MORE EFFICIENT ESTIMATORS FOR CASE-COHORT STUDIES WITH UNIVARIATE AND MULTIVARIATE FAILURE TIMES

Soyoung Kim

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.

Chapel Hill
2013

Approved by:

Dr. Jianwen Cai

Dr. David Couper

Dr. Ka He

Dr. Wenbin Lu

Dr. Donglin Zeng

Dr. Haibo Zhou

Abstract

SOYOUNG KIM : More efficient estimators for case-cohort studies with univariate and multivariate failure times
(Under the direction of Dr. Jianwen Cai)

Case-cohort study design is generally used to reduce cost in large cohort studies when the disease rate is low. The case-cohort design consists of a random sample of the entire cohort, named subcohort, and all the subjects with the disease of interest. When the rate of disease is not low or the number of cases are not small, the generalized case-cohort study which selects subset of all cases is used. In this dissertation, we study more efficient estimators of multiplicative hazards models and additive hazards models for the traditional case-cohort study as well as the generalized case-cohort study.

We first study more efficient estimators for the traditional case-cohort studies with rare diseases. When several diseases are of interest, several case-cohort studies are usually conducted using the same subcohort. When these case-cohort data are analyzed, the common practice is to analyze each disease separately ignoring data collected in subjects with the other diseases. This is not an efficient use of the data. In this study, we propose more efficient estimators by using all available information. We consider both joint analysis of the multiple diseases and separate analysis for each disease. We propose an estimating equation approach with a new weight function. We establish that the proposed estimator is consistent and asymptotically normally distributed. Simulation studies show that the proposed methods using all available information gain efficiency. For comparing the effect of the exposure on different diseases, tests based on the joint analysis are more powerful than those based on the separate analysis assuming independence. We apply our proposed method to the data from the Busselton Health Study.

We extend this approach to the stratified case-cohort design with non-rare diseases.

We also consider the additive hazards regression model for the stratified case-cohort studies. Additive hazards model is more appropriate when risk difference is of interest. 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. We propose an estimating equation approach for parameter estimation in additive hazards regression model by making full use of available information. Asymptotic properties of the proposed estimators were developed and simulation studies were conducted. We apply our proposed methods to data from the Atherosclerosis Risk in Communities (ARIC) study.

Acknowledgments

I would like to thank my dissertation advisor, Dr. Jianwen Cai for supporting me during past three years. I have learned the lesson and gained the experience throughout my dissertation research process. She has set an example of excellence as a researcher, mentor, instructor, and role model.

I also would like to thank the committee members, Dr. David Couper, Dr. Ka He, Dr. Wenbin Lu, Dr. Donglin Zeng, and Dr. Haibo Zhou for all of their guidance through this process.

Table of Contents

List of Tables

Chapter 1

Introduction

In large epidemiologic cohort studies, several thousands of subjects are usually followed for many years and such studies can be expensive. Most of the cost and effort involve the assembly of the covariate information for all cohort members. However, if the disease is rare, much of the covariate information on disease free subjects is largely redundant [Prentice, 1986]. In order to reduce the high cost, Prentice [1986] proposed the case-cohort design. Under the case-cohort study design, the covariate histories are collected only for subjects in a randomly selected sample, named subcohort, from the entire cohort and all the cases (i.e. the subjects with the event of interest). In this dissertation, we develop statistical methods for case-cohort study design with univariate and multivariate failure time data.

One important advantage of the case-cohort study design is that the same subcohort can be used for studying different diseases, whereas for other designs such as the nested case-control design, new matching of cases and controls needs to be done for different diseases [Wacholder et al., 1991; Langholz and Thomas, 1990].

For example, in the Busselton Health Study [Cullen, 1972] two case-cohort studies were conducted. The purpose of this study is to investigate the effect of serum ferritin on coronary heart disease and stroke, respectively. Serum ferritin was measured on a random sample of the cohort as well as all subjects with coronary heart disease and/or stroke. The existing methods do not use the covariate information collected on subjects with stroke when studying the serum ferritin effect on coronary heart disease and vice versa. This is not an efficient use of available resources and new statistical methods which use all available exposure information is needed.

The case-cohort study design was originally proposed to reduce the cost in the cohort study when the disease of interest is rare. Consequently, the traditional case-cohort sampling involves all the cases (i.e. the subjects with the event of interest). In recent years, in order to preserve the raw material collected in the study, case-cohort study design is also used in situations when the disease is not rare. In such studies, it is not desirable to conduct the traditional case-cohort studies which collect the expansive covariate information on all cases. Sampling only a fraction of the cases is more practical [Breslow and Wellner, 2007; Cai and Zeng, 2007; Kang and Cai, 2009]. Existing methods do not make the full use of all available information about all diseases from the generalized case-cohort studies and a correlate of the exposure available for all cohort members. It is desirable to develop new statistical methods which use all available information.

There are two principal frameworks for modeling risks: the multiplicative and additive risks model. Much work for the case-cohort studies were on multiplicative risks models using proportional hazards models. However, the multiplicative risks model is not always applicable in biomedical studies. Furthermore, the researchers could be interested in the risk difference attributed to the exposure. The risk difference is more related to public health since it translates directly into the number of disease cases that would be avoided by eliminating a particular exposure [Kulich and Lin, 2000b]. Under such situation, the additive hazards model would be more appropriate. It will be important to develop statistical methods for the additive hazards model using all available information from case-cohort or generalized case-cohort studies.

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 both univariate and multivariate survival data from cohort studies, case-cohort studies, and case-control studies. The rest of this chapter is organized as follows. We review literature on statistical methods in cohort studies for univariate failure time in section 2.1 and multivariate failure time in section 2.2. In section 2.3, we review literature on statistical methods for case-cohort studies.

2.1 Univariate failure time from cohort studies

In subsection 2.1.1, we first review the Cox proportional hazards model, the most popular model for survival analysis with a single failure time. We review the literature on survival analysis for additive hazards models from cohort studies in subsection 2.1.2.

2.1.1 The Cox proportional hazards model

The Cox proportional hazards model [Cox, 1972] is the most commonly used method in survival analysis to examine the relationship between the effects of covariates and the failure time. The Cox proportional hazards model specifies the hazard rate for failure time T for a given covariate vector Z. Specifically, the Cox model is given by

$$\lambda\{t|Z\} = \lambda_0(t)e^{\beta_0^T Z(t)}, \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β_0 is a p-dimensional fixed and unknown parameter vector.

Let T_i be the failure time, C_i denote the potential censoring time, and $X_i = \min(T_i, C_i)$ denote the observed time for subject i. Let $Y_i(t) = I(X_i \geq t)$ be an at risk indicator and $\Delta_i = I(T_i \leq C_i)$ be failure indicator where $I(.)$ is the indicator function for subject i. Let $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ denote the observed counting process for failure for subject i. Suppose that there are n independent subjects and τ denotes the end of study time.

The partial likelihood score function introduced by Cox [1975] is given by

$$U_1(\beta) = \sum_{i=1}^{n} \left\{ Z_i(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\} \Delta_i,$$

or equivalently using counting process form

$$U_2(\beta) = \sum_{i=1}^{n} \int_0^{\tau} Z_i(u) dN_i(u) - \int_0^{\tau} \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} d\bar{N}_i(u),$$

where

$$\bar{N}_i(t) = \sum_{i=1}^{n} N_i(t), \quad S^{(0)}(\beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)}, \quad S^{(1)}(\beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) Z_i(t) e^{\beta' Z_i(t)}.$$

The regression parameter β can be estimated by solving the score equation $U_2(\beta) = 0$. We denote the solution by $\hat{\beta}$. Under some regularity conditions, $\hat{\beta}$ has been shown to be consistent and follow a normal distribution with mean β_0 and covariance matrix Σ given by

$$\Sigma = \int_0^{\tau} v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt,$$

where $v(\beta_0, t) = s^{(2)}(\beta_0, t)/s^{(0)}(\beta_0, t) - \{s^{(1)}(\beta_0, t)/s^{(0)}(\beta_0, t)\}^{\otimes 2}$, $s^{(d)}(\beta_0, t) = E[S^{(d)}(\beta_0, t)]$ for $d = 0, 1, 2$, and $S^{(2)}(\beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) Z_i(t)^{\otimes 2} e^{\beta' Z_i(t)}$. The asymptotic variance Σ can be estimated by $\widehat{\Sigma} = -\{\partial U_2(\beta)/\partial \beta|_{\beta = \hat{\beta}}\}$ [Andersen and Gill, 1982].

2.1.2 Additive hazards model

Another framework commonly used for regression with censored failure time is the additive hazards model. Much work has been conducted under the assumption of multiplicative hazards models. However, epidemiologists often are interested in the risk difference. Risk difference is another measure of association. It is very relevant to public health decisions, because it translates directly to the expected number of disease cases that would be prevented in the population by removing a certain exposure [Kulich and Lin, 2000b]. When the risk difference is of interest, the additive hazards model is very useful.

The additive hazards model takes the following form:

$$\lambda_a(t; Z) = \lambda_{a0}(t) + \beta_{a0}' Z(t), \tag{2.2}$$

where $Z(.)$ is a p-vector of possibly time-varying covariates, β_{a0} is a p-vector of regression parameters, and $\lambda_{a0}(t)$ is an unspecified baseline hazard function. The regression parameter of the additive hazards model represents the risk difference for one unit change in the covariate while adjusting for the other covariates in the model. Lin and Ying [1994] proposed estimators under model (2.2) and studied the asymptotic properties of the estimators. Mimicking the partial likelihood score function for the proportional hazards model, the estimating function to estimate β_{a0} in (2.2) is given by

$$U_a(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i(t) - \bar{Z}_a(t) \right\} \left\{ dN_i(t) - Y_i(t) \beta_a' Z_i(t) dt \right\},$$

where $\bar{Z}_a(t) = \sum_{j=1}^{n} Y_j(t) Z_j(t) / \sum_{j=1}^{n} Y_j(t)$. The estimator $\hat{\beta}_a$ is defined as the solution to $U(\beta) = 0$ and takes the explicit form

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

Under some regularity conditions, Lin and Ying [1994] showed the random vector $n^{-1/2}(\hat{\beta}_a - \beta_0)$ converges weakly to a p-variate normal distribution with mean zero and with a covariance matrix which can be consistently estimated by $A^{-1}BA^{-1}$, where

$$A = n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t) - \bar{Z}_a(t)\right\}^{\otimes 2} dt \quad \text{and} \quad B = n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i(t) - \bar{Z}_a(t)\right\}^{\otimes 2} dN_i(t).$$

Lin and Ying [1994] also proposed the estimator for the baseline cumulative hazard function:

$$\hat{\Lambda}_{a0}(\hat{\beta}_a, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\hat{\beta}_a' Z_i(u)du\}}{\sum_{j=1}^n Y_j(t)}.$$

They also showed that $n^{1/2}\{\hat{\Lambda}_{a0}(\hat{\beta}, .) - \Lambda_0(.)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at $(t,s)(t \geq s)$ can be consistently estimated by

$$\int_0^s \frac{n\sum_{i=1}^n dN_i(u)}{\{\sum_{j=1}^n Y_j(u)\}^2} + C'(t)A^{-1}BA^{-1}C(s) - C'(t)A^{-1}D(s) - C'(s)A^{-1}D(t),$$

where $C(t) = \int_0^t \bar{Z}_a(u)du$ and $D(t) = \int_0^t \frac{\sum_{i=1}^n \{Z_i(u) - \bar{Z}_a(u)\}dN_i(u)}{\sum_{j=1}^n Y_j(u)}$.

2.2 Multivariate failure time from cohort studies

In section 2.2, we review the literature on survival analysis for multivariate failure time data. Several approaches dealing with multiple failure times or recurrent event data have been proposed. We review literature on statistical methods for multiplicative hazards models in subsection 2.2.1 and additive hazards models in subsection 2.2.2.

2.2.1 Multiplicative risk models

There are in general two types of commonly used models dealing with correlated failure times: 1) marginal models, 2) frailty models. The marginal model approach does not specify the form of the dependence among correlated failure times while the frailty model approach formulates the exact nature of dependence among correlated failure times through an unobservable random variable.

Marginal models

Several authors have studied marginal models. Wei et al. [1989] proposed semiparametric methods for each marginal distribution of failure times with Cox-type proportional hazards form.

Let $Z_{ik}(t) = (Z_{i1k}(t), \ldots, Z_{ipk}(t))'$ be the covariate vector for the ith subject and the kth failure type and $\beta_k = (\beta_{1k}, \ldots, \beta_{pk})'$ be the failure-specific regression parameter. Under the failure-specific model, the hazard function for the ith subject and the kth failure type is given by

$$\lambda_k\{t|Z_{ik}(t)\} = \lambda_{0k}(t)e^{\beta_k Z_{ik}(t)}, \tag{2.3}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for $k = 1, \ldots, K$.

The kth failure-specific partial likelihood function [Cox, 1972, 1975] is

$$L_k(\beta_k) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta_k' Z_{ik}(X_{ik})\}}{\sum_{l \in \mathcal{R}_k(X_{ik})} \exp\{\beta_k' Z_{lk}(X_{ik})\}} \right]^{\Delta_{ik}},$$

where $\mathcal{R}_k(t) = \{l : X_{lk} \geq t\}$ 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 $\hat{\beta}_k$ is defined as the solution to the partial likelihood equation $\partial \log L_k(\beta_k)/\partial\beta_k = 0$ and they are generally correlated. Under some regularity conditions, it is shown that $\hat{\beta}_k$ is a consistent estimator for β_k and $n^{1/2}(\hat{\beta}'_1 - \beta_1', \ldots, \hat{\beta}'_K - \beta_K')$ converges in distribution to a zero-mean normal random vector with covariance matrix Q where

$$Q = \begin{bmatrix} H_{11}(\beta_1, \beta_1) & \cdots & H_{1K}(\beta_1, \beta_K) \\ \vdots & & \vdots \\ H_{K1}(\beta_K, \beta_1) & \cdots & H_{KK}(\beta_K, \beta_K) \end{bmatrix},$$

with

$$
\begin{aligned}
&H_{kl}(\beta_k, \beta_l) = A_k^{-1}(\beta_k)E\{w_{1k}(\beta_k)w_{1l}(\beta_l)'\}A_l^{-1}(\beta_l),\\
&A_k(\beta_k) = \int_0^\tau v_k(\beta_k, t)s_k^{(0)}(\beta_k, t)\lambda_{0k}(t)dt,\\
&v_k(\beta, t) = s_k^{(2)}(\beta, t)/s_k^{(0)}(\beta, t) - \{s_k^{(1)}(\beta_k, t)/s_k^{(0)}(\beta_k, t)\}^{\otimes 2},\\
&s_k^{(d)}(\beta, t) = E[Y_{1k}(t)Z_{1k}(t)^{\otimes d}\exp\{\beta' Z_{1k}(t)\}],\\
&w_{ik}(\beta_k) = \int_0^\infty \{Z_{ik}(t) - s_k^{(1)}(\beta_k, t)/s_k^{(0)}(\beta_k, t)\}dM_{ik}(t), \text{ and}\\
&M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(t)\lambda_{ik}(u)du.
\end{aligned}
$$

Spiekerman and Lin [1998] and Clegg et al. [1999] extended the models proposed by Wei et al. [1989] to formulate the general form by allowing for exchangeable failure time of each distinct failure type in the cluster. Suppose that there are J clusters, K distinct failure types, each of which consists of L exchangeable failure times. Let T_{jkl} denote the failure time and C_{jkl} the censoring time, and $X_{jkl} = \min(T_{jkl}, C_{jkl})$ the observed time for component l of disease k in cluster j. Let $Y_{jkl}(t) = I(X_{jkl} \geq t)$ be an at risk indicator, $\Delta_{jkl} = I(T_{jkl} \leq C_{jkl})$ be failure indicator where $I(.)$ is the indicator function and $N_{jkl}(t) = I(X_{jkl} \leq t, \Delta_{jkl} = 1)$ be the observed counting process for failure for component l of disease k in cluster j. Specifically, the following model for the lth component of the kth type of failure is considered:

$$
\lambda_{kl}\{t|Z_{kl}(t)\} = \lambda_{0k}(t)e^{\beta_0^T Z_{kl}(t)},
$$

where $\lambda_{0k}(t)(k = 1, \ldots, K)$ are unspecified baseline hazard functions.

The pseudo-partial score function is given by

$$
U_{SL}(\beta) = \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L \int_0^\tau \{Z_{jkl}(u) - \frac{S_k^{(1)}(\beta, u)}{S_k^{(0)}(\beta, u)}\}dN_{jkl}(u),
$$

where Z_{jkl} is the covariate vector for the lth component of the kth failure type in the jth cluster and $S_k^{(d)}(\beta, t) = J^{-1}\sum_{j=1}^J \sum_{l=1}^L Y_{jkl}(t)Z_{jkl}(t)^{\otimes d}e^{\beta' Z_{jkl}(t)}$ $d = 0, 1$.

The maximum pseudo-partial-likelihood estimator $\hat{\beta}_{SL}$ for β_0 is defined as the solution to $U_{SL}(\beta) = 0$ and $n^{1/2}(\hat{\beta}_{SL} - \beta_0)$ is shown to converge weakly to a p-variate normal vector with mean 0 and covariance matrix $\Omega = A^{-1}B_{SL}A^{-1}$ where $A = \sum_{k=1}^{K} A_k$ and $B_{SL} = E[\int_0^\tau \{Z_{jkl}(u) - \frac{S_k^{(1)}(\beta_0,u)}{S_k^{(0)}(\beta_0,u)}\} dM_{jkl}(u)]^{\otimes 2}$. In addition, Spiekerman and Lin [1998] showed the uniform convergence and joint weak convergence of the Aalen-Breslow type estimators for the cumulative baseline hazard functions $\hat{\Lambda}_{0k}(t,\beta)$ where

$$\hat{\Lambda}_{0k}(t,\hat{\beta}) = \int_0^t \frac{dN_{.k.}(u)}{nS_k^{(0)}(\hat{\beta},u)}.$$

Frailty models

Marginal model approaches are appropriate when the main interest is to estimate the effects of risk factors while the correlation among failure times is considered as a nuisance. However, when the correlation among failure times is of interest, an alternative approach is needed. Frailty models have been proposed under such situation. Frailty model specifies the intra-subject correlation explicitly through an unobservable random variable (frailty). Specifically, the failure times given the frailty are assumed independent and the conditional hazard given the frailty W_i is assumed to follow the following model:

$$\lambda_{ik}(t|W_i) = W_i \lambda_0(t) \exp\{\beta_0^T Z_{ik}(t)\},$$

where W_i, $i = 1, \ldots, n$, are assumed to be independent and follow a probability distribution is often assumed for the frailty distribution. Other distributions such as the positive stable distribution, the inverse Gaussian distribution, or the log-normal distribution have also been proposed.

Frailty models have been studied by many authors. In approaches for nonparametric maximum likelihood, Klein [1992] proposed the estimation of the frailty by using an EM algorithm based on a partial likelihood. As an alternative of a partial likelihood, a penalized likelihood procedure is used by Therneau and Grambsch [2000] who showed an exact connection between the shared gamma frailty model and a penalized likelihood procedure.

Ripatti and Palmgren [2000] generalized the results of Therneau and Grambsch [2000] by assuming the frailties from a log-normal distribution and thus they got a flexible specification of variance components which can explain negative dependencies.

2.2.2 Additive risk models

The previous subsection discussed multiplicative hazards models. In this subsection, we will review additive hazards models with multivariate failure time data from cohort studies.

When all the failure times are independent, several authors have studied additive hazards models from cohort studies. Martinussen and Scheike [2002] and Lin et al. [1998] has applied the additive hazards model to interval censored data. Moreover, the additive hazards model has been applied to measurement error problems by Kulich and Lin [2000b], to frailty models by Lin and Ying [1997], and to cumulative incidence rates by Shen and Cheng [1999].

For correlated or clustered data, marginal additive hazards models are proposed by Yin and Cai [2004]. They proposed the additive hazards model

$$\Lambda_{jki}(t; Z_{jki}) = \lambda_{0k}(t) + \beta_{0k}' Z_{ikl}(t),$$

where $Z_{jki}(t)$ is a possibly time-varying covariate vector for failure type k of subject i in cluster j. An estimating function for β_{0k} is

$$U_k^A(\beta) = \sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_{jki}(t) - \bar{Z}_k^A(t)\right\}\left\{dN_{jki}(t) - Y_{jki}(t)\beta' Z_{jki}(t)dt\right\},$$

where $\bar{Z}_k^A(t) = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n} Y_{jki}(t)Z_{jki}(t)}{\sum_{j=1}^{J}\sum_{i=1}^{n} Y_{jki}(t)}$. The estimator $\hat{\beta}_k^A$ is defined as the solution to $U_k^A(\beta) = 0$, which is given by

$$\hat{\beta}_k^A = \left[\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^{\tau} Y_{jki}(t)\{Z_{jki}(t) - \bar{Z}_k^A(t)\}^{\otimes 2}dt\right]^{-1}\left[\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^{\tau}\{Z_{jki}(t) - \bar{Z}_k^A(t)\}dN_{jki}(t)\right].$$

Under some regularity conditions, $n^{1/2}(\hat{\beta}_1^{A'} - \beta_{01}', \ldots, \hat{\beta}_K^{A'} - \beta_{0K}')'$ was shown to converge in distribution to a zero-mean $(p \times K)$-dimensional normal random vector with a covariance vector $D^A_{jk}(\beta'_{0j}, \beta'_{0k}) = A_j^{-1} E[U^A_{1j}(\beta_{0j}) U^A_{1k}(\beta_{0k})]\ A_j^{-1}$ where $A_j = E[\sum_{l=1}^{L} \int_0^\tau Y_{1jl}(t)\{Z_{1jl}(t) - \bar{Z}_j(t)\}^{\otimes 2} dt]$. Under the working independence assumption, the baseline cumulative hazard function for failure type k can be estimated by

$$\hat{\Lambda}^A_{0k}(t; \hat{\beta}_k) = \int_0^\tau \frac{\sum_{j=1}^J \sum_{i=1}^n \{dN_{jki}(u) - Y_{jki}(u) \hat{\beta}_k^{A\prime} Z_{jki}(u) du\}}{\sum_{j=1}^J \sum_{i=1}^n Y_{jki}(u)}.$$

Under some regularity conditions, as $n \to \infty$, $n^{1/2}[\{\hat{\Lambda}^A_{01}(t) - \Lambda_{01}(t)\}, \ldots, \{\hat{\Lambda}^A_{0K}(t) - \Lambda_{0K}(t)\}]$ was shown to converge weakly to a zero-mean Gaussian random field. For a specific subject with the covariate vector $Z_0(t)$, the cumulative hazard function can be estimated by $\hat{\Lambda}^A(t; \hat{\beta}^A_k, Z_0) = \hat{\Lambda}^A_{0k}(t; \hat{\beta}^A_k) + \int_0^t \hat{\beta}_k^{A\prime} Z_0(u) du$. To ensure monotonicity, a minor modification was made, i.e. $\hat{\Lambda}^*_{0k}(t) = \max_{s \le t} \hat{\Lambda}^A_{0k}(s)$ for $k = 1, \ldots, K$. By similar arguments as in Lin and Ying [1994], it can be shown that $\hat{\Lambda}^*_{0k}(t)$ and $\hat{\Lambda}^A_{0k}(t)$ are asymptotically equivalent.

Pipper and Martinusse [2004] also considered marginal additive hazards models for clustered data. By using Lin and Ying [1994]'s estimators, they provided estimating equations for the regression parameters and association parameters for marginal additive hazards models. Further, Yin [2007] developed a test for checking the additive structure using clustered data. By relaxing the linear assumption about covariate effects, Zeng and Cai [2010] proposed a general class of additive transformation risk models for clustered failure time data.

2.3 Case-cohort studies

2.3.1 Case-cohort studies vs nested case-control studies

In large epidemiologic cohort studies, several thousands of subjects are followed and thus such studies can be expensive. To reduce the cost in large cohort studies, several study designs have been proposed. Among different sampling schemes, nested case-control study design and case-cohort study design are widely used when the disease rate is low. In this subsection, we will review the literature on nested case-control study design and case-cohort study design.

Thomas [1977] originally suggested nested case-control study design which involves selection of a number of controls from those at risk at the failure time of each case. Prentice and Breslow [1978] further developed the conceptual foundations of the nested case-control design by deriving the conditional likelihood. However, there are some limitations in the nested case-control studies: inefficiency for the alignment of each selected control subject to its matched case and a strict application which involves the selection of a new set of controls for each distinct disease category.

To address the problems, case-cohort study design was proposed by Prentice [1986] as an alternative to the nested case-control study design. Case-cohort study design involves selection of a random sample, named subcohort, and all cases. The subcohort 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 [Prentice, 1986].

Langholz and Thomas [1990] compared case-cohort studies with nested case-control studies. They showed that the nested case-control approach is better than the case-cohort study if there is moderate random censoring or staggered entry. It also has been shown that case-cohort study design for a single disease outcome has higher efficiency than nested case-control study design; however, the difference is very small. Compared to the nested case-control studies, a major advantage of the case-cohort design is the ability to study several disease outcomes using the same subcohort.

We will review the literature for case-cohort studies with univariate failure time in subsection 2.3.2 and multivariate failure time in subsection 2.3.3.

2.3.2 Univariate failure time

Prentice [1986] proposed a case-cohort design and established asymptotic properties of their proposed estimators. He considered a relative risk regression model [Cox, 1972]:

$$\lambda\{t|Z(u), 0 \leq u < t\} = \lambda_0(t) r\{\beta_0' Z_i(t)\}, \tag{2.4}$$

where $r(x)$ is a fixed function with $r(0) = 1$, β_0 is a p-vector of regression parameters, and

$\lambda_0(t)$ is a baseline hazard function.

Prentice [1986] proposed the pseudolikelihood function for estimation of the relative risk parameter β_0 in case-cohort studies given by

$$\widetilde{L}(\beta) = \prod_{i=1}^{n} \left(r_{ii} / \sum_{l \in \tilde{\mathcal{R}}(t_i)} r_{li} \right)^{\Delta_i}$$

where $r_{li} = Y_l(t_i) r\{\beta_0' Z_l(t_i)\}$, $\tilde{\mathcal{R}}(t_i) = F(t) \cup \tilde{C}$, $F(t) = \{i | N_i(t) \neq N_i(t^-)\}$, and $\tilde{C}$ is a random subcohort.

The maximum pseudolikelihood estimator $\tilde{\beta}_{CC}$ is defined as a solution to $U_{CC}(\tilde{\beta}) = 0$ where

$$U_{CC}(\beta) = \frac{\partial \log \widetilde{L}(\beta)}{\partial \beta} = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} \Delta_i \left(c_{ii} - \sum_{l \in \tilde{\mathcal{R}}(t_i)} b_{li} / \sum_{l \in \tilde{\mathcal{R}}(t_i)} r_{li} \right),$$

$b_{li} = Y_l(t_i) Z_l(t_i) r'\{\beta^T Z_l(t_i)\}$, $c_{li} = b_{li} r^{-1}\{\beta^T Z_l(t_i)\}$, and $r'(u) = dr(u)/du$. Under some regularity conditions, Prentice [1986] reasoned that $n^{-1/2} U_{CC}(\beta)$ converge weakly to a normal variate with mean zero and variance matrix A and that $n^{1/2}(\tilde{\beta}_{CC} - \beta_0)$ converges in distribution to a normal variate with mean zero and variance matrix $S = \Omega^{-1} A \Omega^{-1}$ which can be estimated by $nI(\tilde{\beta})^{-1} \tilde{V}(\tilde{\beta}) I(\tilde{\beta})^{-1}$ where

$$\begin{aligned}
I(\beta) &= -\frac{\partial^2 \log \widetilde{L}(\beta)}{\partial \beta \partial \beta^T}, \\
\tilde{V}(\beta) &= \sum_{j=1}^{n} \Delta_j \{ v_{jj} + 2\delta(t_j) \sum_{\{k | t_k < t_j\}} \Delta_k v_{kj} \}, \\
v_{kj} &= -\sum \left(\frac{B_k + b_{jk} - b_{ik}}{R_k + r_{jk} - r_{ik}} \right)' \left(c_{ij} - \frac{B_j}{R_j} \right) r_{ij} R_j^{-1}, \\
R_j &= \sum_{l \in \tilde{\mathcal{R}}(t_j)} r_{lj}, \quad B_j = \sum_{l \in \tilde{\mathcal{R}}(t_j)} b_{lj}, \quad \delta(t) = 0 \text{ if } \tilde{C} = \tilde{\mathcal{R}}(t) \text{ and } 1 \text{ otherwise.}
\end{aligned}$$

An estimator of the baseline cumulative failure rate function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is

$$\hat{\Lambda}_0(t) = \tilde{n} n^{-1} \int_0^t [\sum_{i=1}^{n} Y_i(u) r\{\beta' Z_i(u)\}]^{-1} d\bar{N}(u)$$

where $\bar{N}(t) = \sum_{i=1}^{n} N_i(t)$.

Self and Prentice [1988] proposed a slightly different estimator from Prentice [1986]. While the "comparison risk set" of Self and Prentice [1988] at time t included only all subcohort members at risk at time t, Prentice [1986] added any subjects out of subcohort but who were observed to fail at time t. Self and Prentice [1988] established asymptotic distribution theory for the pseudolikelihood estimators along with that for the corresponding cumulative failure rate estimators by using a combination of martingale and finite population convergence results. Specifically, they considered the maximum pseudolikelihood estimator $\tilde{\beta}_{SP}$, defined as a solution to $\partial \log L(\tilde{\beta})/\partial\beta = 0$, where

$$\log \widetilde{L}(\beta) = \sum_{i=1}^{n} \int_0^\tau \log r\{\beta' Z_i(t)\} dN_i(t) - \int_0^\tau \log[\sum_{i\in\tilde{C}} Y_l(t) r\{\beta' Z_i(t)\}] d\bar{N}(t),$$

and $\tilde{C}$ is a random subcohort of size $\tilde{n}$. They also considered a natural estimator of the cumulative baseline hazard function which is given by

$$\tilde{\Lambda}_{SP}(t) = \tilde{n} n^{-1} \int_0^t [\sum_{i\in\tilde{C}} Y_i(u) r\{\tilde{\beta}'_{SP} Z_i(u)\}]^{-1} d\bar{N}(u).$$

Under some regularity conditions, they showed that $\tilde{\beta}_{SP}$ is a consistent estimator of β_0 and $n^{-1/2}\tilde{U}(\beta_0)$ converges in distribution to zero mean Gaussian process with covariance matrix $\Sigma(\beta_0) + A(\beta_0)$ where $\Sigma(\beta) = -\lim_{n\to\infty} n^{-1}\partial^2 \log \widetilde{L}(\beta)/\partial\beta^2$ is the variation associated with the cohort and $A(\beta)$ corresponds to the variation introduced by sampling the subcohort. Therefore, $n^{-1/2}(\tilde{\beta}_{SP} - \beta_0)$ was shown to converge in distribution to a zero-mean Gaussian random variable with covariance matrix $\Sigma^{-1}(\beta_0) + \Sigma^{-1}(\beta_0)A(\beta_0)\Sigma^{-1}(\beta_0)$ by Taylor series expansions. Moreover, $n^{-1/2}(\tilde{\beta}_{SP} - \beta_0)$ and $n^{-1/2}(\tilde{\Lambda}_{SP} - \Lambda_0)$ were shown to converge weakly and jointly to Gaussian random variables with mean zero. They also proposed the estimator of the limiting covariance matrix between $n^{-1/2}\{\tilde{\Lambda}_{SP}(u) - \Lambda_0(u)\}$ and $n^{-1/2}\{\tilde{\Lambda}_{SP}(t) - \Lambda_0(t)\}$.

Self and Prentice [1988] showed that Prentice [1986]'s estimator $\tilde{\beta}$ and their estimator $\tilde{\beta}_{SP}$ are asymptotically equivalent by showing that an individual's contributions to $S^{(1)}$ and $S^{(0)}$ are asymptotically negligible. Even though Prentice [1986]'s variance estimator is somewhat different from Self and Prentice [1988]'s one, two estimators converge to the same form asymptotically.

Alternative variance estimators which can be computed easily using the existing software are proposed since the variance estimators by Prentice [1986] and Self and Prentice [1988] are complicated. Wacholder et al. [1989] developed bootstrap variance estimates. Barlow [1994] proposed a robust estimator of the variance. By using time-varying weights, he proposed a pseudolikelihood function which are different from those of Prentice [1986] and Self and Prentice [1988]. The weight $w_i(t)$ of subject i at time t is defined as

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

where $m(t)$ is the number of disease-free individuals in the cohort at risk at time t and $\tilde{m}(t)$ is the number of disease-free individuals in subcohort at risk at time t. The conditional probability of failure at failure 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 $r_i(t) = \exp\{\beta_0^T Z_i(t)\}$. Prentice [1986]'s likelihood used an indicator function as a weight, i.e., $w_i(t) = 1$ if $dN_i(t) = 1$ or $i \in \tilde{C}$, otherwise the weight is zero. Whereas Self and Prentice [1988]'s likelihood used a denominator summed over subcohort members only, Barlow [1994]'s pseudolikelihood preserved the correct expectation for the denominator at each failure time.

The estimator $\hat{\beta}_B$ proposed by Barlow [1994] is defined as the solution to the estimating equation defined by the derivative of the logarithm of the pseudolikelihood $\sum_t \sum_i dN_i(t)$ $\log(p_i(t))$. The robust variance estimator using infinitesimal jackknife estimator is

$$V\hat{a}r(\hat{\beta}_B) = I^{-1}(\hat{\beta}_B)\hat{V}(\hat{\beta}_B)I^{-1}(\hat{\beta}_B) = \frac{1}{n}\sum_i \hat{e}_i\hat{e}_i,$$

where $\hat{e}_i = \hat{\beta} - \hat{\beta}_{-i}$ and $\hat{\beta}_{-i}$ is an estimate of β without observation i. Barlow [1994] proposed to estimate $\hat{e}_i$ by $I^{-1}(\hat{\beta})\hat{c}_i(\tau)$ where $I(\hat{\beta}_B) = \sum_t \sum_i \hat{p}_i(t)[z_i(t) - \hat{E}(t)][z_i(t) - \hat{E}(t)]'$ is the information matrix, $\hat{E}(t) = \sum_{k=1}^{n} \hat{p}_i(t) Z_i(t)$ is the estimator for the conditional expectation of the covariate at time t, and $\hat{c}_i(\tau) = \int_0^\tau Y_i(t)[dN_i(t) - \hat{p}_i(t)][z_i(t) - \hat{E}(t)]d\bar{N}(t)$ is the estimated influence of an individual observation on the overall score for subject i at time τ.

Stratified case-cohort studies were discussed in Prentice [1986]. Borgan et al. [2000] developed methods for analysis of such exposure stratified case-cohort samples. Suppose that the baseline data are available for the full cohort and can be partitioned into Q strata. A stratified relative risk regression model is considered:

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

A pseudolikelihood function for β over strata is

$$\tilde{L}_q(\beta) = \prod_{t_j} \left(\frac{\exp\{\beta' Z_{i_j}(t_j)\} w_{i_j}(t_j)}{\sum_{k \in \tilde{\mathcal{R}}(t_j)} Y_k(t_j) \exp\{\beta' Z_k(t_j)\} w_k(t_j)} \right),$$

where t_j is failure time, $\tilde{\mathcal{R}}(t_j)$ is case-cohort set, and $w_{i_j}(t_j)$ is weight for the case i_j at time t_j. They proposed three types of estimators for the stratified case-cohort design:

$$\begin{aligned}
I &: \tilde{\mathcal{R}}(t_j) = \tilde{C}, w_k(t_j) = n_{s(k)}/m_{s(k)}, \\
II &: \tilde{\mathcal{R}}(t_j) = \tilde{C} \cup F, w_k(t_j) = n^0_{s(k)}/m^0_{s(k)} \text{ if } k \in \tilde{C} \backslash F, w_k(t_j) = 1 \text{ if } k \in F, \\
III &: w_k(t_j) = n_{s(k)}/m_{s(k)}, \tilde{\mathcal{R}}(t_j) = \tilde{C} \text{ if } i_j \in \tilde{C}, \tilde{\mathcal{R}}(t_j) = \tilde{C} \cup i_j \backslash \{J_{s(i_j)}\} \text{ if } i_j \notin \tilde{C},
\end{aligned}$$

where $\tilde{C}$ is the subcohort set, F is a set of all cases, n_l and m_l are the number of subjects in the cohort and subcohort in stratum l, respectively, n_l^0 and m_l^0 are the number of cases in the cohort and subcohort in stratum l, respectively, and $s(k)$ is the sampling stratum of individual k. If the case occurs outside the subcohort, subcohort member $J_{s(i_j)}$ swaps place with the case so that the case i_j is inside $\tilde{\mathcal{R}}(t_j)$ while the "swapper" $J_{s(i_j)}$ is removed from this set. They showed that all of the proposed analysis methods were more efficient than a randomly sampled case-cohort study. Breslow and Wellner [2007] generalized asymptotic results of Borgan et al. [2000] by using weighted likelihood estimation in two-phase stratified sample.

Chen [2001b] proposed a unified approach which includes 1) nested case-control sampling, 2) case-cohort sampling, and 3) classical case-control designs and allow the presence of staggered entry. The estimating equation to cover three samplings is given by

$$U_{Ch}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left[Z_i(t) - \frac{\sum_{j=1}^{n} w_{ij} Z_j(t) \exp\{\beta' Z_j(t)\} Y_j(t)}{\sum_{j=1}^{n} w_{ij} \exp\{\beta' Z_j(t)\} Y_j(t)} \right] dN_i(t),$$

where w_{ij} is a weight function for the respective design. They also developed the weight function based on estimating each missing covariate by a local average. Samuelsen et al. [2007] extended the class of designs proposed by Chen [2001b] to accommodate stratified designs.

All work that we discussed in this subsection so far was about proportional hazards models for case-cohort studies. Other type of models have also been studied. The accelerated failure time model and the proportional odds regression model for case-cohort are proposed [Kong and Cai, 2009; Chen, 2001a]. Kulich and Lin [2000a] applied additive hazards models to case-cohort studies. The model they considered is in the same form as (2.2). The subcohort can be selected by Bernoulli sampling with arbitrary selection probabilities or by stratified simple random sampling. Using Bernoulli sampling, they proposed a weighted estimating function:

$$U_H(\beta) = \sum_{i=1}^{n} \rho_i \int_0^{\tau} \left\{ Z_i(t) - \bar{Z}_H(t) \right\} \left\{ dN_i(t) - Y_i(t) \beta^T Z_i(t) dt \right\},$$

where $\bar{Z}_H(t) = \frac{\sum_{j=1}^{n} \rho_j Y_j(t) Z_j(t)}{\sum_{j=1}^{n} \rho_j Y_j(t)}$ and the weight function ρ_i has the following form: $\rho_i = \Delta_i + (1 - \Delta_i)\xi_i \hat{\alpha}^{-1}$, and $\hat{\alpha} = \sum_{i=1}^{n} \xi_i (1 - \Delta_i) / \sum_{i=1}^{n} (1 - \Delta_i)$. The estimator $\hat{\beta}_H$ is defined as a solution to $U_H(\beta) = 0$. An estimator for the cumulative baseline hazard function $\Lambda_0(t)$ is

$$\hat{\Lambda}_{0H}(t) = \int_0^{\tau} \frac{\sum_{i=1}^{n} dN_i(u)}{\sum_{j=1}^{n} \rho_j Y_j(u)} - \int_0^{\tau} \hat{\beta}_H \bar{Z}_H(u) du.$$

Under some regularity conditions, $n^{1/2}(\hat{\beta}_H - \beta_0)$ was shown to converge in distribution to a zero-mean normal random vector with covariance matrix $D_A^{-1}(\Sigma_A + \Sigma_H)D_A^{-1}$, where $D_A = \mathrm{E}\left[\int_0^\tau \{Z_1(t) - e(t)\}^{\otimes 2} Y_1(t)dt\right]$, $\Sigma_A(\beta) = \mathrm{E}\left[\int_0^\tau \{Z_1(t) - e(t)\}^{\otimes 2} dN_1(t)\right]$, $\Sigma_H(\beta) = \mathrm{E}\{(1-\tilde{\alpha})\tilde{\alpha}^{-1}(1-\Delta_1)[\int_0^\tau \{Z_1(t) - e(t)\}dM_1(t)]^{\otimes 2}\}$, $e(t) = E\{Z_1(t)Y_1(t)\}/E\{Y_1(t)\}$, and $M_i(t) = N_i(t) - \int_0^t Y_i(s)s\Lambda_0(s) - \int_0^t \beta_0^T Z_i(s)Y_i(s)ds$. They also showed that $n^{1/2}(\hat{\Lambda}_{0H}(t) - \Lambda_{0H}(t))$ converges weakly on $[0,\tau]$ to a zero-mean Gaussian process whose covariance function at (s,t) is

$$h^T(s)D_A^{-1}(\Sigma_A + \Sigma_H)D_A^{-1}h(t) + R_1(s,t) - h^T(s)D_A^{-1}R_2(t) - h^T(t)D_A^{-1}R_2(s),$$

where $R_1(s,t) = \mathrm{E}[\{\Delta_1 + (1-\Delta_1)/\tilde{\alpha}\} \int_0^s \pi_0^{-1}(u)dM_1(u) \int_0^t \pi_0^{-1}(v)dM_1(v)]$, $R_2(t) = \mathrm{E}[\int_0^t \{Z_1(u) - e(u)\}\pi_0^{-1}(u)dN_1(u)]$, $h(t) = \int_0^t e(u)du$, and $\pi_0(u) = \Pr(X_1 \geq t)$.

2.3.3 Multivariate failure time

Clustered failure time and multiple outcomes have been studied for the case-cohort design. In this subsection, we will review the related literature.

Lu and Shih [2006] considered the clustered failure time data. Conventional case-cohort studies for univariate failure time data cannot be directly applied to clustered failure time data since failure times within a cluster are correlated. Lu and Shih [2006] considered marginal proportional hazards model (2.4). Suppose there are J independent clusters, and each cluster contains n correlated subjects. The estimating function proposed by Lu and Shih [2006] is given by

$$U_{LS}(\beta) = \sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau \left[Z_{ji}(t) - E_{LS}(\beta,t)dN_{ji}(t)\right],$$

where $E_{LS}(\beta,t) = S_{LS}^{(1)}(\beta,t)/S_{LS}^{(0)}(\beta,t)$, $S_{LS}^{(d)}(\beta,t) = J^{-1}\sum_j^J \sum_{i=1}^n H_j H_{ji} Y_{ji}(t) e^{\{\beta^T Z_{ji}(t)\}}$ $\times Z_{ji}(t)^{\otimes d}$, H_j indicates whether or not cluster j is selected into the subcohort, and H_{ji} is the indicator for subject (j,i) being sampled as a potential individual in the subcohort. $\tilde{\beta}_{LS}$ can be estimated by solving $U_{LS}(\beta) = 0$. Under some regularity conditions, $\tilde{\beta}_{LS}$ was shown to be a consistent estimator of β_0. They showed that $n^{1/2}(\tilde{\beta}_{LS} - \beta_0)$ converges in distribution to a normal distribution with mean zero and with covariance matrix $A_{LS}^{-1}(\beta_0)\Omega_{LS}(\beta_0)A_{LS}^{-1}(\beta_0)$ where $\Omega_{LS}(\beta)$ consists of the variations associated with the cohort and subcohort sampling and $A_{LS}(\beta) = -\lim_{J\to\infty} \partial U_{LS}(\beta)/\partial\beta$.

Zhang et al. [2011] extended Lu and Shih [2006]'s method by proposing Bernoulli sampling and using different risk sets. Since information on all failures in the full cohort is available, failures outside the subcohort can also contribute to the risk set for independent subjects. Thus, they constructed the risk sets using the information in the subcohort as well as the information collected on future deaths whereas Lu and Shih [2006] used only subcohort subjects to construct the risk set.

Kang and Cai [2009] considered case-cohort studies with multiple disease outcomes. The marginal hazards function [Cox, 1972] is assumed to follow the model:

$$\lambda_{ik}\{t|Z_{ik}(t)\} = Y_{ik}(t)\lambda_{0k}(t)e^{\beta_0^T Z_{ik}(t)},$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease outcome k. The pseudo-partial likelihood score equation proposed by Kang and Cai [2009] is given by

$$\widehat{U}_{KC}(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \{Z_{ik}(t) - \frac{\widehat{S}_k^{(1)}(\beta,t)}{\widehat{S}_k^{(0)}(\beta,t)}\}dN_{ik}(t),$$

where $\widehat{S}_k^{(d)}(\beta,t) = n^{-1}\sum_{i=1}^n \rho_{ik}(t)Y_{ik}(t)Z_{ik}^{\otimes d}(t)e^{\beta^T Z_{ik}(t)}$ for $d = 0$,1 and 2, $\rho_{ik}(t) = \Delta_{ik} + (1-\Delta_{ik})\xi_i\hat{\alpha}_k^{-1}(t)$, and $\hat{\alpha}_k(t) = \sum_{i=1}^n(1-\Delta_{ik})\xi_i Y_{ik}(t)/\{\sum_{i=1}^n(1-\Delta_{ik})Y_{ik}(t)\}$. Moreover, Kang and Cai [2009] proposed a weighted estimating equation approach for estimating the parameters in the marginal hazards regression models for the multivariate failure time data from the generalized case-cohort study with multiple disease outcomes. The weighted estimating function follows as

$$\widetilde{U}_{KC}(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau w_{ik}(t)\{Z_{ik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)}\}dN_{ik}(t),$$

where $\widetilde{S}_k^{(d)}(\beta,t) = n^{-1}\sum_{i=1}^n w_{ik}(t)Y_{ik}(t)Z_{ik}^{\otimes d}(t)e^{\beta^T Z_{ik}(t)}$ for $d = 0$,1 and 2 , $w_{ik}(t) = (1-\Delta_{ik})\xi_i\hat{\alpha}_k^{-1}(t) + \Delta_{ik}\xi_i + \Delta_{ik}(1-\xi_i)\eta_{ik}\hat{q}_k^{-1}(t)$, $\hat{q}_k(t) = \sum_{i=1}^n \Delta_{ik}(1-\xi_i)\eta_{ik}Y_{ik}(t)/\{\sum_{i=1}^n \Delta_{ik}(1-$

$\xi_i)Y_{ik}(t)\}$, η_{ik} is an indicator for subject i outside the subcohort by random sampling. The estimator $\tilde{\beta}_{KC}$ is defined as solution to the equations $\widetilde{U}_{KC}(\beta) = 0$. A Breslow-Aalen-type estimator of the baseline cumulative hazard function is $\tilde{\Lambda}_{0k}(\tilde{\beta}_{KC}, t)$, where

$$\tilde{\Lambda}_{0k}(\beta, t) = \int_0^t \frac{\sum_{i=1}^n w_{ik}(u) dN_{ik}(u)}{n\tilde{S}_k^{(0)}(\beta, u)}$$

Under some regularity conditions, they showed that $\tilde{\beta}_{KC}$ is a consistent estimator of β_0 and $n^{1/2}(\tilde{\beta}_{KC} - \beta_0)$ is asymptotically normally distributed with mean zero and with variance matrix in the form

$$\Sigma_{KC}(\beta_0) = A(\beta_0)^{-1}\left\{Q(\beta_0) + \frac{1-\alpha}{\alpha}V_1(\beta_0) + (1-\alpha)\sum_{k=1}^K pr(\Delta_{1k}=1)(\frac{1-q_k}{q_k})V_{2k}(\beta_0)\right\}A(\beta_0)^{-1},$$

where

$$\begin{aligned}
A(\beta) &= \sum_{k=1}^K \int_0^\tau v_k(\beta,t)s_k^{(0)}(\beta,t)\lambda_{0k}(t)dt, Q(\beta) = E\{\sum_{k=1}^K M_{\tilde{Z},1k}(\beta)\}^{\otimes 2},\\
V_1(\beta) &= var\left(\sum_{k=1}^K (1-\Delta_{1k})\int_0^t \left[R_{1k}(\beta,t) - \frac{Y_{1k}(t)E\{(1-\Delta_{1k})R_{1k}(\beta,t)\}}{E\{(1-\Delta_{1k})Y_{1k}(\beta,t)\}}d\Lambda_{0k}(t)\right]\right),\\
V_{2k}(\beta) &= var\left[dM_{\tilde{Z},1k}(\beta) - \int_0^t Y_{1k}(t)\frac{E\{dM_{\tilde{Z},1k}(\beta)|\Delta_{1k}=1,\xi_1=0\}}{E\{Y_{1k}(t)|\Delta_{1k}=1\}}\right],\\
\tilde{Z}_{ik}(\beta,t) &= Z_{ik}(t) - e_k(\beta,t), M_{\tilde{Z},ik}(\beta) = \int_0^\tau \tilde{Z}_{ik}(\beta,t)dM_{ik}(t),\\
R_{ik}(\beta,t) &= Y_{ik}(t)\tilde{Z}_{ik}(\beta,t)e^{\beta^T Z_{ik}(t)}.
\end{aligned}$$

Competing risks have also been considered for case-cohort studies with multiple diseases. Sorensen and Andersen [2000] studied competing risks models for case-cohort studies assuming proportional hazards models and considered correlation between estimated effects of exposures on the different outcomes due to re-use of the same subcohort. By studying competing risks data for case-cohort studies, the asymptotic correlation was established.

Despite progress of case-cohort studies for proportional hazards models with multiple diseases outcomes, additive hazards models for the case-cohort design with multiple diseases have been limited. The only reference is by Sun et al. [2004] which extended Kulich and Lin

[2000a]'s method to competing risks analysis for the additive hazards model. Further study on the additive hazards models for the case-cohort design with multiple diseases outcomes is needed.

In this dissertation, we will study the following three topics: (1) more efficient estimators for case-cohort studies, (2) Generalized case-cohort studies with multiple events, and (3) Additive hazards models for traditional and generalized case-cohort studies. The proposal is presented in the next chapter.

Chapter 3

More efficient estimators for case-cohort studies with rare events

3.1 Introduction

For large epidemiologic cohort studies, assembling some types of covariate information, e.g. measuring genetic information or chemical exposures from stored blood samples, for all cohort members may entail enormous cost. With cost in mind, Prentice [1986] proposed the case-cohort study design, which requires covariate information only for a random sample of the cohort, named the subcohort, as well as for all subjects with the disease of interest. One important advantage of the case-cohort study design is that the same subcohort can be used for studying different diseases, whereas for designs such as the nested case-control design, new matching of cases and controls is needed for different diseases [Langholz and Thomas, 1990; Wacholder et al., 1991].

Many methods have been proposed for case-cohort data under the proportional hazards model. Prentice [1986] and Self and Prentice [1988] studied a pseudo-likelihood approach, which is a modification of the partial likelihood method [Cox, 1975] that weights the contributions of the cases and subcohort differently. To improve the efficiency of the pseudo-likelihood estimator, Chen and Lo [1999] and Chen [2001b] studied different classes of estimating equations and used a local type of average as weight, respectively. Borgan et al. [2000] proposed using time-varying weights, and Kulich and Lin [2004] developed a class of weighted estimators by using all available covariate data for the full cohort. Breslow and Wellner [2007] considered the semiparametric model using inverse probability weighted methods with two-phase stratified samples. Various other semiparametric survival models have also been modified to accommodate case-cohort studies [e.g. Chen, 2001a; Chen and Zucker, 2009; Kong et al., 2004; Kulich and Lin, 2000a; Lu and Tsiatis, 2006].

Taking advantage of the case-cohort design, several diseases are often studied using the same subcohort. In such situations, the information on the expensive exposure measure is available on the subcohort as well as any subjects with any of the diseases of interest. For example, in the Busselton Health Study, two case-cohort studies were conducted to investigate the effect of serum ferritin on coronary heart disease and on stroke, respectively [Knuiman et al., 2003]. Serum ferritin was measured on the subcohort, a random sample of the cohort, as well as in all subjects with coronary heart disease and/or stroke. Typically, the coronary heart disease analysis would not include any exposure information collected on stroke patients not in the subcohort, and vice versa. In this paper, we develop more efficient estimators for a single disease outcome, which can effectively use all available exposure information. Because it is often of interest to compare the effect of a risk factor on different diseases, we propose a more efficient version of the Kang and Cai [2009] test of association across multiple diseases.

3.2 Model definitions and assumptions

Suppose that there are n independent subjects in a cohort study with K diseases of interest. Let T_{ik} denote the potential failure time and C_{ik} denote the potential censoring time for disease k of subject i. Let $X_{ik} = \min(T_{ik}, C_{ik})$ denote the observed time, $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ the indicator for failure, and $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)$ and $Y_{ik}(t) = I(X_{ik} \geq t)$ the counting and at-risk processes for disease k of subject i, respectively, where $I(\cdot)$ is the indicator function. Let $Z_{ik}(t)$ be a $p \times 1$ vector of possibly time-dependent covariates for disease k of subject i at time t. The time-dependent covariates are assumed to be external [Kalbfleisch and Prentice, 2002]. Let τ denote the end of study time. We assume that T_{ik} is independent of C_{ik} given the covariates Z_{ik} and follows the multiplicative intensity process [Cox, 1972]

$$\lambda_{ik}\{t \mid Z_{ik}(t)\} = Y_{ik}(t)\lambda_{0k}(t)e^{\beta_0^T Z_{ik}(t)}, \tag{3.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease k of subject i and β_0 is p-dimensional vector of fixed and unknown parameters. Model (3.1) can incorporate disease-specific effect model, $\lambda_{ik}\{t \mid Z_{ik}^*(t)\} = Y_{ik}(t)\lambda_{0k}(t)e^{\beta_k^T Z_{ik}^*(t)}$, as a special case. Specifically, we define $\beta_0^T = (\beta_1^T, \ldots, \beta_k^T, \ldots, \beta_K^T)$ and $Z_{ik}(t)^T = [0_{i1}^T, \ldots, 0_{i(k-1)}^T, \{Z_{ik}^*(t)\}^T, 0_{i(k+1)}^T, \ldots, 0_{iK}^T]$, letting 0^T be a $1 \times p$ zero vector. Then we have $\beta_0^T Z_{ik}(t) = \beta_k^T Z_{ik}^*(t)$.

Assume that there are $\tilde{n}$ subjects in the subcohort. Let ξ_i be an indicator for subcohort membership, i.e. $\xi_i = 1$ denotes that subject i is selected into the subcohort and $\xi_i = 0$ denotes otherwise. Let $\tilde{\alpha} = \text{pr}(\xi_i = 1) = \tilde{n}/n$ denote the selection probability of subject i into the subcohort. The covariates $Z_{ik}(t)$ $(0 \le t \le \tau)$ are measured for subjects in the subcohort and those with any disease of interest.

3.2.1 Estimation for univariate failure time

First, we consider the situation in which only one disease is of interest, but covariate information is available for subjects with other diseases. In the Busselton Health study, for example, this corresponds to the situation in which we are interested in the effect of serum ferritin on coronary heart disease with additional serum ferritin measurements available on subjects outside the subcohort who had stroke.

In this situation, the observable information is $\{X_{ik}, \Delta_{ik}, \xi_i, Z_{ik}(t), 0 \le t \le X_{ik}\}$ when $\xi_i = 1$ or $\Delta_{ik} = 1$, and is $(X_{ik}, \Delta_{ik}, \xi_i)$ when $\xi_i = 0$ and $\Delta_{ik} = 0$ $(k = 1, \ldots, K)$. If we are interested in disease k and ignore the covariate information collected on subjects with other diseases, we can use Borgan et al. [2000]'s estimator with time-varying weights. Specifically, the estimator is the solution to

$$\widehat{U}_k(\beta) \equiv \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_{ik}(t) - \frac{\widehat{S}_k^{(1)}(\beta, t)}{\widehat{S}_k^{(0)}(\beta, t)} \right\} dN_{ik}(t) = 0, \tag{3.2}$$

where $\widehat{S}_k^{(d)}(\beta, t) = n^{-1} \sum_{i=1}^{n} \rho_{ik}(t) Y_{ik}(t) Z_{ik}(t)^{\otimes d} e^{\beta^T Z_{ik}(t)}$ for $d = 0,1$ and 2 with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$, and the time-varying weight $\rho_{ik}(t) = \Delta_{ik} + (1 - \Delta_{ik})\xi_i \hat{\alpha}_k^{-1}(t)$ with $\hat{\alpha}_k(t) = \sum_{i=1}^{n} \xi_i (1 - \Delta_{ik}) Y_{ik}(t) / \{\sum_{i=1}^{n} (1 - \Delta_{ik}) Y_{ik}(t)\}$. Here $\hat{\alpha}_k(t)$, an estimator for the true selection probability $\tilde{\alpha}$, is the proportion of the sampled censored subjects for disease k

among censored subjects who remain in the risk set at time t for disease k. This estimator does not use the covariate information from subjects outside the subcohort who had other diseases.

To use the collected covariate information on subjects who are outside the subcohort and have other diseases, we consider the pseudo-partial likelihood score equations

$$\widetilde{U}_k(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_{ik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)} \right\} dN_{ik}(t) = 0, \tag{3.3}$$

where

$$\begin{aligned} \widetilde{S}_k^{(d)}(\beta,t) &= n^{-1} \sum_{i=1}^{n} \psi_{ik}(t) Y_{ik}(t) Z_{ik}(t)^{\otimes d} e^{\beta^T Z_{ik}(t)} \quad (d = 0,1,2), \\ \psi_{ik}(t) &= \left\{ 1 - \prod_{j=1}^{K} (1 - \Delta_{ij}) \right\} + \prod_{j=1}^{K} (1 - \Delta_{ij}) \xi_i \widetilde{\alpha}_k^{-1}(t), \end{aligned}$$

and $\widetilde{\alpha}_k(t) = \sum_{i=1}^{n} \xi_i \{\prod_{j=1}^{K} (1 - \Delta_{ij})\} Y_{ik}(t) / \sum_{i=1}^{n} \{\prod_{j=1}^{K} (1 - \Delta_{ij})\} Y_{ik}(t)$. Here $\widetilde{\alpha}_k(t)$ is the proportion of sampled subjects among subjects who do not have any diseases and are remaining in the risk set at time t. Our proposed weight for disease k is $\psi_{ik}(t) = 1$ when $\Delta_{ij} = 1$ for some j, and $\psi_{ik}(t) = \widetilde{\alpha}_k^{-1}(t)$ when $\xi_i = 1$ and $\Delta_{ij} = 0$ for all j $(j = 1, \ldots, k)$. This weight takes the failure status of the other diseases into consideration, and thus our proposed estimator will use the available covariate information for other diseases.

3.2.2 Estimation for multivariate failure time

For multivariate failure time data in case-cohort studies, Kang and Cai [2009] proposed the pseudo-likelihood score equations

$$\widehat{U}^M(\beta) \equiv \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^{\tau} \left\{ Z_{ik}(t) - \frac{\widehat{S}_k^{(1)}(\beta,t)}{\widehat{S}_k^{(0)}(\beta,t)} \right\} dN_{ik}(t) = 0, \tag{3.4}$$

with the corresponding solution denoted $\hat{\beta}^M$.

As with Borgan et al. [2000]'s estimator, when calculating the contribution of disease

k in the estimating equation, the quantity $\widehat{S}_k^{(d)}(\beta,t)$ does not use the covariate information collected on subjects with other diseases outside the subcohort. In order to improve efficiency, we consider the pseudo-likelihood score equations with new weights

$$\widetilde{U}^M(\beta) \equiv \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{ Z_{ik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)} \right\} dN_{ik}(t) = 0. \tag{3.5}$$

When there is only a single disease of interest, i.e. $K = 1$, (3.5) reduces to (3.3). Let $\widetilde{\beta}^M$ denote the solution of equation (3.5). We estimate the baseline cumulative hazard function for disease k using a Breslow–Aalen type estimator $\widetilde{\Lambda}_{0k}^M(\widetilde{\beta}^M,t)$, where

$$\widetilde{\Lambda}_{0k}^M(\beta,t) = \int_0^t \frac{\sum_{i=1}^n dN_{ik}(u)}{n\widetilde{S}_k^{(0)}(\beta,u)}. \tag{3.6}$$

3.3 Asymptotic properties

3.3.1 Asymptotic properties of $\tilde{\beta}^M$ and $\tilde{\Lambda}_{0k}^M(\tilde{\beta}^M,t)$

Because the estimators for the univariate failure time are special cases of those for the multivariate failure time, we present results only for the multivariate case. We make the following assumptions:

(a) $(T_i, C_i, Z_i, i = 1,\ldots,n)$ are independently and identically distributed, where $T_i = (T_{i1},\ldots,T_{iK})^T$, $C_i = (C_{i1},\ldots,C_{iK})^T$, and $Z_i = (Z_{i1},\ldots,Z_{iK})^T$;

(b) $\text{pr}\{Y_{ik}(t) = 1\} > 0$ for $t \in [0,\tau]$, $i = 1,\ldots,n$ and $k = 1,\ldots,K$;

(c) $|Z_{ik}(0)| + \int_0^\tau |dZ_{ik}(t)| < D_z < \infty$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$ almost surely, where D_z is a constant;

(d) for $d = 0,1,2$, there exists a neighborhood $\mathcal{B}$ of β_0 such that $s_k^{(d)}(\beta,t)$ are continuous functions and $\sup_{t\in(0,\tau),\beta\in\mathcal{B}} \|S_k^{(d)}(\beta,t) - s_k^{(d)}(\beta,t)\| \to 0$ in probability, where $S_k^{(d)}(\beta,t) = n^{-1}\sum_{i=1}^n Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}$;

(e) the matrix $A_k(\beta_0) = \int_0^\tau v_k(\beta_0,t)s_k^{(0)}(\beta_0,t)\lambda_{0k}(t)dt$ is positive definite for $k = 1,\ldots,K$, where $v_k(\beta,t) = s_k^{(2)}(\beta,t)/s_k^{(0)}(\beta,t) - e_k(\beta,t)^{\otimes 2}$ and $e_k(\beta,t) = s_k^{(1)}(\beta,t)/s_k^{(0)}(\beta,t)$;

(f) for all $\beta \in \mathcal{B}$, $t \in [0, \tau]$, and $k = 1, \dots, K$, $S_k^{(1)}(\beta, t) = \partial S_k^{(0)}(\beta, t)/\partial \beta$, and $S_k^{(2)}(\beta, t) = \partial^2 S_k^{(0)}(\beta, t)/(\partial \beta \partial \beta^T)$, where $S_k^{(d)}(\beta, t)$, $d = 0, 1, 2$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$, and $s_k^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$;

(g) for all $k = 1, \dots, K$, $\int_0^\tau \lambda_{0k}(t) dt < \infty$; and

(h) $\lim_{n \to \infty} \tilde{\alpha} = \alpha$, where $\tilde{\alpha} = \tilde{n}/n$ and α is a positive constant.

We summarize the asymptotic results in the following theorems and provide the proofs in Section 3.3.2.

Theorem 1. *Under regularity conditions (a)–(h), $\tilde{\beta}^M$ converges in probability to β_0 and $n^{1/2}(\tilde{\beta}^M - \beta_0)$ converges in distribution to a mean zero normal distribution with covariance matrix $A(\beta_0)^{-1}\Sigma(\beta_0)A(\beta_0)^{-1}$, where*

$$
\begin{aligned}
&A(\beta) = \sum_{k=1}^K A_k(\beta), \quad \Sigma(\beta) = V_I(\beta) + \frac{1-\alpha}{\alpha} V_{II}(\beta), \\
&V_I(\beta) = E\left\{\sum_{k=1}^K W_{1k}(\beta)\right\}^{\otimes 2}, \quad V_{II}(\beta) = E\left\{\sum_{k=1}^K \int_0^\tau \Omega_{1k}(\beta, t) d\Lambda_{0k}(t)\right\}^{\otimes 2}, \\
&W_{ik}(\beta) = \int_0^\tau \{Z_{ik}(t) - e_{ik}(\beta, t)\} dM_{ik}(t), \\
&\Omega_{ik}(\beta, t) = \prod_{j=1}^K (1 - \Delta_{ij}) \left[Q_{ik}(\beta, t) - \frac{Y_{ik}(t) E\{\prod_{j=1}^K (1 - \Delta_{1j}) Q_{1k}(\beta, t)\}}{E\{\prod_{j=1}^K (1 - \Delta_{1j}) Y_{1k}(t)\}} \right], \\
&Q_{ik}(\beta, t) = Y_{ik}(t)\{Z_{ik}(t) - e_k(\beta, t)\} e^{\beta^T Z_{ik}(t)}.
\end{aligned}
$$

The covariance matrix $\Sigma(\beta_0)$ consists of two parts: $V_I(\beta_0)$ is a contribution to the variance from the full cohort, and $V_{II}(\beta_0)$ is due to sampling the subcohort from the full cohort.

We summarize the asymptotic properties of the proposed baseline cumulative hazard estimator $\tilde{\Lambda}_{0k}^M(\tilde{\beta}^M, t)$ in the next theorem.

Theorem 2. *Under regularity conditions (a)–(h), $\tilde{\Lambda}_{0k}^M(\tilde{\beta}^M, t)$ is a consistent estimator of $\Lambda_{0k}(t)$ in $t \in [0, \tau]$ and $H(t) = \{H_1(t), \dots, H_K(t)\}^T = [n^{1/2}\{\tilde{\Lambda}_{01}^M(\tilde{\beta}^M, t) - \Lambda_{01}(t)\}, \dots, n^{1/2}\{\tilde{\Lambda}_{0K}^M(\tilde{\beta}^M, t) - \Lambda_{0K}(t)\}]^T$ converges weakly to the Gaussian process $\mathcal{H}(t) = \{\mathcal{H}_1(t), \dots,$*

$\mathcal{H}_K(t)\}^T$ *in* $D[0,\tau]^K$ *with mean zero and the following covariance function* $\mathcal{R}_{jk}(t,s)$ *between* $\mathcal{H}_j(t)$ *and* $\mathcal{H}_k(s)$ *for* $j \neq k$

$$\mathcal{R}_{jk}(t,s)(\beta_0) = E\{\eta_{1j}(\beta_0,t)\eta_{1k}(\beta_0,s)\} + \frac{1-\alpha}{\alpha}E\{\zeta_{1j}(\beta_0,t)\zeta_{1k}(\beta_0,s)\},$$

where

$$\eta_{ik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{K} W_{im}(\beta,t) + \int_0^t \frac{1}{s_k^{(0)}(\beta,u)}dM_{ik}(u),$$

$$\zeta_{ik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{K}\int_0^\tau \Omega_{im}(\beta,u)d\Lambda_{0m}(u)$$

$$+\prod_{j=1}^{K}(1-\Delta_{ij})\int_0^t Y_{ik}(u)\left[e^{\beta^T Z_{ik}(u)} - \frac{E\{\prod_{j=1}^{K}(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{1k}(u)\}}{E\{\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(u)\}}\right]\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta,u)},$$

and $l_k(\beta,t)^T = -\int_0^t e_k(\beta,u)d\Lambda_{0k}(u).$

3.3.2 Proofs of Theorems

Under the assumptions in Section 3.3.1, we will provide the proofs for the main theorems. We denote

$$S_k^{(d)}(\beta,t) = n^{-1}\sum_{i=1}^{n} Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}$$

$$W_{ik}(\beta) = \int_0^\tau (Z_{ik}(t) - e_{ik}(\beta,t))dM_{ik}(t),$$

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u)e^{\beta_0 Z_{ik}(u)}\lambda_{0k}(u)du,$$

$$\Omega_{ik}(\beta,t) = \prod_{j=1}^{K}(1-\Delta_{ij})\left[Q_{ik}(\beta,t) - \frac{Y_{ik}(t)E[\prod_{j=1}^{K}(1-\Delta_{1j})Q_{1k}(\beta,t)]}{E[\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t)]}\right],$$

$$Q_{ik}(\beta,t) = Y_{ik}(t)(Z_{ik}(t) - e_k(\beta,t))e^{\beta Z_{ik}(t)}$$

$$\|f\| = \sup_t |f(t)|, \quad \|d\| = \max_i |d_i|, \quad \|\mathbf{D}\| = \max_{ij}|\mathbf{D}_{ij}|$$

where f is a function, d is a vector, and $\mathbf{D}$ is a matrix.

The following lemmas play important roles for proving theorems.

lemma 1. *Let* $\mathcal{H}_n(t)$ *and* $\mathcal{W}_n(t)$ *be two sequences of bounded process. If we assume that the following conditions (1), (2), and (3) hold for some constant* τ *where*

(1) $\sup_{0\le t\le\tau} \parallel \mathcal{H}_n(t) - \mathcal{H}(t) \parallel \to_p 0$ *for some bounded process* $\mathcal{H}(t)$*,*

(2) $\mathcal{H}_n(t)$ *is monotone on* $[0,\tau]$ *and*

(3) $\mathcal{W}_n(t)$ *converges to zero-mean process with continuous sample paths, then*

$$\sup_{0\le t\le\tau} \parallel \int_0^t \{\mathcal{H}_n(s) - \mathcal{H}(s)\} d\mathcal{W}_n(s) \parallel \to_p 0, \ \sup_{0\le t\le\tau} \parallel \int_0^t \mathcal{W}_n(s) d\{\mathcal{H}_n(s) - \mathcal{H}(s)\} \parallel \to_p 0$$

The above lemma is a extension of lemma 1 from Lin et al. [2000]. To prove the asymptotic properties for case-cohort studies, the following lemma will be used frequently and is an extension of the proposition from Kulich and Lin [2000a] and details of proof is given by Lemma 2 in Kang and Cai [2010].

lemma 2. *Let* $B_i(t)$, $i = 1,\dots,n$ *be independent and identically distributed real-valued random process on* $[0,\tau]$ *and denote random process vector,* $\mathbf{B}(t) = [B_1(t),\dots,B_n(t)]$ *with* $EB_i(t) \equiv \mu_B(t)$*, var* $B_i(0) < \infty$*, and var* $B_i(\tau) < \infty$*. Let* $\xi = [\xi_1,\dots,\xi_n]$ *be random vector containing* $\tilde{n}$ *ones and* $n-\tilde{n}$ *zeros with each permutation equally likely. Let* ξ *be independent of* $\mathbf{B}(t)$*. Suppose that almost all paths of* $B_i(t)$ *have finite variation. Then* $n^{-1/2}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_B(t)\}$ *converges weakly in* $l^\infty[0,\tau]$ *to a zero-mean Gaussian process, and* $n^{-1}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_B(t)\}$ *converges in probability to zero uniformly in* t*.*

Since we select subcohort members by using simple random sampling without replacement, the condition of random vector ξ of above lemma is satisfied. For finite sample $n < \infty$, we can express $\mu_B(t) = n^{-1}\sum_{i=1}^n B_i(t)$ and thus $n^{-1/2}\sum_{i=1}^n \xi_i\{B_i(t) - \mu_{B_i}(t)\} = n^{-1/2}\sum_{i=1}^n \xi_i\{B_i(t) - n^{-1}\sum_{i=1}^n B_i(t)\} = n^{-1/2}\sum_{i=1}^n\{\xi_i - \frac{\tilde{n}}{n}\}B_i(t) = n^{-1/2}\tilde{\alpha}\sum_{i=1}^n\{\frac{\xi_i}{\tilde{\alpha}} - 1\}B_i(t)$.

First, we consider the asymptotic properties of time-varying sampling probability estimator $\tilde{\alpha}_k(t) = \sum_{i=1}^n \xi_i\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)/\sum_{i=1}^n\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)$ for true selection probability $\tilde{\alpha}$. For each k, it follows from the Taylor expansion series as

$$\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1} = -\frac{1}{\alpha_*(t)^2}\{\tilde{\alpha}_k(t) - \tilde{\alpha}\},$$

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

Set $B_i(t) = \frac{\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)}{\sum_{i=1}^n\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)}$. Since $\prod_{j=1}^K(1-\Delta_{ij})$ and $Y_{ik}(t)$ are bounded functions in t, $\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)$ is also a bounded function and the finite sum of its has finite variation. Thus, $B_i(t)$ has finite variation

Also it is easy to show $\mathrm{E}[B_i(t)] = n^{-1}\sum_{i=1}^n B_i(t) = n^{-1}\sum_{i=1}^n \frac{\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)}{\sum_{i=1}^n\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)} = n^{-1} = \mu_B$, $\mathrm{Var}[B_i(0)] < \infty$, and $\mathrm{Var}[B_i(\tau)] < \infty$. So,

$$\begin{aligned}
\tilde{\alpha}_k(t) - \tilde{\alpha} &= \frac{\sum_{i=1}^n \xi_i(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)}{\sum_{i=1}^n(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)} - \sum_{i=1}^n \frac{\xi_i}{n} \\
&= \sum_{i=1}^n \xi_i \left[\frac{(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)}{\sum_{i=1}^n(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)} - \frac{1}{n}\right] \\
&= \sum_{i=1}^n \xi_i \left[B_i(t) - \mu_B\right]
\end{aligned}$$

We can express $\tilde{\alpha}_k(t) - \tilde{\alpha} = \sum_{i=1}^n \tilde{\alpha}(\frac{\xi_i}{\tilde{\alpha}} - 1)\frac{(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)}{\sum_{i=1}^n(\prod_{j=1}^K(1-\Delta_{ij}))Y_{ik}(t)}$, and thus $n^{1/2}(\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1})$ can be written as

$$\frac{\tilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{n}{\sum_{i=1}^n \prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)} \cdot n^{-1/2}\{\sum_{i=1}^n(1 - \frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)\} \qquad (3.7)$$

$\frac{1}{n}\sum_{i=1}^n\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)$ converges in probability uniformly to $E[\{\prod_{j=1}^K(1-\Delta_{1j})\}Y_{1k}(t)]$ by Glivenko-Cantelli lemma. Since $\{\prod_{j=1}^K(1-\Delta_{ij})\}Y_{ik}(t)$ is bounded and monotone function in t, $n^{-1/2}\{\sum_{i=1}^n(\frac{\xi_i}{\tilde{\alpha}}-1)\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)\}$ converges weakly to zero-mean Gaussian process in the view of lemma 2. This follows from lemma 2 that $n^{-1}\{\sum_{i=1}^n(\frac{\xi_i}{\tilde{\alpha}}-1)\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)\}$ converges to zero in probability uniformly in t. Thus, $\tilde{\alpha}_k(t)$ and $\tilde{\alpha}$ converge to the same limit uniformly in t. This ensures that $\alpha_*(t)$ also converges to the same limit as $\tilde{\alpha}$ uniformly in t.

By Slutsky's theorem and above results, (3.7) can be written as

$$
\begin{aligned}
n^{1/2}(\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1}) &= \frac{1}{\tilde{\alpha} E(\prod_{j=1}^K (1-\Delta_{1j}) Y_{1k}(t))} n^{-1/2} \left\{ \sum_{i=1}^n (1 - \frac{\xi_i}{\tilde{\alpha}}) \prod_{j=1}^K (1-\Delta_{ij}) Y_{ik}(t) \right\} \\
&+ \left[\frac{\tilde{\alpha}}{\alpha_*(t)^2} \cdot \frac{n}{\sum_{i=1}^n \prod_{j=1}^K (1-\Delta_{ij}) Y_{ik}(t)} - \frac{1}{\tilde{\alpha} E(\prod_{j=1}^K (1-\Delta_{1j}) Y_{1j}(t))} \right] \\
&\times n^{-1/2} \left\{ \sum_{i=1}^n (1 - \frac{\xi_i}{\tilde{\alpha}}) \prod_{j=1}^K (1-\Delta_{ij}) Y_{ik}(t) \right\} \\
&= \frac{1}{\tilde{\alpha} E(\prod_{j=1}^K (1-\Delta_{1j}) Y_{1k}(t))} n^{-1/2} \left\{ \sum_{i=1}^n (1 - \frac{\xi_i}{\tilde{\alpha}}) \prod_{j=1}^K (1-\Delta_{ij}) Y_{ik}(t) \right\} \\
&+ o_p(1) \qquad (3.8)
\end{aligned}
$$

The above properties will be used in some proofs. Here is the proof of theorem 1.

The proof of Theorem 1 We first show the consistency of $\tilde{\beta}^M$. Denote $\widetilde{U}_n^M = n^{-1}\widetilde{U}^M$. By Taylor expansion series, $\tilde{\beta}^M$ can be written as

$$
\tilde{\beta}^M = \beta_0 + \left[-\frac{\partial \widetilde{U}_n^M(\beta_0)}{\partial \beta_0} \right]^{-1} \widetilde{U}_n^M(\beta_0) + o_p(1) \qquad (3.9)
$$

Based on the extension of Fourtz [1977], if the following conditions are satisfied

(I) $\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0,

(II) $\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}$ is negative definite with probability going to one as n $\to \infty$,

(III) -$\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}$ converges to $A(\beta_0)$ in probability uniformly for β in an open neighborhood about β_0,

(IV) $\widetilde{U}_n^M(\beta)$ converges to 0 in probability,

then, we can show that $\tilde{\beta}^M$ converges to β_0 in probability. Note that

$$
\begin{aligned}
\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T} &= -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \widetilde{V}_k(\beta, t) dN_{ik}(t), \\
\text{where } \widetilde{V}_k(\beta, t) &= \frac{\widetilde{S}_k^{(2)}(\beta, t) \widetilde{S}_k^{(0)}(\beta, t) - \widetilde{S}_k^{(1)}(\beta, t)^{\otimes 2}}{\widetilde{S}_k^{(0)}(\beta, t)^2}. \qquad (3.10)
\end{aligned}
$$

Since $\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}$ has the form (3.10) and each component, $\widetilde{S}_k^{(d)}(\beta,t)$ for d=0,1,2 are continuous, (I) is satisfied.

In order to show that conditions (II) and (III) are satisfied, we first will show $\| (-\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}) - A(\beta) \|$ converge to zero in probability uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$.

Let $dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t)e^{\beta_0 Z_{ik}(t)}\lambda_{0k}(t)dt$. We have

$$\begin{aligned}
&\| (-\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}) - A(\beta) \| \\
&= \| \sum_{k=1}^K \int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t) + v_k(\beta,t)\}\frac{1}{n}d\sum_{i=1}^n N_{ik}(t) - \int_0^\tau v_k(\beta_0,t)s_k^{(0)}(\beta_0,t)\lambda_{0k}(t)dt \| \\
&\le \| \sum_{k=1}^K \int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t)\}\frac{1}{n}d\sum_{i=1}^n N_{ik}(t) \| \\
&+ \| \sum_{k=1}^K \int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{i=1}^n N_{ik}(t) - \int_0^\tau v_k(\beta,t)s_k^{(0)}(\beta,t)\lambda_{0k}(t)dt \| \\
&\le \| \sum_{k=1}^K \int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t)\}\frac{1}{n}d\sum_{i=1}^n N_{ik}(t) \| \\
&+ \| \sum_{k=1}^K \int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{i=1}^n \{M_{ik}(t) + Y_{ik}(t)e^{\beta^T Z_{ik}(t)}\Lambda_{0k}(t)\} - v_k(\beta,t)s_k^{(0)}(\beta,t)\lambda_{0k}(t)dt \|
\end{aligned}$$

Since $S_k^{(0)}(\beta,t) = n^{-1}\sum_{i=1}^n Y_{ik}(t)e^{\beta^T Z_{ik}(t)}$, it follows that

$$\begin{aligned}
&\| (-\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T}) - A(\beta) \| \\
\le \quad & \| \sum_{k=1}^K \int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t)\}\frac{1}{n}d\sum_{i=1}^n N_{ik}(t) \| \qquad (3.11) \\
+ \quad & \| \sum_{k=1}^K \int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{i=1}^n M_{ik}(t) \| \qquad (3.12) \\
+ \quad & \| \sum_{k=1}^K \int_0^\tau v_k(\beta,t)\{S_k^{(0)}(\beta,t) - s_k^{(0)}(\beta,t)\}\lambda_{0k}(t)dt \| \qquad (3.13)
\end{aligned}$$

We will show that each of three terms in above inequality converges to zero uniformly in $\beta \in \mathcal{B}$. First, the term in (3.11) will be shown to converges to zero in probability as $n \to \infty$. To show this, first we need to show that

$$\sup_{t\in[0,\tau]\beta\in\mathcal{B}} \| \widetilde{V}_k(\beta,t) - v_k(\beta,t) \| \to 0 \quad \text{as } n \to \infty \quad \text{for } k = 1,\ldots,K.$$

Since $\widetilde{V}_k(\beta,t)$ is a function of $\widetilde{S}_k^{(d)}(\beta,t), d = 0,1,2$, $\sup_{t\in(0,\tau),\beta\in\mathcal{B}} \|S_k^{(d)}(\beta,t) - s_k^{(d)}(\beta,t)\| \to_p 0$ based on condition (d), and $s_k^{(0)}(\beta,t)$ is bounded away from zero base on condition (f), it suffices to show that

$$\sup_{t\in[0,\tau]\beta\in\mathcal{B}} \| \widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t) \| \to 0 \ \text{ as } n \to \infty \ \text{ for } k = 1,\dots,K \text{ and } \ d = 0,1,2.$$

Note that $\widetilde{S}_k^{(d)}(\beta,t) = n^{-1}\sum_{i=1}^n \psi_{ik}(t)Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}$, where $\psi_{ik}(t) = 1 - (\prod_{j=1}^K(1 - \Delta_{ij})) + \prod_{j=1}^K(1-\Delta_{ij})\xi_i\widetilde{\alpha}_k^{-1}(t)$ for $d = 0,1,2$. One can write

$$\begin{aligned}
& \widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t) \\
= \ & n^{-1}\sum_{i=1}^n\{\psi_{ik}(t) - 1\}Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)} \\
= \ & \frac{1}{n}\sum_{i=1}^n\{1 - (\prod_{j=1}^K(1-\Delta_{ij})) + \prod_{j=1}^K(1-\Delta_{ij})\xi_i\widetilde{\alpha}_k^{-1}(t) - 1\}Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)} \\
= \ & \frac{1}{n}\sum_{i=1}^n\{\prod_{j=1}^K(1-\Delta_{ij})\xi_i\widetilde{\alpha}_k^{-1}(t) - (\prod_{j=1}^K(1-\Delta_{ij}))\}Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)} \\
= \ & \frac{1}{n}\sum_{i=1}^n\left(\frac{\xi_i}{\tilde{\alpha}} - 1\right)\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \\
+ \ & \frac{1}{n}\sum_{i=1}^n(\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1})\xi_i\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)
\end{aligned}$$

and then

$$\begin{aligned}
& \| \widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t) \| \\
\le \ & \| \frac{1}{n}\sum_{i=1}^n\left(\frac{\xi_i}{\tilde{\alpha}} - 1\right)\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \| \quad (3.14) \\
+ \ & |\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1}|\frac{1}{n}\sum_{i=1}^n\xi_i\prod_{j=1}^K(1-\Delta_{ij})|Z_{ik}(t)^{\otimes d}|e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \quad (3.15)
\end{aligned}$$

Based on condition (c), the total variation of $\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)$ is finite on $[0,\tau]$. By lemma 2, the term in (3.14) converges to zero in probability uniformly in t. Since it was shown that $(\tilde{\alpha}_k^{-1}(t) - \tilde{\alpha}^{-1})$ converges to zero in probability uniformly in t and $\frac{1}{n}\sum_{i=1}^n\xi_i\prod_{j=1}^K(1-\Delta_{ij})|Z_{ik}(t)^{\otimes d}|e^{\beta^T Z_{ik}(t)}Y_{ik}(t)$ converges to $\tilde{\alpha}\mathrm{E}[\prod_{j=1}^K(1-\Delta_{1j})|Z_{1k}(t)^{\otimes d}|$

$\times e^{\beta^T Z_{1k}(t)} Y_{1k}(t)]$ in probability uniformly in t, the term in (3.15) converges to zero in probability uniformly. Thus, $\widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t)$ converges to zero. Combining this result with condition (d), we can show that

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \parallel \widetilde{S}_k^{(d)}(\beta,t) - s_k^{(d)}(\beta,t) \parallel \to 0 \text{ as } n \to \infty \text{ for } k = 1,\ldots,K \text{ and } d = 0,1,2. \quad (3.16)$$

Since $s_k^{(0)}$ is bounded way from zero based by condition (f), we can also show that $\sup_{t\in[0,\tau],\beta\in\mathcal{B}}$ $\parallel \widetilde{S}_k^{(0)}(\beta,t)^{-1} - s_k^{(0)}(\beta,t)^{-1} \parallel \to 0$ as $n \to \infty$ for $k = 1,\ldots,K$. Combining these results, $\widetilde{V}_k(\beta,t)$ converges to $v_k(\beta,t)$ in probability uniformly in t and β. Moreover, by Lenglart inequality (Andersen and Gill [1982], p1115), there exists n_0 such that for $n \geq n_0$ for any δ, $\eta > 0$,

$$P[n^{-1}\bar{N}_k(\tau) > \eta] \leq \frac{\delta}{\eta} + P[\int_0^\tau S_k^{(0)}(\beta_0;t)\lambda_{0k}(t)dt > \delta],$$

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

Based on condition (d), $P[\int_0^\tau S_k^{(0)}(\beta_0;t)\lambda_{0k}(t)dt > \delta]$ converges to zero as n $\to \infty$ for $\delta > \int_0^\tau s_k^{(0)}(\beta_0;t)\lambda_{0k}(t)dt$ and then $\lim_{\eta\uparrow\infty}\lim_{n\to\infty} P[n^{-1}\bar{N}_k(\tau) > \eta] = 0$. Therefore, the term in (3.11) converges to zero in probability uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$.

For the quantity in (3.12), $\int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{i=1}^n M_{ik}(t)$ is a local square integrable martingale. By the Lenglart inequality(Andersen and Gill [1982], p1115), it can be shown that, for all δ, $\eta > 0$,

$$P\left[\|\frac{1}{n}\int_0^\tau \{v_k(\beta,t)\}_{jj'}\bar{M}_k(t)\| > \eta\right] \leq \frac{\delta}{\eta^2} + P\left[\frac{1}{n}\int_0^\tau \{v_k(\beta,t)\}_{jj'}^2 S_k^{(0)}(\beta;t)\lambda_{0k}(t)dt > \delta\right],$$

where $\bar{M}_k(t) = \sum_{i=1}^n M_{ik}(t)$ and subscript jj' indicates (jj') element of matrix $v_k(\beta,t)$.

Based on boundedness conditions (d), (f), and (g), the second term on right side in the above inequality converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ for any δ as $n \to \infty$. Then it follows that one on the left side converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$. Hence, the quantity in (3.12) converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$.

Due to boundedness of $\sup_{t,\beta}\{v_k(\beta,t)\}$ based on conditions (d) and (e), $\Lambda_{0k}(t)$ for $k = 1,\dots,K$ based on condition (g), and uniform convergence of $\widetilde{S}_k^{(0)}(\beta,t)$ to $s_k^{(0)}(\beta,t)$, the term in (3.13) converges to zero in probability uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$. Therefore, all three terms in (3.11), (3.12), and (3.13) converge to zero in probability uniformly. Consequently, we have

$$-\frac{\partial \widetilde{U}_n^M(\beta)}{\partial \beta^T} \to_p A(\beta) \quad \text{as } n \to \infty \quad \text{uniformly in } \beta \in \mathcal{B}$$

and consequently (II) and (III) are satisfied.

To show that (IV) is satisfied, we will examine the asymptotic behavior of $n^{-1/2}\widetilde{U}_n^M(\beta_0)$.

We can decompose $n^{-1/2}\widetilde{U}^M(\beta_0)$ into two parts such that

$$\begin{aligned}
n^{-1/2}\widetilde{U}^M(\beta_0) &= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{Z_{ik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} dN_{ik}(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{Z_{ik}(t) - \frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)}\right\} dN_{ik}(t) \qquad (3.17) \\
&+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} dN_{ik}(t). \qquad (3.18)
\end{aligned}$$

The quantity in (3.17) is the pseudo partial likelihood score function for full cohort and can be written as

$$\begin{aligned}
(3.17) &= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{Z_{ik}(t) - \frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta_0,t)}\right\} dM_{ik}(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \{Z_{ik}(t) - e_k(\beta_0,t)\} dM_{ik}(t) \\
&+ n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{e_k(\beta_0,t) - \frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)}\right\} dM_{ik}(t).
\end{aligned}$$

We can show that (3.17) was asymptotically equivalent to $n^{-1/2}\sum_{i=1}^n\sum_{k=1}^K W_{ik}(\beta_0)$ where $W_{ik}(\beta) = \int_0^\tau (Z_{ik}(t) - e_{ik}(\beta,t))dM_{ik}(t)$ (Spiekerman and Lin [1998], Clegg et al. [1999]).

Since $dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t)e^{\beta_0 Z_{ik}(t)}d\Lambda_{0k}(t)$, (3.18) can decompose into two parts:

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} dN_{ik}(t) \\
= & \sum_{k=1}^{K}\int_0^\tau \left\{\frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} d\left\{n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)\right\} \qquad (3.19) \\
+ & n^{-1/2}\sum_{k=1}^{K}\int_0^\tau \left\{\frac{S_k^{(1)}(\beta_0,t)}{S_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} \sum_{i=1}^{n} Y_{ik}(t)e^{\beta_0 Z_{ik}(t)}d\Lambda_{0k}(t). \qquad (3.20)
\end{aligned}
$$

Based on the assumed model, $M_{1k}(t),\ldots,M_{nk}(t)$ are identically independently distributed zero-mean random variables for fixed t. $M_{ik}(t)$ is of bounded variation since $M_{ik}^2(0) < \infty$ and $M_{ik}^2(\tau) < \infty$ are satisfied based on conditions (c) and (g). From the example of 2.11.16 of Van Der Vaart and Wellner (1996, p215), $n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{P}_{Mk}(t)$.

To establish that $\mathcal{P}_{Mk}(t)$ has continuous sample paths, we will use Kolmogorov-Centsov theorem. If conditions of Kolmogorov-Centsov theorem $\mathrm{E}[\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^4] \le D_z^*|t-s|^2$ and $\mathrm{E}[\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2] \le \tilde{D}_z|t-s|$ for all $t \ge s$ are satisfied, then we can show that $\mathcal{P}_{Mk}(t)$ has continuous sample paths. Note that $\mathrm{E}[\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2] = \mathrm{E}[\mathcal{P}_{Mk}(t)^2]$ - $2\mathrm{E}[\mathcal{P}_{Mk}(t)\mathcal{P}_{Mk}(s)] + \mathrm{E}[\mathcal{P}_{Mk}(s)^2] = \mathrm{E}[\mathcal{P}_{Mk}(t)^2]$ - $\mathrm{E}[\mathcal{P}_{Mk}(s)^2]$ due to $\mathrm{E}[\mathcal{P}_{Mk}(t)\mathcal{P}_{Mk}(s)]$ $= \mathrm{E}[\mathcal{P}_{Mk}(s)^2]$ for $t \ge s$. Since $\mathrm{E}[\mathcal{P}_{Mk}(t)^2] = \mathrm{E}[n^{-1}\sum_{i=1}^{n} M_{ik}(t)^2] = \mathrm{E}[M_{ik}(t)^2] =$ $\mathrm{E}[\int_0^t Y_{ik}(u)e^{\beta_0^T Z_{ik}(u)}\lambda_{0k}(u)du]$, $\mathrm{E}[\mathcal{P}_{Mk}(t)^2]$ - $\mathrm{E}[\mathcal{P}_{Mk}(s)^2] = \mathrm{E}[\int_s^t Y_{ik}(u)e^{\beta_0^T Z_{ik}(u)}\lambda_{0k}(u)\,du] \le$ $e^{D_z}\mathrm{E}[\int_s^t \lambda_{0k}(u)du] = \tilde{D}_z(\Lambda_{0k}(t) - \Lambda_{0k}(s))$ based on condition (c) where $\tilde{D}_z = e^{D_z}$. There exists constant C such that $\Lambda_{0k}(t) - \Lambda_{0k}(s) \le C(t-s)$ for $t \ge s$ since $\Lambda_{0k}(.)$ is differentiable and $\lambda_{0k}(.)$ is bounded in $[0,\tau]$. Thus $\mathrm{E}\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^2 \le \tilde{D}_{cz}(t-s)$ where $\tilde{D}_{cz} = \tilde{D}_z \times C$. For fixed t, $\mathcal{P}_{Mk}(t)$ is a zero-mean random normal variable. Then, $\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)$ is also a zero-mean random normal variable for fixed t and s. Consequently, $\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2$ is a random chi-square variable for fixed t and s. We can express $\mathrm{E}[\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^4] =$ $\mathrm{Var}\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2 + \mathrm{E}\{(\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s))^2\}^2 = 3\{\mathrm{E}(\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s))^2\}^2$ due to $\mathrm{Var}\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2 = 2\mathrm{E}[\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2]$ from the property of chi-square distribution. Therefore, $E\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^4 = 3\{\mathrm{E}(\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s))^2\}^2 \le D_z^*|t-s|^2$ for some constant D_z^*. Since the two conditions are satisfied, it follows that $\mathcal{P}_{Mk}(t)$ has continuous sample paths from Kolmogorov-Centsov theorem.

Based on conditions (c), (d), and (f), there exists N^* such that $n > N^*$ $S_k^{(1)}(\beta,t)$ and $S_k^{(0)}(\beta,t)$ are of bounded variations and $S_k^{(0)}(\beta,t)$ is bounded away from zero. Thus $\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)}$ is of bounded variation when $n > N^*$. By using $f'(x)/f(x) = [\log f(x)]' \cong F_1^*(t) - F_2^*(t)$ where $F_1^*(t)$ and $F_1^*(t)$ are bounded, monotone and nonnegative functions in t, it can be written as $\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)} = Z_{k1}^*(t) - Z_{k2}^*(t)$ where $Z_{k1}^*(t)$ and $Z_{k2}^*(t)$ are bounded, monotone and nonnegative functions in t. Hence, $\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)}$ is a sum of two monotone functions in t. Similarly, we can show that $\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}$ is of bounded variation due to conditions (c) and (f) and the result of (3.16) by the same manner. Moreover, we can write that $\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}$ is also a sum of two monotone functions in t. Based on condition (d) and the result of (3.16), it can be shown that $\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)}$ and $\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}$ converge to the same limit uniformly. Thus, it follows from lemma 1 that

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

Therefore, the term in (3.19) converges to 0 in probability uniformly in t.

We have shown that $n^{-1/2}\sum_{i=1}^{n} M_{ik}(t)$ converges weakly to a zero-mean Gaussian process with continuous sample paths. To show that $\widetilde{S}_k^{(d)}(\beta,t)$ and $S_k^{(d)}(\beta,t)$ converges to the same limit in probability, we will show that $n^{1/2}\{S_k^{(d)}(\beta,t) - \widetilde{S}_k^{(d)}(\beta,t)\}$ converges to a zero mean

Gaussian process. It can be expressed as

$$
\begin{aligned}
& n^{1/2}\left\{S_k^{(d)}(\beta,t)-\tilde{S}_k^{(d)}(\beta,t)\right\} \\
= \;& n^{-1/2}\left\{\sum_{i=1}^n Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)-\sum_{i=1}^n \psi_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)\right\} \\
= \;& n^{-1/2}\{\sum_{i=1}^n\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \\
- \;& \sum_{i=1}^n\prod_{j=1}^K(1-\Delta_{ij})\xi_i\tilde{\alpha}_k^{-1}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)\} \\
= \;& n^{-1/2}\sum_{i=1}^n(\tilde{\alpha}^{-1}-\tilde{\alpha}_k^{-1}(t))\prod_{j=1}^K(1-\Delta_{ij})\xi_i Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \\
+ \;& n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \\
= \;& n^{-1}\sum_{i=1}^n\left\{\frac{1}{\tilde{\alpha}E(\prod_{j=1}^K(1-\Delta_{ij})Y_{1k}(t))}\frac{1}{\sqrt{n}}\left\{\sum_{p=1}^n(\frac{\xi_p}{\tilde{\alpha}}-1)\prod_{j=1}^K(1-\Delta_{pj})Y_{pk}(t)\right\}\right\} \\
\times \;& \xi_i\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t) \\
+ \;& n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)+o_p(1) \quad \text{(by plugging in (4.6))} \\
= \;& n^{-1/2}\sum_{i=1}^n(\frac{\xi_i}{\tilde{\alpha}}-1)\cdot\frac{\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))}\left\{n^{-1}\sum_{p=1}^n\frac{\xi_p}{\tilde{\alpha}}\prod_{j=1}^K(1-\Delta_{pj})Z_{pk}(t)^{\otimes d}e^{\beta^T Z_{pk}(t)}Y_{pk}(t)\right\} \\
+ \;& n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}Y_{ik}(t)+o_p(1).
\end{aligned}
$$

Since $n^{-1}\sum_{p=1}^n\frac{\xi_p}{\tilde{\alpha}}\prod_{j=1}^K(1-\Delta_{pj})Z_{pk}(t)^{\otimes d}e^{\beta^T Z_{pk}(t)}Y_{pk}(t)$ converges to $E(\prod_{j=1}^K(1-\Delta_{1j})$ $Z_{1k}(t)^{\otimes d}e^{\beta^T Z_{1k}(t)}Y_{1k}(t))$ in probability uniformly in t, it can be written as

$$
\begin{aligned}
& n^{1/2}\left\{S_k^{(d)}(\beta,t)-\tilde{S}_k^{(d)}(\beta,t)\right\} \\
= \;& n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t) \\
\times \;& \left\{Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}-\frac{E(\prod_{j=1}^K(1-\Delta_{1j})Z_{1k}(t)^{\otimes d}e^{\beta^T Z_{1k}(t)}Y_{1k}(t))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))}\right\}+o_p(1) \quad (3.21)
\end{aligned}
$$

By lemma 2, $n^{1/2}\{S_k^{(d)}(\beta,t)-\widetilde{S}_k^{(d)}(\beta,t)\}$ converges weakly to a zero-mean Gaussian process since we set $B_i(t) = \prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t)\{Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)} - \frac{E(\prod_{j=1}^K(1-\Delta_{1j})Z_{1k}(t)^{\otimes d}e^{\beta^T Z_{1k}(t)}Y_{1k}(t))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))}\}$ with $\mathrm{Var}(B_i(0)) < \infty$ and $\mathrm{Var}(B_i(\tau)) < \infty$. Consequently, $\widetilde{S}_k^{(d)}(\beta,t)$ and $S_k^{(d)}(\beta,t)$ converges to the same limit in probability.

To investigate the asymptotic properties of the quantity in (3.20), it can be decomposed into two parts:

$$
\begin{aligned}
& n^{-1/2}\sum_{k=1}^K \int_0^\tau \left\{\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)} - \frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}\right\}\sum_{i=1}^n Y_{ik}(t)e^{\beta_0 Z_{ik}(t)}d\Lambda_{0k}(t) \\
= \;& n^{1/2}\sum_{k=1}^K \int_0^\tau \left\{S_k^{(1)}(\beta,t) - S_k^{(0)}(\beta,t)\cdot\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}\right\}d\Lambda_{0k}(t) \\
= \;& n^{1/2}\sum_{k=1}^K \int_0^\tau \left\{S_k^{(1)}(\beta,t) - \tilde{S}_k^{(1)}(\beta,t)\right\}d\Lambda_{0k}(t) - \left\{S_k^{(0)}(\beta,t) - \tilde{S}_k^{(0)}(\beta,t)\right\}\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}d\Lambda_{0k}(t) \\
= \;& n^{1/2}\sum_{k=1}^K \int_0^\tau \left\{S_k^{(1)}(\beta,t) - \tilde{S}_k^{(1)}(\beta,t)\right\}d\Lambda_{0k}(t) \\
- \;& n^{1/2}\sum_{k=1}^K \int_0^\tau \left\{S_k^{(0)}(\beta,t) - \tilde{S}_k^{(0)}(\beta,t)\right\}e_k(\beta,t)d\Lambda_{0k}(t) + o_p(1) \qquad (3.22)
\end{aligned}
$$

The last quantity in (3.22) holds since $\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}$ converges to $e_k(\beta,t)$ in probability uniformly in t, $n^{1/2}\{S_k^{(d)}(\beta,t) - \widetilde{S}_k^{(d)}(\beta,t)\}$ $d = 0,1$ converges weakly to a zero-mean Gaussian process, and $\Lambda_{0k}(t)$ is bounded on $t \in [0,\tau]$.

Plugging the quantity in (3.21) into equation (3.22), we have

$$
\begin{aligned}
& n^{-1/2}\sum_{k=1}^K \int_0^\tau \left\{\frac{S_k^{(1)}(\beta,t)}{S_k^{(0)}(\beta,t)} - \frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}\right\}\sum_{i=1}^n Y_{ik}(t)e^{\beta_0 Z_{ik}(t)}d\Lambda_{0k}(t) \\
= \;& n^{-1/2}\sum_{k=1}^K\sum_{i=1}^n \int_0^\tau (1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})Y_{ik}(t) \\
\times \;& [Z_{ik}(t)e^{\beta^T Z_{ik}(t)} - \frac{E(\prod_{j=1}^K(1-\Delta_{1j})Z_{1k}(t)e^{\beta^T Z_{1k}(t)}Y_{1k}(t))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))} \\
- \;& \{e^{\beta^T Z_{ik}(t)} - \frac{E(\prod_{j=1}^K(1-\Delta_{1j})e^{\beta^T Z_{1k}(t)}Y_{1k}(t))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))}\}e_k(\beta,t)]d\Lambda_{0k}(t) + o_p(1) \\
= \;& n^{-1/2}\sum_{k=1}^K\sum_{i=1}^n \int_0^\tau (1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})\{Y_{ik}(t)[Z_{ik}(t) - e_k(\beta,t)]e^{\beta^T Z_{1k}(t)}d\Lambda_{0k}(t)
\end{aligned}
$$

$$
\begin{aligned}
- & \quad Y_{ik}(t) \cdot \frac{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t)[Z_{1k}(t)-e_k(\beta,t)]e^{\beta^T Z_{1k}(t)})}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t))} d\Lambda_{0k}(t)\} + o_p(1) \\
= & \quad n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau (1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^{K}(1-\Delta_{ij})\{Q_{ik}(\beta,t) - \frac{Y_{ik}(t)E(\prod_{j=1}^{K}(1-\Delta_{1j})Q_{1k}(\beta,t))}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t))}\} \\
\times & \quad d\Lambda_{0k}(t) + o_p(1) \qquad (3.23)
\end{aligned}
$$

where $Q_{ik}(\beta,t) = Y_{ik}(t)(Z_{ik}(t) - e_k(\beta,t))e^{\beta Z_{ik}(t)}$.

We have shown that the term in (3.17) is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta,t)$ and the term in (3.18) is asymptotically equivalent to $n^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^{K}(1-\Delta_{ij})\{Q_{ik}(\beta,t) - \frac{Y_{ik}(t)E(\prod_{j=1}^{K}(1-\Delta_{1j})Q_{1k}(\beta,t))}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t))}\}d\Lambda_{0k}(t)$.

Therefore, $n^{-1/2}\widetilde{U}^M(\beta_0)$ is asymptotically equivalent to

$$
n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0) + n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau (1-\frac{\xi_i}{\tilde{\alpha}})\Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t), \qquad (3.24)
$$

where $\Omega_{ik}(\beta,t) = \prod_{j=1}^{K}(1-\Delta_{ij})\{Q_{ik}(\beta,t) - \frac{Y_{ik}(t)E(\prod_{j=1}^{K}(1-\Delta_{1j})Q_{1k}(\beta,t))}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t))}\}$.

By Spiekerman and Lin [1998] and Clegg et al. [1999], the first term of (3.24) converges weakly to a zero-mean normal vector with covariance matrix $V_I(\beta_0) = E[\sum_{k=1}^{K} W_{1k}(\beta_0)]^{\otimes 2}$. The second term of (3.24) is asymptotically zero-mean normal vector with covariance matrix $\frac{1-\alpha}{\alpha}V_{II}(\beta_0) = \frac{1-\alpha}{\alpha}E\left[\sum_{k=1}^{K}\int_0^\tau \Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)\right]^{\otimes 2}$ by Hájek [1960]'s central limit theorem for finite sampling.

In addition, $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0)$ and $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau(1-\frac{\xi_i}{\tilde{\alpha}})\Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)$ are independent since

$$
\begin{aligned}
& Cov\left(n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0), n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\frac{\xi_i}{\tilde{\alpha}})\int_0^\tau \Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)\right) \\
= & \quad E\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0)\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\frac{\xi_i}{\tilde{\alpha}})\int_0^\tau \Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)\right\} \\
= & \quad E\left\{E\left(n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0)\sum_{i=1}^{n}\sum_{k=1}^{K}(1-\frac{\xi_i}{\tilde{\alpha}})\int_0^\tau \Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)|\mathcal{F}(\tau)\right)\right\} \\
= & \quad E\left\{n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} W_{ik}(\beta_0)\sum_{i=1}^{n}\sum_{k=1}^{K}E\left(1-\frac{\xi_i}{\tilde{\alpha}}|\mathcal{F}(\tau)\right)\int_0^\tau \Omega_{ik}(\beta_0,t)d\Lambda_{0k}(t)\right\} = 0,
\end{aligned}
$$

where $\{\mathcal{F}(t), t \geq 0\}$ is filtration.

Combining all the above results, $n^{-1/2}\widetilde{U}^M(\beta_0)$ converges weakly to zero-mean normal vector with covariance matrix $\Sigma(\beta_0) = V_I(\beta_0) + \frac{1-\alpha}{\alpha}V_{II}(\beta_0)$. Consequently, $n^{-1}\widetilde{U}^M(\beta_0)$ converges to zero in probability. Therefore, $\tilde{\beta}^M$ converges to β_0 in probability and is a consistent estimator of β_0 by satisfying conditions (I), (II), (III), and (IV) (theorem 2 of Fourtz [1977]).

In addition to the consistency of $\tilde{\beta}^M$, it follows from Taylor expansion, it can be written as

$$n^{1/2}(\tilde{\beta}^M - \beta_0) = [A(\beta_0)]^{-1} n^{-1/2}\widetilde{U}^M(\beta_0). \tag{3.25}$$

Therefore, $n^{1/2}(\tilde{\beta}^M - \beta_0)$ converges weakly zero-mean normal vector with covariance matrix $A(\beta_0)^{-1}\Sigma(\beta_0)A(\beta_0)^{-1}$.

The proof of Theorem 2 Note that

$$\widetilde{\Lambda}^M_{0k}(\tilde{\beta}^M, t) = \int_0^t \frac{\sum_{i=1}^n dN_{ik}(u)}{n\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)} = \int_0^t \frac{\sum_{i=1}^n dM_{ik}(u)}{n\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)} + \int_0^t \frac{S_k^{(0)}(\beta_0, u)d\Lambda_{0k}(u)}{\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)}$$

We can decompose $n^{1/2}\{\widetilde{\Lambda}^M_{0k}(\tilde{\beta}^M, t) - \Lambda_{0k}(t)\}$ into four parts:

$$\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}^M_{0k}(\tilde{\beta}^M, t) - \Lambda_{0k}(t)\} \\
= \;& n^{1/2}\int_0^t \left(\frac{1}{n\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)} - \frac{1}{n\tilde{S}_k^{(0)}(\beta_0, u)}\right) d\sum_{i=1}^n M_{ik}(u) \\
+ \;& n^{1/2}\int_0^t \left(\frac{1}{\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)} - \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)}\right) S_k^{(0)}(\beta_0, u)d\Lambda_{0k}(u) \\
+ \;& n^{-1/2}\int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} d\sum_{i=1}^n M_{ik}(u) \\
+ \;& n^{1/2}\int_0^t \left(\frac{S_k^{(0)}(\beta_0, u) - \tilde{S}_k^{(0)}(\beta_0, u)}{\tilde{S}_k^{(0)}(\beta_0, u)}\right) d\Lambda_{0k}(u)
\end{aligned} \tag{3.26}$$

By Taylor expansion, it can be written as

$$\frac{1}{\tilde{S}_k^{(0)}(\tilde{\beta}^M, u)} - \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} = -\frac{\tilde{S}_k^{(1)}(\beta^*, u)}{\tilde{S}_k^{(0)}(\beta^*, u)^2}(\tilde{\beta}^M - \beta_0)$$

where β^* is on the line segment between $\tilde{\beta}^M$ and β_0. Plugging into the first term in (3.26), we have

$$\int_0^t \left(-\frac{\tilde{S}_k^{(1)}(\beta^*, u)}{\tilde{S}_k^{(0)}(\beta^*, u)^2}\right)(\tilde{\beta}^M - \beta_0)\left\{n^{-1/2} d\sum_{i=1}^n M_{ik}(u)\right\}, \tag{3.27}$$

where β^* is on the line segment between $\tilde{\beta}^M$ and β_0. Due to consistency of $\tilde{\beta}^M$, β^* also converges to β_0 in probability uniformly. Since $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(1)}(\beta^*, u)$ are of bounded variations and $\tilde{S}_k^{(0)}(\beta^*, u)$ is bounded away from 0, $\frac{\tilde{S}_k^{(1)}(\beta^*, u)}{\tilde{S}_k^{(0)}(\beta^*, u)^2}$ is of bounded variation and can be written as sum of two monotone functions in t. In addition, it is shown consistency of $\tilde{\beta}^M$, weak convergence of $n^{-1/2} d\sum_{i=1}^n M_{ik}(u)$ to zero-mean Gaussian process with continuous sample paths, and the uniform convergence of $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(1)}(\beta^*, u)$. Therefore, by lemma 1, the quantity in (4.23) converges to zero in probability uniformly in t.

The second term in (3.26), by Taylor expansion series, can be written as

$$n^{1/2}\int_0^t \left(-\frac{\tilde{S}_k^{(1)}(\beta^*, u)}{\tilde{S}_k^{(0)}(\beta^*, u)^2}\right)(\tilde{\beta}^M - \beta_0) S_k^{(0)}(\beta_0, t) d\Lambda_{0k}(u)$$

Since $\tilde{\beta}^M$ and β^* converge to β_0 in probability uniformly in t, $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(0)}(\beta_0, u)$ converges to $s_k^{(0)}(\beta_0, t)$ in probability uniformly. Also, $\tilde{S}_k^{(1)}(\beta^*, u) \to^p s_k^{(1)}(\beta_0, t)$. Since $d\Lambda_{0k}(u)$ is bounded, we can show that

$$n^{1/2}\int_0^t \left(-\frac{\tilde{S}_k^{(1)}(\beta^*, u)}{\tilde{S}_k^{(0)}(\beta^*, u)^2}\right)(\tilde{\beta}^M - \beta_0) S_k^{(0)}(\beta_0, u) d\Lambda_{0k}(u) \quad = \quad n^{1/2} l_k(\beta, t)^T (\tilde{\beta}^M - \beta_0) + o_p(1),$$

where $l_k(\beta, t)^T = \int_0^t -e_k(\beta, u) d\Lambda_{0k}(u)$ and $e_k(\beta_0, u) = s_k^{(1)}(\beta_0, u)/s_k^{(0)}(\beta_0, u)$.

Since $\tilde{S}_k^{(0)}(\beta_0, u)$ converges to $s_k^{(0)}(\beta_0, u)$ in probability uniformly and $s_k^{(0)}(\beta_0, u)$ is bounded away from 0, we have $\tilde{S}_k^{(0)}(\beta_0, u)^{-1} \to^p s_k^{(0)}(\beta_0, u)^{-1}$. In addition,$n^{-1/2} d\sum_{i=1}^n M_{ik}(u)$

converges to zero-mean Gaussian process with continuous sample paths. Hence, the third term in (3.26) can be written as

$$\int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\left\{n^{-1/2}d\sum_{i=1}^n M_{ik}(u)\right\} = \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}\left\{n^{-1/2}d\sum_{i=1}^n M_{ik}(u)\right\} + o_p(1)$$

Due to uniform convergence of $\tilde{S}_k^{(0)}(\beta_0,u)^{-1}$ to $s_k^{(0)}(\beta_0,u)^{-1}$ where $s_k^{(0)}(\beta_0,u)$ is bounded away from 0 and plug (3.21) into the last term in (3.26), we have

$$\begin{aligned}
& n^{1/2}\int_0^t \left(\frac{S_k^{(0)}(\beta_0,u)-\tilde{S}_k^{(0)}(\beta_0,u)}{\tilde{S}_k^{(0)}(\beta_0,u)}\right) d\Lambda_{0k}(u) \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij}) \\
\times & \int_0^t Y_{ik}(u)\left\{e^{\beta^T Z_{ik}(u)} - \frac{E(\prod_{j=1}^K(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{1k}(u))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(u))}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)} + o_p(1)
\end{aligned}$$

Combining all the results, we have

$$\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^M(\tilde{\beta}^M,t)-\Lambda_{0k}(t)\} \\
= & n^{1/2}l_k(\beta,t)^T(\tilde{\beta}^M-\beta_0) + \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}\left\{n^{-1/2}d\sum_{i=1}^n M_{ik}(u)\right\} \\
+ & \frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\prod_{j=1}^K(1-\Delta_{ij})\int_0^t Y_{ik}(u)\left\{e^{\beta^T Z_{ik}(u)} - \frac{E(\prod_{j=1}^K(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{1k}(u))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(u))}\right\} \\
\times & \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)} + o_p(1) \qquad (3.28)
\end{aligned}$$

Recall (3.25):

$$\begin{aligned}
& n^{1/2}(\tilde{\beta}^M-\beta_0) \\
= & A(\beta_0)^{-1}\left\{n^{-1/2}\sum_{i=1}^n\sum_{m=1}^K W_{im}(\beta_0) + n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})\sum_{m=1}^K\int_0^\tau \Omega_{im}(\beta_0,t)d\Lambda_{0m}(t)\right\} + o_p(1),
\end{aligned}$$

where $W_{ik}(\beta) = \int_0^\tau \{Z_{ik}(t)-e_k(\beta,t)\}\, dM_{ik}(t)$ and
$\Omega_{ik}(\beta,t) = \prod_{j=1}^K(1-\Delta_{ij})\{Q_{ik}(\beta,t) - \frac{Y_{ik}(t)E(\prod_{j=1}^K(1-\Delta_{1j})Q_{1k}(\beta,t))}{E(\prod_{j=1}^K(1-\Delta_{1j})Y_{1k}(t))}\}$.

Using the above equation, we have

$$
\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^{M}(\tilde{\beta}^{M},t)-\Lambda_{0k}(t)\} \\
= \ & n^{-1/2}\sum_{i=1}^{n}[\left\{l_k(\beta,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}W_{im}(\beta_0)+\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}dM_{ik}(u)\right\} \\
+ \ & (1-\frac{\xi_i}{\tilde{\alpha}})\{l_k(\beta,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}\int_0^{\tau}\Omega_{im}(\beta_0,t)d\Lambda_{0m}(t) \\
+ \ & \prod_{j=1}^{K}(1-\Delta_{ij})\int_0^t Y_{ik}(u)\left\{e^{\beta^T Z_{ik}(u)}-\frac{E(\prod_{j=1}^{K}(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{1k}(u))}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(u))}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)}\}] \\
+ \ & o_p(1) \\
= \ & n^{-1/2}\sum_{i=1}^{n}\eta_{ik}(\beta_0,t)+n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\zeta_{ik}(\beta_0,t)+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
& \eta_{ik}(\beta_0,t)=l_k(\beta,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}W_{im}(\beta_0)+\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}dM_{ik}(u) \quad \textit{and} \\
& \zeta_{ik}(\beta_0,t)=l_k(\beta,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}\int_0^{\tau}\Omega_{im}(\beta_0,t)d\Lambda_{0m}(t) \\
+ \ & \prod_{j=1}^{K}(1-\Delta_{ij})\int_0^t Y_{ik}(u)\left\{e^{\beta^T Z_{ik}(u)}-\frac{E(\prod_{j=1}^{K}(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{1k}(u))}{E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(u))}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)}.
\end{aligned}
$$

Let $H(t)=(H^{(1)}(t)+H^{(2)}(t))$ where $H^{(1)}(t)=(H_1^{(1)}(t),\ldots,H_K^{(1)}(t))^T$, $H^{(2)}(t)=(H_1^{(2)}(t),\ldots,H_K^{(2)}(t))^T$, $H_k^{(1)}(t)^T=n^{-1/2}\sum_{i=1}^{n}\eta_{ik}(\beta_0,t)$, and $H_k^{(2)}(t)^T=n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\zeta_{ik}(\beta_0,t)$. Then, by theorem 2 of Spiekerman and Lin [1998], $H^{(1)}(t)=(H_1^{(1)}(t),\ldots,$ $H_K^{(1)}(t))^T$ converges weakly to Gaussian process $\mathcal{H}^{(1)}(t)=(\mathcal{H}_1^{(1)}(t),\ldots,\mathcal{H}_K^{(1)}(t))^T$ whose mean is zero and covariance functions between $\mathcal{H}_j^{(1)}(t)$ and $\mathcal{H}_k^{(1)}(s)$ is $E\{\eta_{1j}(\beta_0,t),\eta_{1k}(\beta_0,s)\}$ for $t,s\in[0,\tau]$ in $D[0,\tau]^K$.

We will show weak convergence of $H^{(2)}(t)$ to a zero-mean Gaussian process $\mathcal{H}^{(2)}(t)$. $s_k^{(0)}(\beta,t)$ and $E(\prod_{j=1}^{K}(1-\Delta_{1j})Y_{1k}(t))$ are bounded away from zero, $l_k(\beta,t)^T$, $e^{\beta^T Z_{1k}(t)}Y_{1k}(t)$, $E(\prod_{j=1}^{K}(1-\Delta_{1j})e^{\beta^T Z_{1k}(t)}Y_{1k}(t))$, and $d\Lambda_{0k}(t)$ are of bounded variations based on conditions (b), (c), (d), and (f); $A(\beta_0)$ is positive definite based on (e). Hence, it follows from Cramer-Wold device and lemma 2 that the finite dimensional distribution of

$H^{(2)}(t)$ is asymptotically same as that of $\mathcal{H}^{(2)}(t)$ for any finite number of time point $(t_1, \ldots, t_L)$. Moreover, we need to show $H^{(2)}(t)$ has tightness. It suffices to show the marginal tightness of $H_k^{(2)}(t)$ for each k since space $D[0,\tau]^K$ is equipped with the uniform metric. By applying lemma 2, the marginal tightness follows to $H_k^{(2)}(t)$. Combining all the results, $H^{(2)}(t) = (H_1^{(2)}(t), \ldots, H_K^{(2)}(t))^T$ converges weakly to Gaussian process $\mathcal{H}^{(2)}(t) = (\mathcal{H}_1^{(2)}(t), \ldots, \mathcal{H}_K^{(2)}(t))^T$ whose mean is zero and covariance functions between $\mathcal{H}_j^{(2)}(t)$ and $\mathcal{H}_k^{(2)}(s)$ is $\frac{1-\alpha}{\alpha} E\{\zeta_{1j}(\beta_0, t), \zeta_{1k}(\beta_0, s)\}$ for $t, s \in [0,\tau]$ in $D[0,\tau]^K$.

$H^{(1)}(t)$ and $H^{(2)}(s)$ are independent since

$$\begin{aligned}
& Cov(H^{(1)}(t), H^{(2)}(s)) \\
= \;& Cov(n^{-1/2}\sum_{i=1}^{n}\eta_{ik}(\beta_0, t), n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\zeta_{ik}(\beta_0, s)) \\
= \;& E(n^{-1}\sum_{i=1}^{n}\eta_{ik}(\beta_0, t)\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\zeta_{ik}(\beta_0, s)) \\
= \;& E(E\{n^{-1}\sum_{i=1}^{n}\eta_{ik}(\beta_0, t)\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\zeta_{ik}(\beta_0, s)|\mathcal{F}(t)\}) \\
= \;& E(n^{-1}\sum_{i=1}^{n}\eta_{ik}(\beta_0, t)\sum_{i=1}^{n}E\{(1-\frac{\xi_i}{\tilde{\alpha}})|\mathcal{F}(t)\}\zeta_{ik}(\beta_0, s)) \\
= \;& 0
\end{aligned}$$

Therefore, $H(t) = (H^{(1)}(t) + H^{(2)}(t))$ converges weakly to zero-mean Gaussian process $\mathcal{H}(t) = (\mathcal{H}^{(1)}(t) + \mathcal{H}^{(2)}(t))$ in $D[0,\tau]^K$ whose covariance function between $\mathcal{H}_j^{(2)}(t)$ and $\mathcal{H}_k^{(2)}(s)$ is $E\{\eta_{1j}(\beta_0, t), \eta_{1k}(\beta_0, s)\} + \frac{1-\alpha}{\alpha} E\{\zeta_{1j}(\beta_0, t), \zeta_{1k}(\beta_0, s)\}$.

3.4 Simulations

We conducted simulation studies to examine the performance of the proposed methods and to compare them with the Borgan et al. [2000] method for univariate outcomes and the Kang and Cai [2009] method for multiple outcomes. We also compared separate analysis with joint analysis. Suppose case-cohort studies have been conducted for diseases 1 and 2. Then covariate information is collected for the subcohort and all the subjects with disease 1 and/or 2. We generated bivariate failure times from the Clayton–Cuzick model [Clayton and Cuzick, 1985] with the conditional survival function

$$S(t_1, t_2 \mid Z_1, Z_2) = \left\{ \exp^{\{\int_0^{t_1} \lambda_{01}(t) \exp^{\beta_1 Z_1} dt\}/\theta} + \exp^{\{\int_0^{t_2} \lambda_{02}(t) \exp^{\beta_2 Z_2} dt\}/\theta} - 1 \right\}^{-\theta},$$

where $\lambda_{0k}(t)$ and β_k $(k = 1, 2)$ are the baseline hazard function and the effect of a covariate for disease k, respectively, and θ is the association parameter between the failure times of the two diseases. Kendall's tau is $\tau_\theta = (2\theta + 1)^{-1}$. Smaller Kendall's tau values represent lower correlation between T_1 and T_2. Values of 0·1, 4, and 10 are used for θ, with corresponding Kendall's tau values 0·83, 0·11, and 0·05, respectively. We set the baseline hazard functions $\lambda_{01}(t) \equiv 2$ and $\lambda_{02}(t) \equiv 4$. We consider the situation $Z_1 = Z_2 = Z$, where Z is generated from a Bernouilli distribution with $\mathrm{pr}(Z = 1) = 0{\cdot}5$. Censoring times are simulated from a uniform distribution $[0, u]$, where u depends on the specified level of the censoring probability. We set the event proportions of approximately 8% and 20% for $k = 1$, and 14% and 35% for $k = 2$. The corresponding u values are 0·08 and 0·22, respectively, for $\beta_1 = 0{\cdot}1$; they are 0·06 and 0·16 for $\beta_1 = \log 2$. The sample size of the full cohort is set to be $n = 1000$. We create the subcohort by simple random sampling and consider subcohort sizes of 100 and 200. For each configuration, 2000 simulations were conducted.

In the first set of simulations, we consider the case that disease 1 is of primary interest. We compare the performance of our proposed estimator with the estimator of Borgan et al. [2000]. Table 3.1 summarizes the results. We see that both methods are approximately unbiased. The average of the estimated standard error of the proposed estimator is close to the empirical standard deviation, and the coverage rate of the 95% confidence interval is close to the nominal level. As expected, the variation of the estimators in general decreases as the subcohort size increases. Our proposed estimators have smaller variance relative to the estimators of Borgan et al. [2000] in all cases. This shows that the extra information collected on subjects with the other disease helps to increase efficiency. The efficiency gain is larger in situations with larger event proportions, smaller subcohort sizes and lower correlation. We also considered disease 2 with $\beta_2 = \log 2$ and conducted additional simulations to compare our proposed estimator with those of Prentice [1986], Self and Prentice [1988],

Table 3.1: Simulation result for a single disease outcome: $\beta_1 = \log(2) = 0.693$

Event proportion	Size of subcohort	τ_θ	The proposed method $\tilde{\beta}_1$	SE	SD	CR	Borgan et al.'s method $\hat{\beta}_1$	SE	SD	CR	SRE
8%	100	0·83	0·706	0·32	0·32	94	0·705	0·33	0·33	94	1·04
		0·11	0·718	0·31	0·32	94	0·719	0·33	0·33	94	1·07
		0·05	0·708	0·32	0·32	94	0·705	0·33	0·33	94	1·06
	200	0·83	0·715	0·28	0·28	95	0·716	0·28	0·28	95	1·02
		0·11	0·704	0·28	0·28	95	0·705	0·28	0·29	95	1·03
		0·05	0·697	0·28	0·27	95	0·698	0·28	0·28	95	1·05
20%	100	0·83	0·703	0·25	0·25	94	0·704	0·26	0·27	95	1·13
		0·11	0·694	0·23	0·23	94	0·694	0·26	0·27	95	1·31
		0·05	0·700	0·23	0·23	94	0·701	0·26	0·26	95	1·29
	200	0·83	0·693	0·20	0·20	95	0·692	0·21	0·21	95	1·10
		0·11	0·696	0·19	0·19	95	0·699	0·21	0·21	95	1·17
		0·05	0·694	0·19	0·19	95	0·695	0·21	0·21	95	1·26

SE, average standard errors; SD, sample standard deviation; CR, coverage rate (%) of the nominal 95% confidence intervals; SRE= $\mathrm{SD}_c^2/\mathrm{SD}_p^2$, sample relative efficiency, where SD_c and SD_p are the sample standard deviation for the Borgan et al. [2000]'s method and the proposed method, respectively.

Kalbfleisch and Lawless [1988], and Barlow [1994]. Similar results were obtained but are not presented in the paper due to space limitations.

In the second set of simulations, we are interested in the joint analysis of the two diseases. We fit the following models:

$$\lambda_{ik}(t \mid Z_i) = Y_{ik}(t)\lambda_{0k}(t)e^{\beta_k Z_i} \quad (k = 1, 2;\ i = 1, \ldots, n).$$

We compare the performance of the proposed estimator with the estimator of Kang and Cai [2009]. Table 3.2 provides summary statistics for the estimator of β_1 for different combinations of event proportion, subcohort sample size, and correlation. The estimates from both methods are nearly unbiased, and their estimated standard errors are close to the empirical standard deviations. Our method is more efficient than that of Kang and Cai [2009]. The efficiency gain is very limited when the event proportion is small. Higher efficiency gains are associated with smaller subcohort sizes. Estimates for β_2 are not shown in Table 3.2, but the overall performance is similar to that of β_1.

We also compared separate analysis of the two diseases with the joint analysis using the

Table 3.2: Simulation result for multiple disease outcomes: $[\beta_1,\, \beta_2] = [0.1,\, 0.7]$

Event proportion	Size of subcohort	τ_θ	The proposed method $\tilde{\beta}_1^M$	SE	SD	CR	Kang & Cai's method $\hat{\beta}_1^M$	SE	SD	CR	SRE $\hat{\beta}_1$
$[8\%, 14\%]$	100	0·83	0·099	0·31	0·30	95	0·101	0·32	0·31	95	1·07
		0·11	0·101	0·30	0·30	95	0·098	0·32	0·32	95	1·13
		0·05	0·109	0·30	0·31	94	0·111	0·32	0·33	94	1·11
	200	0·83	0·106	0·26	0·27	95	0·105	0·27	0·27	95	1·04
		0·11	0·096	0·26	0·26	94	0·096	0·27	0·27	94	1·05
		0·05	0·098	0·26	0·27	94	0·098	0·27	0·27	94	1·05
$[20\%, 35\%]$	100	0·83	0·098	0·23	0·24	94	0·094	0·26	0·27	94	1·24
		0·11	0·099	0·22	0·22	94	0·097	0·26	0·26	95	1·42
		0·05	0·095	0·22	0·22	94	0·101	0·26	0·27	95	1·44
	200	0·83	0·103	0·19	0·19	94	0·104	0·20	0·21	95	1·19
		0·11	0·098	0·18	0·18	95	0·097	0·20	0·20	95	1·29
		0·05	0·098	0·18	0·18	95	0·100	0·20	0·20	96	1·31

SE, average standard errors; SD, sample standard deviation; CR, coverage rate (%) of the nominal 95% confidence intervals; SRE= $\mathrm{SD}_e^2/\mathrm{SD}_p^2$, sample relative efficiency, where SD_e and SD_p are the sample standard deviation for the Kang and Cai [2009]'s method and the proposed method, respectively.

Table 3.3: Comparison between separate and joint analysis: $\beta_1 = \log 2$, $\Pr(\Delta = 1)$=0.2

Size of subcohort	τ_θ	Separate analysis: The proposed weight $\tilde{\beta}_1$	SE	SD	Separate analysis: Borgan at al.'s method $\hat{\beta}_1$	SE	SD
100	0·83	0·713	0·244	0·245	0·716	0·263	0·265
	0·11	0·702	0·226	0·236	0·705	0·262	0·270
	0·05	0·700	0·226	0·232	0·710	0·263	0·268
200	0·83	0·703	0·196	0·194	0·704	0·206	0·206
	0·11	0·697	0·186	0·193	0·699	0·205	0·213
	0·05	0·698	0·186	0·187	0·702	0·206	0·209
Size of subcohort	τ_θ	**Joint analysis: The proposed weight** $\tilde{\beta}_1^M$	**SE**	**SD**	**Joint analysis: Kang and Cai's method** $\hat{\beta}_1^M$	**SE**	**SD**
100	0·83	0·711	0·243	0·245	0·713	0·262	0·264
	0·11	0·701	0·226	0·235	0·701	0·261	0·267
	0·05	0·700	0·225	0·231	0·707	0·262	0·266
200	0·83	0·703	0·195	0·194	0·703	0·205	0·205
	0·11	0·696	0·186	0·193	0·697	0·205	0·212
	0·05	0·698	0·186	0·187	0·700	0·205	0·209

SE, average standard errors; SD, sample standard deviation.

Table 3.4: Type I error and power (%) in separate and joint analyses: Pr(Δ = 1)=0.2

		Type I error ($\beta_1 = \beta_2 = \log 2$)				Power ($\beta_1 = 0{\cdot}1$, $\beta_2 = 0{\cdot}7$)			
Size of		Separate analysis		Joint analysis		Separate analysis		Joint analysis	
subcohort	τ_θ	P	BR	P	KC	P	BR	P	KC
100	0·83	0·6	0·6	6·3	6·7	49	42	90	78
	0·11	0·8	1·7	5·9	5·9	56	42	83	61
	0·05	1·2	2·1	5·1	5·6	59	43	81	61
200	0·83	0·2	0·3	5·2	5·8	80	72	98	94
	0·11	1·6	1·9	5·4	5·4	77	65	89	78
	0·05	1·8	2·5	5·3	5·4	79	68	90	79

P, the proposed weight; BR, the method of Borgan et al. [2000]; KC, the method of Kang and Cai [2009].

proposed method. Data were generated satisfying the following model:

$$\lambda_k(t \mid Z_1, Z_2) = \lambda_{0k}(t)e^{\beta_k Z + \beta_3 Z^*} \quad (k = 1, 2),$$

where β_1 represents the effect of Z on the risk of disease 1, β_2 represents the effect of Z on the risk of disease 2, and β_3 represents the common effect of Z^* for both diseases. We set $\beta_1 = \beta_2 = \log 2$ and $\beta_3 = 0{\cdot}1$. Table 3.3 summarizes the results for β_1. The sample standard deviations of Kang & Cai's estimator in the joint analysis are slightly smaller than Borgan's estimator in the separate analysis. The sample standard deviations of the proposed estimators are similar in the joint and separate analyses, and they are smaller than Kang & Cai's and Borgan's estimators, respectively. Conclusions for the estimator of β_2 are similar. We also conducted hypothesis tests for $H_0 : \beta_1 = \beta_2$. Table 3.4 presents the Type I error rates and power of the tests at the 0·05 significance level. The tests under the separate analysis treat the two estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$, as from two independent samples. Type I error rates from separate analyses are much lower than 5% while those from the joint analysis are close to 5%. The settings for power analysis are the same as before except that $\beta_1 = 0{\cdot}1$ and $\beta_2 = 0{\cdot}7$. Tests based on the proposed methods are more powerful than those based on Kang & Cai's and Borgan's methods, and the joint analysis produces more powerful tests than the separate analysis.

3.5 Data analysis

We apply the proposed method to analyze data from the Busselton Health Study [Cullen, 1972; Knuiman et al., 2003], conducted in the south-west of Western Australia, and intended to evaluate the association between coronary heart disease and stroke and their risk factors. General health information for adult participants was obtained by questionnaire every three years from 1966 to 1981. This study population consists of 1612 men and women aged 40-89 who participated in 1981 and were free of coronary heart disease or stroke at that time. Coronary heart disease event is defined as hospital admission, any procedure, or death related to coronary heart disease. Stroke event is defined as hospital admission, any procedure, or death from stroke. The outcomes of interest were time to the first coronary heart disease event and time to the first stroke event. The event time for a subject was considered censored if the subject was free of that event type by December 31, 1998 or lost to follow-up during the study period.

One of the main interests of the study was to compare the effect of serum ferritin on coronary heart disease with its effect on stroke. To reduce cost and preserve stored serum, case-cohort sampling was used. Serum ferritin was measured for all the subjects with coronary heart disease and/or stroke as well as those in the subcohort. We conduct a joint analysis of the two diseases. In our analysis, the full cohort consists of 1210 subjects with viable blood serum samples, which includes 174 subjects with only coronary heart disease, 75 with only stroke, and 43 with both diseases. The subcohort consisted of 334 disease-free subjects, 61 with only coronary heart disease, 36 with only stroke, and 19 with both diseases. The total number of assayed sera samples was 626. If a subject was censored and free of both events at the censoring time, then the censoring times for the two disease events were the same. Two subjects died due to both coronary heart disease and stroke, for whom the times for both events were the same. No other subjects died at the first diagnosis of either disease. For this study, it is reasonable to assume, as in the original study [Knuiman et al., 2003], that censoring was conditionally independent of the event processes.

Table 3.5: Analysis results for the Busselton Health Study

	Proposed method				Kang & Cai method			
Variables	$\tilde{\beta}_M$	SE	HR	95% CI	$\hat{\beta}_M$	SE	HR	95% CI
log(ferritin) on CHD	0·145	0·0897	1·16	(0·97, 1·38)	0·092	0·0949	1·10	(0·91, 1·32)
log(ferritin) on Stroke	0·172	0·1219	1·19	(0·93, 1·51)	0·186	0·1304	1·20	(0·93, 1·56)
Age	0·071	0·0069	1·07	(1·06, 1·09)	0·069	0·0070	1·07	(1·06, 1·09)
Triglycerides	0·239	0·0484	1·27	(1·16, 1·40)	0·232	0·0541	1·26	(1·13, 1·40)
BPT	0·423	0·1633	1·53	(1·11, 2·10)	0·408	0·1727	1·50	(1·07, 2·11)

CHD, coronary heart disease; BPT, Blood pressure treatment; SE, standard error; HR, hazard ratio; CI, confidence interval.

We fit the following model

$$\lambda_k(t \mid Z_1, Z_2, Z_3, Z_4) = \lambda_{0k}(t)e^{\beta_{1k}Z_1+\beta_{2k}Z_2+\beta_{3k}Z_3+\beta_{4k}Z_4} \quad (k = 1, 2),$$

where Z_1, Z_2, Z_3, and Z_4 denote the logarithm of serum ferritin level, age in years, triglycerides in millimoles per liter, and whether subjects had blood pressure treatment, respectively. We then tested $H_0 : \beta_{21} = \beta_{22}, \beta_{31} = \beta_{32}, \beta_{41} = \beta_{42}$ based on the proposed method, and the p-value is 0·138. Therefore, we fit the final model

$$\lambda_k(t \mid Z_1, Z_2, Z_3, Z_4) = \lambda_{0k}(t)e^{\beta_{1k}Z_1+\beta_2 Z_2+\beta_3 Z_3+\beta_4 Z_4} \quad (k = 1, 2).$$

Table 3.5 summarizes the results of the final fit. With a 1 unit increase in the logarithm of the serum ferritin level, the hazard ratio for coronary heart disease risk is increased by 16% and for stroke risk by 19%. When we tested $H_0 : \beta_{11} = \beta_{12}$, H_0 was not rejected with the p-value = 0·823. We also fit the same model using Kang and Cai [2009]'s method. The standard errors for the effects of the logarithm of the serum ferritin level are slightly larger, 0·0949 for coronary heart disease and 0·1304 for stroke.

3.6 Concluding remarks

When disease rates are low, the efficiency gain of the proposed method is not large. When the event rates are low, the number of cases is small, and consequently, the amount of extra information is small. In the case of common diseases, sampling all cases in the traditional case-cohort design with multiple diseases limits applications [Breslow and Wellner, 2007]. Instead, a generalized case-cohort design [Cai and Zeng, 2007] in which cases are sampled can be considered. Extending the proposed weights to this general case merits further investigation.

In our proposed estimation framework, time-dependent covariates can be allowed. However, estimation generally requires one to know the entire history of time-dependent covariates. In many follow-up studies, this may not be true. One commonly used approach for handling time-dependent covariates is to consider the last-value-carry-forward, but this could introduce bias. A more sensible approach is to consider the joint modeling of survival times and longitudinal covariates via shared random effects, which has not been studied for case-cohort data.

When studying multiple diseases, different diseases may be competing risks for the same subject. In a competing risks situation, a subject can only experience at most one event; in the situation we considered, a subject can still experience the other events. Consequently, in the competing risks situation, a subject is at risk for all types of events simultaneously and will not be at risk for any other events as soon as one event occurs. Our approach in this paper can be adapted to competing risks by modifying the at-risk process and the weight function, but analysis will be based on the cause-specific hazards as studied in Sorensen and Andersen [2000].

The current method is based on estimating equations, which improves the estimation efficiency by incorporating a refined weight function for the risk set. However, it is not semiparametric efficient. To derive the most efficient estimator, we need to specify the joint distribution of the correlated failure times from the same subject and consider nonparametric maximum likelihood estimation based on the joint likelihood function for case-cohort sampling. This may be very challenging, especially when expensive covariates are continuous. This is an interesting topic which warrants future research.

Chapter 4

Stratified case-cohort studies with nonrare events

4.1 Introduction

Case-cohort study design is an economical means for large cohort studies since it can be expensive to assemble covariate information for all cohort members [Prentice, 1986]. To conduct case-cohort design, there are two sampling steps. First, a random sample from the full cohort, named the subcohort, is selected via simple random sampling. Second, we sample subjects having diseases of interest outside the subcohort. The covariate information on the exposure is obtained for the subcohort members as well as sampled cases or failures.

In biomedical studies some covariate information is often available for all subjects in the cohort such as age or gender and these covariate information could be used to define strata variables. Under the situation, the subcohort can be selected via stratified random sampling based on strata variables, which could lead to more powerful and efficient case-cohort study than unstratified case-cohort study using simple random sampling of the subcohort [Borgan et al., 2000]. Kulich and Lin [2004] and Samuelsen et al. [2007] proposed stratified case-cohort design by using the covariate data outside the case-cohort sample and using local averaging method, respectively.

For case-cohort studies with a single disease outcome as well as multivariate disease outcomes, extensive progress has been made. From unstratified case-cohort data, Prentice [1986] proposed a pseudo-likelihood approach, Self and Prentice [1988] proposed the inference of a slightly modified pseudo-likelihood estimator, and Barlow [1994] developed a robust estimator of the variance with a time-varying weight. For multivariate failure time outcomes, Lu and Shih [2006] proposed estimation for case-cohort studies for clustered failure time. In order to be able to compare the effects of a risk factor on different diseases, Kang and Cai [2009] developed the estimation procedure based on the joint analysis in generalized case-cohort studies. By using stratum variables, stratified case-cohort design with a single disease outcome has been studied [Borgan et al., 2000; Kulich and Lin, 2000a, 2004].

Aforementioned methods were considered in traditional case-cohort design when diseases are infrequent. However, in many biomedical studies, the disease rate may not be low or the number of cases is large. Under the situation, Cai and Zeng [2007] proposed the generalized case-cohort design by selecting a subset of all cases or failures. When stratum variables are available for all cohort members, Kang and Cai [2010] considered stratified generalized case-cohort design by using stratified random sampling of the subcohort and cases. For example, the Atherosclerosis Risk in Communities (ARIC) study is to investigate the association between high-sensitivity C-reactive protein (hs-CRP) and incident diabetes events. Since the disease rate of incident diabetes is 11.2% and frozen biologic specimen from which hs-CRP can be measured should be conserved, selecting all subjects with incident diabetes is prohibited. To preserve frozen biologic specimen and save cost, the generalized case-cohort study was conducted by sampling a subset of diabetes cases. Based on age (≤ 55, > 55), gender, and race, the subcohort and a subset of diabetes cases were selected via stratified sampling.

When it is of interest to study the effect of one risk factor on multiple diseases, several case-cohort studies were conducted separately. For example, another case-cohort study for association between hs-CRP and incident coronary heart diseases (CHD) had been conducted in the ARIC study [Ballantyne et al., 2004]. In this study, hs-CRP information was available on the subcohort as well as all incident coronary heart diseases. When constructing estimating equations for diabetes in generalized stratified case-cohort studies, hs-CRP information for subjects collected from CHD cases was not used. This motivates us to consider a different approach which can utilize all available exposure information in generalized stratified case-cohort studies.

In this paper, we develop estimation procedure for generalized stratified case-cohort study design with a single disease outcome as well as multiple disease outcomes by using all available exposure information. In section 4.2, we propose models and estimation procedures for the proposed methods. Section 4.3 summarizes asymptotic properties to be proved for the proposed estimators and section 4.4 reports some simulation results. In section 4.5, we apply our proposed methods to data from the ARIC study. In section 4.6, concluding remarks are provided.

4.2 Model and estimation

4.2.1 Model

Suppose that there are n independent subjects and K diseases of interest in a cohort which can divided into L mutually exclusive strata using information available for all the cohort members. Suppose that the total size of cohort n is partitioned into n_l intervals for $l = 1, \ldots, L$. Let T_{lik} be the failure time, C_{lik} the potential censoring time, and $Z_{lik}(t)$ be a $p \times 1$ possibly time-dependent covariates vector for disease k of subject i in stratum l. Let $X_{lik} = \min(T_{lik}, C_{lik})$ denote the observed time, $\Delta_{lik} = I(T_{lik} \leq C_{lik})$ the indicator for failure, $N_{lik}(t) = I(X_{lik} \leq t, \Delta_{lik} = 1)$ the counting process for the observed failure time, and $Y_{lik}(t) = I(X_{lik} \geq t)$ the at risk indicator for disease k of subject i in stratum l, where $I(.)$ is the indicator function.

We assume that all the time-dependent covariates are external [Kalbfleisch and Prentice, 2002] and T_{lik} is independent of C_{lik} for given possibly time-dependent covariates $Z_{lik}(t)$. Let τ denote the end of study time. For disease k of subject i in stratum l, the hazard function $\lambda_{lik}(.)$ associated with $Z_{lik}(t)$ is given by

$$\lambda_{lik}\{t|Z_{lik}(t)\} = Y_{lik}(t)\lambda_{0k}(t)e^{\beta_0^T Z_{lik}(t)}, \tag{4.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease k of subject i in stratum l and β_0 is p-dimensional fixed and unknown parameters. Model(4.1) can incorporate disease-type-specific effect model $\lambda_{lik}\{t|Z^*_{lik}(t)\} = Y_{lik}(t)\ \lambda_{0k}(t)e^{\beta_k^T Z^*_{lik}(t)}$ as a special case. Specifically, we define $\beta_0^T = (\beta_1^T, \ldots, \beta_k^T, \ldots, \beta_K^T)$ and $Z_{lik}(t)^T = (0^T_{li1}, \ldots, 0^T_{li(k-1)}, \{Z^*_{lik}(t)\}^T, 0^T_{li(k+1)}, \ldots, 0^T_{liK})$ where 0^T is $1 \times p$ zero vector. We have $\beta_0^T Z_{lik}(t) = \beta_k^T Z^*_{lik}(t)$.

Since obtaining Z for all the subjects in the cohort can be very expensive, a generalized case-cohort design is often used where a subcohort and a sample of disease cases from each stratum are selected to measure Z's values. Let V_{ik} denote the discrete random variable for indicating stratum for subject i with disease k. The stratum variable is assumed to be independent of T_{lik} given $Z_{lik}(t)$ [Kulich and Lin, 2004].

Under generalized case-cohort design with stratified sampling, subjects in the subcohort are assumed to be selected by stratified random sampling. Specifically, we select a fixed size $\tilde{n}_l$ subjects from n_l subjects in stratum l into the subcohort by using simple random sampling without replacement. Let the total size of the subcohort be $\tilde{n} = \sum_{l=1}^L \tilde{n}_l$ and $\tilde{\alpha}_l = \Pr(\xi_{li} = 1) = \tilde{n}_l/n_l$ be the selection probability of subject i in stratum l into the subcohort, where $\xi_{li} = 1$ denotes that subject i in stratum l is selected into the subcohort and $\xi_{li} = 0$ denotes otherwise. After sampling the subcohort, stratified random samples of cases outside of the subcohort for each disease outcome are drawn. Specifically, for disease k in stratum l, we select $\tilde{m}_{lk}$ cases outside of the subcohort using simple random sampling without replacement. Let $\tilde{\gamma}_{lk} = \Pr\ (\eta_{lik} = 1|\Delta_{lik} = 1, \xi_{li} = 0) = \tilde{m}_{lk}/(n_{lk} - \tilde{n}_{lk})$ denote the selection probability of subjects among non-subcohort members with disease k in stratum l, where η_{lik} is the indicator for whether subject i with disease k in stratum l among non-subcohort members is sampled, n_{lk} and $\tilde{n}_{lk}$ denote the number of subjects with disease k in the cohort and in the subcohort in stratum l, respectively. For $k \neq k'$ or $l \neq l'$, $(\eta_{l1k}, \ldots, \eta_{ln_lk})$ is independent of $(\eta_{l'1k'}, \ldots, \eta_{l'n_lk'})$; however, $(\eta_{l1k}, \ldots, \eta_{ln_lk})$ are correlated because of the sampling scheme.

4.2.2 Estimation

The observable information for subject i is $(X_{lik}, \Delta_{lik}, Z_{lik}(t), V_{ik}, 0 \le t \le X_{lik})$ when $\xi_{li} = 1$ or $\eta_{lik} = 1$ and $(X_{lik}, \Delta_{lik}, V_{ik})$ when $\xi_{li} = 0$ and $\eta_{lik} = 0$ $(k = 1, \ldots, K)$. If we ignore the covariate information available for the sampled subjects with other diseases outside of the subcohort, the relative risk parameter β_0 can be estimated by the weighted estimating equation, $\hat{U}^{KC}(\beta) = 0$ [Kang and Cai, 2010] where

$$\hat{U}^{KC}(\beta) = \sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^{\tau} w_{lik}(t)\left\{Z_{lik}(t) - \frac{\hat{S}_k^{(1)}(\beta,t)}{\hat{S}_k^{(0)}(\beta,t)}\right\}dN_{lik}(t), \tag{4.2}$$

$\hat{S}_k^{(d)}(\beta,t) = n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l} w_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}$ for $d = 0, 1$ and 2 and $w_{lik}(t)$ is a time-varying weight function which has the follow form: $w_{lik}(t) = (1 - \Delta_{lik})\xi_{li}\hat{\alpha}_{lk}^{-1}(t) + \Delta_{lik}\xi_{li} + \Delta_{lik}(1-\xi_{li})\eta_{lik}\hat{\gamma}_{lk}^{-1}(t)$, where $\hat{\alpha}_{lk}(t) = \sum_{i=1}^{n_l}(1-\Delta_{lik})\xi_{li}Y_{lik}(t)/\{\sum_{i=1}^{n_l}(1-\Delta_{lik})Y_{lik}(t)\}$, $\hat{\gamma}_{lk}(t) = \sum_{i=1}^{n_l}\Delta_{lik}(1-\xi_{li})\eta_{lik}Y_{lik}(t)/\{\sum_{i=1}^{n_l}\Delta_{lik}(1-\xi_{li})Y_{lik}(t)\}$. Note that we can set $K = 1$ in (4.2) if we are interested in only one disease. If $\hat{\gamma}_{lk}(t)$ for all k is 1, then the generalized stratified case-cohort design is reduced to the traditional stratified case-cohort design whose the weight function $\rho_{lik}(t) = (1 - \Delta_{lik})\xi_{li}\hat{\alpha}_{lk}^{-1}(t) + \Delta_{lik}$.

Note that $\hat{S}_k^{(d)}(\beta,t)$ only uses the covariate information collected for the subcohort and the subset of subjects with disease k outside of the subcohort. In other words, covariate information collected on the subset of subjects with other diseases outside the subcohort is ignored when calculating $\hat{S}_k^{(d)}(\beta,t)$ in the estimating equation. To make use of available information about other diseases, we propose the proposed weight with two types of diseases (i.e. $K = 2$). The key idea in the proposed weight function with two types of diseases is that the weight for one type of the disease uses the covariate information collected on the selected subjects with the other type of the disease. Specifically, subcohort subjects without any disease (i.e. $\prod_{j=1}^{2}(1 - \Delta_{lij})\xi_{li} = 1$) are weighted by $\tilde{\alpha}_{lk}(t)^{-1}$, the inverse of the estimated selection probabilities, while subjects with disease 1 or disease 2 in the subcohort (i.e. $\{1 - \prod_{j=1}^{2}(1 - \Delta_{lij})\}\xi_{li} = 1$) are weighted by 1. To use the information collected on the sampled subjects with disease 2, the sampled non-subcohort subjects with disease 1 (i.e. $\Delta_{li1}(1 - \xi_{li})\eta_{li1} = 1$) can be decomposed into two groups: those with only disease

1 (i.e. $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li1}=1$) and those with both disease 1 and disease 2 (i.e. $\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}=1$). The sampled subjects in the first group (i.e. $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})=1$) are weighted by $\tilde{\gamma}_{l1k}(t)^{-1}$, the inverse of their estimated sampling probabilities. Similarly, the sampled non-subcohort subjects with disease 2 can also be decomposed into two groups: those with only disease 2 (i.e. $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li2}=1$) and those with both disease (i.e. $\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}=1$). Those with only disease 2 are weighted by $\tilde{\gamma}_{l2k}(t)^{-1}$, the inverse of their estimated sampling probabilities. For those sampled non-subcohort subjects with both diseases, they can be weighted by either $\tilde{\gamma}_{l3k}^{-1}(t)$ or $\tilde{\gamma}_{l4k}^{-1}(t)$, the inverse of the estimated sampling probability based on disease 1 and disease 2, respectively. We take the average of $\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}\tilde{\gamma}_{l3k}^{-1}(t)$ and $\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}\tilde{\gamma}_{l4k}^{-1}(t)$ as the weight for this group. Therefore, the proposed weight with two types of diseases has the following form:

$$
\begin{aligned}
\pi_{lik}(t) &= \Pi_{j=1}^{2}(1-\Delta_{lij})\xi_{li}\tilde{\alpha}_{lk}^{-1}(t)+\left\{1-\Pi_{j=1}^{2}(1-\Delta_{lij})\right\}\xi_{li}\\
&+ \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li1}\tilde{\gamma}_{l1k}^{-1}(t)+(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\eta_{li2}\tilde{\gamma}_{l2k}^{-1}(t)\\
&+ \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}\tilde{\gamma}_{l3k}^{-1}(t)+\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}\tilde{\gamma}_{l4k}^{-1}(t),
\end{aligned}
\tag{4.3}
$$

where

$$
\begin{aligned}
&\tilde{\alpha}_{lk}(t)=\frac{\sum_{i=1}^{n_l}\Pi_{j=1}^{2}(1-\Delta_{lij})\xi_{li}Y_{lik}(t)}{\sum_{i=1}^{n_l}\Pi_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)},\ \tilde{\gamma}_{l1k}(t)=\frac{\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li1}Y_{lik}(t)}{\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)}\\
&\tilde{\gamma}_{l2k}(t)=\frac{\sum_{i=1}^{n_l}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\eta_{li2}Y_{lik}(t)}{\sum_{i=1}^{n_l}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)},\ \tilde{\gamma}_{l3k}(t)=\frac{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}Y_{lik}(t)}{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)}\\
&\tilde{\gamma}_{l4k}(t)=\frac{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}Y_{lik}(t)}{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)}.
\end{aligned}
$$

Note that if all cases outside the subcohort are selected, the weight functions in (4.28) reduce to $\phi_{lik}(t)=\prod_{j=1}^{K}(1-\Delta_{lij})\xi_{li}\tilde{\alpha}_{lk}^{-1}(t)+\{1-\prod_{j=1}^{K}(1-\Delta_{lij})\}$.

Using the proposed weight functions in (4.28), we propose the following weighted estimating functions for the estimation of the regression coefficient:

$$
\tilde{U}^{G}(\beta)=\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^{\tau}\pi_{lik}(t)\left\{Z_{lik}(t)-\frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)}\right\}dN_{lik}(t), \tag{4.4}
$$

where $\tilde{S}_k^{(d)}(\beta,t) = n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}$ for $d = 0,1$ and 2. The solution to $\tilde{U}^G(\beta) = 0$ is defined to be the estimator $\tilde{\beta}^G$ for the regression parameter β_0.

A Breslow-Aalen type estimator of the baseline hazard cumulative hazard function is $\tilde{\Lambda}_{0k}(\tilde{\beta}^G, t)$ which is given by

$$\tilde{\Lambda}_{0k}(\beta,t) = \int_0^t \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(u)dN_{lik}(u)}{n\tilde{S}_k^{(0)}(\beta,u)}.$$

4.3 Asymptotic properties

In this section, we summarize the asymptotic properties for the proposed methods. We will show the asymptotic properties of the proposed estimator only for two types of diseases. The other situations can be proved similarly. We make the following assumptions:

(a) $\{T_{li}, C_{li}, Z_{li}\}$, $i = 1,\dots,n$ and $l = 1,\dots,L$ are independent and identically distributed where $T_{li} = (T_{li1},\dots,T_{liK})^T$, $C_{li} = (C_{li1},\dots,C_{liK})^T$, and $Z_{li} = (Z_{li1},\dots,Z_{liK})^T$;

(b) $\mathrm{P}\{Y_{lik}(t) = 1\} > 0$ for $t \in [0,\tau]$, $i = 1,\dots,n$, $l = 1,\dots,L$ and $k = 1,2$;

(c) $|Z_{lik}(0)| + \int_0^\tau |dZ_{lik}(t)| < D_z < \infty$, $i = 1,\dots,n$, $l = 1,\dots,L$ and $k = 1,2$ almost surely and D_z is a constant;

(d) (Asymptotic stability) For $d = 0,1,2$, there exists a neighborhood $\mathcal{B}$ of β_0 such that $s_k^{(d)}(\beta,t)$ are continuous functions and $\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\|S_k^{(d)}(\beta,t) - s_k^{(d)}(\beta,t)\| \xrightarrow{p} 0$ where $S_k^{(d)}(\beta,t) = n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}$;

(e) The matrix $A_k(\beta_0) = \int_0^\tau v_k(\beta_0,t)s_k^{(0)}(\beta_0,t)\lambda_{0k}(t)dt$ is positive definite for $k = 1,2$ where $v_k(\beta,t) = s_k^{(2)}(\beta,t)/s_k^{(0)}(\beta,t) - e_k(\beta,t)^{\otimes 2}$ and $e_k(\beta,t) = s_k^{(1)}(\beta,t)/s_k^{(0)}(\beta,t)$;

(f) (Asymptotic regularity) For all $\beta \in \mathcal{B}$, $t \in [0,\tau]$, and $k = 1,2$, $S_k^{(1)}(\beta,t) = \frac{\partial}{\partial\beta}S_k^{(0)}(\beta,t)$, and $S_k^{(2)}(\beta,t) = \frac{\partial^2}{\partial\beta\partial\beta'}S_k^{(0)}(\beta,t)$, where $S_k^{(d)}(\beta,t)$, $d = 0,1,2$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0,\tau]$ and are bounded on $\mathcal{B}\times[0,\tau]$, $s_k^{(0)}$ is bounded away from zero on $\mathcal{B}\times[0,\tau]$;

(g) (Finite interval) For all $k = 1,2$, $\int_0^\tau \lambda_{0k}(t)dt < \infty$;

To show the desired asymptotic properties for generalized case-cohort samples, the following conditions are also needed:

(h) For all $l = 1, \ldots, L$, $\lim_{n_l \to \infty} \tilde{\alpha}_l = \alpha_l$, where $\tilde{\alpha}_l = \tilde{n}_l / n_l$ and α is a positive constant.

(i) $\lim_{n_l \to \infty} \tilde{\gamma}_{l1} = \lim_{n_l \to \infty} \tilde{\gamma}_{l3} = \gamma_{l1}, \lim_{n_l \to \infty} \tilde{\gamma}_{l2} = \lim_{n_l \to \infty} \tilde{\gamma}_{l4} = \gamma_{l2}$ where $\tilde{\gamma}_{l1} = \Pr[\eta_{li1} = 1 | \Delta_{li1}(1 - \Delta_{li2}) = 1, \xi_{li} = 0] = \tilde{m}_{l,10}/(n_{l,10} - \tilde{n}_{l,10})$, $\tilde{m}_{l,10}$ denotes the number of sampled non-subcohort subjects in the lth stratum with only disease 1, but not disease 2 (i.e. $\Delta_{li1}(1 - \Delta_{li2}) = 1$), $n_{l,10}$ and $\tilde{n}_{l,10}$ denote the number of subjects with only diseases 1, but not disease 2 (i.e. $\Delta_{li1}(1 - \Delta_{li2}) = 1$) in the cohort and the subcohort in lth stratum, respectively, $\tilde{\gamma}_{l2} = \Pr[\eta_{li2} = 1 | (1 - \Delta_{li1})\Delta_{li2} = 1, \xi_{li} = 0] = \tilde{m}_{l,01}/(n_{l,01} - \tilde{n}_{l,01})$, $\tilde{m}_{l,01}$ denotes the number of sampled non-subcohort subjects in the lth stratum with only disease 2, but not disease 1 (i.e. $(1 - \Delta_{li1})\Delta_{li2} = 1$), $n_{l,01}$ and $\tilde{n}_{l,01}$ denote the number of subjects with only diseases 2, but not disease 1 (i.e. $(1 - \Delta_{li1})\Delta_{li2} = 1$) in the cohort and the subcohort in lth stratum, respectively, $\tilde{\gamma}_{l3} = \Pr[\eta_{li1} = 1 | \Delta_{li1}\Delta_{li2} = 1, \xi_{li} = 0] = \tilde{m}_{l,111}/(n_{l,11} - \tilde{n}_{l,11})$, $\tilde{m}_{l,111}$ denotes the number of non-subcohort subjects with both disease 1 and disease 2 who are sampled with respect to disease 1 in the lth stratum, $n_{l,11}$ and $\tilde{n}_{l,11}$ denote the number of subjects with both diseases 1 and disease 2 in the cohort and the subcohort in lth stratum, respectively, $\tilde{\gamma}_{l4} = \Pr[\eta_{li2} = 1 | \Delta_{li1}\Delta_{li2} = 1, \xi_{li} = 0] = \tilde{m}_{l,112}/(n_{l,11} - \tilde{n}_{l,11})$, $\tilde{m}_{l,112}$ denotes the number of non-subcohort subjects with both disease 1 and disease 2 who are sampled with respect to disease 2 in the lth stratum, and γ_{lk} is a positive constant on (0,1] for all $k = 1, 2$ and $l = 1, \ldots, L$.

(j) $\lim_{n \to \infty} n_l / n_= q_l$, where q_l is a positive constant on (0,1) for all $l = 1, \ldots, L$.

We summarize the asymptotic properties of $\tilde{\beta}^G$ in the following theorem.

Theorem 3. *Under the regularity conditions (a)-(j), $\tilde{\beta}^G$ converges in probability to β_0 and $n^{1/2}(\tilde{\beta}^G - \beta_0)$ is asymptotically normally distributed with mean zero and with the covariance matrix*

$A(\beta_0)^{-1}\Sigma_G(\beta_0)A(\beta_0)^{-1}$, *where*

$$
\begin{aligned}
& A(\beta) = \sum_{k=1}^{2} A_k(\beta),\ \Sigma_G(\beta) = \sum_{l=1}^{L} q_l[V_{I,l}(\beta) + \frac{1-\alpha_l}{\alpha_l}V_{II,l}(\beta) + (1-\alpha_l)\sum_{k=1}^{2} V_{III,lk}(\beta)], \\
& V_{I,l}(\beta) = E_l[\sum_{k=1}^{2} Q_{l1k}(\beta)]^{\otimes 2}, \\
& V_{II,l}(\beta) \\
= & \ Var_l[\prod_{j=1}^{2}(1-\Delta_{l1j})\sum_{k=1}^{2}\int_0^{\tau}[R_{l1k}(\beta,t) - \frac{Y_{l1k}(t)E_l(\prod_{j=1}^{2}(1-\Delta_{l1j})R_{l1k}(\beta,t))}{E_l(\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t))}]d\Lambda_{0k}(t)], \\
& V_{III,lk}(\beta) \\
= & \ Pr(\Theta_{l10})\frac{1-\gamma_{l1}}{\gamma_{l1}} Var_l(Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l10},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l10}\}}|\Theta_{l10},\xi_{l1}=0) \\
+ & \ Pr(\Theta_{l01})\frac{1-\gamma_{l2}}{\gamma_{l2}} Var_l(Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l01},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l01}\}}|\Theta_{l01},\xi_{l1}=0) \\
+ & \ \frac{1}{4}Pr(\Theta_{l11})\frac{1-\gamma_{l1}}{\gamma_{l1}} Var_l(Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0) \\
+ & \ \frac{1}{4}Pr(\Theta_{l11})\frac{1-\gamma_{l2}}{\gamma_{l2}} Var_l(Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0), \\
& Q_{lik}(\beta) = \int_0^{\tau}\{Z_{lik} - e_k(\beta,t)\}dM_{lik}(t), \Theta_{ljk} = \{(\Delta_{l11} = j \ \ and\ \Delta_{l12} = k)\}, \\
& R_{lik}(\beta,t) = Y_{lik}(t)[Z_{lik}(t) - e_k(\beta,t)]e^{\beta^T Z_{lik}(t)}.
\end{aligned}
$$

Note that $\Sigma_G(\beta)$ consists of three parts. The first part $V_{I,l}(\beta)$ is a contribution to the variance from the full cohort, the second part $V_{II,l}(\beta)$ and the last part $V_{III,lk}(\beta)$ are due to sampling the subcohort from the full cohort and due to sampling a portion of cases in non-subcohort. For cohort studies, the second and last part vanish and their variance is only first part $V_{I,l}(\beta)$. If traditional stratified case-cohort studies are conducted, then the last part goes to zero and so the first and second parts are remained. For unstratified generalized case-cohort studies (i.e. $L = 1$ and $q_l = 1$), variance consists of $V_{I,1}(\beta)$, $V_{II,1}(\beta)$, and $V_{III,1k}(\beta)$.

We summarize the asymptotic property of the proposed baseline cumulative hazard estimator $\tilde{\Lambda}_{0k}(\tilde{\beta}^G, t)$.

Theorem 4. *Under the regularity conditions (a)-(j), $\tilde{\Lambda}_{0k}(\tilde{\beta}^G, t)$ is a consistent estimator of $\Lambda_{0k}(t)$ in $t \in [0,\tau]$ and $P(t) = [P_1(t), P_2(t)]^T = [n^{1/2}(\tilde{\Lambda}_{01}(\tilde{\beta}^G, t) - \Lambda_{01}(t)), n^{1/2}(\tilde{\Lambda}_{02}(\tilde{\beta}^G, t) - \Lambda_{02}(t))]^T$ converges weakly to the Gaussian process $\mathcal{P}(t) = \{\mathcal{P}_1(t), \mathcal{P}_2(t)\}^T$ in $D[0,\tau]^K$*

with mean zero and the following covariance function $\mathcal{P}_{jk}(t,s)$ *between* $\mathcal{P}_j(t)$ *and* $\mathcal{P}_k(s)$ *for* $j \neq k$.

$$\begin{aligned}\mathcal{P}_{jk}(t,s)(\beta_0) &= \sum_{l=1}^{L} q_l[E_l\{v_{l1j}(\beta_0,t)v_{l1k}(\beta_0,s)\} \\ &+ \frac{1-\alpha_l}{\alpha_l}E_l\{\zeta_{l1j}(\beta_0,t)\zeta_{l1k}(\beta_0,s)\} + E_l\{\varphi_{l1j}(\beta_0,t)\varphi_{l1k}(\beta_0,s)\}],\end{aligned}$$

where

$$v_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{2} Q_{lim}(\beta,t) + \int_0^t \frac{1}{s_k^{(0)}(\beta,u)}dM_{lik}(u),$$

$$\zeta_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{2}\int_0^{\tau}\Omega_{lim}(\beta,u)d\Lambda_{0m}(u)$$
$$+\prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t Y_{lik}(u)\{e^{\beta^T Z_{lik}(u)} - \frac{E_l(\prod_{j=1}^{K}(1-\Delta_{l1j})e^{\beta^T Z_{l1k}(u)}Y_{l1k}(u))}{E_l(\prod_{j=1}^{K}(1-\Delta_{l1j})Y_{l1k}(u))}\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta,u)},$$

$$\varphi_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{2}(1-\xi_{li})[\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l10})$$
$$+(1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l01}) + \frac{1}{2}\Delta_{li1}\Delta_{li2}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l11})$$
$$+(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l11})\}] + (1-\xi_{li})[\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l10})$$
$$+(1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l01}) + \frac{1}{2}\Delta_{li1}\Delta_{li2}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l11})$$
$$+(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l11})\}],$$

$$B_{lik}^{(1)}(\beta,t|\Theta_{ljm}) = \int_0^t \frac{1}{s_k^{(0)}(\beta,u)}[dM_{lik}(u) - Y_{lik}(u)\frac{E\{dM_{l1k}(u)|\Theta_{ljm},\xi_{li}=0\}}{E\{Y_{l1k}(u)|\Theta_{ljm}\}}],$$

$$B_{lik}^{(2)}(\beta,t|\Theta_{ljm}) = Q_{lik}(\beta) - \int_0^t Y_{lik}(u)\frac{E\{dQ_{l1k}(\beta,u)|\Theta_{ljm},\xi_{li}=0\}}{E\{Y_{l1k}(\beta,u)|\Theta_{ljm}\}},$$
$$l_k(\beta,t)^T = -\int_0^t e_k(\beta,u)d\Lambda_{0k}(u), \Theta_{ljk} = \{\Delta_{li1} = j \ \ and \ \ \Delta_{li2} = k\}.$$

The proofs for Theorem 3 and 4 are provided in Appendix.

We summarize the asymptotic efficiency for unstratified case-cohort studies (i.e. $L = 1$) with two types of diseases (i.e. $K = 2$). Note that the covariance matrix for $\tilde{\beta}^G$,

$A(\beta_0)^{-1}\Sigma_G(\beta_0)A(\beta_0)^{-1}$ involves the first derivative of the weighted estimating functions $A(\beta_0)$ and the asymptotic variance of the weighted estimating functions, $\Sigma_G(\beta_0)$.

Theorem 5. *Under the condition* $E[w_{111}^2 - \pi_{11}^2]E[w_{112}^2 - \pi_{11}^2] > (\rho E[w_{111}w_{112} - \pi_{11}^2])^2$*, the asymptotic variance for our proposed method* $A(\beta_0)^{-1}\Sigma_G(\beta_0)A(\beta_0)^{-1}$ *is smaller than that for Kang and Cai [2010]'s method, where*

$$
\begin{aligned}
E_1[w_{111}^2 - \pi_{11}^2] &= (1-\alpha_1)p_2\left[\frac{1+\alpha_1}{\alpha_1}(1-p_1) + \frac{3p_1}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{12}}\right],\\
E_1[w_{112}^2 - \pi_{11}^2] &= (1-\alpha_1)p_1\left[\frac{1+\alpha_1}{\alpha_1}(1-p_2) + \frac{3p_2}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{11}}\right],\\
E_1[w_{111}w_{112} - \pi_{11}^2] &= (1-\alpha_1)\left[p_1\{1-\frac{1}{\gamma_{11}}\} + p_2\{1-\frac{1}{\gamma_{12}}\} + p_1p_2\{\frac{3}{4}(\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}) - 1\}\right],\\
\rho &= Corr(Q_{111}(\beta), Q_{112}(\beta)), \alpha_1 = pr(\xi_{11}=1),\ p_1 = pr(\Delta_{111}=1),\ p_2 = pr(\Delta_{112}=1),\\
\gamma_{11} &= pr(\eta_{111}=1|\Delta_{111}(1-\xi_{11})=1),\ \gamma_{12} = pr(\eta_{112}=1|\Delta_{112}(1-\xi_{11})=1).
\end{aligned}
$$

Specifically, smaller α_1 induces larger $(1+\alpha_1)/\alpha_1$, which dominates other contributions in $E[w_{111}^2 - \pi_{11}^2]E[w_{112}^2 - \pi_{11}^2]$. The quantity $(\rho E[w_{111}w_{112} - \pi_{11}^2])^2$ depends on the selection probability of a subset of cases γ_{11} and γ_{12} for fixed the disease rates p_1 and p_2. This indicates that in situations where the subcohort size is smaller and the selected case proportion is higher, the proposed method produce more efficient estimates over that of Kang & Cai's method.

If we consider the simple situation such as $p_1 = p_2 = p$, $\gamma_{11} = \gamma_{12} = \gamma$, and $\rho = 1$, specific conditions to lead larger power are $0 < p < \frac{2}{3}$, $\frac{1}{2} - \frac{3}{4}p < \gamma < 1 - \frac{3}{2}p$, $0 < \alpha_1 < \gamma(1 - \frac{3}{2}p - \gamma)^{-1}$.

4.3.1 Proofs of Theorems

Under the assumptions in Section 4.3, we will outline the proofs for the main theorems.

Before we prove theorems, we consider the asymptotic properties of time-varying sampling probability estimator $\tilde{\alpha}_{lk}(t) = \sum_{i=1}^{n_l}\xi_{li}[\prod_{j=1}^{2}(1-\Delta_{lij})]Y_{lik}(t)/\sum_{i=1}^{n_l}[\prod_{j=1}^{2}(1-\Delta_{lij})]Y_{lik}(t)$. For each k, by the Taylor expansion series of $\tilde{\alpha}_{lk}^{-1}(t)$ around $\tilde{\alpha}_l$,

$$\tilde{\alpha}_{lk}^{-1}(t) - \tilde{\alpha}_l^{-1} = -\frac{1}{\alpha_l^*(t)^2}\{\tilde{\alpha}_{lk}(t) - \tilde{\alpha}_l\}, \tag{4.5}$$

where $\alpha_l^*(t)$ is on the line segment between $\tilde{\alpha}_{lk}(t)$ and $\tilde{\alpha}_l$.

We can express $\tilde{\alpha}_{lk}(t) - \tilde{\alpha}_l = \sum_{i=1}^{n_l} \tilde{\alpha}_l \{\frac{\xi_{li}}{\tilde{\alpha}_l} - 1\} \frac{\{\prod_{j=1}^{2}(1-\Delta_{lij})\}Y_{lik}(t)}{\sum_{i=1}^{n_l}(\prod_{j=1}^{2}(1-\Delta_{lij}))Y_{lik}(t)}$, and thus (4.5) can be written as

$$
\begin{aligned}
& n_l^{1/2}(\tilde{\alpha}_{lk}^{-1}(t) - \tilde{\alpha}_l^{-1}) \\
= \quad & \frac{\tilde{\alpha}_l}{\alpha^*(t)^2} \cdot \frac{n_l}{\sum_{i=1}^{n_l} \prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)} n_l^{-1/2} \{\sum_{i=1}^{n_l}(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}) \prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\}
\end{aligned}
$$

By Glivenko-Cantelli lemma, $n_l^{-1}\sum_{i=1}^{n_l}[\prod_{j=1}^{2}(1-\Delta_{lij})]Y_{lik}(t)$ converges in probability uniformly to $E_l[(\prod_{j=1}^{2}(1-\Delta_{l1j}))Y_{l1k}(t)]$, where $E_l[(\prod_{j=1}^{2}(1-\Delta_{l1j}))Y_{l1k}(t)]$ is bounded away from zero by condition (b). In view of lemma 2, $n_l^{-1}\{\sum_{i=1}^{n_l}(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1)\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\}$ converges to zero in probability uniformly in t since $\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)$ is bounded and monotone in t. Therefore, $\tilde{\alpha}_{lk}(t) - \tilde{\alpha}_l = \frac{\sum_{i=1}^{n_l}\tilde{\alpha}_l\{\frac{\xi_{li}}{\tilde{\alpha}_l}-1\}\{\prod_{j=1}^{2}(1-\Delta_{lij})\}Y_{lik}(t)}{\sum_{i=1}^{n_l}(\prod_{j=1}^{2}(1-\Delta_{lij}))Y_{lik}(t)}$ converges to zero in probability uniformly in t. Thus, $\tilde{\alpha}_{lk}(t)$ and $\tilde{\alpha}_l$ converge to the same limit in probability uniformly in t, which ensures $\alpha_l^*(t)$ also converges to the same limit as $\tilde{\alpha}_l$. By Slutsky's theorem and above results, we get

$$
\begin{aligned}
n_l^{1/2}[\tilde{\alpha}_{lk}^{-1}(t) - \tilde{\alpha}_l^{-1}] \quad = \quad & \frac{1}{\tilde{\alpha}_l E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]} n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\right\} \\
+ \quad & \left[\frac{\tilde{\alpha}_l}{\alpha_l^*(t)^2} \cdot \frac{n_l}{\sum_{i=1}^{n_l}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)} - \frac{1}{\tilde{\alpha}_l E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1j}(t)]}\right] \\
\times \quad & n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\right\} \\
= \quad & \frac{1}{\tilde{\alpha}_l E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]} n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\right\} \\
+ \quad & o_p(1) \qquad (4.6)
\end{aligned}
$$

Due to the sampling of cases, we consider the asymptotic properties for time-varying sampling probability estimators $\tilde{\gamma}_{l1k}(t)$, $\tilde{\gamma}_{l2k}(t)$, $\tilde{\gamma}_{l3k}(t)$, and $\tilde{\gamma}_{l4k}(t)$ in (4.3). For each k, by Taylor expansion of $\tilde{\gamma}_{l1k}(t)$ around $\tilde{\gamma}_{l1}$, it can be written as

$$
\tilde{\gamma}_{l1k}^{-1}(t) - \tilde{\gamma}_{l1}^{-1} = -\frac{1}{\gamma_{l1}^*(t)^2}(\tilde{\gamma}_{l1k}(t) - \tilde{\gamma}_{l1}), \qquad (4.7)
$$

where $\gamma_{l1}^*(t)$ is on the line segment between $\tilde{\gamma}_{l1k}(t)$ and $\tilde{\gamma}_{l1}$. By similar arguments for $\tilde{\alpha}_l$, (4.7) can be written as

$$\begin{aligned} & n_l^{1/2}(\tilde{\gamma}_{l1k}^{-1}(t) - \tilde{\gamma}_{l1}^{-1}) \\ = & \frac{\tilde{\gamma}_{l1}}{\gamma_{l1k}^*(t)^2} \frac{n_l}{\sum_{i=1}^{n_l} \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)} n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \frac{\eta_{li1}}{\tilde{\gamma}_{l1}})\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t). \end{aligned}$$

By Glivenko-Cantelli lemma, $n_l^{-1}\sum_{i=1}^{n_l} \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)$ converges to $(1-\alpha_l)E_l[\Delta_{l11}$ $(1-\Delta_{l12})Y_{l1k}(t)]$ which is bounded away from 0 based on condition (b). In view of lemma 2, $n_l^{-1}\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1k}})\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)$ converges to 0 in probability uniformly in t since $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)$ is bounded and monotone function in t. Thus, $\tilde{\gamma}_{l1k}(t) - \tilde{\gamma}_{l1} = \frac{\sum_{i=1}^{n_l}\tilde{\gamma}_{l1}(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)}{\sum_{i=1}^{n_l}\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)}$ converges to zero in probability uniformly in t. Hence, $\tilde{\gamma}_{l1k}(t)$ and $\tilde{\gamma}_{l1}$ converge to same limit uniformly in t. This means that $\gamma_{l1k}^*(t)$ also converges to same limit uniformly as $\tilde{\gamma}_{l1}$. Combining above the results, it follows from Slutsky's theorem that

$$\begin{aligned} & n_l^{1/2}(\tilde{\gamma}_{l1k}^{-1}(t) - \tilde{\gamma}_{l1}^{-1}) \\ & = \frac{1}{\tilde{\gamma}_{l1}(1-\alpha_l)E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]} n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1k}})\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\right\} \\ & + \left[\frac{\tilde{\gamma}_{l1}}{\tilde{\gamma}_{l1k}^2(t)}\frac{n_l}{\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)} - \frac{1}{\tilde{\gamma}_{l1}(1-\alpha_l)E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]}\right] \\ & \times n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1}})\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\right\} \\ & = \frac{1}{\tilde{\gamma}_{l1}(1-\alpha_l)E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]} \\ & \times n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1}})\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\right\} + o_p(1). \qquad (4.8) \end{aligned}$$

Similarly, we can show that

$$\begin{aligned} n_l^{1/2}(\tilde{\gamma}_{l2k}^{-1}(t) - \tilde{\gamma}_{l2}^{-1}) &= \frac{1}{\tilde{\gamma}_{l2}(1-\alpha_l)E_l[(1-\Delta_{l11})\Delta_{l12}Y_{l1k}(t)]} \\ & \times \ n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li2}}{\tilde{\gamma}_{l2}})(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\right\} + o_p(1), \quad (4.9) \end{aligned}$$

$$
\begin{aligned}
n_l^{1/2}(\tilde{\gamma}_{l3k}^{-1}(t) - \tilde{\gamma}_{l3}^{-1}) &= \frac{1}{\tilde{\gamma}_{l3}(1-\alpha_l)E_l[\Delta_{l11}\Delta_{l12}Y_{l1k}(t)]} \\
&\times \; n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l3}})\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\right\} + o_p(1), \quad (4.10) \\
n_l^{1/2}(\tilde{\gamma}_{l4k}^{-1}(t) - \tilde{\gamma}_{l4}^{-1}) &= \frac{1}{\tilde{\gamma}_{l4}(1-\alpha_l)E_l[\Delta_{l11}\Delta_{l12}Y_{l1k}(t)]} \\
&\times \; n_l^{-1/2}\left\{\sum_{i=1}^{n_l}(1-\frac{\eta_{li2}}{\tilde{\gamma}_{l4}})\Delta_{1i}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\right\} + o_p(1). \quad (4.11)
\end{aligned}
$$

The above properties will be used in the proofs. The following is the proof of theorem 3.

Proof of Theorem 3

We first show the consistency of $\tilde{\beta}^G$. Denote $\widetilde{U}_n^G = n^{-1}\widetilde{U}^G$. By Taylor expansion series, $\tilde{\beta}^G$ can be written as

$$
\tilde{\beta}^G = \beta_0 + \left[-\frac{\partial \widetilde{U}_n^G(\beta_0)}{\partial \beta_0}\right]^{-1} \widetilde{U}_n^G(\beta_0) + o_p(1) \qquad (4.12)
$$

Based on the extension of Fourtz [1977], if (I), (II), (III),and (IV) conditions are satisfied

(I) $\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0

(II) $\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ is negative definite with probability going to one as n $\to \infty$

(III) -$\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ converges to $A(\beta_0)$ in probability uniformly for β in an open neighborhood about β_0

(IV) $\widetilde{U}_n^G(\beta)$ converges to 0 in probability

then, we can show that $\tilde{\beta}^G$ converges to β_0 in probability. Note that

$$
\begin{aligned}
\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T} &= -\frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\widetilde{V}_k(\beta,t)dN_{lik}(t) \quad \text{where} \\
\widetilde{V}_k(\beta,t) &= \frac{\widetilde{S}_k^{(2)}(\beta,t)\widetilde{S}_k^{(0)}(\beta,t) - \widetilde{S}_k^{(1)}(\beta,t)^{\otimes 2}}{\widetilde{S}_k^{(0)}(\beta,t)^2}. \qquad (4.13)
\end{aligned}
$$

By continuity of each component in (4.13) and condition (f), (I) is satisfied.

In order to show that conditions (II),(III) are satisfied, we will show $\parallel (-\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}) - A(\beta) \parallel$ converges to zero in probability uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$, where $A(\beta)$ =

$\sum_{k=1}^{2}\int_0^\tau v_k(\beta,t)s_k^{(0)}(\beta,t),\lambda_{0k}(t)dt.$

By Andersen and Gill [1982], it can be written as

$$
\begin{aligned}
&\| (-\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}) - A(\beta) \| \\
&= \| \sum_{k=1}^{2}\int_0^\tau \widetilde{V}_k(\beta,t)\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)N_{lik}(t) - \sum_{k=1}^{2}\int_0^\tau v_k(\beta_0,t)s_k^{(0)}(\beta_0,t)\lambda_{0k}(t)dt \| \\
&\leq \| \sum_{k=1}^{2}\int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t)\}\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}N_{lik}(t) \| \\
&+ \ \| \sum_{k=1}^{2}\int_0^\tau \{\widetilde{V}_k(\beta,t) - v_k(\beta,t)\}\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\pi_{lik}(t)-1\}N_{lik}(t) \| \\
&+ \| \sum_{k=1}^{2}\int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}M_{lik}(t) \| \\
&+ \| \sum_{k=1}^{2}\int_0^\tau v_k(\beta,t)\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\pi_{lik}(t)-1\}M_{lik}(t) \| \\
&+ \| \sum_{k=1}^{2}\int_0^\tau v_k(\beta,t)\{S_k^{(0)}(\beta,t) - s_k^{(0)}(\beta,t)\}\lambda_{0k}(t)dt \| \qquad (4.14)
\end{aligned}
$$

We can show that each term in (4.14) converges to zero uniformly in $\beta \in \mathcal{B}$. To show that the first term in (4.14) converges to zero in probability, we need to show that

$$\sup_{t\in[0,\tau]\beta\in\mathcal{B}} \| \widetilde{V}_k(\beta,t) - v_k(\beta,t) \| \xrightarrow{p} 0 \quad \text{as } n\to\infty \quad \text{for } k=1,2$$

which suffices to show that

$$\sup_{t\in[0,\tau]\beta\in\mathcal{B}} \| \widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t) \| \to 0 \quad \text{as } n \xrightarrow{p} \infty \quad \text{for } k=1,2 \text{ and } \text{d} = 0,1,2$$

It can be written as

$$
\begin{aligned}
&\widetilde{S}_k^{(d)}(\beta,t) - S_k^{(d)}(\beta,t) \\
= \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\xi_{li}}{\tilde{\alpha}_l}-1\right)\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{ik}(t) \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\tilde{\alpha}_{lk}^{-1}(t) - \tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t)
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{i1}}{\tilde{\gamma}_{l1}}-1\right)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\tilde{\gamma}_{l1k}^{-1}(t)-\tilde{\gamma}_{l1}^{-1})\eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1\right)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\tilde{\gamma}_{l2k}^{-1}(t)-\tilde{\gamma}_{l2}^{-1})\eta_{li2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}\left(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1\right)\Delta_{li1}\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}(\tilde{\gamma}_{l3k}^{-1}(t)-\tilde{\gamma}_{l3}^{-1})\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}\left(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1\right)\Delta_{li1}\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}(\tilde{\gamma}_{l4k}^{-1}(t)-\tilde{\gamma}_{l4}^{-1})\eta_{li2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \qquad (4.15)
\end{aligned}
$$

By using the result of (4.6), the first and the second terms in (4.15) can be written as

$$
\begin{aligned}
& \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\xi_{li}}{\tilde{\alpha}_l}-1\right)\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{ik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\tilde{\alpha}_{lk}^{-1}(t)-\tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&= \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\xi_{li}}{\tilde{\alpha}_l}-1\right)\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{ik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{1}{\tilde{\alpha}_l E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}n_l^{-1}\left\{\sum_{m=1}^{n_l}(1-\frac{\xi_{lm}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lmj})Y_{lmk}(t)\right\}\right. \\
&\times \quad \xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t)+o_p(1) \\
&= \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\xi_{li}}{\tilde{\alpha}_l}-1\right)\prod_{j=1}^{2}(1-\Delta_{lij})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
&+ \quad \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})\frac{Y_{lik}(t)}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]} \\
&\times \quad \{n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})Y_{lmk}(t)Z_{lmk}(t)^{\otimes d}e^{\beta^T Z_{lmk}(t)}\}+o_p(1).
\end{aligned}
$$

By lemma 2, $n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})Y_{lmk}(t)Z_{lmk}(t)^{\otimes d}e^{\beta^T Z_{lmk}(t)}$ converges to

$E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)Z_{l1k}(t)^{\otimes d}e^{\beta^T Z_{l1k}(t)}]$ in probability uniformly in t. Thus, the first and second terms are asymptotically equivalent to

$$\frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\xi_{li}}{\tilde{\alpha}_l}-1\right)\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)$$

$$\times \quad [Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)} - \frac{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)Z_{l1k}(t)^{\otimes d}e^{\beta^T Z_{l1k}(t)}]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)]}].$$

Similarly, the third and forth terms can be written as

$$\begin{aligned}
& \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{i1}}{\tilde{\gamma}_{l1}}-1\right)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{\tilde{\gamma}_{l1}(1-\alpha_l)E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]} \\
\times \quad & n_l^{-1}\left\{\sum_{m=1}^{n_l}(1-\frac{\eta_{lm1}}{\tilde{\gamma}_{l1}})\Delta_{lm1}(1-\Delta_{lm2})(1-\xi_{lm})Y_{lmk}(t)\right\} \\
\times \quad & \eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) + o_p(1) \\
= \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{i1}}{\tilde{\gamma}_{l1}}-1\right)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)}Y_{lik}(t) \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1}})(1-\xi_{li})\Delta_{li1}(1-\Delta_{li2})\frac{Y_{lik}(t)}{E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]} \\
\times \quad & \{n_l^{-1}\sum_{m=1}^{n_l}\Delta_{lm1}(1-\Delta_{lm2})\frac{(1-\xi_{lm})}{(1-\alpha_l)}\frac{\eta_{lm1}}{\tilde{\gamma}_{l1}}Y_{lmk}(t)Z_{lmk}(t)^{\otimes d}e^{\beta^T Z_{lmk}(t)}\} + o_p(1)
\end{aligned}$$

It follows from lemma 2, $n_l^{-1}\sum_{m=1}^{n_l}\Delta_{lm1}(1-\Delta_{lm2})\frac{(1-\xi_{lm})}{(1-\alpha_l)}\frac{\eta_{lm1}}{\tilde{\gamma}_{l1}}Y_{lmk}(t)Z_{lmk}(t)^{\otimes d}e^{\beta^T Z_{lmk}(t)}$ converges to $E_l[Y_{l1k}(t)Z_{l1k}(t)^{\otimes d}e^{\beta^T Z_{l1k}(t)}|\Theta_{l10},\xi_{1l}=0]$ in probability uniformly in t. Thus, the third and fourth terms can be written as

$$\frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{i1}}{\tilde{\gamma}_{l1}}-1\right)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)$$

$$\times \quad [Z_{lik}(t)^{\otimes d}e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t)Z_{l1k}(t)^{\otimes d}e^{\beta^T Z_{l1k}(t)}|\Theta_{l10},\xi_{1l}=0]}{E_l[Y_{l1k}(t)|\Theta_{l10}]}] + o_p(1).$$

By using similar arguments, the fifth to the last terms can be written as

$$\frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1\right)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)$$

$$
\begin{aligned}
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l01}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l01}]}] \\
+ & \quad \frac{1}{2n} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big((\frac{\eta_{li1}}{\tilde{\gamma}_{l3}} - 1) + (\frac{\eta_{li2}}{\tilde{\gamma}_{l4}} - 1) \Big) \Delta_{li1} \Delta_{li2} (1 - \xi_{li}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l11}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l11}]}] + o_p(1).
\end{aligned}
$$

Combining all the results, we have

$$
\begin{aligned}
& \quad n^{1/2} \{ \widetilde{S}_k^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \} \\
= & \quad n^{-1/2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1 \Big) \prod_{j=1}^{2} (1 - \Delta_{lij}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[\prod_{j=1}^{2} (1 - \Delta_{l1j}) Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)}]}{E_l[\prod_{j=1}^{2} (1 - \Delta_{l1j}) Y_{l1k}(t)]}] \\
+ & \quad n^{-1/2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}} - 1 \Big) \Delta_{li1} (1 - \Delta_{li2}) (1 - \xi_{li}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l10}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l10}]}] \\
+ & \quad n^{-1/2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}} - 1 \Big) (1 - \Delta_{li1}) \Delta_{li2} (1 - \xi_{li}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l01}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l01}]}] \\
+ & \quad n^{-1/2} \frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}} - 1 \Big) \Delta_{li1} \Delta_{li2} (1 - \xi_{li}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l11}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l11}]}] \\
+ & \quad n^{-1/2} \frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \Big(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}} - 1 \Big) \Delta_{li1} \Delta_{li2} (1 - \xi_{li}) Y_{lik}(t) \\
\times & \quad [Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l[Y_{l1k}(t) Z_{l1k}(t)^{\otimes d} e^{\beta^T Z_{l1k}(t)} | \Theta_{l11}, \xi_{1l} = 0]}{E_l[Y_{l1k}(t) | \Theta_{l11}]}] + o_p(1)
\end{aligned}
$$

By lemma 2 and condition (h) and (i), for $d = 0, 1,$ and 2, $n^{1/2} \{ \widetilde{S}_k^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \}$ converges weakly to zero-mean Gaussian process. Hence, $\widetilde{S}_k^{(d)}(\beta, t) - S_k^{(d)}(\beta, t)$ converges to zero in probability uniformly in t based on the condition (d) and then it can be shown that

$$
\sup_{t \in [0, \tau] \beta \in \mathcal{B}} \| \widetilde{S}_k^{(d)}(\beta, t) - s_k^{(d)}(\beta, t) \| \to 0 \text{ as } n \to \infty \text{ for } k = 1, 2 \text{ and } d = 0, 1, 2 \tag{4.16}
$$

Combining all the results, $\widetilde{V}_k(\beta, t)$ converges to $v_k(\beta, t)$ in probability uniformly in t and β since $s_k^{(0)}(\beta, t)$ is bounded way from zero by condition (f).

By Lenglart inequality(Andersen and Gill [1982], p1115), there exists n_0 such that for $n \geq n_0$ for any δ, $\eta > 0$,

$$P[n^{-1}\bar{N}_k(\tau) > \eta] \leq \frac{\delta}{\eta} + P[\int_0^\tau S_k^{(0)}(\beta_0; t)\lambda_{0k}(t)dt > \delta],$$

where $\bar{N}_k(t) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} N_{lik}(t)$.

Based on condition (d), $P[\int_0^\tau S_k^{(0)}(\beta_0; t)\lambda_{0k}(t)dt > \delta]$ converges to zero as n $\to \infty$ for $\delta > \int_0^\tau s_k^{(0)}(\beta_0; t)\lambda_{0k}(t)dt$ and then $\lim_{\eta\uparrow\infty}\lim_{n\to\infty} P[n^{-1}\bar{N}_k(\tau) > \eta] = 0$. Therefore, the first term in (4.14) converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$. It follows from lemma 2 that the second and fourth terms in (4.14) can be shown to converge to zero in probability uniformly in t.

The third term in (4.14), $\int_0^\tau v_k(\beta, t)\frac{1}{n}d\sum_{l=1}^{L}\sum_{i=1}^{n_l} M_{lik}(t)$ is a local square integrable martingale. By the Lenglart inequality(Andersen and Gill [1982], p1115), it can be shown that for all δ, $\eta > 0$,

$$P\left[\|\frac{1}{n}\int_0^\tau \{v_k(\beta, t)\}_{jj'}\bar{M}_k(t)\| > \eta\right] \leq \frac{\delta}{\eta^2} + P\left[\frac{1}{n}\int_0^\tau \{v_k(\beta, t)\}_{jj'}^2 S_k^{(0)}(\beta; t)\lambda_{0k}(t)dt > \delta\right],$$

where $\bar{M}_k(t) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} M_{lik}(t)$ and subscript jj' indicates (jj') element of matrix $v_k(\beta, t)$.

Due to boundedness conditions (d),(f),and (g), the second term on right side of the above inequality converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ for any δ as $n \to \infty$. Then it follows that one on the left side of inequality converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$. Hence, the third term in (4.14) converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \to \infty$.

Due to the boundedness of $\sup_{t,\beta}\{v_k(\beta, t)\}$, $\Lambda_{0k}(t)$ for $k = 1, 2$ based on conditions (d),(e), and (g) and uniform convergence of $\tilde{S}_k^{(0)}$ to $s_k^{(0)}$, the last term in (4.14) converges to zero in probability uniformly $\beta \in \mathcal{B}$ as $n \to \infty$. All five terms in (4.14) converge to zero in probability uniformly. Thus, by condition (e), we have

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

and, consequently, (II) and (III) are satisfied.

If we show that $n^{-1/2}\widetilde{U}_n^G(\beta)$ is asymptotically normally distributed, it can be shown that $\widetilde{U}_n^G(\beta)$ converges to zero in probability. Then, (IV) also will be satisfied. Therefore, we can show that $\tilde{\beta}^G$ converges to β_0 in probability and is a consistent estimator of β_0 by satisfying (I),(II),(III), and (IV)(Fourtz [1977] theorem 2). We will show the asymptotic properties of $n^{-1/2}\widetilde{U}_n^G(\beta_0)$. We can decompose $n^{-1/2}\widetilde{U}^G(\beta_0)$ into two parts such that

$$
\begin{aligned}
n^{-1/2}\widetilde{U}^G(\beta_0) &= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\left\{Z_{ik}(u) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dN_{lik}(t) \\
= \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\left\{Z_{ik}(u) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\left\{Z_{lik}(u) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}Y_{lik}(t)e^{\beta_0^T Z_{lik}(t)}d\Lambda_{0k}(t) \quad (4.17)
\end{aligned}
$$

The second term in (4.17) vanishes to zero since it follows from condition (g) that

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\left\{Z_{lik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}Y_{lik}(t)e^{\beta_0^T Z_{lik}(t)}d\Lambda_{0k}(t) \\
= \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{k=1}^{2}\int_0^\tau \{\sum_{i=1}^{n_l}\pi_{lik}(t)Z_{lik}(u)Y_{lik}(t)e^{\beta_0^T Z_{lik}(t)} \\
- \quad & \sum_{i=1}^{n_l}\pi_{lik}(t)\frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}Y_{lik}(t)e^{\beta_0^T Z_{lik}(t)}\}d\Lambda_{0k}(t) \\
= \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{k=1}^{2}\int_0^\tau \left\{\widetilde{S}_k^{(1)}(\beta_0,t) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\widetilde{S}^{(0)}(\beta_0,t)\right\}d\Lambda_{0k}(t).
\end{aligned}
$$

Then, it can be written as

$$n^{-1/2}\widetilde{U}^G(\beta_0) = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\left\{Z_{lik}(t) - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)$$

$$
\begin{aligned}
&= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\pi_{lik}(t)\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}+\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}-\frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\left\{\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}-\frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{lik}(t)-1)\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{lik}(t)-1)\left\{\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}-\frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t) \qquad (4.18)
\end{aligned}
$$

By Spiekerman and Lin [1998], it can be shown that the first part in (4.18) was asymptotically equivalent to $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)$.

Next we will show that the second and last terms in (4.18) converge to zero in probability, uniformly in t. First, the second term in (4.18) can be written as

$$
\sum_{k=1}^{2}\int_0^{\tau}\left\{\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}-\frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\}\{n^{-1/2}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}M_{lik}(t)\}
$$

Note that $M_{l1k}(t),\ldots,M_{lnk}(t)$ is identically and independently distributed zero-mean random variables and $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}M_{lik}(t)$ is a sum of i.i.d. zero-mean random variables for fixed t. $M_{lik}(t)$ is of bounded variation since $M_{lik}^2(0)<\infty$ and $M_{lik}^2(\tau)<\infty$ are satisfied based on conditions (c) and (g). From the example of 2.11.16 of van der Vaart and Wellner [1996](p215), $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}M_{lik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{P}_{Mk}(t)$.

To establish the existence of stochastic processes with continuous sample paths, we will use Kolmogorov-Centsov theorem. If conditions of Kolmogorov-Centsov theorem $E\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^4\le D_z^*|t-s|^2$ and $E\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^2\le\tilde{D}_z|t-s|$ for all $t\ge s$ are satisfied, then we can show that $\mathcal{P}_{Mk}(t)$ has continuous sample paths. Since $\mathrm{E}\mathcal{P}_{Mk}(t)^2=\mathrm{E}[n^{-1}\sum_{i=1}^{n_l}M_{lik}(t)^2]$ $=\mathrm{E}M_{lik}(t)^2=\mathrm{E}[\int_0^t Y_{ik}(u)e^{\beta_0^T Z_{lik}(u)}\lambda_{0k}(u)du]$, $\mathrm{E}\{\mathcal{P}_{Mk}(t)-\mathcal{P}_{Mk}(s)\}^2=\mathrm{E}\mathcal{P}_{Mk}(t)^2$ - $2\mathrm{E}\mathcal{P}_{Mk}(t)\mathcal{P}_{Mk}(s)+\mathrm{E}\mathcal{P}_{Mk}(s)^2=\mathrm{E}\mathcal{P}_{Mk}(t)^2$ - $\mathrm{E}\mathcal{P}_{Mk}(s)^2=\mathrm{E}[\int_s^t Y_{ik}(u)e^{\beta_0^T Z_{lik}(u)}\lambda_{0k}(u)du]\le$ $e^{D_z}E[\int_s^t\lambda_{0k}(u)du]=\tilde{D}_z(\Lambda_{0k}(t)-\Lambda_{0k}(s))$ based on condition (c). There exists constant C

such that $\Lambda_{0k}(t) - \Lambda_{0k}(s) \le C(t-s)$ for $t \ge s$ since $\Lambda_{0k}(.)$ is differentiable and $\lambda_{0k}(.)$ is bounded in $[0,\tau]$. Thus $\mathrm{E}\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^2 \le \tilde{D}_{cz}(t-s)$. For fixed t, $\mathcal{P}_{Mk}(t)$ is random normal variable. Therefore, we have $E\{\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s)\}^4 = \mathrm{Var}(\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s))^2 + \mathrm{E}\{(\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s))^2\}^2 = 3\ \{E(\mathcal{P}_{Mk}(t) - \mathcal{P}_{Mk}(s))^2\}^2 \le D^*_{cz}|t-s|^2$.

Since two conditions are satisfied, it follows that $\mathcal{P}_{Mk}(t)$ has continuous sample path from Kolmogorov-Centsov theorem. Based on conditions (c), (d), and (f), it can be shown that $\widetilde{S}_k^{(1)}(\beta,t)$ and $\widetilde{S}_k^{(0)}(\beta,t)$ are of bounded variations and specially $\widetilde{S}_k^{(0)}(\beta,t)$ is bounded away from zero. Thus $\frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)}$ is of bounded variation and can be written as $\frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)} = G_{k1} - G_{k1}$ where both G_{k1} and G_{k2} are nonnegative, monotone functions in t, and bounded. Therefore, $\frac{\widetilde{S}_k^{(1)}(\beta,t)}{\widetilde{S}_k^{(0)}(\beta,t)}$ is the sum of two monotone functions. By the result in (4.16), it can be shown that $\sup_{t\in[0,\tau]\beta\in\mathcal{B}} \| \frac{\tilde{S}_k^{(1)}(\beta,t)}{\tilde{S}_k^{(0)}(\beta,t)} - \frac{s_k^{(1)}(\beta,t)}{s_k^{(0)}(\beta,t)} \| \xrightarrow{p} 0$. By lemma 1, the second term in (4.18) converges to zero in probability uniformly in t as $n \to \infty$.

By using similar arguments, the last term in (4.18) can be shown to converge to zero in probability uniformly in t.

The third term in (4.18) can be decomposed such that

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\pi_{lik}(t)-1)\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
= \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
+ \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\tilde{\alpha}_{lk}^{-1}(t)-\tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
+ \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}(1-\Delta_{li2})(1-\xi_{l1})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
+ \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}(1-\Delta_{li2})(1-\xi_{l1})(\tilde{\gamma}_{l1k}^{-1}(t)-\tilde{\gamma}_{l1}^{-1})\eta_{li1}\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
+ \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (1-\Delta_{li1})\Delta_{li2}(1-\xi_{l1})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t) \\
+ \ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (1-\Delta_{li1})\Delta_{li2}(1-\xi_{l1})(\tilde{\gamma}_{l2k}^{-1}(t)-\tilde{\gamma}_{l2}^{-1})\eta_{li2}\left\{Z_{lik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{lik}(t)
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\Delta_{li1}\Delta_{li2}(1-\xi_{l1})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ \quad n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\Delta_{li1}\Delta_{li2}(1-\xi_{l1})(\tilde{\gamma}_{l3k}^{-1}(t)-\tilde{\gamma}_{l3}^{-1})\eta_{li1}\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ \quad n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\Delta_{li1}\Delta_{li2}(1-\xi_{l1})(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)\\
&+ \quad n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\Delta_{li1}\Delta_{li2}(1-\xi_{l1})(\tilde{\gamma}_{l4k}^{-1}(t)-\tilde{\gamma}_{l4}^{-1})\eta_{li2}\left\{Z_{lik}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\}dM_{lik}(t)
\end{aligned}
\tag{4.19}
$$

By using the result in (4.6), the second term in (4.19) is asymptotically equivalent to

$$
\begin{aligned}
&= \quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})(1-\frac{\xi_{li}}{\tilde{\alpha}_l})Y_{lik}(t)\frac{1}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}\\
&\times \quad \left\{n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})(Z_{lmk}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)})dM_{lmk}(t)\right\}\\
&= \quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})(1-\frac{\xi_{li}}{\tilde{\alpha}_l})Y_{lik}(t)\frac{1}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}\\
&\times \quad \left\{n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})(Z_{lmk}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)})(dN_{lmk}(t)-Y_{lmk}(t)e^{\beta^T Z_{lmk}(t)}d\Lambda_{0k}(t))\right\}.
\end{aligned}
\tag{4.20}
$$

Since the term related with $dN_{lmk}(t)$ in (4.20) does not contribute to $\prod_{j=1}^{2}(1-\Delta_{lmj})$, we have

$$
\begin{aligned}
& \quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)Y_{lik}(t)\frac{1}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}\\
&\times \quad \left\{n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})(Z_{lmk}(t)-\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)})Y_{lmk}(t)e^{\beta^T Z_{lmk}(t)}d\Lambda_{0k}(t)\right\}\\
&= \quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)Y_{lik}(t)\\
&\times \quad \frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)R_{l1k}(t)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}d\Lambda_{0k}(t)+o_p(1).
\end{aligned}
$$

If follows from the result of (4.8) and lemma 2 that the fourth term in (4.19) can be written as

$$
\begin{aligned}
&= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}(1-\Delta_{li2})(1-\xi_{l1})\eta_{li1}\frac{1}{\tilde\gamma_{l1}(1-\alpha_l)E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]}\\
&\times\ n_l^{-1}\left\{\sum_{m=1}^{n_l}(1-\frac{\eta_{lm1}}{\tilde\gamma_{l1}})\Delta_{lm1}(1-\Delta_{lm2})(1-\xi_{lm})Y_{lmk}(t)\right\}\{Z_{lik}(t)-e_k(\beta_0,t)\}\,dM_{lik}(t)\\
&= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}(1-\Delta_{li2})(1-\frac{\eta_{li1}}{\tilde\gamma_{l1}})Y_{lik}(t)\\
&\times\ \frac{E_l[\Delta_{l11}(1-\Delta_{l12})(Z_{l1k}(t)-e_k(\beta_0,t))dM_{l1k}(t)]}{E_l[\Delta_{l11}(1-\Delta_{l12})Y_{l1k}(t)]}+o_p(1)
\end{aligned}
$$

By similar arguments, the sixth, eighth, and tenth terms are asymptotically equivalent to

$$
\begin{aligned}
&n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(1-\frac{\eta_{li2}}{\tilde\gamma_{l2}})Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\Theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l01}]}\\
&+\ n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}\Delta_{li2}(1-\xi_{li})(1-\frac{\eta_{li1}}{\tilde\gamma_{l3}})Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l11}]}\\
&+\ n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \Delta_{li1}\Delta_{li2}(1-\xi_{li})(1-\frac{\eta_{li2}}{\tilde\gamma_{l4}})Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l11}]}.
\end{aligned}
$$

Combining all the results, the third term in (4.18) is asymptotically equivalent to

$$
\begin{aligned}
&n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde\alpha_l}-1)[R_{lik}(t)\\
&-\ Y_{lik}(t)\frac{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)R_{l1k}(t)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)]]}d\Lambda_{0k}(t)\\
&+\ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde\gamma_{l1}}-1)[Q_{lik}(\beta_0)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta_0,t)|\Theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l10}]}]\\
&+\ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde\gamma_{l2}}-1)[Q_{lik}(\beta_0)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta_0,t)|\Theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l01}]}]\\
&+\ n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li1}}{\tilde\gamma_{l3}}-1)[Q_{lik}(\beta_0)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta_0,t)|\Theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l11}]}]\\
&+\ n^{-1/2}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li2}}{\tilde\gamma_{l4}}-1)[Q_{lik}(\beta_0)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta_0,t)|\Theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\Theta_{l11}]}]
\end{aligned}
$$

Therefore, $n^{-1/2}\widetilde{U}^G(\beta_0)$ is asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^\tau \prod_{j=1}^{2}(1-\Delta_{lij})\{R_{lik}(t) \\
- \quad & \frac{Y_{lik}(t)E_l[\prod_{j=1}^2(1-\Delta_{l1j})R_{l1k}(t)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)]}\}d\Lambda_{0k}(t) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\xi_{li})\{\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)[Q_{li1}(\beta_0) \\
- \quad & \int_0^\tau \frac{Y_{lik}(t)\mathrm{E}_l[dQ_{l1k}(\beta,t)|\Theta_{l10},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1k}(t)|\Theta_{l10}]}] \\
+ \quad & (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)[Q_{lik}(\beta)-\int_0^\tau \frac{Y_{lik}(t)\mathrm{E}_l[dQ_{l1k}(\beta_0,t)|\Theta_{l01},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1k}(t)|\Theta_{l01}]}] \\
+ \quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)[Q_{lik}(\beta_0)-\int_0^\tau \frac{Y_{lik}(t)\mathrm{E}_l[dQ_{l1k}(\beta_0)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1k}(t)|\Theta_{l11}]}] \\
+ \quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)[Q_{lik}(\beta)-\int_0^\tau \frac{Y_{lik}(t)\mathrm{E}[dQ_{l1k}(\beta_0)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1k}(t)|\Theta_{l11}]}]\}. \qquad (4.21)
\end{aligned}
$$

By Spiekerman and Lin [1998], it can be shown that the first term in (4.21) converges to weakly to a zero-mean normal vector with covariance $V_{I,l}(\beta_0)=E_l[\sum_{k=1}^2 Q_{l1k}(\beta_0)]^{\otimes 2}$.

The second term in (4.21) is asymptotically zero-mean normal vector with covariance matrix $\frac{1-\alpha_l}{\alpha_l}V_{II,l}(\beta_0)=\frac{1-\alpha_l}{\alpha_l}\mathrm{Var}_l(\prod_{j=1}^2(1-\Delta_{l1j})\sum_{k=1}^2\int_0^\tau[R_{l1k}(\beta_0,t)-\frac{Y_{l1k}E(\prod_{j=1}^2(1-\Delta_{l1j})R_{l1k}(\beta_0,t))}{E_l(\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t))}]$ $d\Lambda_{0k}(t))$ by Hájek [1960]'s central limit theorem for finite sampling.

It follows from Hájek [1960]'s central limit theorem for finite sampling and Cramer-Wold devices that the third term to the last term in (4.21) converges to weakly a zero-mean normal vector with covariance $(1-\alpha_l)\sum_{k=1}^2 V_{III,lk}(\beta_0)$ where

$$
\begin{aligned}
& V_{III,lk}(\beta) \\
= \quad & \Pr(\Theta_{l10})\frac{1-\gamma_{l1}}{\gamma_{l1}}\mathrm{Var}(Q_{l1k}(\beta)-\int_0^\tau \frac{Y_{l1k}(t)E\{dQ_{l1k}(\beta,t)|\Theta_{l10},\xi_{l1}=0\}}{E\{Y_{l1k}(t)|\Theta_{l10}\}}|\Theta_{l10},\xi_{l1}=0) \\
+ \quad & \Pr(\Theta_{l10})\frac{1-\gamma_{l2}}{\gamma_{l2}}\mathrm{Var}(Q_{l1k}(\beta)-\int_0^\tau \frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l01},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l01}\}}|\Theta_{l10},\xi_{l1}=0) \\
+ \quad & \frac{1}{4}\Pr(\Theta_{l11})\frac{1-\gamma_{l1}}{\gamma_{l1}}\mathrm{Var}(Q_{l1k}(\beta)-\int_0^\tau \frac{Y_{l1k}(t)E\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{1}=0\}}{E\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0) \\
+ \quad & \frac{1}{4}\Pr(\Theta_{l11})\frac{1-\gamma_{l2}}{\gamma_{l2}}\mathrm{Var}_l(Q_{l1k}(\beta)-\int_0^\tau \frac{Y_{l1k}(t)E\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0\}}{E\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0)
\end{aligned}
$$

since four components are independent.

In addition, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)$ and $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)$ where $L_{lik}(t) = R_{lik}(t) - \frac{Y_{lik}(t)E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})R_{l1k}(t)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}$ are independent since

$$
\begin{aligned}
&Cov_l\left(n_l^{-1/2}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0), n_l^{-1/2}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)\right)\\
=\ & E\left\{n_l^{-1}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)\right)\\
=\ & E\left\{E\left(n_l^{-1}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)|\mathcal{F}(\tau)\right)\right\}\\
=\ & E\left\{n_l^{-1}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2}E_l\left((\frac{\xi_{li}}{\tilde{\alpha}_l}-1)|\mathcal{F}(\tau)\right)\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)\right\} = 0
\end{aligned}
$$

By using similar arguments, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(\beta_0)$ and the third to the last term in (4.21) are independent. Since ξ_{li} and η_{lik} for $k=1,2$ are independent, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^{\tau}\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)d\Lambda_{0k}(t)$ and the third to the last term in (4.21) are independent. Combining all the results, $n^{-1/2}\widetilde{U}^G(\beta_0)$ converges weakly to zero-mean normal vector with covariance matrix $\Sigma_G(\beta)$ where

$$
\begin{aligned}
&\Sigma_G(\beta) = \sum_{l=1}^{L}q_l\{V_{I,l}(\beta) + \frac{1-\alpha_l}{\alpha_l}V_{II,l}(\beta) + (1-\alpha_l)\sum_{k=1}^{2}V_{III,lk}(\beta)\},\\
&V_{I,l}(\beta) = E_l[\sum_{k=1}^{2}Q_{l1k}(\beta)]^{\otimes 2},\\
&V_{II,l}(\beta)\\
=\ & \mathrm{Var}_l(\prod_{j=1}^{2}(1-\Delta_{l1j})\sum_{k=1}^{2}\int_0^{\tau}[R_{l1k}(\beta,t) - \frac{Y_{l1k}(t)E_l(\prod_{j=1}^{2}(1-\Delta_{l1j})R_{l1k}(\beta,t))}{E_l(\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t))}]d\Lambda_{0k}(t)),\\
&V_{III,lk}(\beta)\\
=\ & \Pr(\Theta_{l10})\frac{1-\gamma_{l1}}{\gamma_{l1}}\mathrm{Var}_l[Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E\{dQ_{l1k}(\beta,t)|\Theta_{l10},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l10}\}}|\Theta_{l10},\xi_{l1}=0]\\
+\ & \Pr(\Theta_{l01})\frac{1-\gamma_{l2}}{\gamma_{l2}}\mathrm{Var}[Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E\{dQ_{l1k}(\beta,t)|\Theta_{l01},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l01}\}}|\Theta_{l01},\xi_{l1}=0]\\
+\ & \frac{1}{4}\Pr(\Theta_{l11})\frac{1-\gamma_{l1}}{\gamma_{l1}}\mathrm{Var}_l[Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0]\\
+\ & \frac{1}{4}\Pr(\Theta_{l11})\frac{1-\gamma_{l2}}{\gamma_{l2}}\mathrm{Var}_l[Q_{l1k}(\beta) - \int_0^{\tau}\frac{Y_{l1k}(t)E_l\{dQ_{l1k}(\beta,t)|\Theta_{l11},\xi_{l1}=0\}}{E_l\{Y_{l1k}(t)|\Theta_{l11}\}}|\Theta_{l11},\xi_{l1}=0].
\end{aligned}
$$

Hence, $\widetilde{U}^G(\beta_0)$ converges to zero in probability and, consequently, (IV) is satisfied. Therefore, (I), (II), (III), and (IV) are satisfied, which implies that $\tilde{\beta}^G$ converges to β_0 in probabilty by the extension of Fourtz [1977]. By consistency of $\tilde{\beta}^G$ and Talor expansion of $\widetilde{U}^G(\beta_0)$, $\tilde{\beta}^G - \beta_0$ is asymptotically normally distributed with mean zero and with variance matrix $A^{-1}\Sigma_G(\beta_0)A^{-1}$ where $A = \sum_{k=1}^{2} A_k$.

Proof of Theorem 4 Note that

$$\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^G, t) = \int_0^t \frac{\sum_{i=1}^{n_l} \pi_{lik}(u) dN_{lik}(u)}{n\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} = \int_0^t \frac{\sum_{i=1}^{n_l} \pi_{lik}(u) dM_{lik}(u)}{n\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} + \int_0^t \frac{\tilde{S}_k^{(0)}(\beta_0, u) d\Lambda_{0k}(u)}{\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)}$$

$n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^G, t) - \Lambda_{0k}(t)\}$ can be decomposed into five parts:

$$\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^G, t) - \Lambda_{0k}(t)\} \\
= \; & n^{1/2} \int_0^t \left(\frac{1}{n\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} - \frac{1}{n\tilde{S}_k^{(0)}(\beta_0, u)} \right) d \sum_{l=1}^{L} \sum_{i=1}^{n_l} M_{lik}(u) \\
+ \; & n^{1/2} \int_0^t \left(\frac{1}{n\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} - \frac{1}{n\tilde{S}_k^{(0)}(\beta_0, u)} \right) d \sum_{l=1}^{L} \sum_{i=1}^{n_l} \{\pi_{lik}(u) - 1\} M_{lik}(u) \\
+ \; & n^{1/2} \int_0^t \left(\frac{1}{\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} - \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} \right) \tilde{S}_k^{(0)}(\beta_0, u) d\Lambda_{0k}(u) \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} d\{n^{-1/2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} M_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} \{n^{-1/2} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \{\pi_{lik}(u) - 1\} dM_{lik}(u)\}
\end{aligned} \tag{4.22}$$

By Taylor expansion, it can be written as

$$\frac{1}{\tilde{S}_k^{(0)}(\tilde{\beta}^G, u)} - \frac{1}{\tilde{S}_k^{(0)}(\beta_0, u)} = -\frac{\tilde{S}_k^{(1)}(\beta^*, u)^T}{\tilde{S}_k^{(0)}(\beta^*, u)^2} (\tilde{\beta}^G - \beta_0)$$

where β^* is on the line segment between $\tilde{\beta}^G$ and β_0. Plugging into the first term in (4.22), we have

$$\int_0^t \left(-\frac{\tilde{S}_k^{(1)}(\beta^*, u)^T}{\tilde{S}_k^{(0)}(\beta^*, u)^2} \right) (\tilde{\beta}^G - \beta_0) \left\{ n^{-1/2} d \sum_{l=1}^{L} \sum_{i=1}^{n_l} M_{lik}(u) \right\}, \tag{4.23}$$

where β^* is on the line segment between $\tilde{\beta}^G$ and β_0. Due to consistency of $\tilde{\beta}^G$, β^* also converges to β_0 in probability uniformly. Since $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(1)}(\beta^*, u)$ are of bounded variations and $\tilde{S}_k^{(0)}(\beta^*, u)$ is bounded away from 0, $\frac{\tilde{S}_k^{(1)}(\beta^*,u)^T}{\tilde{S}_k^{(0)}(\beta^*,u)^2}$ is of bounded variation and can be written as sum of two monotone functions in t. In addition, we have shown consistency of $\tilde{\beta}^G$, weak convergence of $n^{-1/2}d\sum_{l=1}^{L}\sum_{i=1}^{n_l} M_{lik}(u)$ to zero-mean Gaussian process with continuous sample paths, and the uniform convergence of $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(1)}(\beta^*, u)$. Therefore, by lemma 1, (4.23) converges to zero in probability uniformly in t.

By similar arguments for the first term, the second term in (4.22) can be shown to converges to zero in probability uniformly in t.

It follows from Taylor expansion that the third term in (4.22) can be written as

$$n^{1/2}\int_0^t \left(-\frac{\tilde{S}_k^{(1)}(\beta^*, u)^T}{\tilde{S}_k^{(0)}(\beta^*, u)^2}\right)(\tilde{\beta}^G - \beta_0)\tilde{S}_k^{(0)}(\beta_0, u)d\Lambda_{0k}(u) \tag{4.24}$$

Note that $\tilde{\beta}^G$ and β^* converge to β_0 in probability uniformly, $\tilde{S}_k^{(0)}(\beta^*, u)$ and $\tilde{S}_k^{(0)}(\beta_0, u)$ converge to $s_k^{(0)}(\beta_0, u)$ in probability uniformly where $s_k^{(0)}(\beta_0, u)$ is bounded away from zero, $\tilde{S}_k^{(1)}(\beta^*, u)$ converges to $s_k^{(1)}(\beta_0, t)$ in probability uniformly, and $d\Lambda_{0k}(u)$ is bounded. It follows from the above results, (4.24) is asymptotically equivalent to

$$n^{1/2}l_k(\beta_0, t)^T(\tilde{\beta}^G - \beta_0),$$

where $l_k(\beta_0, t)^T = \int_0^t -e_k(\beta, u)d\Lambda_{0k}(u)$ and $e_k(\beta_0, u) = s_k^{(1)}(\beta_0, u)/s_k^{(0)}(\beta_0, u)$.

Since $\tilde{S}_k^{(0)}(\beta_0, u)$ converges to $s_k^{(0)}(\beta_0, u)$ in probability uniformly and $s_k^{(0)}(\beta_0, u)$ is bounded away from 0, we have $\tilde{S}_k^{(0)}(\beta_0, u)^{-1} \xrightarrow{p} s_k^{(0)}(\beta_0, u)^{-1}$. In addition, $n^{-1/2}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}$ $M_{lik}(u)$ converges to a zero-mean Gaussian process with continuous sample paths. Hence, the fourth term in (4.22) is asymptotically equivalent to

$$\int_0^t \frac{1}{s_k^{(0)}(\beta_0, u)}\left\{n^{-1/2}d\sum_{l=1}^{L}\sum_{i=1}^{n_l} M_{lik}(u)\right\}.$$

The last term in (4.22) can be written as

$$
\begin{aligned}
& \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\pi_{lik}(u)-1\}dM_{lik}(u)\} \\
= \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\prod_{j=1}^{2}(1-\Delta_{lij})\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\tilde{\alpha}_{lk}(u)^{-1}-\tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)(1-\xi_{li})\Delta_{li1}(1-\Delta_{li2})\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\tilde{\gamma}_{l1k}(u)^{-1}-\tilde{\gamma}_{l1}^{-1})\eta_{li1}(1-\xi_{li})\Delta_{li1}(1-\Delta_{li2})\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)(1-\xi_{li})(1-\Delta_{li1})\Delta_{li2}\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{(\tilde{\gamma}_{l2k}(u)^{-1}-\tilde{\gamma}_{l2}^{-1})\eta_{li2}(1-\xi_{li})(1-\Delta_{li1})\Delta_{li2}\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\frac{1}{2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)(1-\xi_{li})\Delta_{li1}\Delta_{li2}\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\frac{1}{2}(\tilde{\gamma}_{l3k}(u)^{-1}-\tilde{\gamma}_{l3}^{-1})\eta_{li1}(1-\xi_{li})\Delta_{li1}\Delta_{li2}\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\frac{1}{2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)(1-\xi_{li})\Delta_{li1}\Delta_{li2}\}dM_{lik}(u)\} \\
+ \; & \int_0^t \frac{1}{\tilde{S}_k^{(0)}(\beta_0,u)}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\{\frac{1}{2}(\tilde{\gamma}_{l4k}(u)^{-1}-\tilde{\gamma}_{l4}^{-1})\eta_{li2}(1-\xi_{li})\Delta_{li1}\Delta_{li2}\}dM_{lik}(u)\} \quad (4.25)
\end{aligned}
$$

Due to uniform convergence of $\tilde{S}_k^{(0)}(\beta_0,u)^{-1}$ to $s_k^{(0)}(\beta_0,u)^{-1}$ where $s_k^{(0)}(\beta_0,u)$ is bounded away from zero, the asymptotic properties of (4.6), the first and second terms are asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t Y_{lik}(u)\{e^{\beta^T Z_{lik}(u)} \\
- \; & \frac{E[\prod_{j=1}^{2}(1-\Delta_{1j})e^{\beta^T Z_{lik}(u)}Y_{l1k}(u)]}{E[\prod_{j=1}^{2}(1-\Delta_{1j})Y_{l1k}(u)]}\}\cdot\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)}.
\end{aligned}
$$

By using uniform convergence of $\tilde{S}_k^{(0)}(\beta_0,u)^{-1}$ to $s_k^{(0)}(\beta_0,u)^{-1}$ where $s_k^{(0)}(\beta_0,u)$ is bounded away from zero and the asymptotic properties of (4.8), the third and fouth terms are asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_1}-1) \\
\times & \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l10},\xi_1=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}].
\end{aligned}
$$

By similar arguments, the last term in (4.22) is asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij}) \\
\times & \int_0^t Y_{lik}(u)\{e^{\beta^T Z_{lik}(u)}-\frac{E_l[\prod_{j=1}^2(1-\Delta_{1j})e^{\beta^T Z_{lik}(u)}Y_{l1k}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{1j})Y_{l1k}(u)]}\}\cdot\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)} \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1) \\
\times & \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1) \\
\times & \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1) \\
\times & \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1) \\
\times & \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}]
\end{aligned}
$$

Combining all the results, we have

$$
\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^G,t)-\Lambda_{0k}(t)\} \\
= & n^{1/2}l_k(\beta_0,t)^T(\tilde{\beta}^G-\beta_0)+\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}\left\{n^{-1/2}d\sum_{l=1}^{L}\sum_{i=1}^{n_l}M_{lik}(u)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij}) \\
&\times \quad \int_0^t Y_{lik}(u)\{e^{\beta^T Z_{lik}(u)} - \frac{E_l[\prod_{j=1}^2(1-\Delta_{1j})e^{\beta^T Z_{lik}(u)}Y_{l1k}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{1j})Y_{l1k}(u)]}\}\cdot\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)} \\
&+ \quad \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_1}-1) \\
&\times \quad \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}] \\
&+ \quad (1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_2}-1) \\
&\times \quad \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}] \\
&+ \quad \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_1}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}] \\
&+ \quad \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_2}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}]] \\
&+ \quad o_p(1)
\end{aligned}
$$

Note that

$$
\begin{aligned}
& n^{1/2}(\tilde{\beta}^G - \beta_0) = A(\beta_0)^{-1}[n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1) \\
\times \quad & \int_0^\tau \prod_{j=1}^{2}(1-\Delta_{lij})\{\bar{Z}_{lik}(t) - \frac{Y_{lik}(t)E_l[\prod_{j=1}^2(1-\Delta_{1j})\bar{Z}_{1k}(t)]}{E[\prod_{j=1}^2(1-\Delta_{1j})Y_{l1k}(t)]}\}dM_{lik}(t) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\xi_{li})\{\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1) \\
\times \quad & [Q_{lik}(\beta) - \int_0^\tau \frac{Y_{lik}(t)\mathrm{E}[dQ_{l1k}(\beta,t)|\Theta_{l10},\xi_{l1}=0]}{\mathrm{E}[Y_{l1k}(t)|\Theta_{l10}]}] \\
+ \quad & (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)[Q_{lik}(\beta) - \int_0^\tau \frac{Y_{lik}(t)\mathrm{E}[dQ_{l1k}(\beta,t)|\Theta_{l01},\xi_{l1}=0]}{\mathrm{E}[Y_{l1k}(t)|\Theta_{l01}]}] \\
+ \quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)[Q_{lik}(\beta) - \int_0^\tau \frac{Y_{lik}(t)\mathrm{E}[dQ_{l1k}(\beta)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}[Y_{l1k}(t)|\Theta_{l11}]}] \\
+ \quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)[Q_{lik}(\beta) - \int_0^\tau \frac{Y_{lik}(t)\mathrm{E}[dQ_{l1k}(\beta)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}[Y_{l1k}(t)|\Theta_{l11}]}]\}] + o_p(1)
\end{aligned}
$$

Using the above equation, we have

$$
\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^{M}(\tilde{\beta}^{G},t)-\Lambda_{0k}(t)\} \\
= & \ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left\{l_k(\beta_0,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}Q_{lim}(\beta_0)+\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}dM_{lik}(u)\right\} \\
+ & \ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\{l_k(\beta_0,t)^T A(\beta_0)^{-1}\sum_{m=1}^{K}\int_0^\tau [R_{lim}(u) \\
- & \ \frac{Y_{lim}(u)E_l[\prod_{j=1}^2(1-\Delta_{l1j})R_{l1m}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1m}(u)]}]d\Lambda_{0m}(u) \\
+ & \ \prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t Y_{lik}(u)\left\{e^{\beta^T Z_{lik}(u)}-\frac{E_l(\prod_{j=1}^2(1-\Delta_{l1j})e^{\beta^T Z_{l1k}(u)}Y_{l1k}(u))}{E_l(\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(u))}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)}\} \\
+ & \ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[l_k(\beta_0,t)^T A(\beta_0)^{-1}n^{-1/2}\sum_{m=1}^{2}(1-\xi_{li}) \\
\times & \ \{\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)[Q_{lim}(\beta)-\int_0^\tau \frac{Y_{lim}(u)\mathrm{E}_l[dQ_{l1m}(\beta_0,u)|\Theta_{l10},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1m}(u)|\Theta_{l10}]}] \\
+ & \ (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)[Q_{lim}(\beta)-\int_0^\tau \frac{Y_{lim}(u)\mathrm{E}_l[dQ_{l1m}(\beta_0,t)|\Theta_{l01},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1m}(t)|\Theta_{l01}]}] \\
+ & \ \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)[Q_{lim}(\beta)-\int_0^\tau \frac{Y_{lim}(u)\mathrm{E}_l[dQ_{l1m}(\beta_0)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1m}(u)|\Theta_{l11}]}] \\
+ & \ \frac{1}{2}\Delta_{li1}\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)[Q_{lim}(\beta_0)-\int_0^\tau \frac{Y_{lim}(u)\mathrm{E}_l[dQ_{l1m}(\beta_0)|\Theta_{l11},\xi_{l1}=0]}{\mathrm{E}_l[Y_{l1m}(u)|\Theta_{l11}]}]\}] \\
+ & \ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)} \\
\times & \ [dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}] \\
+ & \ (1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)} \\
\times & \ [dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}] \\
+ & \ \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}] \\
+ & \ \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l3}}-1)\int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{lik}(u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}]] \\
+ & \ o_p(1) \\
= & \ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\nu_{lik}(\beta_0,t)+n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\zeta_{lik}(\beta_0,t)+n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\varphi_{lik}(\beta_0,t)+o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
&\nu_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{K} Q_{lim}(\beta) + \int_0^t \frac{1}{s_k^{(0)}(\beta,u)} dM_{lik}(u),\\
&\zeta_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\\
\times\quad & \sum_{m=1}^{2}\int_0^{\tau}[R_{lim}(u) - \frac{Y_{lim}(u)E_l[\prod_{j=1}^2(1-\Delta_{l1j})R_{l1m}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1m}(u)]}]d\Lambda_{0m}(u)\\
+\quad & \prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t Y_{lik}(u)\left\{e^{\beta^T Z_{lik}(u)} - \frac{E_l(\prod_{j=1}^2(1-\Delta_{1j})e^{\beta^T Z_{1k}(u)}Y_{l1k}(u))}{E_l(\prod_{j=1}^2(1-\Delta_{1j})Y_{l1k}(u))}\right\}\frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0,u)},\\
&\varphi_{lik}(\beta,t) = l_k(\beta,t)^T A(\beta)^{-1}\sum_{m=1}^{2}(1-\xi_{li})[\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l10})\\
+\quad & (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l01})\\
+\quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l11}) + (\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)B_{lim}^{(1)}(\beta,t|\Theta_{l11})\}]\\
+\quad & (1-\xi_{li})[\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l10}) + (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l01})\\
+\quad & \frac{1}{2}\Delta_{li1}\Delta_{li2}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l11}) + (\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)B_{lik}^{(2)}(\beta,t|\Theta_{l11})\}]\\
&B_{lik}^{(1)}(\beta,t|\Theta_{ljm}) = \int_0^t \frac{1}{s_k^{(0)}(\beta_0,u)}\left[dM_{lik}(u) - Y_{lik}(u)\frac{E_l\{dM_{l1k}(u)|\Theta_{l10},\xi_{li}=0\}}{E_l\{Y_{l1k}(u)|\Theta_{ljm}\}}\right]\\
&B_{lik}^{(2)}(\beta,t|\Theta_{ljm}) = Q_{lik}(\beta) - \int_0^t Y_{lik}(u)\frac{E_l\{dQ_{l1k}(\beta,u)|\Theta_{ljm},\xi_{li}=0\}}{E_l\{Y_{l1k}(\beta,u)|\Theta_{ljm}\}},\\
&l_k(\beta,t)^T = -\int_0^t e_k(\beta,u)d\Lambda_{0k}(u)
\end{aligned}
$$

Let $P(t) = (P^{(1)}(t) + P^{(2)}(t) + P^{(3)}(t))$ where $P^{(1)}(t) = (P_1^{(1)}(t), P_2^{(1)}(t))^T$, $P^{(2)}(t) = (P_1^{(2)}(t), P_2^{(2)}(t))^T$, $P^{(3)}(t) = (P_1^{(3)}(t), P_2^{(3)}(t))^T$, $P_k^{(1)}(t)^T = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\nu_{lik}(\beta_0,t)$, $P_k^{(2)}(t)^T = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\zeta_{lik}(\beta_0,t)$, and $P_k^{(3)}(t)^T = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\varphi_{lik}(\beta_0,t)$. Then, by theorem 4 of Spiekerman and Lin [1998], $P^{(1)}(t) = (P_1^{(1)}(t), P_2^{(1)}(t))^T$ converges weakly to a zero-mean Gaussian process $\mathcal{P}^{(1)}(t) = (\mathcal{P}_1^{(1)}(t), \mathcal{P}_2^{(1)}(t))^T$ and covariance functions between $\mathcal{P}_j^{(1)}(t)$ and $\mathcal{P}_k^{(1)}(s)$ is $\sum_{l=1}^{L} q_l E_l\{\nu_{l1j}(\beta_0,t), \nu_{l1k}(\beta_0,s)\}$ for $t,s \in [0,\tau]$ in $D[0,\tau]^K$.

We will show weak convergence of $P^{(2)}(t)$ to a zero-mean Gaussian process $\mathcal{P}^{(2)}(t)$. Note that $s_k^{(0)}(\beta,t)$ and $E_l(\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t))$ are bounded away from zero, $l_k(\beta_0,t)^T$, $e^{\beta^T Z_{lik}(t)}Y_{lik}(t)$, $E_l(\prod_{j=1}^2(1-\Delta_{l1j})e^{\beta^T Z_{l1k}(t)}Y_{l1k}(t))$, and $d\Lambda_{0k}(t)$ are of bounded variations based on conditions (b), (c), (d), and (g); $A(\beta_0)$ is positive definite based on (e). Hence, it follows from Cramer-Wold device and lemma 2 that the finite dimensional distribution of $P^{(2)}(t)$ is asymptotically same as that of $\mathcal{P}^{(2)}(t)$ for any finite number of time point $(t_1, \ldots, t_L)$. The next thing that we will show is that $P^{(2)}(t)$ has tightness. It suffices to show the marginal tightness of $P_k^{(2)}(t)$ for each k since space $D[0,\tau]^K$ is equipped with the uniform metric. By applying lemma 2, the marginal tightness follows to $P_k^{(2)}(t)$. Combining all the results, $P^{(2)}(t) = (P_1^{(2)}(t), P_2^{(2)}(t))^T$ converges weakly to a zero-mean Gaussian process $\mathcal{P}^{(2)}(t) = (\mathcal{P}_1^{(2)}(t), \mathcal{P}_2^{(2)}(t))^T$ and covariance functions between $\mathcal{P}_j^{(2)}(t)$ and $\mathcal{P}_k^{(2)}(s)$ is $\sum_{l=1}^L q_l \frac{1-\alpha_l}{\alpha_l} E_l\{\zeta_{l1j}(\beta_0, t), \zeta_{l1k}(\beta_0, s)\}$ for $t, s \in [0,\tau]$ in $D[0,\tau]^K$.

Similarly, $P^{(3)}(t)$ can be shown to converge weakly to a zero-mean Gaussian process where covariance function $\mathcal{P}_j^{(3)}(t)$ and $\mathcal{P}_k^{(3)}(s)$ is $\sum_{l=1}^L q_l E_l\{\varphi_{l1j}(\beta_0, t), \varphi_{l1k}(\beta_0, s)\}$ where

$$
\begin{aligned}
& E_l\{\varphi_{l1j}(\beta_0,t), \varphi_{l1k}(\beta_0,s)\} \\
= \;& I(j=k) pr(\Theta_{l10}) (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lik}^{(2)}(\beta, t|\Theta_{l10}) B_{lik}^{(2)}(\beta, s|\Theta_{l10}) | \Theta_{l10}, \xi_{l1} = 0] \\
+ \;& I(j=k) pr(\Theta_{l01}) (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lik}^{(2)}(\beta, t|\Theta_{l01}) B_{lik}^{(2)}(\beta, s|\Theta_{l01}) | \Theta_{l01}, \xi_{l1} = 0] \\
+ \;& I(j=k) \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lik}^{(2)}(\beta, t|\Theta_{l11}) B_{lik}^{(2)}(\beta, s|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& I(j=k) \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lik}^{(2)}(\beta, t|\Theta_{l11}) B_{lik}^{(2)}(\beta, s|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& pr(\Theta_{l10}) (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lij}^{(2)}(\beta_0, t|\Theta_{l10}), l_k(\beta_0, s)^T A(\beta_0)^{-1} B_{lij}^{(1)}(\beta, s|\Theta_{l10}) | \Theta_{l10}, \xi_{l1} = 0] \\
+ \;& pr(\Theta_{l01}) (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lij}^{(2)}(\beta_0, t|\Theta_{l10}), l_k(\beta_0, s)^T A(\beta_0)^{-1} B_{lij}^{(1)}(\beta, s|\Theta_{l10}) | \Theta_{l10}, \xi_{l1} = 0] \\
+ \;& \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lij}^{(2)}(\beta_0, t|\Theta_{l11}), l_k(\beta_0, s)^T A(\beta_0)^{-1} B_{lij}^{(1)}(\beta, s|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lij}^{(2)}(\beta_0, t|\Theta_{l11}), l_k(\beta_0, s)^T A(\beta_0)^{-1} B_{lij}^{(1)}(\beta, s|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& pr(\Theta_{l10}) (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lik}^{(2)}(\beta_0, s|\Theta_{l10}), l_j(\beta_0, t)^T A(\beta_0)^{-1} B_{lik}^{(1)}(\beta, t|\Theta_{l10}) | \Theta_{l10}, \xi_{l1} = 0] \\
+ \;& pr(\Theta_{l01}) (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lik}^{(2)}(\beta_0, s|\Theta_{l10}), l_j(\beta_0, t)^T A(\beta_0)^{-1} B_{lik}^{(1)}(\beta, t|\Theta_{l10}) | \Theta_{l10}, \xi_{l1} = 0] \\
+ \;& \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l1}}{\gamma_{l1}}) Cov_l[B_{lik}^{(2)}(\beta_0, s|\Theta_{l11}), l_j(\beta_0, t)^T A(\beta_0)^{-1} B_{lik}^{(1)}(\beta, t|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& \frac{pr(\Theta_{l11})}{4} (\frac{1-\gamma_{l2}}{\gamma_{l2}}) Cov_l[B_{lik}^{(2)}(\beta_0, s|\Theta_{l11}), l_j(\beta_0, t)^T A(\beta_0)^{-1} B_{lik}^{(1)}(\beta, t|\Theta_{l11}) | \Theta_{l11}, \xi_{l1} = 0] \\
+ \;& \sum_{m=1}^{2} [pr(\Theta_{l10}) (\frac{1-\gamma_{l1}}{\gamma_{l1}}) l_k(\beta_0, t)^T A(\beta_0)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\times & \quad Cov_l[B_{lim}^{(1)}(\beta,t|\Theta_{l10}),B_{lim}^{(1)}(\beta,s|\Theta_{l10})]A(\beta_0)^{-1}l_j(\beta_0,s)^T \\
+ & \quad pr(\Theta_{l01})(\frac{1-\gamma_{l2}}{\gamma_{l2}})l_k(\beta_0,t)^T A(\beta_0)^{-1} \\
\times & \quad Cov_l[B_{lim}^{(1)}(\beta,t|\Theta_{l01}),B_{lim}^{(1)}(\beta,s|\Theta_{l01})]A(\beta_0)^{-1}l_j(\beta_0,s)^T \\
+ & \quad \frac{pr(\Theta_{l11})}{4}(\frac{1-\gamma_{l1}}{\gamma_{l1}})l_k(\beta_0,t)^T A(\beta_0)^{-1} \\
\times & \quad Cov_l[B_{lim}^{(1)}(\beta,t|\Theta_{l11}),B_{lim}^{(1)}(\beta,s|\Theta_{l11})]A(\beta_0)^{-1}l_j(\beta_0,s)^T \\
+ & \quad \frac{pr(\Theta_{l11})}{4}(\frac{1-\gamma_{l2}}{\gamma_{l2}})l_k(\beta_0,t)^T A(\beta_0)^{-1} \\
\times & \quad Cov_l[B_{lim}^{(1)}(\beta,t|\Theta_{l11}),B_{lim}^{(1)}(\beta,s|\Theta_{l11})]A(\beta_0)^{-1}l_j(\beta_0,s)^T].
\end{aligned}
$$

By the conditional expectation arguments, all terms are mutually independent. Therefore, $P(t) = P^{(1)}(t) + P^{(2)}(t) + P^{(3)}(t)$ converges to a zero-mean Gaussian process $\mathcal{G}(t) = \mathcal{P}^{(1)}(t) + \mathcal{P}^{(2)}(t) + \mathcal{P}^{(3)}(t)$.

Proof of Theorem 5

We will compare the asymptotic variance for the proposed method and the existing method. Consider the unstratified generalized case-cohort study (i.e. $L = 1$). From Theorem 3, the covariance matrix for $\tilde{\beta}^G$ involves the first derivative of the weighted estimating functions, $A(\beta_0)$ and the asymptotic variance of the the weighted estimating functions, $\Sigma_G(\beta_0)$. The first derivative of the proposed weighted estimating functions, $A(\beta_0)$ is the same as that of (4.2). Therefore, we only need to compare the asymptotic variance of the proposed weighted estimating functions $\widetilde{U}^G(\beta)$ with that of Kang & Cai's weighted estimating functions $\widehat{U}^{KC}(\beta)$. Note that $n^{-1/2}\widetilde{U}^G(\beta)$ can be decomposed into four parts:

$$
\begin{aligned}
= & \quad n^{-1/2}\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^\tau \left\{Z_{1ik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{1ik}(t) \\
+ & \quad n^{-1/2}\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^\tau \left\{\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} dM_{1ik}(t) \\
+ & \quad n^{-1/2}\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^\tau (\pi_{1ik}(t)-1)\left\{Z_{1ik}(t) - \frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)}\right\} dM_{1ik}(t) \\
+ & \quad n^{-1/2}\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^\tau (\pi_{1ik}(t)-1)\left\{\frac{s_k^{(1)}(\beta_0,t)}{s_k^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_k^{(1)}(\beta_0,t)}{\widetilde{S}_k^{(0)}(\beta_0,t)}\right\} dM_{1ik}(t) \qquad (4.26)
\end{aligned}
$$

Since the second and the fourth terms in (4.26) converge to zero in probability uniformly in t, resepctively and the first term in (4.26) converges to the same limit as that for $\widehat{U}^{KC}(\beta)$, we only need to compare asymptotic properties of the third term with the proposed weight in (4.26) with that with the existing weight. Therefore, we will compare asymptotic variance of the third term with the proposed weight with that with existing weight:

$$Var\left[\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^{\tau}(w_{1ik}(t)-1)\{Z_{1ik}(t)-e_k(t)\}\,dM_{1ik}(t)\right]$$
$$-Var\left[\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{1ik}(t)-1)\{Z_{1ik}(t)-e_k(t)\}\,dM_{1ik}(t)\right]$$

It is sufficient to show the difference between the second moments since the first moment of $E[\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(w_{lik}(t)-1)\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)]$ and $E[\sum_{i=1}^{n}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{lik}(t)-1)\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)]$ are zero. Hence, we get

$$E\left[\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^{\tau}(w_{1ik}(t)-1)\{Z_{1ik}(t)-e_k(t)\}\,dM_{1ik}(t)\right]^2$$
$$-E\left[\sum_{i=1}^{n_1}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{1ik}(t)-1)\{Z_{1ik}(t)-e_k(t)\}\,dM_{1ik}(t)\right]^2 \qquad (4.27)$$

Note that the weight functions $w_{1ik}(t)$ and $\pi_{1ik}(t)$ converge to time-invariant weights, w_{1ik} and π_{1i}, respectively, where $w_{1ik}=(1-\Delta_{1ik})\xi_{1i}\alpha_1^{-1}+\Delta_{1ik}\xi_{1i}+\Delta_{1ik}(1-\xi_{1i})\eta_{1ik}\gamma_{1k}^{-1}$ and $\pi_{1ik}=\Pi_{j=1}^{2}(1-\Delta_{1ij})\xi_{1i}\alpha^{-1}+\{1-\Pi_{j=1}^{2}(1-\Delta_{1ij})\}\xi_{1i}+\Delta_{1i1}(1-\Delta_{1i2})(1-\xi_{1i})\eta_{1i1}\gamma_{11}^{-1}+(1-\Delta_{1i1})\Delta_{1i2}(1-\xi_{1i})\eta_{1i2}\gamma_{12}^{-1}+\frac{1}{2}\Delta_{1i1}\Delta_{1i2}(1-\xi_{1i})\eta_{1i1}\gamma_{11}^{-1}+\frac{1}{2}\Delta_{1i1}\Delta_{1i2}(1-\xi_{1i})\eta_{1i2}\gamma_{12}^{-1}$.

Hence, (4.27) is asymptotically equivalent to

$$\begin{aligned}
& E\left[\sum_{k=1}^{2}(w_{11k}-1)Q_{11k}(t)\right]^2-E\left[\sum_{k=1}^{2}(\pi_{11}-1)Q_{11k}(t)\right]^2\\
= \ & E[w_{111}^2-\pi_{11}^2]E[Q_{111}^2(t)]+E[w_{112}^2-\pi_{11}^2]E[Q_{112}^2(t)]\\
+ \ & 2\{E[w_{111}w_{112}-\pi_{11}^2]\}E[Q_{111}(t)Q_{112}(t)]\\
\geq \ & 2\sqrt{E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]E[Q_{111}^2(t)]E[Q_{112}^2(t)]}\\
+ \ & 2\{E[w_{111}w_{112}-\pi_{11}^2]\}E[Q_{111}(t)Q_{112}(t)],
\end{aligned}$$

$$
\begin{aligned}
\geq\ & 2\sqrt{E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]}\sqrt{E[Q_{111}^2(t)]E[Q_{112}^2(t)]} \\
+\ & 2\rho\{E[w_{111}w_{112}-\pi_{11}^2]\}\sqrt{E[Q_{111}^2(t)]E[Q_{112}^2(t)]} \\
=\ & 2[\sqrt{E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]}+\rho\{E[w_{111}w_{112}-\pi_{11}^2]\}]\sqrt{E[Q_{111}^2(t)]E[Q_{112}^2(t)]}.
\end{aligned}
$$

Since $\sqrt{E[Q_{111}^2(t)]E[Q_{112}^2(t)]}$ is always positive, our prposed weight is more efficient than the existing weight if

$$
\begin{aligned}
& [\sqrt{E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]}+\rho\{E[w_{111}w_{112}-\pi_{11}^2]\}]>0 \\
=\ & \sqrt{E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]}>-\rho\{E[w_{111}w_{112}-\pi_{11}^2]\} \\
=\ & E[w_{111}^2-\pi_{11}^2]E[w_{112}^2-\pi_{11}^2]>[\rho\{E[w_{111}w_{112}-\pi_{11}^2]\}]^2.
\end{aligned}
$$

To get the simple form of $E[w_{111}^2-\pi_{11}^2]$, $E[w_{112}^2-\pi_{11}^2]$, and $[\rho\{E[w_{111}w_{112}-\pi_{11}^2]\}]$, we denote

$$
\begin{aligned}
& \rho=\mathrm{Corr}(Q_{111}(t),Q_{112}(t)),\alpha_1=\lim_{n\to\infty}\mathrm{pr}(\xi_{1i}=1), \\
& p_1=\lim_{n\to\infty}\mathrm{pr}(\Delta_{1i1}=1),\ p_2=\lim_{n\to\infty}\mathrm{pr}(\Delta_{1i2}=1), \\
& \gamma_{11}=\lim_{n\to\infty}\mathrm{pr}(\eta_{1i1}=1|\Delta_{1i1}(1-\xi_{1i})=1),\ \gamma_{12}=\lim_{n\to\infty}\mathrm{pr}(\eta_{1i2}=1|\Delta_{1i2}(1-\xi_{1i})=1).
\end{aligned}
$$

Then we get

$$
\begin{aligned}
& E[w_{111}^2-\pi_{11}^2] \\
=\ & \frac{1-p_1}{\alpha_1}+\alpha p_1+\frac{(1-\alpha_1)p_1}{\gamma_{11}}-[\frac{(1-p_1)(1-p_2)}{\alpha_1}+\alpha_1\{1-(1-p_1)(1-p_2)\} \\
+\ & \frac{(1-\alpha_1)p_1(1-p_2)}{\gamma_{11}}+\frac{(1-\alpha_1)p_1p_2}{4\gamma_{11}}+\frac{(1-\alpha_1)p_2(1-p_1)}{\gamma_{12}}+\frac{(1-\alpha_1)p_1p_2}{4\gamma_{12}}] \\
=\ & \frac{p_2-p_1p_2}{\alpha_1}-\alpha(p_2-p_1p_2)+\frac{(1-\alpha_1)p_1p_2}{\gamma_{11}}-\frac{(1-\alpha_1)p_1p_2}{4\gamma_{11}} \\
-\ & \frac{(1-\alpha_1)(1-p_1)p_2}{\gamma_{12}}-\frac{(1-\alpha_1)p_1p_2}{4\gamma_{12}} \\
=\ & (\frac{1}{\alpha_1}-\alpha_1)p_2(1-p_1)+(1-\alpha_1)\left[\frac{3p_1p_2}{4\gamma_{11}}-\frac{p_2}{\gamma_{12}}+\frac{3p_1p_2}{4\gamma_{12}}\right] \\
=\ & \frac{1}{\alpha_1}\{(1-\alpha_1)(1+\alpha_1)p_2(1-p_1)\}+\frac{3(1-\alpha_1)p_1p_2}{4}\{\frac{1}{\gamma_{11}}+\frac{1}{\gamma_{12}}\}-\frac{(1-\alpha_1)p_2}{\gamma_{12}}
\end{aligned}
$$

$$= \quad (1-\alpha_1)p_2\left[\frac{1+\alpha_1}{\alpha_1}(1-p_1) + \frac{3p_1}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{12}}\right].$$

Similarly, $E[w_{112}^2 - \pi_{11}^2]$ can be written as

$$E[w_{112}^2 - \pi_{11}^2] = (1-\alpha_1)p_1\left[\frac{1+\alpha_1}{\alpha_1}(1-p_2) + \frac{3p_2}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{11}}\right].$$

Also, $E[w_{111}w_{112} - \pi_{11}^2]$ can be written as

$$\begin{aligned}
& E[w_{111}w_{112} - \pi_{11}^2] \\
= \quad & \frac{(1-p_1)(1-p_2)}{\alpha_1} + (1-p_1)p_2 + p_1(1-p_2) + \alpha_1 p_1 p_2 + (1-\alpha_1)p_1p_2 - [\frac{(1-p_1)(1-p_2)}{\alpha_1} \\
+ \quad & \alpha_1\{1-(1-p_1)(1-p_2)\} + \frac{(1-\alpha_1)p_1(1-p_2)}{\gamma_{11}} + \frac{(1-\alpha_1)p_1p_2}{4\gamma_{11}} \\
+ \quad & \frac{(1-\alpha_1)(1-p_1)p_2}{\gamma_{12}} + \frac{(1-\alpha_1)p_1p_2}{4\gamma_{12}}] \\
= \quad & p_1 + p_2 - p_1p_2 - \alpha_1(p_1 + p_2 - p_1p_2) - (1-\alpha_1)\left[\frac{p_1}{\gamma_{11}} - \frac{3p_1p_2}{4\gamma_{11}} + \frac{p_2}{\gamma_{12}} - \frac{3p_1p_2}{4\gamma_{12}}\right] \\
= \quad & (1-\alpha_1)\left[p_1 + p_2 - p_1p_2 + \frac{3p_1p_2}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{p_1}{\gamma_{11}} - \frac{p_2}{\gamma_{12}}\right] \\
= \quad & (1-\alpha_1)\left[p_1\{1 - \frac{1}{\gamma_{11}}\} + p_2\{1 - \frac{1}{\gamma_{12}}\} + p_1p_2\{\frac{3}{4}(\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}) - 1\}\right]
\end{aligned}$$

Thus, we have

$$\begin{aligned}
& E[w_{111}^2 - \pi_{11}^2]E[w_{112}^2 - \pi_{11}^2] > [\rho\{E[w_{111}w_{112} - \pi_{11}^2]\}]^2 \\
= \quad & p_1p_2\left[\frac{1+\alpha_1}{\alpha_1}(1-p_1) + \frac{3p_1}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{12}}\right]\left[\frac{1+\alpha_1}{\alpha_1}(1-p_2) + \frac{3p_2}{4}\{\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}\} - \frac{1}{\gamma_{11}}\right] \\
> \quad & \rho^2\left[p_1\{1 - \frac{1}{\gamma_{11}}\} + p_2\{1 - \frac{1}{\gamma_{12}}\} + p_1p_2\{\frac{3}{4}(\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{12}}) - 1\}\right]^2.
\end{aligned}$$

Therefore, if the condition $E[w_{111}^2 - \pi_{11}^2]E[w_{112}^2 - \pi_{11}^2] > (\rho E[w_{111}w_{112} - \pi_{11}^2])^2$ is satisfied, then the asymptotic variance for our proposed method is smaller than that for Kang and Cai [2010]'s method. Specifically, smaller α_1 induces larger $(1+\alpha_1)/\alpha_1$, which dominates other contributions in $E[w_{111}^2 - \pi_{11}^2]E[w_{112}^2 - \pi_{11}^2]$. The quantity $(\rho E[w_{111}w_{112} - \pi_{11}^2])^2$ depends on the selection probability of a subset of cases γ_{11} and γ_{12} for fixed the disease rates p_1 and p_2. This indicates that selecting small size of the subcohort and larger portion of cases improves the efficiency for Kang & Cai's method.

4.4 Simulations

We conducted simulation studies to investigate the finite sample properties of the proposed methods, compare it with Kang and Cai [2010]'s method, and compare the performance of stratified sampling with unstratified sampling. Consider the situation that stratum variables are available and two generalized case-cohort studies have been conducted for nonrare disease 1 and nonrare disease 2, respectively. In this situation, covariate information are collected for the subcohort and a portion of the subjects outside the subcohort with disease 1 and disease 2. We generated multivariate failure time data from Clayton-Cuzick model (Clayton and Cuzick [1985]). The bivariate survival function for the bivariate survival time (T_1, T_2) given (Z_{l1}, Z_{l2}) has the following form:

$$F(t_1, t_2|Z_{l1}, Z_{l2}) = \{S_1(t_1; Z_{l1})^{-1/\theta} + S_2(t_2; Z_{l2})^{-1/\theta} - 1\}^{-\theta},$$

where $\lambda_{0k}(t)$ and β_k $k = 1,2$ are the baseline hazard function and the effect of covariate for disease k, respectively, θ is the association parameter between the failure times of the two diseases, and $S_k(t; Z_l) = Pr(T_k > t|Z_{lk}) = e^{-\int_0^{t_k} \lambda_{0k}(t)e^{\beta_k Z_{lk}}dt}$. Exponential distribution with failure rate $\lambda_{0k}e^{\beta_k Z_{lk}}$ is considered for the marginal distribution of T_k $k = 1,2$. Two failure times, T_1 and T_2 are independent as $\theta \to \infty$. The relationship between Kendall's tau, τ_θ, and θ is $\tau_\theta = \frac{1}{2\theta+1}$. Smaller Kendall's tau represents less correlation between T_1 and T_2. Values of 0.1, 0.67 and 4 are used for θ and the corresponding Kendall's tau is 0.83, 0.43 and 0.11, respectively. We set the baseline hazard function $\lambda_{01} = 2$ for the first failure event type $k = 1$ and $\lambda_{02} = 6$ for the second failure event type $k = 2$. For covariates, we consider the situation $Z_{l1} = Z_{l2} = Z$ where Z is generated from Bernouilli distribution with $\text{pr}(Z = 1) = 0.5$. To consider stratified subcohort sampling from two strata defined by V_i, we define two parameters: $\eta = \text{Pr}(V = 1|Z = 1)$ and $\nu = \text{Pr}(V = 0|Z = 0)$ where η is sensitivity and ν is the specificity for Z. Unstratified sampling with same probability, i.e., $\eta = 0.5$ and $\nu = 0.5$ is a special case. Larger values η and ν values than 0.5

indicate that V is highly correlated with Z. For stratified case-cohort studies, we set the values $[\eta, \nu] = [0.7, 0.7]$ and $[\eta, \nu] = [0.9, 0.9]$. Thus, a stratum variable is simulated with $\Pr(V = 1) = (1 - \nu)\Pr(Z = 0) + \eta\Pr(Z = 1) = 0.5$. Censoring time is simulated from uniform distribution $[0, u]$ where u depends on the specified level of the censoring probability. We set the event proportions of approximately 8% and 15% for $k = 1$ and 22% and 36% for $k = 2$. The corresponding u values are 0.08 and 0.16, for $\beta = 0.1$; 0.06 and 0.11, for $\beta = \log(2)$.

The sample size of the full cohort is set to be $n = 1000$. We select the subcohort and a subset of cases by unstratified sampling as well as stratified sampling and consider the subcohort size of 200. We select the subcohort $\tilde{n}_l = \tilde{n} \times q_l$ from each stratum. By using simple random sampling, we select non-subcohort cases size of $\tilde{m}_{lk} = (n_{lk} - \tilde{n}_{lk}) \times \gamma_k$ for $k = 1, 2$ and $l = 0, 1$. We consider the same sample size for two sets of event proportions. For event proportion $[8\%, 22\%]$, γ_k is set to be $[1, 0.57]$; for event proportion $[15\%, 36\%]$, γ_k is set to be $[0.53, 0.44]$. For each configuration, 2000 simulations were conducted.

In the first set of simulation, we consider the simulations of two stratified generalized case-cohort studies with non-rare diseases. Our main interest is to estimate the effect of Z on disease 1. We will examine the performance of our proposed estimator based on (4.4) with $K = 1$ which uses the additional information collected on the sampled subjects with disease 2 and compare the stratified sampling with the unstratified sampling. We will also compare our results with those using Kang and Cai [2010]'s method for disease 1 which are based on (4.2) with $K = 1$.

Table 4.1 summarizes the results. For different combinations of event proportion, the subcohort sample size, correlation, and sampling methods, Table 4.1 shows the average of the estimates for β_2, the average of the proposed estimated standard error (SE), empirical standard deviation (SD), and sample relative efficiency (SRE). The subscripts for SE, SD refer to the proposed method (P) and Kang and Cai [2010]'s method (K). To compare the stratified sampling with unstratified sampling, the sample relative efficiency in the proposed method (SRE_p) is defined as SD_p^2 for unstratified sampling over SD_p^2 for stratified sampling. The sample relative efficiency (SRE_k) in Kang and Cai [2010]'s method is defined as SD_k^2 for unstratified sampling over SD_k^2 for stratified sampling. $STR1$ and $STR2$ represent

Table 4.1: Simulation result with a single disease outcome ($K = 1$): $\beta_1 = \log(2) = 0.693$

Model				The Proposed method					Kang and Cai's method					
	P_2	$\tilde{\gamma}_2$	τ_θ	$\tilde{\beta}_2^G$	SE_p	SD_p	CR_p	SRE_p	$\hat{\beta}_2$	SE_k	SD_k	CR_k	SRE_k	SRE
UNS	22%	0.57	0.83	0.704	0.221	0.225	95	1.00	0.704	0.227	0.229	95	1.00	1.03
			0.43	0.704	0.221	0.229	94	1.00	0.705	0.226	0.233	94	1.00	1.04
			0.11	0.702	0.221	0.215	96	1.00	0.704	0.227	0.220	96	1.00	1.04
	36%	0.44	0.83	0.706	0.193	0.194	95	1.00	0.701	0.199	0.197	95	1.00	1.03
			0.43	0.697	0.193	0.195	94	1.00	0.698	0.199	0.200	94	1.00	1.04
			0.11	0.696	0.192	0.194	96	1.00	0.699	0.199	0.199	96	1.00	1.05
STR1	22%	0.57	0.83	0.695	0.215	0.223	95	1.02	0.693	0.223	0.228	95	1.00	1.05
$[\eta, \nu]$			0.43	0.707	0.216	0.214	95	1.14	0.709	0.223	0.217	96	1.15	1.03
$= [0.7, 0.7]$			0.11	0.704	0.217	0.215	95	1.00	0.704	0.223	0.218	95	1.01	1.03
	36%	0.44	0.83	0.702	0.189	0.192	95	1.02	0.696	0.197	0.195	95	1.02	1.04
			0.43	0.697	0.190	0.186	95	1.11	0.697	0.197	0.192	95	1.08	1.07
			0.11	0.700	0.189	0.185	95	1.09	0.702	0.197	0.194	96	1.05	1.09
STR2	22%	0.57	0.83	0.701	0.196	0.198	95	1.29	0.698	0.209	0.204	95	1.26	1.06
$[\eta, \nu]$			0.43	0.703	0.200	0.193	96	1.40	0.703	0.209	0.196	97	1.41	1.03
$= [0.9, 0.9]$			0.11	0.700	0.202	0.194	96	1.23	0.700	0.209	0.198	96	1.23	1.04
	36%	0.44	0.83	0.706	0.174	0.165	97	1.38	0.703	0.191	0.171	97	1.33	1.08
			0.43	0.694	0.179	0.167	97	1.36	0.694	0.192	0.172	97	1.35	1.06
			0.11	0.700	0.182	0.167	97	1.35	0.699	0.192	0.173	97	1.32	1.08

SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; $SRE_p = SD_p^2$ for unstratified sampling/SD_p^2 for stratified sampling, sample relative efficiency in the proposed method; $SRE_k = SD_k^2$ for unstratified sampling/SD_k^2 for stratified sampling, sample relative efficiency in Kang & Cai's method; UNS, unstratified sampling; $STR1$, stratified sampling with $[\eta, \nu] = [0.7, 0.7]$; $STR2$, stratified sampling with $[\eta, \nu] = [0.9, 0.9]$.

stratified sampling with sensitivity and specificity with $[\eta, \nu] = [0.7, 0.7]$ $[\eta, \nu] = [0.9, 0.9]$, respectively.

From the results, we see that both methods are approximately unbiased. The average of the proposed estimated standard error is close to the empirical standard deviation and the range of the 95% confidence interval coverage rate is on 94%-97%. In general, the estimates for stratified sampling of the subcohort and cases have smaller variance than those for unstratified sampling. To compare the stratified sampling with unstratified sampling, all the sample relative efficiency (SRE_p and SRE_k) for models with stratified sampling ($STR1$ and $STR2$) are more than 1, which indicates that stratified sampling is more efficient than unstratified sampling. This shows that stratum variable available on all the subjects helps to gain efficiency. When correlation between stratum variables and covariates is larger, more efficiency gain is obtained. Also, estimated standard errors for the proposed method are smaller than those for Kang and Cai [2010]'s method. From sample relative efficiency (SRE), all SREs are larger than 1. Hence, our proposed method gains the efficiency. The results for β_1 are not shown, but they are similar with β_2.

In the second set of simulations, we are interested in the joint modeling of the two diseases (i.e. $K = 2$). These correspond to (4.2) for Kang and Cai [2010]'s method and (4.4)) for the proposed method. We examine the performance of our proposed estimator and compare it to those from Kang and Cai [2010]. Our main interests are to estimate the effect of Z on disease 1 (β_1) and disease 2 (β_2) and compare them. Table 4.2 provides summary statistics for the estimate of β_1 for different combinations of event proportion, subcohort sample size, correlation, and sampling methods. The simulation results suggest that the estimates for both methods are approximately unbiased and their estimated standard errors are close to the empirical standard deviations. The range of the coverage rate of the nominal 95% confidence interval is 94%-97%. All sample relative efficiency (SRE_p and SRE_k) for models with stratified sampling ($STR1$ and $STR2$) are more than 1 which implies that stratified sampling is more efficient than that of unstratified sampling and higher efficiency gain is related with higher sensitivity and specificity. The variance for the propose method are smaller than those for Kang and Cai [2010]'s method, which indicates that our proposed

Table 4.2: Simulation result with multiple disease outcomes (K = 2): β = [0.1, 0.7]

Model	P_1		The Proposed method					Kang and Cai's method					
	$[\tilde{\gamma}_1, \tilde{\gamma}_2]$	τ_θ	$\tilde{\beta}_2^G$	SE_p	SD_p	CR_p	SRE_p	$\hat{\beta}_2$	SE_k	SD_k	CR_k	SRE_k	SRE
UNS	[8%,22%]	0.83	0.710	0.201	0.205	95	1.00	0.706	0.206	0.208	95	1.00	1.03
	[1,0.57]	0.43	0.705	0.202	0.201	96	1.00	0.703	0.206	0.204	95	1.00	1.03
		0.11	0.706	0.201	0.206	94	1.00	0.707	0.206	0.209	95	1.00	1.03
	[15%,36%]	0.83	0.712	0.178	0.185	94	1.00	0.704	0.182	0.188	94	1.00	1.04
	[0.53, 0.44]	0.43	0.710	0.178	0.181	95	1.00	0.709	0.182	0.186	95	1.00	1.05
		0.11	0.712	0.176	0.180	95	1.00	0.713	0.182	0.187	94	1.00	1.08
STR1	[8%,22%]	0.83	0.704	0.196	0.196	95	1.10	0.701	0.201	0.199	96	1.09	1.04
$[\eta, \nu]$	[1,0.57]	0.43	0.707	0.197	0.192	96	1.09	0.706	0.202	0.196	96	1.08	1.04
$= [0.7, 0.7]$		0.11	0.700	0.196	0.198	94	1.09	0.700	0.201	0.203	95	1.07	1.05
	[15%,36%]	0.83	0.716	0.175	0.174	95	1.13	0.709	0.181	0.177	95	1.13	1.04
	[0.53, 0.44]	0.43	0.705	0.176	0.175	95	1.07	0.703	0.182	0.178	95	1.08	1.04
		0.11	0.709	0.174	0.173	95	1.08	0.709	0.181	0.179	95	1.09	1.07
STR2	[8%,22%]	0.83	0.711	0.179	0.172	96	1.42	0.709	0.189	0.177	96	1.38	1.07
$[\eta, \nu]$	[1,0.57]	0.43	0.701	0.182	0.171	96	1.38	0.700	0.189	0.174	97	1.38	1.03
$= [0.9, 0.9]$		0.11	0.698	0.183	0.174	96	1.40	0.697	0.189	0.177	96	1.39	1.04
	[15%,36%]	0.83	0.705	0.164	0.155	97	1.42	0.701	0.179	0.161	97	1.37	1.08
	[0.53, 0.44]	0.43	0.700	0.168	0.151	97	1.44	0.699	0.179	0.157	97	1.40	1.07
		0.11	0.706	0.169	0.153	97	1.37	0.706	0.179	0.156	97	1.43	1.04

SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; SRE_p = SD_p^2 for unstratified sampling/SD_p^2 for stratified sampling, sample relative efficiency in the proposed method; SRE_k = SD_k^2 for unstratified sampling/SD_k^2 for stratified sampling, sample relative efficiency in Kang & Cai's method; UNS, unstratified sampling; $STR1$, stratified sampling with $[\eta, \nu] = [0.7, 0.7]$; $STR2$, stratified sampling with $[\eta, \nu] = [0.9, 0.9]$.

Table 4.3: Type I error and power (%) in separate and joint analyses: $[\eta, \nu] = [0.7, 0.7]$

Event			Type I error ($\beta_1 = \beta_2 = \log 2$)				Power ($\beta_1 = 0.1, \beta_2 = 0.7$)			
			S		J		S		J	
proportion	$\tilde{\gamma}$	τ_θ	PR	KC	PR	KC	PR	KC	PR	KC
[8%, 22%]	[1,0.57]	0.83	1.0	0.9	4.0	5.0	40	44	63	67
		0.43	1.7	2.5	4.3	5.2	39	43	54	56
		0.11	3.1	2.9	4.5	5.8	41	43	51	54
[15%, 36%]	[0.53, 0.44]	0.83	0.5	1.6	4.2	5.5	52	49	73	69
		0.43	1.5	1.7	4.6	4.6	50	49	67	63
		0.11	2.7	2.7	5.1	5.1	52	49	62	58

S, Separate analysis; J, Joint analysis; KC, Kang and Cai [2010]'s method; PR, proposed method.

method are more efficient than those for Kang and Cai [2010]'s method.

We also conducted simulation studies to examine the Type I error rates and powers in comparing the effect of the risk factor on the two diseases. We conducted the test based on the joint analysis with stratified sampling with $[\eta, \nu] = [0.7, 0.7]$ for the two diseases. We also conducted tests using the coefficient estimate from separate analysis for each of the two diseases assuming independence of the sample. Estimating equations (4.2) and (4.4) with $K = 1$ are used for the separate analysis and estimating equations (4.2) and (4.4) with $K = 2$ are used for the joint analysis. Table 4.3 summarizes the results for Type I error rates and powers. Type I error rates are obtained by testing $H_0 : \beta_1 = \beta_2$ under setting $\beta_1 = \beta_2 = \log(2)$ at the significant level .05. The settings for the simulation for the power are the same as before except that $\beta_1 = 0.1$ and $\beta_2 = 0.7$. The tests under separate analysis treat the two estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$, as from two independent samples. The results indicate that tests based on our proposed weight function are more powerful than those based on Kang and Cai [2010]'s weight function and the powers of joint analysis are larger than those based on the separate analysis. Note that Type I error rates of separate analysis are much less than .05 while the joint analysis methods have Type I error rates close to .05.

4.5 Data analysis

We applied the proposed method to a data set from the ARIC study which is a population-based cohort study [Duncan et al., 2003; Ballantyne et al., 2004]. This study consists of 15,792 men and women 45 - 64 years of age from four U.S. communities recruited during 1987 to 1989. All subjects were followed for incident diabetes. The incident diabetes are defined as a reported physician diagnosis, use of antidiabetes medications, a fasting (≥ 8 hours) glucose ≥ 7.0 mmol/l, or a nonfasting glucose of ≥ 11.1 mmol/l. Subjects are regarded as censored if they are alive and event-free at the end of 1998 or lost to follow-up.

Our main interest is to investigate the association between high-sensitivity C-reactive protein (hs-CRP), which is a biomarker of inflammation, and incident diabetes events. In order to measure hs-CRP, a case-cohort study was conducted to reduce the cost and save blood specimen. Hs-CRP is also available on subjects for incident coronary heart disease (CHD) from another case-cohort study in the ARIC study [Ballantyne et al., 2004]. Using available hs-CRP from another case-cohort, we excluded subjects with prevalent CHD and prevalent diabetes at baseline, transient ischemic attach or stroke, had missing follow-up visits; were minority race group; had no valid diabetes determination at follow-ups, missing CHD information, and baseline measurements. The full cohort consist of 10,279 subjects.

To preserve frozen biologic specimens and reduce costs, generalized case-cohort design is used by selecting a subset of incident diabetes events since the rate of diabetes during follow-up is 11.2%. The subcohort and cases of incident diabetes are randomly selected via stratified sampling where the strata variables are age at baseline (≤ 55 and > 55), sex, and race (black and white). Age, gender, race, parental history of diabetes, hypertension, and center are confounding factors and are adjusted in the model. The risk factor, hs-CRP, is used as a categorical variables with 4 levels based on quartiles. In table 4.4, hs-CRP (C2), hs-CRP (C3), and hs-CRP (C4) are indicator variables for hs-CRP values in the second, third, fourth quartiles, respectively. The hs-CRP values in the first quartile is used as the reference group in our analysis.

By using available hs-CRP information collected from subjects who have CHD, we can

Table 4.4: Results for the effect of hs-CRP from the ARIC Study

	Proposed method				The existing method			
Variables	$\tilde{\beta}^{Gk}$	SE	HR	95% CI	$\hat{\beta}^{Gk}$	SE	HR	95% CI
hs-CRP(C4)	1.00	0.214	2.71	(1.78 , 4.12)	1.02	0.220	2.78	(1.80 , 4.28)
hs-CRP(C2)	0.21	0.239	1.23	(0.77 , 1.97)	0.23	0.243	1.26	(0.78 , 2.02)
hs-CRP(C3)	0.73	0.213	2.07	(1.36 , 3.14)	0.75	0.220	2.12	(1.38 , 3.26)
Age	0.01	0.011	1.00	(0.98 , 1.03)	0.01	0.012	1.01	(0.98 , 1.03)
African	0.56	0.278	1.74	(1.01 , 3.01)	0.55	0.287	1.73	(0.98 , 3.03)
Male	0.31	0.120	1.37	(1.08 , 1.73)	0.33	0.131	1.40	(1.08 , 1.81)
PHD	0.61	0.153	1.84	(1.36 , 2.48)	0.63	0.160	1.88	(1.37 , 2.57)
HYP	0.56	0.155	1.75	(1.29 , 2.37)	0.56	0.161	1.75	(1.28 , 2.40)
Center (F)	0.15	0.228	1.16	(0.74 , 1.82)	0.18	0.237	1.19	(0.75 , 1.90)
Center (J)	-0.11	0.325	0.89	(0.47 , 1.69)	-0.09	0.334	0.92	(0.48 , 1.76)
Center (M)	-0.04	0.225	0.96	(0.62 , 1.49)	-0.02	0.233	0.98	(0.62 , 1.56)

hs-CRP, high-sensitivity C-reactive protein; PHD, parental history of diabetes; HYP, hypertension; SE, standard error estimate; HR, hazard ratio estimate; CI, confidence interval

apply our proposed method to this data set. The total sample size is 1,576 subjects including 572 noncases, 581 diabetes cases, 423 CHD cases. The subcohort size is 669 which consists of 96 diabetes cases and 572 non-cases. To study the effect of hs-CRP for diabetes, we fit the model using (4.1) and compare the results for the proposed method in (4.4) and Kang and Cai [2010]'s method in (4.2) when $K = 1$.

Table 4.4 represents the estimates, standard errors, hazard ratios, 95% confidence intervals for two methods. The hazard ratio comparing the fourth with the first hs-CRP quartile group is 2.71 and confidence interval indicates that it is of statistical significance. Moreover, the hazard ratio comparing the third with the first hs-CRP quartile group is also statistically significant, but the hazard ratio for the second versus the first quartile group is not statistically significant. The regression coefficient estimates for the proposed method are similar with those for the existing method, but all the standard errors are smaller than those of the existing method and consequently the 95% confidence intervals are narrower.

4.6 Concluding Remarks

We proposed more efficient estimators for stratified generalized case-cohort design than those for [Kang and Cai, 2010] by using available stratum variables and exposure information for the other diseases. For a single disease outcome and multiple disease outcomes, weighted estimating equations with the proposed weight function were proposed. We have shown that our proposed estimators are consistent and asymptotically normally distributed under some regularity conditions. The asymptotic relative efficiency of the proposed was derived and we can calculate the efficiency gain in practice. Based on simulation results, our proposed methods improve efficiency and stratified sampling of the subcohort and cases produces more efficiency gain than unstratified sampling.

In this paper, we proposed the new weight function for the generalized case-cohort study with two types of diseases. We can extend the general weight function with K diseases:

$$\begin{aligned}\pi_{lik}(t) &= \prod_{j=1}^{K}(1-\Delta_{lij})\xi_{li}\tilde{\alpha}_{lk}^{-1}(t) + \left\{1-\prod_{j=1}^{K}(1-\Delta_{lij})\right\}\xi_{li} \\ &+ (1-\xi_{li})[\sum_{m\in M(1)}\frac{1}{N(M)}\{\prod_{j\in M}\Delta_{lij}\prod_{j'\in A-M}(1-\Delta_{lij'})\}\eta_{lim}\tilde{\gamma}_{l,jj',k}^{-1}(t)],\end{aligned}$$

where $\tilde{\alpha}_{lk}(t) = \sum_{i=1}^{n_l}\prod_{j=1}^{K}(1-\Delta_{lij})\xi_{li}Y_{lik}(t)/\sum_{i=1}^{n_l}\prod_{j=1}^{K}(1-\Delta_{lij})Y_{lik}(t)$, A is set with $\{1,2,\ldots,K\}$, M are all possible subsets of A except for $\varnothing$, $N(M)$ is the number of elements in M, $M(1)$ is one of elements in the set M and $\tilde{\gamma}_{l,jj',k}(t)$ is the selection probability of cases among non-subcohort members in each part. Therefore, the situation that there are K diseases can be proved by using similar arguments.

In practice, full cohort size and the disease rates are fixed. Using the formula in Theorem 5, we can calculate the efficiency gain for different combinations of α_1, γ_{11}, and γ_{12}. However, if the conditions are not satisfied, variance for our proposed method could be smaller than that for Kang and Cai [2010]. Therefore, our proposed method is not always efficient. We need to derive the most efficient estimator by specifying the joint distribution of the correlated failure times from the same subject. This would be worthwhile, especially for data with expensive covariates. This could be interesting future research.

In some data, proportional hazard assumptions are not appropriate and some investigators could be interested in another association between risk factor and disease outcomes. Hence, alternatives of proportional hazard models are other types of models such as additive hazards models, proportional odds model, the accelerated failure time model, and the semiparametric transformation model. In addition to proportional hazards models, we can adapt our approaches to the stratified case-cohort study with the above models.

Chapter 5

Additive hazards model for stratified case-cohort design

5.1 Introduction

There are two main principal frameworks to investigate the associations between risk factors and the disease outcome: Cox [1972]'s proportional hazards model and the additive hazards model. Most of the authors have studied multiplicative hazards models for relative risk using proportional hazards models in which covariate effects can be expressed as hazard ratios. However, the proportional hazards assumption might not be appropriate for some data. In addition, epidemiologists are often interested in the risk difference attributed to the exposure and the risk difference is useful in public health decision since it can translate directly into the number of disease cases [Kulich and Lin, 2000b]. Therefore, additive hazards models have been a useful and important alternative to Cox [1972]'s proportional hazards model.

There are some work for additive hazards models. Lin and Ying [1994] proposed semi-parametric estimation for univariate failure time data and studied asymptotic properties of the estimators. Yin and Cai [2004] extended this approach to the multivariate failure time data. By using Lin and Ying [1994]'s estimators, Pipper and Martinusse [2004] also considered marginal additive hazards models for clustered data.

All the aforementioned work deals with all the subjects in the full cohort. In large cohort studies, obtaining expensive covariate information on all members in the entire cohort could be costly and it could be infeasible due to limited financial resource. In order to reduce cost, the case-cohort study is proposed by Prentice [1986]. Under the case-cohort design, covariate information can be collected only from the subcohort which is a random sample from whole cohort and all the subjects who have diseases of interest. The important advantage for the case-cohort study is that the same subcohort can be used when several types of diseases are of interest [Wacholder et al., 1991].

A few methods for case-cohort studies with additive hazards models have been studied. For univariate failure time, Kulich and Lin [2000a] applied additive hazards models to the case-cohort study and derived the large-sample theory of the proposed estimators. Sun et al. [2004] extended this approach to competing risks analysis in the case-cohort study. For multiple disease events, Kang et al. [2012] proposed marginal additive hazards model for case-cohort studies and consider stratified sampling for selection of the subcohort.

Taking advantage of the case-cohort design, several diseases are usually studied using the same subcohort. In such situation, the information on the expensive exposure measure are available on the subcohort as well as on any subjects with any of the diseases under the study. For example, one of the goals in the Atherosclerosis Risk in Communities (ARIC) study is to investigate the association between the genetic variation in PTGS1 and coronary heart disease (CHD) as well as stroke and to compare the effects of the genetic variation on CHD and stroke [Lee et al., 2008]. In this study, the case-cohort design with stratified sampling for the subcohort are used. To examine the relationship between the genetic variation and CHD as well as stroke, two case-cohort studies were conducted separately. We are interested in examining the effect of PTGS1 on the CHD and stroke.

The genetic variation in PTGS1 was collected from the subcohort and all subjects with CHD and/or stroke. Typically, when analysis for CHD was conducted, the available information for stroke were ignored. This is not efficient use of the available information. In addition, it is often of interest to compare the effects of risk factors on multiple diseases. Kang et al. [2012] considered the joint modeling with additive hazards models. However, they also did not fully use all the available information. These motivate us to consider a more efficient estimator which uses all the available information for the additive hazards model with stratified case-cohort design.

In this paper, we propose estimation procedure in the additive hazards model for traditional and generalized stratified case-cohort design with univariate failure time as well as multivariate failure time. In Section 5.2, we propose models and estimation procedures for the proposed methods. Section 5.3 summarizes asymptotic properties for the proposed estimators and Section 5.4 reports some simulation results. In Section 5.5, we analyze data from the ARIC study by using the proposed method. Concluding remarks are provided in Section 5.6.

5.2 Model

Suppose that a cohort study consists of n independent subjects with K diseases of interest and can be divided into L mutually exclusive strata based on available information V from all cohort members. Let T_{lik} denote the potential failure time and C_{lik} the potential censoring time for disease k of subject i within stratum l. We assume that T_{lik} is independent of C_{lik} given covariates. Let $Z_{lik}(t)$ be a $p \times 1$ possibly time-dependent covariates vector for diseases k of subject i within stratum l at time t. We assume that time-dependent covariates are external; that is, they are not influenced by the disease processes [Kalbfleisch and Prentice, 2002]. Let $X_{lik} = \min(T_{lik}, C_{lik})$ denote the observed time, $\Delta_{lik} = I(T_{lik} \leq C_{lik})$ the indicator for failure, $N_{lik}(t) = I(X_{lik} \leq t, \Delta_{lik} = 1)$ the counting process, and $Y_{lik}(t) = I(X_{lik} \geq t)$ the at risk indicator for disease k of subject i within stratum l, where $I(.)$ is the indicator function. Let V_i denote a discrete random variable for subject i as a stratum variable. The stratum variable is assumed to be independent of T_{lik} given $Z_{lik}(t)$, i.e., V_i affects T_{lik} only through $Z_{lik}(t)$ [Kulich and Lin, 2004]. Let τ denote the end of study time.

Consider the following additive hazards model for T_{lik} given $Z_{lik}(t)$

$$\lambda_{lik}\{t|Z_{lik}(t)\} = \lambda_{0k}(t) + \beta_0^T Z_{lik}(t), \tag{5.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease k of subject i and β_0 is

p-dimensional fixed and unknown parameters. Model (5.1) can incorporate disease-type-specific effect model $\lambda_{ik}\{t|Z^*_{lik}(t)\} = \lambda_{0k}(t) + \beta_k^T Z^*_{lik}(t)$ as a special case. Specifically, we define $\beta_0^T = (\beta_1^T, \ldots, \beta_k^T, \ldots, \beta_K^T)$ and $Z_{lik}(t)^T = (0_{i1}^T, \ldots, 0_{i(k-1)}^T, \{Z^*_{lik}(t)\}^T, 0_{li(k+1)}^T, \ldots, 0_{liK}^T)$ where 0^T is a $1 \times p$ zero vector. We have $\beta_0^T Z_{lik}(t) = \beta_k^T Z^*_{lik}(t)$.

First, we consider the traditional case-cohort design with stratified sampling and refer to this design as traditional stratified case-cohort design. Suppose that the total size of cohort n is partitioned into n_l intervals for $l = 1, \ldots, L$. Under traditional stratified case-cohort design, we assume that subjects in the subcohort are selected by stratified random sampling. Specifically, we select a fixed size $\tilde{n}_l$ subjects from the n_l subjects in stratum l into the subcohort by using simple random sampling and the total subcohort size is $\tilde{n} = \Sigma_{l=1}^L \tilde{n}_l$.

Let ξ_{li} be an indicator for subcohort membership for subject i in stratum l. Each subject in stratum l has the same probability $\tilde{\alpha}_l = \Pr(\xi_{li} = 1) = \tilde{n}_l / n_l$ into the subcohort. $Z_{lik}(t)(0 \le t \le \tau)$ are measured for subjects in the subcohort and those with any disease of interest.

In many biomedical and clinical studies with common diseases or the large number of cases, selecting all cases is not feasible due to limited resources. Under this situation, it is appropriate to consider the stratified case-cohort design which has flexibility to select a different portion of all cases among the non-subcohort members in a different stratum. We refer to this design as generalized stratified case-cohort design.

Under the generalized stratified case-cohort design, after selection of subcohort, we select a fixed number $\tilde{m}_{lk}$ of the type k disease cases among non-subcohort members in stratum l by simple random sampling. Denote by $\tilde{m}_k = \Sigma_{l=1}^L \tilde{m}_{lk}$ the total size of the type k disease cases. Let η_{lik} be the indicator for whether subject i in stratum l is sampled for non-subcohort disease k. Let $\tilde{\gamma}_{lk} =$ Pr $(\eta_{lik} = 1 | \Delta_{lik} = 1, \xi_{li} = 0) = \tilde{m}_{lk} / (n_{lk} - \tilde{n}_{lk})$ denote the selection probability of subjects among non-subcohort members in stratum l with disease k, where n_{lk} and $\tilde{n}_{lk}$ denote the number of disease k in the cohort and in the subcohort within stratum l, respectively. Due to sampling scheme, the elements in $(\eta_{l1k}, \ldots, \eta_{ln_lk})$ are correlated, however, $(\eta_{l1k}, \ldots, \eta_{ln_lk})$ is independent of $(\eta_{l'1k'}, \ldots, \eta_{l'n_{l'}k'})$ for $k \ne k'$ or $l \ne l'$.

5.2.1 Estimation for univariate failure time

Consider the situation with only one rare disease of interest, but with covariate information available for subjects with other diseases. Under this situation, the observable information is $(X_{lik}, \Delta_{lik}, \xi_{li}, Z_{lik}(t), 0 \le t \le X_{lik}, V_i)$ when $\xi_{li} = 1$ or $\Delta_{lik} = 1$ and is $(X_{lik}, \Delta_{lik}, \xi_{li}, V_i)$ when $\xi_{li} = 0$ and $\Delta_{lik} = 0$. In the situation that covariate information are not available for subjects with other diseases, Kulich and Lin [2000a] proposed the additive hazards model for traditional case-cohort studies for a single disease using stratified simple random sampling. For example, if we are interested in disease k and ignore the covariate information collected on subjects with the other disease, the true regression parameter β_0 in (5.1) can be estimated by solving the estimating equation [Kulich and Lin, 2000a]:

$$U_k^A(\beta) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} \rho_{ik} \int_0^{\tau} \{Z_{lik}(t) - \bar{Z}_k(t)\}\{dN_{lik}(t) - \beta^T Z_{lik}(t) Y_{lik}(t) dt\} = 0, \qquad (5.2)$$

where

$$\bar{Z}_k(t) = \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l} \rho_{lik} Z_{lik}(t) Y_{lik}(t)}{\sum_{l=1}^{L}\sum_{i=1}^{n_l} \rho_{lik} Y_{lik}(t)}$$

and $\rho_{lik} = \Delta_{lik} + (1 - \Delta_{lik})\xi_{li}\hat{\alpha}_{lk}^{-1}$ with $\hat{\alpha}_{lk} = \sum_{i=1}^{n_l} \xi_{li}(1 - \Delta_{lik}) / \sum_{i=1}^{n_l}(1 - \Delta_{lik})$. Here $\hat{\alpha}_{lk}$, an estimator for the true selection probability $\tilde{\alpha}$, is the proportion of the sampled subjects in the subcohort without disease k among all subjects in stratum l without disease k. This approach for the weight function was first proposed by Kalbfleisch and Lawless [1988] and Borgan et al. [2000] proposed the time-varying weight version $\rho_{lik}(t)$ where $\rho_{lik}(t) = \Delta_{lik} + (1 - \Delta_{lik})\xi_{li}\hat{\alpha}_{lk}^{-1}(t)$ with $\hat{\alpha}_{lk}(t) = \sum_{i=1}^{n_l} \xi_{li}(1 - \Delta_{lik})Y_{lik}(t) / \sum_{i=1}^{n_l}(1 - \Delta_{lik})Y_{lik}(t)$. $\hat{\beta}$ is defined as the solution to (5.2) and has the following explicit form:

$$\hat{\beta}^A = \left[\sum_{l=1}^{L}\sum_{i=1}^{n_l} \rho_{lik} \int_0^{\tau} \{Z_{lik}(t) - \bar{Z}_k(t)\}^{\otimes 2} Y_{lik}(t) dt\right]^{-1} \sum_{l=1}^{L}\sum_{i=1}^{n_l} \int_0^{\tau} \{Z_{lik}(t) - \bar{Z}_k(t)\} dN_{lik}(t),$$

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

To make full use of collected covariate information on subjects with other diseases, we consider the following weighted estimating equation:

$$U_k^I(\beta) = \sum_{l=1}^{L}\sum_{i=1}^{n_l}\int_0^\tau \psi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}\{dN_{lik}(t) - \beta^T Z_{lik}(t)Y_{lik}(t)dt\} = 0, \qquad (5.3)$$

where

$$\bar{Z}_k^I(t) = \sum_{l=1}^{L}\sum_{i=1}^{n_l}\psi_{lik}(t)Z_{lik}(t)Y_{lik}(t) / \sum_{l=1}^{L}\sum_{i=1}^{n_l}\psi_{lik}(t)Y_{lik}(t)$$

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

$$\psi_{lik}(t) = \{1 - \prod_{j=1}^{K}(1-\Delta_{lij})\} + \prod_{j=1}^{K}(1-\Delta_{lij})\xi_{li}\widetilde{\alpha}_{lk}^{-1}(t) \qquad (5.4)$$

where $\widetilde{\alpha}_{lk}(t) = \sum_{i=1}^{n_l}\xi_i\{\prod_{j=1}^{K}(1-\Delta_{lij})\}Y_{lik}(t)/\sum_{i=1}^{n_l}\{\prod_{j=1}^{K}(1-\Delta_{lij})\}Y_{lik}(t)$. The explicit form of $\tilde{\beta}^I$ which is defined by the solution of the estimating equation (5.3) is following:

$$\tilde{\beta}^I = \left[\sum_{l=1}^{L}\sum_{i=1}^{n_l}\int_0^\tau \psi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}^{\otimes 2}Y_{lik}(t)dt\right]^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\int_0^\tau\{Z_{lik}(t) - \bar{Z}_k^I(t)\}dN_{lik}(t).$$

In the situation that two case-cohort studies were conducted using the same subcohort for disease 1 and disease 2, respectively, covariate information are available for the subcohort members as well as subjects with disease 1 and/or disease 2. If we are interested in estimating the covariate effect for disease 1, the time-varying weight function from the existing method is $\rho_{li1}(t) = \Delta_{li1} + (1-\Delta_{li1})\xi_{li}\hat{\alpha}_{1l}^{-1}(t) = 1$ when $\Delta_{li1} = 1$ and $\rho_{li1} = \hat{\alpha}_{l1}^{-1}(t)$ when $\Delta_{li1} = 0$ and $\xi_{li} = 1$, regardless of disease 2 information. Therefore, the existing weight function does not use information collected on subjects with disease 2. On the other hand, our proposed weight function for disease 1 is $\psi_{li1}(t) = \{1 - \prod_{j=1}^{2}(1-\Delta_{lij})\} + \prod_{j=1}^{2}(1-\Delta_{lij})\xi_{li}\widetilde{\alpha}_{1l}^{-1}(t) = 1$ when $\Delta_{li1} = 1$ or $\Delta_{li2} = 1$ and $\psi_{li1}(t) = \widetilde{\alpha}_{1l}^{-1}(t)$ when $\Delta_{li1} = 0$, $\Delta_{li2} = 0$, and $\xi_{li} = 1$. This weight function takes disease 2 information into consideration. Note that $\widetilde{\alpha}_{lk}(t)$, which is an estimator of the true sampling probability $\tilde{\alpha}$, is the proportion of sampled subjects among those who do not have any diseases in stratum l and are remaining in the risk set at time t. When estimating the effect of risk factors on a disease, the proposed weight uses covariate information collected on subjects with other failure events.

Let $\Lambda_{0k}(t) = \int_0^\tau \lambda_{0k}(s)ds$. We propose to estimate $\Lambda_{0k}(t)$ by a Breslow-Aalen type estimator $\tilde{\Lambda}_{0k}^I(\tilde{\beta}^I, t)$, where

$$\tilde{\Lambda}_{0k}^I(\beta, t) = \int_0^\tau \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \psi_{lik}(u)\{dN_{lik}(u) - Y_{lik}(u)\beta^T Z_{lik}(u)du\}}{\sum_{l=1}^L \sum_{i=1}^{n_l} \psi_{lik}(u)Y_{lik}(u)}. \tag{5.5}$$

If the disease of interest is common, then the generalized case-cohort design is more appropriate than the traditional case-cohort design. We can extend our approach to the generalized stratified case-cohort design. We consider the following weight function $\pi_{lik}(t)$ with two types of diseases (i.e. $K = 2$):

$$\begin{aligned}
\pi_{lik}(t) &= \Pi_{j=1}^2(1-\Delta_{lij})\xi_{li}\tilde{\alpha}_{lk}^{-1}(t) + \left\{1 - \Pi_{j=1}^2(1-\Delta_{lij})\right\}\xi_{li} \\
&+ \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li1}\tilde{\gamma}_{l1k}^{-1}(t) + (1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\eta_{li2}\tilde{\gamma}_{l2k}^{-1}(t) \\
&+ \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}\tilde{\gamma}_{l3k}^{-1}(t) + \frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}\tilde{\gamma}_{l4k}^{-1}(t),
\end{aligned} \tag{5.6}$$

where

$$\begin{aligned}
\tilde{\alpha}_{lk}(t) &= \sum_{i=1}^{n_l} \Pi_{j=1}^2(1-\Delta_{lij})\xi_{li}Y_{lik}(t)/\{\sum_{i=1}^{n_l}\Pi_{j=1}^2(1-\Delta_{lij})Y_{lik}(t)\} \\
\tilde{\gamma}_{l1k}(t) &= \sum_{i=1}^{n_l} \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\eta_{li1}Y_{lik}(t)/\{\sum_{i=1}^{n_l}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\} \\
\tilde{\gamma}_{l2k}(t) &= \sum_{i=1}^{n_l} (1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\eta_{li2}Y_{lik}(t)/\{\sum_{i=1}^{n_l}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\} \\
\tilde{\gamma}_{l3k}(t) &= \sum_{i=1}^{n_l} \Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li1}Y_{lik}(t)/\{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\} \\
\tilde{\gamma}_{l4k}(t) &= \sum_{i=1}^{n_l} \Delta_{li1}\Delta_{li2}(1-\xi_{li})\eta_{li2}Y_{lik}(t)/\{\sum_{i=1}^{n_l}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\}.
\end{aligned}$$

For the generalized case-cohort data, we can construct the weighted estimating functions and the estimator for the baseline cumulative hazard function by replacing the weight function $\psi_{lik}(t)$ with $\pi_{lik}(t)$ in (5.3). The explicit form of $\tilde{\beta}_G^I$ which is defined by the solution of the estimating equation with a weight function $\pi_{lik}(t)$ is following:

$$\tilde{\beta}_G^I = \left[\sum_{l=1}^{L}\sum_{i=1}^{n_l}\int_0^\tau \pi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}^{\otimes 2} Y_{lik}(t)dt\right]^{-1}$$
$$\times \quad \sum_{l=1}^{L}\sum_{i=1}^{n_l}\int_0^\tau \pi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}dN_{lik}(t).$$

5.2.2 Estimation for multivariate failure time

Suppose that there are $n = \Sigma_{l=1}^{L} n_l$ independent subjects with K diseases of interest. Let independent failure time vector be $T_{li} = (T_{li1}, \ldots, T_{lik})$ and the observed time vector be $X_{li} = (X_{li1}, \ldots, X_{lik},)$ $i = 1, \ldots, n$. Thus, for subject i in stratum l complete observations are $(X_{lik}, \Delta_{lik},\ \xi_{li}, Z_{lik}(t), 0 \le t \le \tau, k = 1, \ldots, K, V_i)$ when $\xi_{li} = 1$ or $\Delta_{lik} = 1$ and $(X_{lik}, \Delta_{lik}, \xi_i, k = 1, \ldots, K, V_i)$ when $\xi_{li} = 0$ and $\Delta_{lik} = 0$.

For traditional stratified case-cohort data with K rare diseases, we consider the estimating equation

$$U^{II}(\beta) = \sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^\tau \psi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}\{dN_{lik}(t) - \beta^T Z_{lik}(t)Y_{lik}(t)dt\} = 0, \quad (5.7)$$

with $\psi_{lik}(t)$ defined as in (5.4).

The estimator of the hazards regression parameter β_0, $\tilde{\beta}^{II}$, is defined as the solution to (5.7) which has the following explicit form:

$$\tilde{\beta}^{II} = \left[\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^\tau \psi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^I(t)\}^{\otimes 2} Y_{lik}(t)dt\right]^{-1} \sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^\tau \{Z_{lik}(t) - \bar{Z}_k^I(t)\}dN_{lik}(t).$$

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

$$\tilde{\Lambda}_{0k}^{II}(\beta, t) = \int_0^\tau \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\psi_{lik}(u)\{dN_{lik}(u) - Y_{ik}(u)\beta^T Z_{lik}(u)du\}}{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\psi_{lik}(u)Y_{lik}(u)}. \quad (5.8)$$

Under the generalized case-cohort design, the estimating equation and estimator of the baseline cumulative hazard function are the same as those in (5.7) and (5.8) replacing $\psi_{lik}(t)$

by $\pi_{lik}(t)$ defined as in (5.6). The estimator $\tilde{\beta}_G^{II}$ has the explicit form:

$$\begin{aligned}\tilde{\beta}_G^{II} &= \left[\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^{\tau}\pi_{ik}(t)\{Z_{lik}(t)-\bar{Z}_k^I(t)\}^{\otimes 2}Y_{lik}(t)dt\right]^{-1}\\ &\times \sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{K}\int_0^{\tau}\pi_{lik}(t)\{Z_{lik}(t)-\bar{Z}_k^I(t)\}dN_{lik}(t).\end{aligned}$$

5.3 Asymptotic properties

5.3.1 Asymptotic properties of $\tilde{\beta}_G^{II}$ and $\tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_G^{II}, t)$

In this section, we will study the asymptotic properties of the proposed methods. Since the estimators for the univariate failure time are a special case of those for the multivariate failure time and the traditional case-cohort study is a special case of the generalized case-cohort study, we will only present the results for the multivariate case for the generalized case-cohort study. We make the following assumptions:

(a) $\{T_{li}, C_{li}, Z_{li}\}$, $i = 1, \ldots, n$ and $l = 1, \ldots, L$ are independent and identically distributed where $T_{li} = (T_{li1}, \ldots, T_{liK})^T$, $C_{li} = (C_{li1}, \ldots, C_{liK})^T$, and $Z_{li} = (Z_{li1}, \ldots, Z_{liK})^T$;

(b) $\mathrm{P}\{Y_{lik}(t) = 1\} > 0$ for $t \in [0, \tau]$, $i = 1, \ldots, n_l$, $k = 1, 2$, and $L = 1, \ldots, L$;

(c) $|Z_{lik}(0)| + \int_0^{\tau} |dZ_{lik}(t)| < D_z < \infty$, $i = 1, \ldots, n_l$, $k = 1, 2$, and $L = 1, \ldots, L$ almost surely and D_z is a constant;

(d) The matrix A_k is positive definite for $k = 1, 2$ where $A_k = \Sigma_{l=1}^{L} q_l E_l(\int_0^{\tau} Y_{l1k}(t)\{Z_{l1k}(t)^{\otimes 2} - [E\{Y_{l1k}(t)Z_{l1k}(t)\}/E\{Y_{l1k}(t)\}]^{\otimes 2}\}dt)$ where $q_l = \lim_{n\to\infty} n_l/n$;

(e) For all $k = 1, 2$, $\int_0^{\tau} \lambda_{0k}(t)dt < \infty$;

To show the desired asymptotic properties for generalized case-cohort samples, the following conditions are also needed:

(f) For all $l = 1, \ldots, L$, $\lim_{n\to\infty} \tilde{\alpha}_l = \alpha_l$, where $\tilde{\alpha}_l = \tilde{n}_l/n_l$ and α_l is a positive constant.

(g) $\lim_{n\to\infty} \tilde{\gamma}_{l1k} = \lim_{n\to\infty} \tilde{\gamma}_{l3k} = \gamma_{l1}, \lim_{n\to\infty} \tilde{\gamma}_{l2k} = \lim_{n\to\infty} \tilde{\gamma}_{l4k} = \gamma_{l2}$ where $\tilde{\gamma}_{l1k} = \Pr[\eta_{li1} = 1|\Delta_{li1} = 1, \Delta_{li2} = 0, \xi_{li} = 0] = \tilde{m}_{l10}/(n_{1l0} - \tilde{n}_{l10})$, $\tilde{m}_{ljk}$ denotes the number of sampled diseased subjects in non-subcohort with ($\Delta_{l1} = j$ and $\Delta_{l2} = k$), n_{ljk} and $\tilde{n}_{ljk}$ denote the number of subjects with diseases ($\Delta_{l1} = j$ and $\Delta_{l2} = k$) in the cohort and the subcohort in stratum l, respectively, $\tilde{\gamma}_{l2k} = \tilde{m}_{l01}/(n_{l01} - \tilde{n}_{l01})$, $\tilde{\gamma}_{l3k} = \tilde{m}_{11}^{l1}/(n_{l11} - \tilde{n}_{l11})$, $\tilde{\gamma}_{l4k} = \tilde{m}_{l11}^{2}/(n_{l11} - \tilde{n}_{l11})$, and γ_{lk} is a positive constant on (0,1] for all $k = 1, 2$ and $l = 1, \ldots, L$.

(h) $\lim_{n\to\infty} n_{lk}/n_l = p_{lk}$, where p_{lk} is a positive constant on [0,1] for all $k = 1, 2$ and $l = 1, \ldots, L$.

(i) $\lim_{n\to\infty} n_l/n = q_l$, where q_l is a positive constant on [0,1] for all $l = 1, \ldots, L$.

The following theorems summarize the main results. Here is the asymptotic properties for $\tilde{\beta}_G^{II}$.

Theorem 6. *Under the regularity conditions (a)-(i), $\tilde{\beta}_G^{II}$ converges in probability to β_0 and $n^{1/2}(\tilde{\beta}_G^{II} - \beta_0)$ converges in distribution to a mean zero normal distribution with covariance matrix $A(\beta_0)^{-1}\sum_{l=1}^{L}\Sigma^{GII}(\beta_0)A(\beta_0)^{-1}$, where*

$$A(\beta) = \sum_{k=1}^{K} A_k(\beta), \quad \Sigma^{GII}(\beta) = \sum_{l=1}^{L} q_l[V_{I,l}^a(\beta) + \frac{1-\alpha_l}{\alpha_l}V_{II,l}^a(\beta) + (1-\alpha_l)\sum_{k=1}^{2} V_{III,lk}^a(\beta)],$$

$$V_{I,l}^a(\beta) = E_l[\sum_{k=1}^{2} Q_{l1k}(\beta)]^{\otimes 2},$$

$$V_{II,l}^a(\beta) = Var_l\left[\prod_{j=1}^{2}(1-\Delta_{lij})\sum_{k=1}^{2}\int_0^\tau [B_{l1k}(\beta,t) - Y_{lik}(t)\frac{E[\prod_{j=1}^{2}(1-\Delta_{lij})B_{l1k}(\beta_0,t)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{lij})Y_{l1k}(t)]}dt]\right],$$

$$V_{III,lk}^a(\beta) = Pr[\theta_{l10}]\frac{1-\gamma_{l1}}{\gamma_{l1}}Var_l\left[Q_{l1k}(\beta) - \int_0^\tau Y_{l1k}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l10}]}|\theta_{l10},\xi_{l1}=0\right]$$
$$+Pr[\theta_{l01}]\frac{1-\gamma_{l2}}{\gamma_{l2}}Var_l\left[Q_{l1k}(\beta) - \int_0^\tau Y_{l1k}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l01}]}|\theta_{l01},\xi_{l1}=0\right]$$
$$+\frac{Pr[\theta_{l11}]}{4}[\frac{1-\gamma_{l1}}{\gamma_{l1}} + \frac{1-\gamma_{l2}}{\gamma_{l2}}]Var_l\left[Q_{l1k}(\beta) - \int_0^\tau Y_{l1k}(t)\frac{E[dQ_{l1k}(\beta,t)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]}|\theta_{l11},\xi_{l1}=0\right],$$

$$Q_{lik}(t,\beta) = \int_0^t \{Z_{lik}(t) - e_k(t)\}dM_{lik}(t),$$

$$B_{lik}(t,\beta) = \{Z_{l1k}(t) - e_k(t)\}Y_{l1k}(t)(\lambda_{0k}(t) + \beta^T Z_{lik}(t)),$$

$$e_k(t) = \frac{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(t)Z_{l1k}(t)]}{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(t)]}, \ \theta_{ljk} = \{\Delta_{li1} = j \ and \ \Delta_{li2} = k\}.$$

Note that $\Sigma_l^{GII}(\beta_0)$ consists of three parts. The first part $V_{I,l}^a(\beta_0)$ is a contribution to the variance from the full cohort, the second part $V_{II,l}^a(\beta_0)$ is due to sampling subcohort from the full cohort, and the last one $V_{III,l}^a(\beta_0)$ is due to sampling a fraction of all cases. If we select all cases, which is the traditional stratified case-cohort study, the last variance goes to zero.

We summarize the asymptotic properties of the proposed baseline cumulative hazard estimator $\tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_G^{II}, t)$ in the next theorem.

Theorem 7. *Under the regularity conditions (a)-(i), $\tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_G^{II}, t)$ is a consistent estimator of $\Lambda_{0k}(t)$ in $t \in [0,\tau]$ and $G(t) = \{G_1(t), \ldots, G_K(t)\}^T = [n^{1/2}\{\tilde{\Lambda}_{01}^{II}(\tilde{\beta}_G^{II}, t) - \Lambda_{01}(t)\}, n^{1/2}\{\tilde{\Lambda}_{02}^{II}(\tilde{\beta}_G^{II}(t) - \Lambda_{02}(t))\}]^T$ converges weakly to the Gaussian process $\mathcal{G}(t) = \{\mathcal{G}_1(t), \mathcal{G}_2(t)\}^T$ in $D[0,\tau]^K$ with mean zero and the following covariance function $\mathcal{G}_{jk}(t,s)$ between $\mathcal{G}_j(t)$ and $\mathcal{G}_k(s)$ for $j \neq k$.*

$$\begin{aligned}\mathcal{G}_{jk}(t,s)(\beta_0) &= \sum_{l=1}^{L} q_l[E_l\{\mu_{l1j}(\beta_0,t)\mu_{l1k}(\beta_0,s)\} + \frac{1-\alpha_l}{\alpha_l}E_l\{w_{l1j}(\beta_0,t)w_{l1k}(\beta_0,s)\} \\ &+ E_l\{\nu_{l1j}(\beta_0,t)\nu_{l1k}(\beta_0,s)\}],\end{aligned}$$

where the explicit forms of μ_{lik},w_{lik}, and $\nu_{lik}(\beta,t)$ are given in Appendix.

The proof of $\tilde{\Lambda}_{0k}^{G}(\tilde{\beta}_G^{II}, t)$ is provided in Appendix. The proof uses Taylor expansion, Kolmogorov-Centsov theorem, weak convergence of the baseline cumulative hazard estimator from full cohort studies with multivariate failure time, Hájek [1960]'s central limit theorem for finite population sampling, and Cramer-Wold device.

5.3.2 Proofs of Theorems

Proof of Theorem 6

We first show the consistency of $\tilde{\beta}_G^{II}$. Denote $\widetilde{U}_n^G = n_l^{-1}\widetilde{U}^G$. By Taylor expansion series, $\tilde{\beta}_G^{II}$ can be written as

$$\tilde{\beta}_G^{II} = \beta_0 + \left[-\frac{\partial \widetilde{U}_n^G(\beta_0)}{\partial \beta_0}\right]^{-1} \widetilde{U}_n^G(\beta_0) + o_p(1) \tag{5.9}$$

Based on the extension of Fourtz [1977], if the following conditions are satisfied

(I) $\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0,

(II) $\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ is negative definite with probability going to one as n $\rightarrow \infty$,

(III) -$\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T}$ converges to $A(\beta_0)$ in probability uniformly for β in an open neighborhood about β_0,

(IV) $\widetilde{U}_n^G(\beta)$ converges to 0 in probability,

then, we can show that $\tilde{\beta}_G^{II}$ converges to β_0 in probability. One can write

$$
\begin{aligned}
-\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T} &= \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\{Z_{lik}(t) - \bar{Z}_k^{II}(t)\}Z_{lik}(t)Y_{lik}(t)dt \\
&= \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)Z_{lik}(t)\}dt \\
&= \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \qquad (5.10)
\end{aligned}
$$

Since (5.10) is constant with respect to β , (I) is satisfied. In order to show that (II) and (III) are satisfied, we need to show uniform convergence of $\bar{Z}_k^{II}(t)$ to $e_k(t)$ such that $\sup_{t\in[0,\tau]} \parallel \bar{Z}_k^{II}(t) - e_k(t) \parallel \xrightarrow{p} 0$ as $n \rightarrow \infty$ for $k = 1, 2$. It is sufficient to show that

$$
\sup_{t\in[0,\tau]} \parallel n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d} - n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)Z_{lik}(t)^{\otimes d} \parallel \xrightarrow{p} 0 \text{as } n \rightarrow \infty \text{ for } d = 0, 1.
$$

It can be written as

$$
\begin{aligned}
& n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d} - n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)Z_{lik}(t)^{\otimes d} \\
= \; & n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\xi_{li}}{\tilde{\alpha}_l} - 1]\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)Z_{lik}(t)^{\otimes d} \\
- \; & n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\alpha}_l^{-1} - \tilde{\alpha}_{lk}(t)^{-1}]\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)Z_{lik}(t)^{\otimes d} \\
+ \; & n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\eta_{li1}}{\tilde{\gamma}_{l1k}} - 1]\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}
\end{aligned}
$$

$$
\begin{aligned}
&- \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\gamma}_{l1}^{-1}-\tilde{\gamma}_{l1k}(t)^{-1}]\eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&+ \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1](1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&- \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\gamma}_{l2}^{-1}-\tilde{\gamma}_{l2k}(t)^{-1}]\eta_{li2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&+ \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1]\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&- \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\tilde{\gamma}_{l3}^{-1}-\tilde{\gamma}_{l3k}(t)^{-1}]\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&+ \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1]\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\\
&- \quad n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\tilde{\gamma}_{l4}^{-1}-\tilde{\gamma}_{l4k}(t)^{-1}]\eta_{li2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}
\end{aligned}
$$

Then, one can write

$$
\begin{aligned}
& \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d}-n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
\leq & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\xi_{li}}{\tilde{\alpha}_l}-1]\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\alpha}_l^{-1}-\tilde{\alpha}_{lk}(t)^{-1}]\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\eta_{li1}}{\tilde{\gamma}_{l1k}}-1]\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\gamma}_{l1}^{-1}-\tilde{\gamma}_{l1k}(t)^{-1}]\eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1](1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}[\tilde{\gamma}_{l2}^{-1}-\tilde{\gamma}_{l2k}(t)^{-1}]\eta_{li2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1]\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\tilde{\gamma}_{l3}^{-1}-\tilde{\gamma}_{l3k}(t)^{-1}]\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|\\
+ & \quad \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}[\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1]\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}\ \|
\end{aligned}
$$

$$+ \quad \| n^{-1} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \frac{1}{2} [\tilde{\gamma}_{l4}^{-1} - \tilde{\gamma}_{l4k}(t)^{-1}] \eta_{li2} \Delta_{li1} \Delta_{li2} (1 - \xi_{li}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} \| \tag{5.11}$$

Based on condition (c), the total variation of $\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)Z_{lik}(t)^{\otimes d}$, $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}$, $(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}$, and $\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)Z_{lik}(t)^{\otimes d}$ are finite on $[0, \tau]$. By applying lemma 2, the first, third, fifth, seventh, and ninth terms in (5.11) converge to zero in probability uniformly in t.

Note that $\tilde{\alpha}_l^{-1} - \tilde{\alpha}_{lk}(t)^{-1}$ converges to zero in probability uniformly in t by lemma 2 since $\prod_{j=1}^{2}(1 - \Delta_{lij})Y_{lik}(t)$ is bounded variation and $E_l[\prod_{j=1}^{2}(1 - \Delta_{l1j})Y_{l1k}(t)]$ is bounded away from zero. Similarly, $\tilde{\gamma}_{l1}^{-1} - \tilde{\gamma}_{l1k}(t)^{-1}$, $\tilde{\gamma}_{l2}^{-1} - \tilde{\gamma}_{l2k}(t)^{-1}$, $\tilde{\gamma}_{l3}^{-1} - \tilde{\gamma}_{l3k}(t)^{-1}$, and $\tilde{\gamma}_{l4}^{-1} - \tilde{\gamma}_{l4k}(t)^{-1}$ can be shown to converge to zero in probability uniformly in t, respectively. By lemma 2, $\frac{1}{n_l}\sum_{i=1}^{n_l}\xi_{li}\prod_{j=1}^{2}(1 - \Delta_{lij})Y_{lik}(t) \parallel Z_{lik}(t)^{\otimes d} \parallel$ converges to $\alpha E_l[\prod_{j=1}^{2}(1 - \Delta_{l1j})Y_{l1k}(t) \parallel Z_{l1k}(t)^{\otimes d} \parallel]$ in probability uniformly in t. Thus, the second, fourth, sixth, eighth, and tenth terms in (5.11) converge to zero in probability uniformly in t, respectively. Combining all the above results, $n^{-1}(\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d} - n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)Z_{lik}(t)^{\otimes d})$ converges to zero in probability uniformly in t as $n \to \infty$ for $d = 0, 1$.

Since $Y_{lik}(t)Z_{lik}(t)^{\otimes d}$ is bounded variation based on condition (c), $n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}Y_{lik}(t)$ $Z_{lik}(t)^{\otimes d}$ converges to $\sum_{l=1}^{L}q_l E_l[Y_{l1k}(t)Z_{lik}(t)^{\otimes d}]$. Therefore, it can be shown that

$$\sup_{t\in[0,\tau]} \| n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)Z_{lik}(t)^{\otimes d} - \sum_{l=1}^{L}q_l E_l[Y_{l1k}(t)Z_{lik}(t)^{\otimes d}] \| \xrightarrow{p} 0 \text{ as } n \to \infty \text{ for } d = 0, 1.$$

Since $\sum_{l=1}^{L}q_l E_l[Y_{l1k}(t)]$ is bounded away from zero based on condition (b), $\bar{Z}_k^{II}(t)$ can be shown to converge to $e_k(t)$ in probability uniformly in t as $n \to \infty$ for $k = 1, 2$. One can write

$$\begin{aligned} -\frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta^T} &= \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\pi_{lik}(t)Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\ &= \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\ &+ \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\pi_{lik}(t) - 1)Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt. \end{aligned} \tag{5.12}$$

Note that the first term in (5.12) converges to $A = \sum_{l=1}^{L} q_l E_l[\sum_{k=1}^{2} \int_0^\tau \pi_{l1k}(t) Y_{l1k}(t)\{Z_{l1k}(t)^{\otimes 2} - e_k(t)^{\otimes 2}\}dt]$ where $q_l = \lim_{n\to\infty} n_l/n$ in probability as $n \to \infty$ by the uniform convergence of $\bar{Z}_k^{II}(t)$ to $e_k(t)$.

Now we will show that the second term converges to zero in probability uniformly in t. The second term in (5.12) can be written as

$$
\begin{aligned}
& \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\frac{\xi_{li}}{\tilde{\alpha}_l} - 1)\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau [\tilde{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1}]\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\frac{\eta_{li1}}{\tilde{\gamma}_{l1}} - 1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau [\tilde{\gamma}_{l1k}(t)^{-1} - \tilde{\gamma}_{l1}^{-1}]\eta_{li1}(1-\xi_{li})\Delta_{li1}(1-\Delta_{li2})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\frac{\eta_{li2}}{\tilde{\gamma}_{l2}} - 1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau [\tilde{\gamma}_{l2k}(t)^{-1} - \tilde{\gamma}_{l2}^{-1}]\eta_{li2}(1-\xi_{li})(1-\Delta_{li1})\Delta_{li2}Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{2n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}} - 1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{2n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau [\tilde{\gamma}_{l3k}(t)^{-1} - \tilde{\gamma}_{l3}^{-1}]\eta_{li1}(1-\xi_{li})\Delta_{li1}\Delta_{li2}Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{2n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau (\frac{\eta_{li2}}{\tilde{\gamma}_{l4}} - 1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt \\
+ \quad & \frac{1}{2n}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau [\tilde{\gamma}_{l4k}(t)^{-1} - \tilde{\gamma}_{l4}^{-1}]\eta_{li2}(1-\xi_{li})\Delta_{li1}\Delta_{li2}Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - \bar{Z}_k^{II}(t)^{\otimes 2}\}dt
\end{aligned}
\tag{5.13}
$$

By the uniform convergence of $\bar{Z}_k^{II}(t)$ to $e_k(t)$, the first term in (5.13) is asymptotically equivalent to $n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}dt$. Similarly, the third term, the fifth term, seventh term, and ninth term in (5.13) are asymptotically equivalent to $n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2} - e_k(t)^{\otimes 2}\}dt$, $n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}dt$, $(2n)^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}dt$, and $(2n)^{-1}\sum_{l=1}^{L}$

$\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)(1-\xi_{li})\Delta_{li1}\Delta_{li2}Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}dt$, respectively.

Based on condition (c), $\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}$, $\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}$, $(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}$, $\Delta_{li1}\Delta_{li2}(1-\xi_{li})Y_{lik}(t)\{Z_{lik}(t)^{\otimes 2}-e_k(t)^{\otimes 2}\}$ are of bounded variations and they are independent and identically distributed. It follows from lemma 2 that the first term, the third term, the fifth term, seventh term, and ninth term can be shown to converge to zero in probability uniformly in t, respectively.

Since $\tilde{\alpha}_{lk}(t)^{-1}-\tilde{\alpha}_l^{-1}$, $\tilde{\gamma}_{l1k}(t)^{-1}-\tilde{\gamma}_{l1}^{-1}$, $\tilde{\gamma}_{l2k}(t)^{-1}-\tilde{\gamma}_{l2}^{-1}$, $\tilde{\gamma}_{l3k}(t)^{-1}-\tilde{\gamma}_{l3}^{-1}$, and $\tilde{\gamma}_{l4k}(t)^{-1}-\tilde{\gamma}_{l4}^{-1}$ converge to zero in probability uniformly respectively and $\bar{Z}_k^{II}(t)$ converges to $e_k(t)$ in probability uniformly in t, we can show that the second, fourth, sixth, eighth, and tenth terms converge to zero in probability uniformly in t respectively.

Combining all the results, we have

$$-\frac{\partial\widetilde{U}_n^G(\beta)}{\partial\beta^T}\xrightarrow{p}A \text{ as } n\to\infty$$

, and, thus, (II) and (III) are satisfied.

Now, $n^{1/2}\widetilde{U}_n^G(\beta)$ can be decomposed into four parts:

$$\begin{aligned} n^{1/2}\widetilde{U}_n^G(\beta) &= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\{Z_{lik}(t)-\bar{Z}_k^{II}(t)\}dM_{lik}(t)\\ &= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \pi_{lik}(t)\{Z_{lik}(t)-e_k(t)+e_k(t)-\bar{Z}_k^{II}(t)\}dM_{lik}(t)\\ &= n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\ &+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \{\pi_{lik}(t)-1\}\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\ &+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \{e_k(t)-\bar{Z}_k^{II}(t)\}dM_{lik}(t)\\ &+ n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau \{\pi_{lik}(t)-1\}\{e_k(t)-\bar{Z}_k^{II}(t)\}dM_{lik}(t) \qquad (5.14)\end{aligned}$$

Since the first term in (5.14) is the pseudo partial likelihood score function for the full likelihood, it is asymptotically zero-mean normal with covariance $V_{I,l}(\beta_0)=\sum_{l=1}^{L}q_lE_l[\sum_{k=1}^{2}Q_{l1k}(\beta_0)]^{\otimes 2}$

where $Q_{lik}(t, \beta) = \int_0^t \{Z_{lik}(t) - e_k(t)\} dM_{lik}(t)$ [Yin and Cai, 2004].

The third term can be shown to converge to zero. Note that for fixed t, $M_{l1k}(t), \ldots, M_{lnk}(t)$ are identically and independent distributed zero-mean random variables and $\sum_{i=1}^{n_l} M_{lik}(t)$ is sum of identically and independently distributed zero-mean random variables.

Since $M_{lik}^2(0) < \infty$ and $M_{lik}^2(\tau) < \infty$ are satisfied based on condition (c) and (e), $M_{lik}(t)$ is of bounded variation and therefore it can be written as a difference of two monotone functions in t. From the example of 2.11.16 of van der Vaart and Wellner [1996](p215), $n_l^{-1/2} \sum_{i=1}^{n_l} M_{lik}(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{P}_{M,lk}(t)$.

To establish the existence of stochastic processes with continuous sample paths, we will use Kolmogorov-Centsov theorem. If conditions of Kolmogorov-Centsov theorem $E\{\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s)\}^4 \leq C_z^* |t-s|^2$ and $E\{\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s)\}^2 \leq C|t-s|$ for all $t \geq s$ are satisfied, then we can show that $\mathcal{P}_{M,lk}(t)$ has continuous sample paths. Note that $\mathrm{E}\mathcal{P}_{M,lk}(t)^2 = \mathrm{E}[n_l^{-1} \sum_{i=1}^{n_l} M_{lik}(t)^2] = \mathrm{E}M_{lik}(t)^2 = \mathrm{E}[\int_0^t Y_{lik}(u)(\lambda_{0k}(u)du + \beta_0^T Z_{lik}(u))du]$, and $\mathrm{E}\{\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s)\}^2 = \mathrm{E}\mathcal{P}_{M,lk}(t)^2$ - $\mathrm{E}\mathcal{P}_{M,lk}(s)^2 = \mathrm{E}[\int_s^t Y_{lik}(u)(\lambda_{0k}(u)du + \beta_0^T Z_{lik}(u))du]$. Based on condition (c), (e), $\lambda_{0k}(.)$ and $\beta_0^T Z_{lik}(.)$ are of bounded variations on $[0, \tau]$. Thus, it follows from mean value theorem that there exists a constant C such that $\mathrm{E}[\int_s^t Y_{lik}(u)(\lambda_{0k}(u)du + \beta_0^T Z_{lik}(u))du] \leq C(t-s)$ for $s \leq t$. Hence, $\mathrm{E}[\{\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s)\}^2] \leq C(t-s)$ and $\mathrm{E}[\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s)\}^4] = \mathrm{Var}(\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s))^2 + \mathrm{E}\{(\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s))^2\}^2 = 3 \{E(\mathcal{P}_{M,lk}(t) - \mathcal{P}_{M,lk}(s))^2\}^2 \leq C^* |t-s|^2$ for some constant C^*. Since two conditions are satisfied, it follows that $\mathcal{P}_{M,lk}(t)$ has continuous sample path from Kolmogorov-Centsov theorem. Based on conditions (b) and (c), it can be shown that $n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(t) Y_{lik}(t) Z_{lik}(t)$ and $n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(t) Y_{lik}(t)$ are of bounded variations and specially $n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(t) Y_{lik}(t)$ is bounded away from zero. Thus $\bar{Z}_k^{II}(t)$ is of bounded variation and can be written as $\bar{Z}_k^{II}(t) = G_{k1} - G_{k1}$ where both G_{k1} and G_{k2} are nonnegative, monotone functions in t, and bounded. Therefore, $\bar{Z}_k^{II}(t)$ is the sum of two monotone functions. By Lemma 1, the third term in (5.14) converges to zero in probability uniformly in t as $n \to \infty$.

By similar arguments, the fourth term in (5.14) converges to zero in probability uniformly since $\pi_{lik}(t) - 1$ is of bounded variation.

Now, the second term in (5.14) can be written as

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\{\pi_{lik}(t)-1\}\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
= \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\prod_{j=1}^{2}(1-\Delta_{lij})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\tilde{\alpha}_{lk}(t)^{-1}-\tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\tilde{\gamma}_{l1k}(t)^{-1}-\tilde{\gamma}_{l1k}^{-1})\eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}(\tilde{\gamma}_{l2k}(t)^{-1}-\tilde{\gamma}_{l2k}^{-1})\eta_{li2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\int_0^{\tau}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\int_0^{\tau}(\tilde{\gamma}_{l3k}(t)^{-1}-\tilde{\gamma}_{l3k}^{-1})\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\int_0^{\tau}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)\\
+ \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\int_0^{\tau}(\tilde{\gamma}_{l4k}(t)^{-1}-\tilde{\gamma}_{l4k}^{-1})\eta_{li2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)
\end{aligned}
\tag{5.15}
$$

Using the result of (4.6), the second term in (5.15) can be written as

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^{\tau}\frac{1}{\tilde{\alpha}_l E_l(\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t))}\frac{1}{n_l}\left\{\sum_{m=1}^{n_l}(1-\frac{\xi_{lm}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lmj})Y_{lmk}(t)\right\}\\
\times \;& \xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})\{Z_{lik}(t)-e_k(t)\}dM_{lik}(t)+o_p(1)\\
= \;& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\prod_{j=1}^{2}(1-\Delta_{lij})(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\\
\times \;& \sum_{k=1}^{2}\int_0^{\tau}\frac{Y_{lik}(t)n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})\{Z_{lmk}(t)-e_k(t)\}dM_{lmk}(t)}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t))]}+o_p(1)
\end{aligned}
$$

It follows from Glivenko-Cantelli lemma and Lemma 2 that $n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})\{Z_{lmk}(t)-e_k(t)\}dM_{lmk}(t)$ can be written as

$$
\begin{aligned}
& n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})\{Z_{lmk}(t)-e_k(t)\}dM_{lmk}(t) \\
= \; & n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})\{Z_{lmk}(t)-e_k(t)\}\{dN_{lmk}(t)-Y_{lmk}(t)(\lambda_{0k}(t)+\beta_0^T Z_{lmk}(t))dt\} \\
= \; & -n_l^{-1}\sum_{m=1}^{n_l}\frac{\xi_{lm}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lmj})\{Z_{lmk}(t)-e_k(t)\}Y_{lmk}(t)(\lambda_{0k}(t)+\beta_0^T Z_{lmk}(t))dt \\
\rightarrow \; & -E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})\{Z_{l1k}(t)-e_k(t)\}Y_{l1k}(t)(\lambda_{0k}(t)+\beta^T Z_{l1k}(t))]dt
\end{aligned}
$$

Since only censored observations contribute to this term, the last equality holds.

Therefore, the second term on the right-side of (5.15) is asymptotically equivalent to

$$
n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\sum_{k=1}^{2}\int_0^\tau \frac{Y_{lik}(t)E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})B_{lik}(t,\beta)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j}Y_{l1k}(t))]}dt
$$

where $B_{lik}(t,\beta)=\{Z_{lik}(t)-e_k(t)\}Y_{lik}(t)[\lambda_{0k}(t)+\beta^T Z_{lik}(t)]$ and $e_k(t)=\frac{\sum_{l=1}^{L}q_l E_l[Y_{l1k}(t)Z_{l1k}(t)]}{\sum_{l=1}^{L}q_l E_l[Y_{l1k}(t)]}$. The first term on the right-side of (5.15) is asymptotically equivalent to

$$
\begin{aligned}
& -n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\prod_{j=1}^{2}(1-\Delta_{lij})\{Z_{lik}(t)-e_k(t)\}Y_{lmk}(t)(\lambda_{0k}(t)+\beta_0^T Z_{lmk}(t))dt \\
= \; & -n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\int_0^\tau(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\prod_{j=1}^{2}(1-\Delta_{lij})B_{lik}(t,\beta)dt
\end{aligned}
$$

Combining these results, it can be shown that the first and second terms in (5.15) are asymptotically equivalent to

$$
n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\prod_{j=1}^{2}(1-\Delta_{lij})(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\int_0^\tau\left[B_{lik}(t,\beta)-Y_{lik}(t)\frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})B_{l1k}(t,\beta)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}\right]dt
$$

Using the result of (4.8), Glivenko-Cantelli lemma and Lemma 2, it can be shown that the fourth, the sixth, the eighth, and tenth terms in (5.15) are asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1) \\
\times & \int_0^\tau Y_{lik}(t)\frac{E_l[\{Z_{l1k}(t)-e_k(t)\}dM_{l1k}(t)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l10}]} \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1) \\
\times & \int_0^\tau Y_{lik}(t)\frac{E_l[\{Z_{l1k}(t)-e_k(t)\}dM_{l1k}(t)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l01}]} \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)\int_0^\tau Y_{lik}(t)\frac{E_l[\{Z_{l1k}(t)-e_k(t)\}dM_{l1k}(t)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]} \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)\int_0^\tau Y_{lik}(t)\frac{E_l[\{Z_{l1k}(t)-e_k(t)\}dM_{l1k}(t)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]}.
\end{aligned}
$$

Combining all results, the term in (5.14) is asymptotically equivalent to

$$
\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(t,\beta) \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\prod_{j=1}^{2}(1-\Delta_{lij})(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\int_0^\tau\left[B_{lik}(\beta_0,t)-Y_{lik}(t)\frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})B_{l1k}(t,\beta)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}\right]dt \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}(1-\Delta_{li2})(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l1}})\left[Q_{lik}(\beta)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l10}]}\right] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\Delta_{li1})\Delta_{li2}(1-\frac{\eta_{li2}}{\tilde{\gamma}_{l2}})\left[Q_{lik}(\beta)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l01}]}\right] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}})\left[Q_{lik}(\beta)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]}\right] \\
+ & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\frac{\eta_{li2}}{\tilde{\gamma}_{l4}})\left[Q_{lik}(\beta)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]}\right] \qquad (5.16)
\end{aligned}
$$

By Hájek [1960]'s central limit theorem and conditions (c) and (f), the second term in (5.16) is asymptotically zero-mean normal random variable with covariance matrix $\sum_{l=1}^{L} q_l \frac{1-\alpha_l}{\alpha_l} V_{II,l}(\beta_0)$ where

$$
V_{II,l}(\beta_0) = Var_l\left[\prod_{j=1}^{2}(1-\Delta_{l1j})\sum_{k=1}^{2}\int_0^\tau[B_{l1k}(\beta_0,t)-Y_{l1k}(t)\frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})B_{l1k}(t,\beta_0)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}]dt\right].
$$

It follows from Lemma 2 and Hájek [1960]'s central limit theorem that the third, fourth, and fifth terms are asymptotically zero-mean normal with covariance matrix $\sum_{l=1}^{L} q_l(1-\alpha_l)\sum_{k=1}^{2} V_{III,lk}(\beta_0)$ where

$$
\begin{aligned}
& V_{III,lk}(\beta_0) \\
= \; & Pr[\theta_{l10}]\frac{1-\gamma_{l1}}{\gamma_{l1}} Var_l\left[Q_{l1k}(\beta_0) - \int_0^\tau Y_{l1k}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l10}]}|\theta_{l10},\xi_{l1}=0\right] \\
& + Pr[\theta_{l01}]\frac{1-\gamma_{l2}}{\gamma_{l2}} Var_l\left[Q_{l1k}(\beta_0) - \int_0^\tau Y_{l1k}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l01}]}|\theta_{l01},\xi_{l1}=0\right] \\
& + \frac{Pr[\theta_{l11}]}{4}[\frac{1-\gamma_{l1}}{\gamma_{l1}} + \frac{1-\gamma_{l2}}{\gamma_{l2}}] \\
\times \; & Var_l\left[Q_{l1k}(\beta_0) - \int_0^\tau Y_{l1k}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]}|\theta_{l11},\xi_{l1}=0\right].
\end{aligned}
$$

In addition, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0)$ and $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t,\beta)dt$ where $L_{lik}(t) = B_{lik}(t,\beta) - Y_{lik}(t)\frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})B_{l1k}(t,\beta)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]}dt$ are independent since

$$
\begin{aligned}
& Cov_l\left(n_l^{-1/2}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0), n_l^{-1/2}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{1-\tilde{\alpha}_l})\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)dM_{lik}(t)\right) \\
= \; & E_l\left\{n_l^{-1/2}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{1-\tilde{\alpha}_l})\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)dM_{lik}(t)\right) \\
= \; & E_l\left\{E\left(n_l^{-1}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2}(\frac{\xi_{li}}{1-\tilde{\alpha}_l})\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)dM_{lik}(t)|\mathcal{F}(\tau)\right)\right\} \\
= \; & E_l\left\{n_l^{-1}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0)\sum_{i=1}^{n_l}\sum_{k=1}^{2} E\left((\frac{\xi_{li}}{1-\tilde{\alpha}_l})|\mathcal{F}(\tau)\right)\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)dM_{lik}(t)\right\} = 0
\end{aligned}
$$

By using same arguments, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2} Q_{lik}(\beta_0)$ and the third to the last term in (5.16) are independent. Since ξ_{li} and η_{lik} $(k=1,2)$ are independent, $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\int_0^\tau\prod_{j=1}^{2}(1-\Delta_{lij})L_{lik}(t)dM_{lik}(t)$ and the third to the last term in (5.16) are independent.

Therefore, $n^{-1/2}\tilde{U}^G(\beta_0)$ converges weakly to zero-mean normal vector with covariance matrix $\Sigma_{II}^G(\beta_0)$ where

$$\Sigma_{II}^G(\beta) = \sum_{l=1}^{L} q_l \left[V_{I,l}(\beta) + \frac{1-\alpha_l}{\alpha_l} V_{II,l}(\beta) + (1-\alpha_l) \sum_{k=1}^{2} V_{III,lk}(\beta) \right],$$

$$V_{I,l}(\beta) = E_l[\sum_{k=1}^{2} Q_{l1k}(\beta)]^{\otimes 2},$$

$$V_{II,l}(\beta) = Var_l \left[\prod_{j=1}^{2}(1-\Delta_{lij}) \sum_{k=1}^{2} \int_0^{\tau} [B_{l1k}(t,\beta) - Y_{l1k}(t) \frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j}) B_{l1k}(t,\beta)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j}) Y_{l1k}(t)]}] dt \right],$$

$$\begin{aligned} & V_{III,lk}(\beta) \\ = \; & Pr[\theta_{l10}] \frac{1-\gamma_{l1}}{\gamma_{l1}} Var_l \left[Q_{l1k}(\beta) - \int_0^{\tau} Y_{l1k}(t) \frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l10}, \xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l10}]} |\theta_{l10}, \xi_{l1}=0 \right] \\ + \; & Pr[\theta_{l01}] \frac{1-\gamma_{l2}}{\gamma_{l2}} Var_l \left[Q_{l1k}(\beta) - \int_0^{\tau} Y_{l1k}(t) \frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l01}, \xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l01}]} |\theta_{l01}, \xi_{l1}=0 \right] \\ + \; & \frac{Pr[\theta_{l11}]}{4} [\frac{1-\gamma_{l1}}{\gamma_{l1}} + \frac{1-\gamma_{l2}}{\gamma_{l2}}] \\ \times \; & Var \left[Q_{l1k}(\beta) - \int_0^{\tau} Y_{l1k}(t) \frac{E_l[dQ_{l1k}(t,\beta)|\theta_{l11}, \xi_{l1}=0]}{E_l[Y_{l1k}(t)|\theta_{l11}]} |\theta_{l11}, \xi_{l1}=0 \right]. \end{aligned}$$

Therefore, $\widetilde{U}_n^G(\beta)$ converges to zero in probability and (iv) is satisfied.

Since all conditions (i), (ii), (iii) and (iv) are satisfied, $\tilde{\beta}_G^{II}$ is a consistent estimator of β_0 by an extension of Fourtz [1977]. By consistency of $\tilde{\beta}_G^{II}$ and Taylor expansion of $\widetilde{U}_{II}^G(\beta)$ such as

$$\widetilde{U}_n^G(\beta) = \widetilde{U}_n^G(\beta_0) + \frac{\partial \widetilde{U}_n^G(\beta)}{\partial \beta}[\tilde{\beta}_G^{II} - \beta_0] + o_p(1),$$

$n^{1/2}(\tilde{\beta}_G^{II} - \beta_0)$ is asymptotically normally distributed with mean zero and with variance matrix $A^{-1}\Sigma_{II}^G(\beta_0)A^{-1}$ where $A = \sum_{k=1}^{2} A_k$.

Now, here is an outline for the proof of Theorem 7.

Proof of Theorem 7

$$\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^{II}, t) = \int_0^t \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l} \pi_{lik}(u)\{dN_{lik}(u) - Y_{lik}(u)\beta^T Z_{lik}(u) du\}}{\sum_{l=1}^{L}\sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)}$$

We can decompose $n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^{II}, t) - \Lambda_{0k}(t)\}$ into three parts:

$$
\begin{aligned}
& n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^{II}, t) - \Lambda_{0k}(t)\} \\
= \;& n^{1/2}\{\widetilde{\Lambda}_{0k}^{II}(\tilde{\beta}^{II}, t) - \widetilde{\Lambda}_{0k}^{II}(\beta_0, t) + \widetilde{\Lambda}_{0k}^{II}(\beta_0, t) - \Lambda_{0k}(t)\} \\
= \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)\{\beta_0 - \tilde{\beta}^{II}\}^T Z_{lik}(u)du}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \\
+ \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u)\{dN_{lik}(u) - Y_{lik}(u)\beta_0 Z_{lik}(u)du\}}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \\
- \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)\lambda_{0k}(u)}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \\
= \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)\{\beta_0 - \tilde{\beta}^{II}\}^T Z_{lik}(u)du}{\sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} + \int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) dM_{lik}(u)}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)}] \\
= \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)\{\beta_0 - \tilde{\beta}^{II}\}^T Z_{lik}(u)du}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} + n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} dM_{lik}(u)}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \\
+ \;& n^{1/2}\int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \{\pi_{lik}(u) - 1\} dM_{lik}(u)}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)}. \qquad (5.17)
\end{aligned}
$$

Due to the uniform convergence of $\bar{Z}_k^{II}(t)$ to $e_k(t)$, the first term in (5.17) is asymptotically equivalent to $n_l^{1/2}(\tilde{\beta}^{II} - \beta_0) l_k(t)$, where $l_k(t) = \int_0^\tau -e_k(u)du$.

Note that $[n^{-1}\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(t) Y_{lik}(t)]^{-1}$ can be written as a sum of two monotone function in t, converges to $[\sum_{l=1}^L q_l E_l[Y_{l1k}(t)]]^{-1}$ where $\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]$ is bounded away from zero, and $n^{-1/2}\sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(t)$ converges to a zero-mean Gaussian process with continuous sample path. By Lemma 1, the second term in (5.17) is asymptotically equivalent to

$$
\int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(t)]} d\{n^{1/2}\sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(u)\}.
$$

The third term in(5.17) can be written as

$$
\begin{aligned}
& \int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \{\pi_{lik}(u) - 1\} dM_{lik}(u)}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \\
= \;& n^{-1/2}\int_0^t \frac{1}{\sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{lik}(u) Y_{lik}(u)} \{\sum_{l=1}^L \sum_{i=1}^{n_l} (1 - \frac{\xi_{li}}{\tilde{\alpha}_l}) \prod_{j=1}^2 (1 - \Delta_{lij}) dM_{lik}(u) \\
+ \;& \sum_{l=1}^L \sum_{i=1}^{n_l} (\tilde{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1}) \xi_{li} \prod_{j=1}^2 (1 - \Delta_{lij}) dM_{lik}(u)
\end{aligned}
$$

$$+ \quad \sum_{l=1}^{L}\sum_{i=1}^{n_l}[(\frac{\eta_{li1}}{\tilde{\gamma}_{l1k}} - 1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li}) + \{\tilde{\gamma}_{l1k}(t)^{-1} - \tilde{\gamma}_{l1}^{-1}\}\eta_{li1}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})]dM_{lik}(u)$$
$$+ \quad \sum_{l=1}^{L}\sum_{i=1}^{n_l}[(\frac{\eta_{li2}}{\tilde{\gamma}_{l2k}} - 1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li}) + \{\tilde{\gamma}_{l2k}(t)^{-1} - \tilde{\gamma}_{l2}^{-1}\}\eta_{li2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})]dM_{lik}(u)$$
$$+ \quad \sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{1}{2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}} - 1)\Delta_{li1}\Delta_{li2}(1-\xi_{li}) + \{\tilde{\gamma}_{l3k}(t)^{-1} - \tilde{\gamma}_{l3}^{-1}\}\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})]dM_{lik}(u)$$
$$+ \quad \sum_{l=1}^{L}\sum_{i=1}^{n_l}[\frac{1}{2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4k}} - 1)\Delta_{li1}\Delta_{li2}(1-\xi_{li}) + \{\tilde{\gamma}_{l4k}(t)^{-1} - \tilde{\gamma}_{l4}^{-1}\}\eta_{li1}\Delta_{li1}\Delta_{li2}(1-\xi_{li})]dM_{lik}(u)\}. \quad (5.18)$$

Since $\{n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(t)Y_{lik}(t)\}^{-1}$ converges to $\sum_{l=1}^{L}q_l E_l[Y_{l1k}(u)]^{-1}$, where $\sum_{l=1}^{L}q_l E_l[Y_{l1k}(u)]$ is bounded away from zero in probability uniformly, the first term in (5.18) is asymptotically equivalent to $n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t \frac{Y_{lik}(t)\{\lambda_{0k}(u)+\beta_0^T Z_{lik}(u)\}du}{\sum_{l=1}^{L}q_l E_l[Y_{l1k}(u)]}$.

By the result of (4.6), the second term in (5.18) can be written as

$$\begin{aligned}
& n^{-1/2}\int_0^t \frac{1}{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(u)Y_{lik}(u)}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\tilde{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})dM_{lik}(u) \\
= \quad & n^{-1/2}\int_0^t \frac{1}{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(u)Y_{lik}(u)}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\frac{1}{\tilde{\alpha}_l E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(t)]} \\
\times \quad & n_l^{-1}\left\{\sum_{m=1}^{n_l}(1-\frac{\xi_{lm}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lmj})Y_{lmk}(t)\right\})\xi_{li}\prod_{j=1}^{2}(1-\Delta_{lij})dM_{lik}(u) \\
= \quad & n^{-1/2}\int_0^t \frac{1}{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(u)Y_{lik}(u)}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1)\prod_{j=1}^{2}(1-\Delta_{lij})\int_0^t \frac{1}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)]} \\
\times \quad & Y_{lik}(u)n_l^{-1}\sum_{i=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lij})(Y_{lik}(t)\{\lambda_{0k}(u)+\beta_0^T Z_{l1k}(u)\}).
\end{aligned}$$

If follows from the uniform convergence of $\{n^{-1}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\pi_{lik}(u)Y_{lik}(u)\}^{-1}$ to $\{\sum_{l=1}^{L}q_l E_l[Y_{l1k}(u)]\}$, $n_l^{-1}\sum_{i=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(u)$ to $E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)]$, $n_l^{-1}\sum_{i=1}^{n_l}\frac{\xi_{li}}{\tilde{\alpha}_l}\prod_{j=1}^{2}(1-\Delta_{lij})Y_{lik}(u)$ $\beta_0^T Z_{l1k}(u)$ to $E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)\beta_0^T Z_{l1k}(u)]$ and Lemma 2 that the second term in (5.18) is asymptotically equivalent to

$$\begin{aligned}
& n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1)\prod_{j=1}^{2}(1-\Delta_{lij}) \\
\times \quad & \int_0^t \frac{Y_{lik}(u)E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)\{\lambda_{0k}(u)+\beta_0^T Z_{l1k}(u)\}]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)]}\cdot\frac{du}{\sum_{l=1}^{L}q_l E_l[Y_{l1k}(u)]}.
\end{aligned}$$

Combining the above results, the first and second term on the right-hand side of (5.18) are asymptotically equivalent to

$$n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^{2}(1-\Delta_{lij})$$
$$\times\quad \int_0^t Y_{lik}(u)\left[\beta_0^T Z_{lik}(u)-\frac{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(u)\beta_0^T Z_{lik}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(u)]}\right]\cdot\frac{du}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}.$$

Similarly, the third to the last term on the right-hand side of (5.18) are asymptotically equivalent to

$$n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})$$
$$\times\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{11k}(u)]}\left[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{l1k}(u)|\theta_{l10},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}\right]$$
$$+\quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})$$
$$\times\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{11k}(u)]}\left[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{l1k}(u)|\theta_{l01},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}\right]$$
$$+\quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})$$
$$\times\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}\left[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{l1k}(u)|\theta_{l11},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}\right]$$
$$+\quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\frac{1}{2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li})$$
$$\times\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}\left[dM_{lik}(u)-Y_{lik}(u)\frac{E_l[dM_{l1k}(u)|\theta_{l11},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}\right].$$

Note that $n_l^{1/2}(\tilde{\beta}^{II}-\beta_0)$ is asymptotically equivalent to

$$A^{-1}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(t,\beta_0)$$
$$+\quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^\tau[B_{lik}(t,\beta_0)-Y_{lik}(t)\frac{E_l[\prod_{j=1}^2(1-\Delta_{lij})B_{l1k}(t,\beta_0)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(t)]}]dt)$$
$$+\quad n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)$$
$$\times\quad \left[Q_{lik}(t,\beta_0)-\int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l10},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l10}]}\right]$$

$$
\begin{aligned}
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1) \\
\times \quad & \left[Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l01},\xi_1=1]}{E_l[Y_{l1k}(t)|\theta_{l01}]}\right] \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1) \\
\times \quad & \left[Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l11},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l11}]}\right] \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1) \\
\times \quad & \left[Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l11},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l11}]}\right]\}
\end{aligned}
$$

Combining all the results, we have

$$
\begin{aligned}
& n^{1/2}\{\tilde{\Lambda}_{0k}^{II}(\tilde{\beta^{II}},t)-\Lambda_{0k}(t)\} \\
= \quad & l_k(t)^T A^{-1}\{n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}Q_{lik}(t,\beta_0) \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\prod_{j=1}^{2}(1-\Delta_{lij})(\frac{\xi_{li}}{\tilde{\alpha}_l}-1)\int_0^{\tau}[B_{lik}(t,\beta_0)-Y_{lik}(t)\frac{E_l[\prod_{j=1}^{2}(1-\Delta_{lij})B_{l1k}(t,\beta_0)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{lij})Y_{l1k}(t)]}]dt \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1) \\
\times \quad & [Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l10},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l10}]}] \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1) \\
\times \quad & [Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l01},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l01}]}] \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1) \\
\times \quad & [Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l11},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l11}]}] \\
+ \quad & n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{k=1}^{2}\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1) \\
\times \quad & [Q_{lik}(t,\beta_0)-\int_0^{\tau}Y_{lik}(t)\frac{E_l[dQ_{l1k}(t,\beta_0)|\theta_{l11},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l11}]}]\}
\end{aligned}
$$

$$
\begin{aligned}
+ &\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]} d\{n^{1/2}\sum_{l=1}^L\sum_{i=1}^{n_l} M_{lik}(u)\} + n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})\prod_{j=1}^2(1-\Delta_{lij}) \\
\times &\quad \int_0^t Y_{lik}(u)\left[\beta_0^T Z_{lik}(u) - \frac{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(u)\beta_0^T Z_{lik}(u)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1k}(u)]}\right]\cdot \frac{du}{\sum_{l=1}^L q_l E_l[Y_{lik}(u)]} \\
+ &\quad n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\Delta_{li1}(1-\Delta_{li2})(1-\xi_{li}) \\
\times &\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}\left[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(t,\beta_0)|\theta_{l10},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}\right] \\
+ &\quad n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li}) \\
\times &\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}\left[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(\beta_0,u)|\theta_{l01},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}\right] \\
+ &\quad n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}\frac{1}{2}(\frac{\eta_{li1}}{\tilde{\gamma}_{l3k}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li}) \\
\times &\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E[Y_{l1k}(u)]}\left[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(\beta_0,u)|\theta_{l11},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{11}]}\right] \\
+ &\quad n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}\frac{1}{2}(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)\Delta_{li1}\Delta_{li2}(1-\xi_{li}) \\
\times &\quad \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]}\left[dM_{lik}(u) - Y_{lik}(u)\frac{E_l[dM_{lik}(\beta_0,u)|\theta_{l11},\xi_{li}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}\right] \\
+ &\quad o_p(1)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\quad n^{1/2}\{\tilde{\Lambda}_{0k}^{II}(\tilde{\beta^{II}},t) - \Lambda_{0k}(t)\} \\
= &\quad n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}\mu_{lik}(\beta_0,t) + n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})w_{lik}(\beta_0,t) + n^{-1/2}\sum_{l=1}^L\sum_{i=1}^{n_l}\nu_{lik}(\beta_0,t) + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
&\mu_{lik}(\beta,t) = l_k(t)^T A^{-1}\sum_{m=1}^2 Q_{lim}(\beta) + \int_0^t \frac{1}{\sum_{l=1}^L q_l E_l[Y_{l1k}(u)]} dM_{lik}(u), \\
&w_{lik}(\beta,t) = l_k(t)^T A^{-1}\sum_{m=1}^2\prod_{j=1}^2(1-\Delta_{lij}) \\
\times &\quad \int_0^\tau [B_{lim}(u,\beta) - \frac{Y_{lim}(u)E_l[\prod_{j=1}^2(1-\Delta_{l1j})B_{l1m}(u,\beta)]}{E_l[\prod_{j=1}^2(1-\Delta_{l1j})Y_{l1m}(u)]}]du
\end{aligned}
$$

$$
\begin{aligned}
&+ \prod_{j=1}^{2}(1-\Delta_{lij}) \int_0^t Y_{lik}(u)\{\beta_0^T Z_{lik}(u) \\
&- \frac{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)\beta_0^T Z_{l1k}(u)]}{E_l[\prod_{j=1}^{2}(1-\Delta_{l1j})Y_{l1k}(u)]} \cdot \frac{du}{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(u)]}\},
\end{aligned}
$$

$$
\begin{aligned}
&\nu_{lik}(\beta,t) = l_k(t)^T A^{-1} \sum_{m=1}^{2} \nu_{lim}^{(1)}(\beta,t) + \nu_{lim}^{(2)}(\beta,t), \\
&\nu_{lik}^{(1)}(\beta,t) = \Delta_{li1}(1-\Delta_{li2})(1-\xi_{li})(\frac{\eta_{li1}}{\tilde{\gamma}_{l1}}-1)\nu_{lik,1}^{(1)}(\beta,t) \\
&+(1-\Delta_{li1})\Delta_{li2}(1-\xi_{li})(\frac{\eta_{li2}}{\tilde{\gamma}_{l2}}-1)\nu_{lik,2}^{(1)}(\beta,t), \\
&+\frac{1}{2}\Delta_{li1}\Delta_{li2}(1-\xi_{li})[(\frac{\eta_{li1}}{\tilde{\gamma}_{l3}}-1)+(\frac{\eta_{li2}}{\tilde{\gamma}_{l4}}-1)]\nu_{lim,3}^{(1)}(\beta,t), \\
&\nu_{lik,1}^{(1)}(\beta,t) = Q_{lik}(\beta,t) - \int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\theta_{l10},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l10}]}, \\
&\nu_{lik,2}^{(1)}(\beta,t) = Q_{ik}(\beta,t) - \int_0^\tau Y_{ik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\theta_{l01},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l01}]}, \\
&\nu_{lik,3}^{(1)}(\beta,t) = Q_{lik}(\beta,t) - \int_0^\tau Y_{lik}(t)\frac{E_l[dQ_{l1k}(\beta,t)|\theta_{l11},\xi_{l1}=1]}{E_l[Y_{l1k}(t)|\theta_{l11}]}, \\
&\nu_{lik}^{(2)}(\beta,t) = (1-\xi_{li})\{\Delta_{li1}(1-\Delta_{li2})(\frac{\eta_{li1}}{\tilde{\gamma}_{1k}}-1)\nu_{lik,1}^{(2)}(\beta,t) + (1-\Delta_{li1})\Delta_{li2}(\frac{\eta_{li2}}{\tilde{\gamma}_{2k}}-1)\nu_{lik,2}^{(2)}(\beta,t), \\
&+\frac{1}{2}\Delta_{li1}\Delta_{li2}\{(\frac{\eta_{li1}}{\tilde{\gamma}_{3k}}-1)+(\frac{\eta_{li2}}{\tilde{\gamma}_{4k}}-1)\}\nu_{lik,3}^{(2)}(\beta,t)\}, \\
&\nu_{lik,1}^{(2)}(\beta,t) = \int_0^t \frac{1}{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(u)]}[dM_{lik}(\beta,u) - Y_{lik}(u)\frac{E_l[dM_{l1k}(\beta,u)|\theta_{l10},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l10}]}], \\
&\nu_{lik,2}^{(2)}(\beta,t) = \int_0^t \frac{1}{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(u)]}[dM_{lik}(\beta,u) - Y_{lik}(u)\frac{E_l[dM_{l1k}(\beta,u)|\theta_{l01},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l01}]}], \\
&\nu_{lik,3}^{(2)}(\beta,t) = \int_0^t \frac{1}{\sum_{l=1}^{L} q_l E_l[Y_{l1k}(u)]}[dM_{lik}(\beta,u) - Y_{lik}(u)\frac{E_l[dM_{l1k}(\beta,u)|\theta_{l11},\xi_{l1}=0]}{E_l[Y_{l1k}(u)|\theta_{l11}]}]
\end{aligned}
$$

Let $G^{(1)}(t) = \{G_1^{(1)}(t), G_2^{(1)}(t)\}^T$ where $G_k^{(1)}(t) = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\mu_{lik}(\beta,t)$, $G^{(2)}(t) = \{G_1^{(2)}(t),$ $G_2^{(2)}(t)\}^T$ where $G_k^{(2)}(t) = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}(1-\frac{\xi_{li}}{\tilde{\alpha}_l})w_{lik}(\beta,t)$, and $G^{(3)}(t) = \{G_1^{(3)}(t), G_2^{(3)}(t)\}^T$ where $G_k^{(3)}(t) = n^{-1/2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\nu_{lik}(\beta,t)$ for $k = 1, 2$.

Then, $G^{(1)}(t)$ converges weakly to a zero-mean Gaussian process, $\mathcal{G}^{(1)}(t) = \{\mathcal{G}_1^{(1)}(t), \mathcal{G}_2^{(1)}(t)\}^T$ in $D[0,\tau]^k$ where the covariance function between $\mathcal{G}_j^{(1)}(t)$ and $\mathcal{G}_k^{(1)}(s)$ is $E_l[\mu_{l1j}(t,\beta_0), \mu_{l1k}(s,\beta_0)]$ by theorem 2 of Yin and Cai [2004].

It can be shown that $G^{(2)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{G}^{(2)}(t) =$ $\{\mathcal{G}_1^{(2)}(t), \mathcal{G}_2^{(2)}(t)\}$ where covariance function $\mathcal{G}_j^{(2)}(t)$ and $\mathcal{G}_k^{(2)}(s)$ is $\frac{1-\alpha_l}{\alpha_l}E_l[w_{l1j}(\beta_0,t), w_{l1k}(\beta_0,s)]$ by Lemma 2, Cramer-Wold device and the marginal tightness of $G_k^{(2)}(t)$ for each k.

Similarly, $G^{(3)}(t)$ converges weakly to a zero-mean Gaussian process where covariance function $\mathcal{G}_j^{(3)}(t)$ and $\mathcal{G}_k^{(3)}(s)$ is

$$
\begin{aligned}
& (1-\alpha_l)[(I(j=k)[Pr[\theta_{l10}](\frac{1-\gamma_{l1}}{\gamma_{l1}})Cov_l[\nu^{(2)}_{l1k,1}(\beta_0,t),\nu^{(2)}_{l1k,1}(\beta_0,s)|\theta_{l10},\xi_{l1}=0] \\
+ & Pr[\theta_{l01}](\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1k,2}(\beta_0,t),\nu^{(2)}_{l1k,2}(\beta_0,s)|\theta_{l01},\xi_{l1}=0] \\
+ & \frac{1}{4}Pr[\theta_{l11}](\frac{1-\gamma_{l1}}{\gamma_{l1}}+\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1k,3}(\beta_0,t),\nu^{(2)}_{l1k,3}(\beta_0,s)|\theta_{l11},\xi_{l1}=0]]) \\
+ & Pr[\theta_{l10}](\frac{1-\gamma_{l1}}{\gamma_{l1}})Cov_l[\nu^{(2)}_{l1j,1}(\beta_0,t),l_k(s)^TA^{-1}\nu^{(1)}_{l1j,1}(\beta_0,t)|\theta_{l10},\xi_{l1}=0] \\
+ & Pr[\theta_{l01}](\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1j,2}(\beta_0,t),l_k(s)^TA^{-1}\nu^{(1)}_{l1j,2}(\beta_0,t)|\theta_{l01},\xi_{l1}=0] \\
+ & \frac{1}{4}Pr[\theta_{l11}](\frac{1-\gamma_{l1}}{\gamma_{l1}}+\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1j,3}(\beta_0,t),l_k(s)^TA^{-1}\nu^{(1)}_{l1j,3}(\beta_0,t)|\theta_{l11},\xi_{l1}=0] \\
+ & Pr[\theta_{l10}](\frac{1-\gamma_{l1}}{\gamma_{l1}})Cov_l[\nu^{(2)}_{l1k,1}(\beta_0,s),l_j(t)^TA^{-1}\nu^{(1)}_{l1k,1}(\beta_0,s)|\theta_{l10},\xi_{l1}=0] \\
+ & Pr[\theta_{l01}](\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1k,2}(\beta_0,s),l_j(t)^TA^{-1}\nu^{(1)}_{l1k,2}(\beta_0,s)|\theta_{l01},\xi_{l1}=0] \\
+ & \frac{1}{4}Pr[\theta_{l11}](\frac{1-\gamma_{l1}}{\gamma_{l1}}+\frac{1-\gamma_{l2}}{\gamma_{l2}})Cov_l[\nu^{(2)}_{l1k,3}(\beta_0,s),l_j(t)^TA^{-1}\nu^{(1)}_{l1k,3}(\beta_0,s)|\theta_{l11},\xi_{l1}=0] \\
+ & \sum_{m=1}^{2}(Pr[\theta_{l10}](\frac{1-\gamma_{l1}}{\gamma_{l1}})l_j(t)^TA^{-1}Cov_l[\nu^{(1)}_{l1m,1}(\beta_0,t),\nu^{(1)}_{l1m,1}(\beta_0,s)|\theta_{l10},\xi_{l1}=0]A^{-1}l_k(s) \\
+ & Pr[\theta_{l01}](\frac{1-\gamma_{l2}}{\gamma_{l2}})l_j(t)^TA^{-1}Cov_l[\nu^{(1)}_{l1m,2}(\beta_0,t),\nu^{(1)}_{l1m,2}(\beta_0,s)|\theta_{l01},\xi_{l1}=0]A^{-1}l_k(s) \\
+ & Pr[\theta_{l11}](\frac{1-\gamma_{l1}}{\gamma_{l1}}+\frac{1-\gamma_{l2}}{\gamma_{l2}})l_j(t)^TA^{-1}Cov_l[\nu^{(1)}_{l1m,3}(\beta_0,t),\nu^{(1)}_{l1m,3}(\beta_0,s)|\theta_{l11},\xi_{l1}=0]A^{-1}l_k(s))]
\end{aligned}
$$

By the conditional expectation arguments, all terms are mutually independent. Therefore, $G(t) = G^{(1)}(t)+G^{(2)}(t)+G^{(3)}(t)$ converges to a zero-mean Gaussian process $\mathcal{G}(t) = \mathcal{G}^{(1)}(t)+\mathcal{G}^{(2)}(t)+\mathcal{G}^{(3)}(t)$.

5.4 Simulations

We conducted simulation studies to examine the performance of the proposed methods and compare the existing methods with the proposed methods. Correlated bivariate failure time data were generated from Clayton-Cuzick model [Clayton and Cuzick, 1985]. The bivariate survival function for the bivariate survival time (T_1, T_2) given (Z_{l1}, Z_{l2}) has the following form:

$$F(t_1, t_2 \mid Z_{l1}, Z_{l2}) = \{e^{\frac{\int_0^{t_1}(\lambda_{01}(t)+\beta_0 Z_{l1})dt}{\theta}} + e^{\frac{\int_0^{t_2}(\lambda_{02}(t)+\beta_0 Z_{l2})dt}{\theta}} - 1\}^{-\theta},$$

where $\lambda_{0k}(t)$ and β_k, $k = 1, 2$ are the baseline hazard function and the effect of covariate for disease k, respectively, l is a dichotomous stratum variable, and θ is the parameter related with correlation between the failure times of the two diseases. Smaller θ indicates higher correlation between the two failure times T_1 and T_2. The relationship between Kendall's tau, τ_θ, and θ is $\tau_\theta = \frac{1}{2\theta+1}$. For θ, we used values of 0.10, 0.67, and 4 and the corresponding Kendall's tau values are 0.83, 0.43, and 0.11, respectively. We set the baseline hazard function $\lambda_{01} = 2$ for the first failure event type $k = 1$ and $\lambda_{02} = 4$ for the second failure event type $k = 2$. The regression parameters are examined at $\beta_0 = 0$ and 0.3.

We generate Z from Bernouilli distribution with $\text{pr}(Z = 1) = 0.5$ under the situation $Z_{l1} = Z_{l2} = Z$. To consider stratified subcohort sampling from two strata defined by V_i, we define two parameters: $\eta = \Pr(V = 1|Z = 1)$ and $\nu = \Pr(V = 0|Z = 0)$ where η is sensitivity and ν is the specificity for Z. Unstratified sampling with same probability, i.e., $\eta = 0.5$ and $\nu = 0.5$ is a special case. Larger values η and ν values than 0.5 indicate that V is highly correlated with Z. For stratified case-cohort studies, we set the values $[\eta, \nu] = [0.7, 0.7]$. Thus, a stratum variable is simulated with $\Pr(V = 1) = (1 - \nu)\Pr(Z = 0) + \eta\Pr(Z = 1) = 0.5$. Censoring times are generated from uniform distribution $[0, u]$ where u depends on the specified level of the censoring probability.

For simulations of the traditional case-cohort study, we set the event proportions of approximately 8% and 20% for $k = 1$ and 14% and 35% for $k = 2$. For the simulations of the generalized case-cohort study, the event proportions are considered as 15% and 25% for $k = 1$ and 26% and 42% for $k = 2$ and we sample half of the cases outsides the subcohort, $[\gamma_1, \gamma_2] = [0.5, 0.5]$. The sample size of the full cohort is set to be $n = 1000$. For stratified sampling, we consider the total subcohort size of 100 and 200 and select the subcohort $\tilde{n}_l = \tilde{n} \times q_l$ from each stratum. For each configuration, we conducted 2000 simulations.

Table 5.1: Simulation result for the traditional case-cohort study: $K = 1$, $\beta_1 = 0$

	Event			The Proposed weight				Kulich and Lin's method						
S	PR	$\tilde{n}$	τ_θ	$\tilde{\beta}_1^I$	SE_p	SD_p	CR_p	$\hat{\beta}_1^A$	SE_k	SD_k	CR_k	SRE	SRE$_p$	SRE$_k$
UN	8%	100	0·10	0·001	0·622	0·621	0·96	0·000	0·643	0·633	0·96	1·05	1·00	1·00
			0·67	-0·005	0·610	0·623	0·95	-0·002	0·639	0·652	0·95	1·10	1·00	1·00
			4	0·012	0·613	0·649	0·95	0·014	0·644	0·677	0·95	1·09	1·00	1·00
		200	0·10	-0·005	0·527	0·536	0·95	-0·006	0·537	0·543	0·95	1·02	1·00	1·00
			0·67	0·007	0·525	0·532	0·94	0·013	0·539	0·546	0·94	1·05	1·00	1·00
			4	-0·002	0·523	0·523	0·95	0·004	0·538	0·542	0·95	1·07	1·00	1·00
	20%	100	0·10	-0·001	0·485	0·505	0·95	-0·005	0·525	0·548	0·95	1·18	1·00	1·00
			0·67	0·022	0·466	0·488	0·94	0·032	0·525	0·539	0·96	1·22	1·00	1·00
			4	0·002	0·453	0·477	0·94	0·010	0·525	0·551	0·95	1·33	1·00	1·00
		200	0·10	0·008	0·385	0·395	0·95	0·007	0·406	0·412	0·95	1·09	1·00	1·00
			0·67	0·001	0·374	0·375	0·95	0·000	0·406	0·402	0·96	1·15	1·00	1·00
			4	-0·007	0·367	0·375	0·95	-0·012	0·405	0·412	0·95	1·20	1·00	1·00
STR	8%	100	0·10	0·006	0·601	0·621	0·95	0·003	0·618	0·631	0·95	1·03	1·00	1·01
			0·67	0·001	0·593	0·594	0·96	-0·002	0·617	0·620	0·96	1·09	1·10	1·10
			4	-0·001	0·596	0·600	0·96	-0·003	0·621	0·626	0·96	1·09	1·17	1·17
		200	0·10	-0·005	0·520	0·522	0·95	-0·004	0·528	0·528	0·95	1·02	1·05	1·06
			0·67	0·021	0·515	0·521	0·95	0·021	0·526	0·534	0·95	1·05	1·04	1·05
			4	-0·008	0·514	0·517	0·95	-0·006	0·526	0·529	0·95	1·05	1·02	1·05
	20%	100	0·10	-0·007	0·470	0·480	0·95	-0·007	0·502	0·512	0·96	1·14	1·11	1·15
			0·67	0·004	0·454	0·460	0·95	-0·004	0·503	0·507	0·96	1·21	1·12	1·13
			4	-0·015	0·442	0·456	0·94	-0·009	0·503	0·509	0·96	1·25	1·09	1·17
		200	0·10	-0·005	0·377	0·383	0·95	-0·004	0·394	0·405	0·94	1·12	1·06	1·03
			0·67	-0·005	0·367	0·357	0·96	-0·005	0·393	0·377	0·97	1·12	1·10	1·14
			4	-0·006	0·361	0·370	0·95	-0·005	0·392	0·403	0·95	1·19	1·03	1·04

S, sampling; PR, proportion; UN, unstratified sampling; STR, stratified sampling; SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; SRE$_p$, sample relative efficiency of proposed estimators with unstratified sampling relative to stratified sampling; SRE$_k$, sample relative efficiency of Kulich and Lin's estimators with unstratified sampling relative to stratified sampling.

Table 5.2: Simulation result for the generalized case-cohort study: $K = 1$, $\beta_1 = 0$

	Event			The Proposed weight				The existing method						
S	PR	$\tilde{n}$	τ_θ	$\tilde{\beta}^I_{G1}$	SE_p	SD_p	CR_p	$\hat{\beta}^G_1$	SE_k	SD_k	CR_k	SRE	SRE$_p$	SRE$_k$
UN	15%	100	0·10	-0·051	1·065	1·075	0·96	-0·047	1·085	1·104	0·95	1·06	1·00	1·00
			0·67	0·000	1·040	1·047	0·95	-0·003	1·083	1·101	0·95	1·10	1·00	1·00
			4	-0·004	1·030	1·052	0·95	-0·012	1·089	1·109	0·96	1·11	1·00	1·00
		200	0·10	0·003	0·822	0·804	0·96	0·016	0·842	0·839	0·96	1·09	1·00	1·00
			0·67	-0·021	0·815	0·823	0·95	-0·019	0·841	0·850	0·96	1·07	1·00	1·00
			4	-0·003	0·810	0·803	0·96	0·002	0·842	0·840	0·95	1·09	1·00	1·00
	25%	100	0·10	-0·004	0·953	0·990	0·94	0·003	0·969	1·008	0·94	1·04	1·00	1·00
			0·67	-0·020	0·926	0·964	0·95	-0·010	0·971	1·014	0·95	1·11	1·00	1·00
			4	0·005	0·888	0·919	0·95	0·002	0·973	0·998	0·95	1·18	1·00	1·00
		200	0·10	0·001	0·714	0·707	0·96	-0·005	0·729	0·725	0·96	1·05	1·00	1·00
			0·67	-0·008	0·703	0·729	0·95	-0·009	0·730	0·763	0·95	1·10	1·00	1·00
			4	0·003	0·684	0·704	0·95	0·004	0·728	0·736	0·95	1·09	1·00	1·00
STR	15%	100	0·10	0·007	1·040	1·029	0·96	0·008	1·057	1·050	0·96	1·04	1·09	1·11
			0·67	0·025	1·019	1·008	0·96	0·011	1·056	1·042	0·96	1·07	1·08	1·11
			4	-0·022	1·006	0·992	0·96	-0·022	1·056	1·045	0·96	1·11	1·12	1·13
		200	0·10	-0·014	0·808	0·803	0·96	-0·015	0·825	0·822	0·95	1·05	1·00	1·04
			0·67	0·005	0·806	0·798	0·96	0·005	0·828	0·831	0·95	1·09	1·06	1·05
			4	-0·002	0·800	0·795	0·95	0·004	0·827	0·830	0·95	1·09	1·02	1·02
	25%	100	0·10	0·000	0·938	0·919	0·96	-0·003	0·954	0·937	0·96	1·04	1·16	1·16
			0·67	0·038	0·913	0·899	0·95	0·033	0·954	0·951	0·95	1·12	1·15	1·14
			4	0·000	0·875	0·831	0·96	0·000	0·949	0·922	0·96	1·23	1·22	1·17
		200	0·10	0·017	0·705	0·678	0·96	0·019	0·720	0·695	0·96	1·05	1·09	1·09
			0·67	-0·018	0·695	0·682	0·96	-0·017	0·720	0·711	0·95	1·09	1·14	1·15
			4	0·002	0·679	0·665	0·96	0·002	0·719	0·705	0·96	1·13	1·12	1·09

S, sampling; PR, proportion; UN, unstratified sampling; STR, stratified sampling; SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; SRE$_p$, sample relative efficiency of proposed estimators with unstratified sampling relative to stratified sampling; SRE$_k$, sample relative efficiency of the existing estimators with unstratified sampling relative to stratified sampling.

Table 5.3: Simulation result for the traditional case-cohort study: $K = 2$, $\beta_0 = 0.3$

	Event			The Proposed weight				Kang & Cai's method						
S	PR	$\tilde{n}$	τ_θ	$\tilde{\beta}^{II}$	SE_p	SD_p	CR_p	$\hat{\beta}^A$	SE_k	SD_k	CR_k	SRE	SRE$_p$	SRE$_k$
UN	[8%, 14%]	100	0·10	0·320	0·803	0·829	0·95	0·317	0·815	0·835	0·96	1·02	1·00	1·00
			0·67	0·280	0·757	0·781	0·96	0·283	0·777	0·793	0·96	1·03	1·00	1·00
			4	0·301	0·740	0·765	0·95	0·297	0·764	0·784	0·95	1·05	1·00	1·00
		200	0·10	0·311	0·647	0·653	0·95	0·311	0·654	0·655	0·95	1·01	1·00	1·00
			0·67	0·323	0·599	0·603	0·95	0·322	0·610	0·610	0·95	1·02	1·00	1·00
			4	0·298	0·580	0·595	0·95	0·298	0·593	0·603	0·95	1·03	1·00	1·00
	[20%, 35%]	100	0·10	0·292	0·680	0·700	0·95	0·295	0·694	0·714	0·95	1·04	1·00	1·00
			0·67	0·300	0·632	0·645	0·95	0·297	0·663	0·665	0·95	1·06	1·00	1·00
			4	0·323	0·596	0·610	0·94	0·319	0·643	0·645	0·95	1·12	1·00	1·00
		200	0·10	0·307	0·514	0·533	0·94	0·309	0·521	0·538	0·94	1·02	1·00	1·00
			0·67	0·309	0·476	0·498	0·95	0·309	0·493	0·513	0·94	1·06	1·00	1·00
			4	0·311	0·445	0·462	0·94	0·316	0·472	0·488	0·94	1·12	1·00	1·00
STR	[8%, 14%]	100	0·10	0·286	0·772	0·797	0·95	0·285	0·781	0·802	0·95	1·02	1·08	1·08
			0·67	0·325	0·724	0·749	0·95	0·327	0·738	0·759	0·95	1·03	1·09	1·09
			4	0·302	0·706	0·710	0·96	0·300	0·723	0·720	0·96	1·03	1·16	1·19
		200	0·10	0·292	0·628	0·637	0·95	0·292	0·632	0·640	0·95	1·01	1·05	1·05
			0·67	0·302	0·580	0·578	0·95	0·301	0·587	0·583	0·95	1·01	1·09	1·09
			4	0·282	0·562	0·569	0·95	0·281	0·570	0·580	0·95	1·04	1·09	1·08
	[20%, 35%]	100	0·10	0·324	0·655	0·650	0·96	0·325	0·664	0·663	0·96	1·04	1·16	1·16
			0·67	0·302	0·610	0·601	0·95	0·304	0·632	0·629	0·95	1·10	1·15	1·12
			4	0·292	0·576	0·598	0·94	0·289	0·611	0·637	0·95	1·13	1·04	1·03
		200	0·10	0·315	0·497	0·490	0·95	0·317	0·502	0·496	0·96	1·02	1·18	1·17
			0·67	0·310	0·460	0·458	0·95	0·311	0·472	0·478	0·95	1·09	1·19	1·15
			4	0·301	0·431	0·445	0·95	0·300	0·450	0·469	0·94	1·11	1·08	1·09

S, sampling; PR, proportion; UN, unstratified sampling; STR, stratified sampling; SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; SRE$_p$, sample relative efficiency of proposed estimators with unstratified sampling relative to stratified sampling; SRE$_k$, sample relative efficiency of the existing estimators with unstratified sampling relative to stratified sampling.

We first considered traditional case-cohort sample with a single disease but with covariates available on subjects with other diseases. We examine the the performance of our proposed estimator and compare our results with those with the time-varying weight [Kulich and Lin, 2000a]. Moreover, we compare the results of unstratified sampling with stratified sampling using the proposed and Kulich and Lin [2000a] estimators, respectively. Table 5.1 reports the summary of $\tilde{\beta}_1^I$ and $\hat{\beta}_1^A$. For different combinations of β, event proportion, subcohort sample size, and correlation, Table 5.1 shows the average of the estimates $\tilde{\beta}_1^I$, the average of the proposed estimated standard error (SE), empirical standard deviation (SD), sample relative efficiency of the proposed estimators relative to estimators of Kulich and Lin [2000a] (SRE), sample relative efficiency of proposed estimators with unstratified sampling relative to with stratified sampling (SRE_p), and sample relative efficiency of estimators of Kulich and Lin [2000a] with unstratified sampling relative to with stratified sampling (SRE_k). The subscripts for SE, SD, SRE refer to the proposed method (P) and the existing traditional case-cohort analysis for additive hazards models, Kulich and Lin [2000a] (K). The simulation results suggest that both methods are approximately unbiased across the setup for $\beta = 0.3$ with both event proportions (8% and 20%) and correlations (0.10, 0.67, and 4). The average of the proposed estimated standard error is close to the empirical standard deviation and it is smaller with lower correlation, larger event proportions or subcohort size, as expected. The 95% confidence interval coverage rate ranges between 94% and 97%. All the sample relative efficiency (SRE), defined as SD_k^2/SD_p^2, are larger than 1 which indicates that the proposed estimates are more efficient than those from Kulich and Lin [2000a]. This shows that the extra information collected on subjects with the other disease helps to gain efficiency. In general, the efficiency is larger in situations with larger event proportion, smaller subcohort size, and smaller correlation. Also, SRE_p and SRE_k for stratified sampling are more than 1 suggesting that stratified sampling is more efficient than unstratified sampling. However, when the disease rate is low and the subcohort size is larger, the proposed method does not improve much efficiency.

In the second set of simulation, we are interested in the non-rare event and we sample half of the cases outside the subcohort. We examine the performance of our proposed estimator

Table 5.4: Simulation result for the generalized case-cohort study: $K = 2$, $\beta_0 = 0.3$

	Event			The Proposed weight				The existing method						
S	PR	$\tilde{n}$	τ_θ	$\tilde{\beta}_G^{II}$	SE_p	SD_p	CR_p	$\hat{\beta}^A$	SE_k	SD_k	CR_k	SRE	SRE$_p$	SRE$_k$
UN	$[15\%, 26\%]$	100	0·10	0·303	0·760	0·760	0·96	0·308	0·781	0·773	0·96	1·04	1·00	1·00
			0·67	0·307	0·720	0·716	0·96	0·311	0·751	0·749	0·96	1·09	1·00	1·00
			4	0·314	0·697	0·711	0·95	0·325	0·736	0·764	0·95	1·16	1·00	1·00
		200	0·10	0·297	0·591	0·607	0·95	0·301	0·609	0·622	0·95	1·05	1·00	1·00
			0·67	0·296	0·559	0·544	0·95	0·298	0·579	0·569	0·96	1·10	1·00	1·00
			4	0·315	0·542	0·526	0·96	0·317	0·564	0·564	0·95	1·15	1·00	1·00
	$[25\%, 42\%]$	100	0·10	0·293	0·677	0·692	0·95	0·292	0·697	0·708	0·95	1·05	1·00	1·00
			0·67	0·301	0·632	0·643	0·95	0·299	0·668	0·682	0·95	1·12	1·00	1·00
			4	0·302	0·592	0·587	0·95	0·313	0·646	0·654	0·95	1·24	1·00	1·00
		200	0·10	0·304	0·512	0·521	0·95	0·302	0·527	0·527	0·96	1·02	1·00	1·00
			0·67	0·285	0·481	0·488	0·95	0·292	0·502	0·512	0·95	1·10	1·00	1·00
			4	0·307	0·451	0·440	0·96	0·308	0·480	0·480	0·95	1·19	1·00	1·00
STR	$[15\%, 26\%]$	100	0·10	0·304	0·737	0·731	0·96	0·306	0·754	0·748	0·96	1·05	1·08	1·07
			0·67	0·312	0·698	0·659	0·96	0·310	0·720	0·693	0·96	1·10	1·18	1·17
			4	0·315	0·678	0·662	0·97	0·319	0·706	0·708	0·96	1·14	1·15	1·17
		200	0·10	0·278	0·579	0·582	0·95	0·281	0·593	0·596	0·95	1·05	1·09	1·09
			0·67	0·295	0·549	0·539	0·96	0·299	0·563	0·559	0·95	1·08	1·02	1·04
			4	0·310	0·530	0·523	0·95	0·314	0·545	0·556	0·94	1·13	1·01	1·03
	$[25\%, 42\%]$	100	0·10	0·288	0·656	0·650	0·96	0·295	0·671	0·664	0·96	1·04	1·13	1·14
			0·67	0·285	0·620	0·602	0·95	0·286	0·646	0·634	0·96	1·11	1·14	1·16
			4	0·295	0·583	0·569	0·95	0·299	0·623	0·617	0·96	1·18	1·06	1·13
		200	0·10	0·304	0·501	0·494	0·95	0·303	0·513	0·507	0·96	1·05	1·11	1·08
			0·67	0·293	0·473	0·450	0·96	0·296	0·489	0·474	0·96	1·11	1·18	1·17
			4	0·295	0·445	0·434	0·96	0·301	0·466	0·466	0·95	1·15	1·03	1·06

S, sampling; PR, proportion; UN, unstratified sampling; STR, stratified sampling; SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the coverage rate of the nominal 95% confidence intervals; SRE = SD_k^2/SD_p^2, sample relative efficiency; SRE$_p$, sample relative efficiency of proposed estimators with unstratified sampling relative to stratified sampling; SRE$_k$, sample relative efficiency of the existing estimators with unstratified sampling relative to stratified sampling.

and compare it to the existing method with a single disease outcome in the generalized case-cohort study. Table 5.2 summarizes the results. Overall performance are similar to Table 5.1: the unbiased estimates for β_1, estimated standard errors close to the empirical standard deviations, the 95% confidence interval coverage rate close to the nominal level. All the sample relative efficiency (SRE) are more than 1 which implies that our proposed method is more efficient than that of Kang et al. [2012]. Moreover, all the sample relative efficiency of stratified sampling with unstratified sampling (SRE_p and SRE_k) are more than 1 which suggest that stratified sampling is more efficient than unstratified sampling.

In the third set of simulation, we consider the joint modeling of the two diseases for case-cohort sample with the rare event. We examine the performance of our proposed estimator and compare it to the existing method with multiple disease outcome. Table 5.3 provides summary statistics for $\tilde{\beta}^{II}$ and $\hat{\beta}^{A}$. We found that biases in the coefficient estimates are small; estimated standard errors close to the empirical standard deviations; the 95% confidence interval coverage rate ranges in 94% and 96%. All the sample relative efficiency (SRE) with more than 1 indicates that our proposed method is more efficient than that of the existing method. For stratified sampling design, all the sample relative efficiency of both proposed and exiting estimators are more than 1. This shows that stratified sampling for the subcohort improve the efficiency for the traditional case-cohort study with multiple outcomes.

Table 5.4 summarizes the simulation results for the joint modeling of two non-rare diseases. We used the selection probability of cases with 0.5 for each disease. Overall, the performance is similar to Table 5.3. For the proposed estimator, sample relative efficiency gain of stratified sampling relative to unstratified sampling ranges in 1% to 18%. For the existing estimator, it ranges in 3% to 17%. They imply that estimators with stratified sampling are more efficient than those with unstratified sampling.

5.5 Data Analysis

We applied the proposed method to data from the Atherosclerosis Risk in Communities (ARIC) study [Lee et al., 2008]. This study is a longitudinal and population-based cohort study consisting of 15,792 men and women aged from 45 to 64 years recruited from four US communities. For this analysis, the follow up for incident coronary heart disease (CHD) event and incident stroke event is through 1998. Incident CHD event is defined as definite or probable myocardial infarction, electrocardiographic evidence of silent myocardial infarction, definite CHD death, or coronary revascularization procedure. Incident stroke was defined as a definite or probable ischemic stroke. We regarded the subject as censored if he or she was free of that event type by December 31, 1998 or lost to follow-up during the study.

The primary aim of this study was to investigate the association between PTGS1 polymorphisms and risk of incident CHD and stroke. Cyclooxygenase-derived prostaglandins can be significant modifiers of risk of cardiovascular diseases event. It has been suggested that variation in the genes encoding cyclooxygenase-derived prostaglandins (PTGS1) play an important role of cardiovascular disease risk [Antman et al., 2005; Camitta et al., 2001; Ulrich et al., 2002].

Using case-cohort design, genomic DNA genotyped for the polymorphisms in PTGS1 were available on all incident CHD, ischemic stroke cases, and the subcohort. The subcohort was selected by using stratified sampling design with three stratum variables: age (≥ 55 or < 55 years), gender, and race (Caucasian or African American). After excluding subjects with missing genotype data and covaroates, a full cohort consisted of a total of 13,731 subjects which includes 900 subjects with only CHD, 188 subjects with only stroke, 61 subjects with both CHD and stroke. The subcohort involved 850 disease-free subjects, 72 subjects with only CHD, 15 subject with only stroke, and 7 subjects with both CHD and stoke. The total size of assayed samples was 1,999. To adjust for confounding and other risk factors, traditional and clinical covariates related to cardiovascular diseases are used: age, gender, race, study center, current smoking status, diabetes, and hypertension.

In order to study the effects of genetic variation (PTGS1) on CHD as well as stroke, we fit the model using (5.1). Since all cases for CHD and stoke are selected and we are interested in comparing the risk effects on CHD and on stroke, we conduct the joint analysis for traditional stratified case-cohort design.

Table 5.5: Analysis results for the effects of PTGS1 G/A+A/A versus G/G ($\times 10^{-6}$)

		Proposed method			Kang & Cai's method		
	Variable	$\tilde{\beta}^{II}$	SE	P-value	$\tilde{\beta}$	SE	P-value
CHD	PTGS1	2.52	2.94	0.196	2.47	2.94	0.201
	African	-10.95	6.68	0.051	-10.44	6.80	0.063
	Age	0.69	0.18	< 0.001	0.69	0.18	< 0.001
	Male	20.03	2.15	< 0.001	19.85	2.16	< 0.001
	Center(F)	-1.38	3.36	0.341	-1.61	3.37	0.317
	Center(J)	-1.94	7.29	0.395	-2.64	7.42	0.361
	Center(M)	-9.05	3.03	< 0.001	-9.00	3.05	< 0.001
	Current smoking	12.93	2.61	< 0.001	12.89	2.63	< 0.001
	Diabetes	22.67	5.32	< 0.001	23.27	5.46	< 0.001
	Hypertension	15.43	2.75	< 0.001	15.55	2.78	< 0.001
Storke	PTGS1	2.76	1.46	0.029	2.97	1.52	0.025
	African	1.37	3.42	0.344	1.54	3.36	0.324
	Age	0.33	0.07	< 0.001	0.34	0.07	< 0.001
	Male	2.42	0.79	< 0.001	2.33	0.81	< 0.001
	Center(F)	-0.42	1.08	0.350	-0.48	1.14	0.337
	Center(J)	0.59	3.70	0.437	0.24	3.68	0.474
	Center(M)	-0.48	0.96	0.310	-0.79	0.98	0.211
	Current smoking	3.94	1.05	< 0.001	4.26	1.09	< 0.001
	Diabetes	9.32	2.30	< 0.001	8.43	2.24	< 0.001
	Hypertension	6.01	1.10	< 0.001	5.90	1.13	< 0.001

Table 5.5 represents the results of additive hazards regression parameters estimates for PTGS1 G/A+A/A versus G/G, estimated standard errors (SE), and P-values. We fit the model allowing for different effects for CHD and stoke. The effects of PTGS1 on both CHD and stroke are not statistically significant with P-value of 0.196 and 0.201, respectively. We also fit the same model using Kang et al. [2012]'s method. Except for the standard errors of African, Center (J) and diabetes on stroke, all the standard errors for the proposed estimator are slightly smaller than those for the estimator of Kang et al. [2012].

5.6 Concluding remarks

By using the new weight function, we have proposed more efficient estimators for the additive hazards model in stratified case-cohort design with rare and non-rare diseases. The new weight functions incorporate the extra information for subjects with other diseases, which can help to increase efficiency relative to existing methods. Moreover, stratified sampling for the subcohort also improved the efficiency relative to unstratified sampling. However, under the situation that the disease rate is low, the proposed method did not improve much efficiency due to small amount of extra information.

In many biomedical and clinical studies, multiple case-cohort studies have been conducted separately. Under the situation, covariate information collected on subjects with the other diseases can be obtained and stratum variables are often available on all the cohort members. By using available information for subjects with other diseases and stratum variables, we are able to estimate the risk effects more efficiently in additive hazards model for case-cohort studies.

It would be worthwhile to consider models with different association between failure time and risk factors. Therefore, we can adapt our approaches to other types of models such as proportional odds model, the accelerated failure time model, and the semiparametric transformation model by using all available information including stratum variables and covariate information for other diseases. They are expected to improve efficiency.

Chapter 6

Summary and Future Research

In this dissertation, we have studied more efficient statistical methods for case-cohort studies with univariate and multivariate failure times. Specially, the following topics are considered: 1) more efficient estimators for the traditional case-cohort study 2) stratified case-cohort study with nonrare diseases, and 3) more efficient estimation in additive hazards models for stratified case-cohort studies.

Case-cohort study design is generally used to reduce cost in large cohort studies. When several diseases are of interest, several case-cohort studies are usually conducted using the same subcohort. When these case-cohort data are analyzed, the common practice is to analyze each disease separately ignoring data collected in subjects with the other diseases. In addition, many baseline covariates are often available for the full cohort. Hence, the main contribution of this dissertation is to provide statistical methods for case-cohort studies which use all available information. We proposed new weights for both rare and nonrare diseases. We developed weighted estimating equations with new weight functions for parameter estimation and studied the cumulative baseline hazard functions.

In Chapter 2.3.3, we considered case-cohort studies with rare diseases. In Chapter 3.6, stratified case-cohort studies with nonrare diseases were considered. In Chapter 4.6, we considered additive hazards models for stratified case-cohort studies.

The asymptotic properties of the proposed estimators were shown to be consistent and asymptotically normally distributed under some regularity conditions. We investigated the finite sample properties of the proposed methods and compared those with the existing methods. The simulation results show that our proposed methods worked properly and were more efficient than the existing methods. We applied our proposed methods to data from the Busselton Health Study and the Atherosclerosis Risk in Communities study.

There are many directions that can be pursued in my future research.

First, I would like to extend the current methodology to competing risks. In the competing risks situation, a subject can only experience at most one event, while in the situation we considered a subject can still experience the other events after experiencing one event. Consequently, in the competing risks situation, a subject is at risk for all types of events simultaneously and will not be at risk for any other events as soon as one event occurs. I will adapt the approach in my dissertation to competing risks by modifying the at-risk process and the weight function.

The second topic is to consider the joint modeling of survival time and longitudinal covariates via shared random effects in case-cohort studies. Our current approaches can allow the time-dependent covariates only when there are no missing data in covariates. However, in many follow-up studies, the entire time-dependent covariate history is not always available. I would like to investigate the joint modeling approach to address the missing covariate data problem.

Last, I would like to apply our proposed approaches to case-cohort studies with models including proportional odds model, the accelerated failure time model, and the semiparametric transformation model. In some data, the proportional or additive hazards assumption may not always be true. Therefore, it is worthwhile to consider modeling association from different aspects.

Bibliography

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, 10:1100–20.

Antman, E. M., DeMets, D., and Loscalzo, J. (2005). Cyclooxygenase inhibition and cardiovascular risk. *Circulation*, 112:759–770.

Ballantyne, C. M., Hoogeveen, R. C., Bang, H., Coresh, J., Folsom, A. R., Heiss, G., and Sharrett, A. R. (2004). Lipoptrtein-associated phospholipase a2, high-sensitivity c-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the atherosclerosis risk in communities (aric) study. *Circulation*, 109:837–842.

Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics*, 50:1064–72.

Borgan, O., Langholz, B., Samuelsen, S. O., G. L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Anal.*, 6:39–58.

Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Statist.*, 34:86–102.

Cai, J. and Zeng, D. (2007). Power calculation for case-cohort studies with nonrare events. *Biometrics*, 63:1288–95.

Camitta, M. G., Gabel, S. A., Chulada, P., Bradbury, J. A., Langenbach, R., Zeldin, D. C., and Murphy, E. (2001). Cyclooxygenase-1 and -2 knockout mice demonstrate increased cardiac ischemia/reperfusion injury but are protected by acute preconditioning. *Circulation*, 104:2453–2458.

Chen, H. Y. (2001a). Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. *J. Am. Statist. Assoc.*, 96:1446–1457.

Chen, K. (2001b). Generalized case-cohort sampling. *J. R. Statist*, B 63:791–809.

Chen, K. and Lo, S. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, 86:755–64.

Chen, Y. and Zucker, D. M. (2009). Case-cohort analysis with semiparametric transformation model. *J. Statist. Plan. and inf.*, 139:3706–3717.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model(with discussion). *J. R. Statist. Soc. A*, 148:82–117.

Clegg, L. X., Cai, J., and Sen, P. K. (1999). A mariginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, 55:805–812.

Cox, D. R. (1972). Regression models and life-tables(with discussion). *J.R. Statist Soc.*, B 34:187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269.

Cullen, K. J. (1972). Mass health examinations in the busselton population, 1996 to 1970. *Aust. J. Med.*, 2:714–8.

Duncan, B. B., Schmidt, M. I., Pankow, J. S., Ballantyne, C. M., Couper, D., Vigo, A., Hoogeveen, R., Folsom, A. R., and Heiss, G. (2003). Low-grade systemic inflammation and the development of type 2 diabetes. *Diabetes*, 52:1799–1805.

Fourtz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Am. Statist. Assoc.*, 72:147–8.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:361–74.

Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multistate models for disease incidence and mortality. *Statist. Med.*, 7:149–60.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Anlaysis of Failure Time data.* John Wiley, New York, 2nd edition.

Kang, S. and Cai, J. (2009). Marginal hazard model for case-cohort studies with multiple disease outcomes. *Biometrika*, 96:887–901.

Kang, S. and Cai, J. (2010). Asymptotic results for fitting marginal hazards models from stratified case-cohort studies with multiple disease outcomes. *J Korean Stat Soc.*, 39:371–385.

Kang, S., Cai, J., and Chambless, L. (2012). Marginal additive hazards model for case-cohort studies with multiple disease outcomes: an application to the atherosclerosis risk in communities (aric) study. *Biostatistics*, 0:1–24.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrika*, 48:795–806.

Knuiman, M. W., Divitini, M. L., Olynyk, J. K., Cullen, D. J., and Bartholomew, H. C. (2003). Serum ferritin and cardiovascular disease: A 17-year following-up study in busselton, western australia. *Am. J. Epidemiol.*, 158:144–9.

Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, 65:135–142.

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. (2000a). Additive hazards regression for case-cohort studies. *Biometrika*, 87:73–87.

Kulich, M. and Lin, D. Y. (2000b). Additive hazards regression with covariate measurement error. *J. Am. Statist. Assoc.*, 95:238–248.

Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in caes-cohort study. *J. Am. Statist. Assoc.*, 99:832–44.

Langholz, B. and Thomas, D. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *Am.J. Epidemiol*, 131:169–76.

Lee, C. R., North, K., Bray, M., J., C. D., Heiss, G., and Zeldin, D. C. (2008). Cyclooxygenase polymorphisms and risk of cardiovascular events: The atherosclerosis risk in communities (aric) study. *Clin. Pharmacol. Ther.*, 83:52–60.

Lin, D. Y., Oakes, D., and Ying, Z. (1998). Additive hazards regression for current status data. *Biometrika*, 85:289–98.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc. B.*, 62:711–730.

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.* Springer, New York.

Lu, S. and Shih, J. H. (2006). Case-cohort designs and anlysis for clustered failure time data. *Biometrics*, 62:1138–48.

Lu, W. and Tsiatis, A. A. (2006). Miscellanea semiparametric transformation models for the case-cohort study. *Biometrika*, 93:207–214.

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89:649–58.

Pipper, C. B. and Martinusse, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazard. *J. R. Statist. Soc. 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. L. and Breslow, N. E. (1978). Retrospective studies and failure time models. *Biometrika*, 65:153–158.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56:1016–1022.

Samuelsen, S. O., Anestad, H., and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scan. J. Statist.*, 34:103–19.

Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.*, 34:103–19.

Shen, Y. and Cheng, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–100.

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

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Am. Statist. Assoc.*, 93:1164–75.

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

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Thomas, D. C. (1977). Addendum to methods of cohort analysis: Appraisal by application to asbestos mining. *J. R. Statist. Soc.*, 140:483–485.

Ulrich, C. M., Bigler, J., Sibert, J., Greene, E., Sparks, R., Carlson, C. S., and Potter, J. D. (2002). Cyclooxygenase 1 (cox1) polymorphisms in africanamerican and caucasian populations. *Hum Mutat*, 20:409–410.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, 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*, 76:117–23.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.*, 84:1065–73.

Yin, G. (2007). Model checking for additive hazards model with multivariate survival data. *J. Multi. Aal.*, 98:1018–1032.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika*, 91:801–818.

Zeng, D. and Cai, J. (2010). Additive transformation models for clustered failure time data. *Lifetime Data Anal*, 16:333–352.

Zhang, H., Schaubel, D. E., and D., K. J. (2011). Proportional hazards regression for the analysis of clustered survival data from case-cohort studies. *Biometrics*, 67:18–28.