–198, New York. Springer.

Martinussen, T. and Scheike, T. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89**, 649–658.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487–493.

O'Neill, T. J. (1986). Inconsistency of the misspecified proportional hazards model. *Statistics & Probability Letter* **4**, 219–222.

Pipper, C. B. and Martinussen, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazards. *Journal of the Royal Statistical Society, Series B* **66**, 207–220.

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

Prentice, R. and Breslow, N. (1978). Retrospective studies and failure time models. *Biometrika* **65**, 153–158.

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

Rothman, K. J. (2002). *Epidemiology: An Introduction.* Oxford University Press.

Samuelsen, S. O. (1997). A pseudolikelihood approach to analysis of nested case-cohort studies. *Biometrika* **84**, 379–394.

Samuelsen, S. O., Ånestad, H. and Skrondal, A. (2005). Stratified case-cohort analysis of general cohort sampling designs. Technical Report ISSN 0806-3842, Department of Mathematics, University of Oslo.

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

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics*. Chapman & Hall, New York.

Shen, Y. and Cheng, S. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics* **55**, 1093–1100.

Shih, J. H. and Chatterjee, N. (2002). Analysis of survival data from case-control family studies. *Biometrics* **58**, 502–509.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

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.

Sullivan, J. L. (1996). Iron versus cholesterol-perspectives on the iron and heart disease debate. *Journal of Clinical Epidemiology* **159**, 1542–1548.

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

Therneau, T. M. and Grambsch, P. M. (2001). *Modeling Survival Data*. Springer, New York.

Thistead, R. (1988). *Elements of Statistical Computing*. Chapman & Hall.

Thomas, D. C. (1977). Addendum to 'methods of cohort analysis: Appraisal by application to asbestos mining'(by f. d. k. liddell, j. c. mcdonald and d. c. thomas). *Journal of the Royal Statistical Society, Series A* **140**, 483–485.

Tsiatis, A. A. (1981). A large sample study of cox's regression mmodel. *The Annals of Statistics* **9**, 93–108.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wacholder, S., Gail, M. and Pee, D. (1991). Efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics* **47**, 63–76.

Wacholder, S., Gail, M. H., Pee, D. and Brookmeyer, R. (1989). Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* **76**, 117–123.

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

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data.

Biometrika **91**, 801–818.