Published in final edited form as:

J Korean Stat Soc. 2010 September ; 39(3): 371–385. doi:10.1016/j.jkss.2010.03.005.

Asymptotic results for fitting marginal hazards models from stratified case-cohort studies with multiple disease outcomes

Sangwook Kang and

Department of Epidemiology and Biostatistics, University of Georgia, Athens, Georgia 30602, United States

Jianwen Cai^{*,1}

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States

Sangwook Kang: skang@uga.edu

Abstract

In stratified case-cohort designs, samplings of case-cohort samples are conducted via a stratified random sampling based on covariate information available on the entire cohort members. In this paper, we extended the work of Kang & Cai (2009) to a generalized stratified case-cohort study design for failure time data with multiple disease outcomes. Under this study design, we developed weighted estimating procedures for model parameters in marginal multiplicative intensity models and for the cumulative baseline hazard function. The asymptotic properties of the estimators are studied using martingales, modern empirical process theory, and results for finite population sampling.

Keywords

marginal hazards model; multivariate failure times; stratified case-cohort design; survival analysis; weighted estimating equations

1 Introduction

The case-cohort study design (Prentice, 1986) has been proposed to reduce the cost and effort arising in large cohort studies with time-to-event data. The amount of reduction could be substantial especially if the main disease of interest is rare and the main covariate of interest (exposure) is expensive to measure since the case-cohort design requires measurement of the exposure only on a subset of the whole cohort. Specifically, the sampling in the case-cohort design is comprised of the following two steps. First, a subset called subcohort from the entire cohort is sampled randomly regardless of failure status. Second, remaining cases outside the subcohort are sampled. Information on the exposure is obtained only on these sampled subjects.

© 2010 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

^{*}Corresponding author: cai@bios.unc.edu (Jianwen Cai).

¹Tel.: +1 919 9667788, fax: +1 919 9663804

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

While the exposure will be available only for the case-cohort sample, some less expensive covariates such as age, gender, or a correlate of the exposure might be easily obtained for all the cohort members. In such cases, the subcohort could be sampled via a stratified simple random sampling based on strata defined by some of these covariates. This stratified case-cohort design could lead to a large efficiency gain compared to the unstratified counterpart since the latter ignores the available information.

For a single disease outcome assuming independent failure times among subjects in Cox models, many statistical methods have been proposed and studied for data from unstratified case-cohort studies (Prentice, 1986; Self & Prentice, 1988; Barlow, 1994; Chen & Lo, 1999; Chen, 2001) and stratified case-cohort designs (Borgan et al., 2000; Kulich & Lin, 2004).

The case-cohort design, among several other study designs which have been proposed for the similar purpose, is known to have advantage since the same subcohort can be utilized for different disease outcomes (Langholz & Thomas, 1990; Wacholder et al., 1991). When more than one disease outcomes from a subject are of interest, failure time data from the same subject constitute multivariate failure time data wherein correlations among the failure times within the same subject should be accounted for. Such multivariate failure time data are frequently encountered in many biomedical studies. One interesting example is a study of relationship between serum ferritin and coronary heart disease and stroke events in the Busselton Health Study (Cullen, 1972). In order to reduce costs and preserve stored serum, case-cohort sampling was used. It is of scientific interest to compare the effects of serum ferritin on coronary heart disease and on stroke. A subject can experience both coronary heart disease and stroke, and times to coronary heart disease and stroke events observed from the same subject are obviously not independent. In this case, methods developed for single disease outcome assuming univariate failure time data can no longer be directly applied.

Statistical methods which address this problem have been somewhat limited. Recently, we proposed weighted estimating equation methods for failure time data with multiple disease outcomes from case-cohort studies assuming marginal hazards models (Kang & Cai, 2009). In that paper, the generalized case-cohort design which allows sampling of cases outside the subcohort was considered. It is more realistic and flexible when considering multiple disease outcomes since not all disease outcomes need to be rare or number of cases need to be small (Breslow & Wellner, 2007).

The main purpose of this article is to extend the study design considered in Kang & Cai (2009) to a stratified case-cohort design, propose estimation procedure under such study designs, and provide a detailed derivation for the asymptotic properties of the proposed estimators. The model and the estimating procedures for regression coefficients and cumulative baseline hazards function are presented in Section 2. The corresponding asymptotic properties are stated and proven in Section 3. A brief summary and discussion is provided in Section 4.

2 Model, study design, and estimating procedure

2.1 Model

Suppose a cohort with n subjects can be divided into L mutually exclusive strata using some information available for all the cohort members. Let T_{lik} be the failure time for the k th type of disease outcome ($k = 1, \dots, K$) of the i th subject ($i = 1, \dots, n_l$) within the l th stratum ($l = 1, \dots, L$). Due to right censoring, what one actually observes for the k th type of disease outcome within the l th stratum is $X_{lik} = \min(T_{lik}, C_{lik})$ where C_{lik} is the potential censoring time. Given p -vector of covariates $Z_{lik}(t)$, T_{lik} and C_{lik} are assumed to be independent. We assume that all the time-dependent covariates in $Z_{lik}(t)$ are “external”, i.e., they are not affected by the disease processes, as described by Kalbfleisch & Prentice (2002). Let $\Delta_{lik} = I(T_{lik} \leq C_{lik})$, $N_{lik}(t) =$

$\Delta_{lik}I(X_{lik} \leq t)$, and $Y_{lik}(t) = I(X_{lik} \geq t)$ where $I(\cdot)$ is an indicator function. Let $\lambda_{lik}(t)$ denote the corresponding marginal hazards function and let τ denote the study end time.

For the k th type of disease outcome of the i th subject within the l th stratum, the marginal hazards function $\lambda_{lik}(t)$ is assumed to be associated with the covariate $Z_{lik}(t)$ by

$$\lambda_{lik}\{t|Z_{lik}(t)\} = Y_{lik}(t) \lambda_{0k}(t) e^{\beta_0^T Z_{lik}(t)}, \quad (1)$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for the k th disease outcome and $\beta_0 = (\beta_{01}, \dots, \beta_{0K})^T$ is a $p \times 1$ vector of regression parameters.

2.2 Study designs

Let V denote the discrete random variable for indicating the stratum. We consider sampling procedures depending on V . We assume that T_{lik} is independent of V_{ik} given $Z_{lik}(\cdot)$, i.e., V_{ik} affects the failure time only through the covariates (Kulich & Lin, 2004).

First, we consider a direct extension of the stratified case-cohort design for a single disease outcome to multiple disease outcomes and refer to this design as the “original” stratified case-cohort design. Under the original stratified case-cohort design for multiple disease outcomes, the subcohort is selected by a stratified random sampling. Specifically, for the l th stratum, we select a fixed size $\tilde{n}_l$ subjects from n_l subjects in the entire cohort via simple random sampling without replacement. Thus, each subject in the l th stratum has the same probability $pr(\xi_{li} = 1) = \tilde{\alpha}_l = \tilde{n}_l/n_l$ of being selected into the subcohort where ξ_{li} is subcohort sampling indicator for the i th subject in the l th stratum. We obtain covariate measurements only on the subcohort members and all the remaining cases outside the subcohort. For the k th type of disease outcome, complete data, $\{X_{lik}, \Delta_{lik}, Z_{lik}(t), 0 \leq t \leq X_{ik}, V_{ik}\}$ are available for the subcohort members ($\xi_{li} = 1$) or cases ($\Delta_{lik} = 1$). Note that, for cases, information on V_{ik} do not need to be available. For the non-subcohort controls ($\xi_{li} = 0$ and $\Delta_{lik} = 0$), only partial data, $\{X_{lik}, \Delta_{lik}, V_{ik}\}$ are available.

Since we consider more than one disease outcome, it might be more realistic that some of diseases outcomes are not rare or the numbers of cases are not small. In this situation, obtaining covariate information on all the cases outside the subcohort might not be feasible. Thus, we consider a stratified case-cohort design which allows the sampling of cases outside the subcohort to be different for different stratum and refer to this design as the “generalized” stratified case-cohort design.

Under the generalized stratified case-cohort design for multiple disease outcomes, sampling of the subcohort members follows the same routine as before: for the l th stratum, sampling a fixed size $\tilde{n}_l$ subjects from n_l subjects in the entire cohort via simple random sampling without replacement. After the sampling of a subcohort, instead of sampling all the cases outside the subcohort, we allow sampling a fraction of cases for each of the disease outcomes. Specifically, for the k th type of disease ($k = 1, \dots, K$) within the l th stratum ($l = 1, \dots, L$), we select a fixed number of m_l^k cases who are outside the subcohort via simple random sampling without replacement. Then, each case outside the subcohort has the same probability $\tilde{q}_{lk} = pr(\eta_{lik} = 1 | \Delta_{lik} = 1, \xi_{li} = 0) = m_l^k / (n_l^k - \tilde{n}_l^k)$ of being sampled where η_{lik} is the case sampling indicator, n_l^k is the number of the k th type of disease cases within the l th stratum in the cohort and $\tilde{n}_l^k$ is the number of k th disease cases within the l th stratum in the subcohort.

Note that due to the sampling scheme, $(\eta_{11k}, \dots, \eta_{l n_l k})$ are correlated, however, $(\eta_{l 1 k}, \dots, \eta_{l n_l k})$ and $(\eta_{l' 1 k}, \dots, \eta_{l' n_{l' k}})$ are not correlated for $k \neq k'$ or $l \neq l'$. We obtain covariate

measurements only on the sampled subject. Thus complete data, $\{X_{lik}, \Delta_{lik}, Z_{lik}(t), 0 \leq t \leq X_{lik}, V_{ik}\}$, are available for the sub-cohort members ($\xi_{li} = 1$) or sampled cases outside the subcohort ($\eta_{lik} = 1$). Only partial data, $\{X_{lik}, \Delta_{lik}, V_{ik}\}$ are available for all others ($\xi_{li} = 0$ and $\eta_{lik} = 0$). Note that the generalized stratified case-cohort design includes the original stratified case-cohort design as a special case since if $\tilde{q}_{lk} = 1$ for all k and l , it reduces to the original stratified case-cohort design which samples all the cases outside the subcohort. Also, if we do not consider the strata for the cohort, i.e., $L = 1$, then it reduces to the generalized case-cohort design considered by Kang & Cai (2009).

2.3 Estimation of regression parameters under the original stratified case-cohort design

For full cohort data, Wei et al. (1989) proposed the following pseudo-likelihood score equations for the estimation of the hazards regression parameter β_0 :

$$U(\beta) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \left\{ Z_{ik}(u) - \frac{S_k^{(1)}(\beta, u)}{S_k^{(0)}(\beta, u)} \right\} dN_{ik}(u), \quad (2)$$

where $S_k^{(d)}(\beta, t) = n^{-1} \sum_{i=1}^n Y_{ik}(t) Z_{ik}(t)^{\otimes d} e^{\beta^T Z_{ik}(t)}$ for $d = 0$ and 1 , and $a^{\otimes 2} = aa^T$, $a^{\otimes 1} = a$, $a^{\otimes 0} = 1$ for a vector a . Since these estimating equations do not have analytical solutions, they need to be solved iteratively, for example, by Newton-Raphson method (Thisted, 1988).

For data from the original stratified case-cohort studies, $S_k^{(d)}(\beta, t)$ ($d=0, 1$) in (2) cannot be calculated due to the incompleteness of the data. In order to handle this problem, we consider the idea of weighting the incomplete data by the inverse selection probability (Horvitz & Thompson, 1951). Specifically, we consider

$\widehat{S}_k^{(d)}(\beta, t) = n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \rho_{lik}(t) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)}$ in place of $S_k^{(d)}(\beta, t)$ for $d = 0$ and 1 where $\rho_{lik}(t)$ is a possibly time-varying weight function, incorporated to account for the sampling scheme and has the following form:

$$\rho_{lik}(t) = \Delta_{lik} + (1 - \Delta_{lik}) \xi_{li} / \widehat{\alpha}_{lk}(t) \quad \text{where} \quad \widehat{\alpha}_{lk}(t) = \frac{\sum_{i=1}^{n_l} (1 - \Delta_{lik}) \xi_{li} Y_{lik}(t)}{\sum_{i=1}^{n_l} (1 - \Delta_{lik}) Y_{lik}(t)}.$$

Then, for the estimation of β_0 , we propose the following pseudo-partial-likelihood score equations $\widehat{U}(\beta) = 0_{p \times 1}$, where

$$\widehat{U}(\beta) = \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \int_0^\tau \left\{ Z_{lik}(u) - \frac{\widehat{S}_k^{(1)}(\beta, u)}{\widehat{S}_k^{(0)}(\beta, u)} \right\} dN_{lik}(u), \quad (3)$$

A time-invariant version of the weight function which uses $\tilde{\alpha}_l$ in place of $\widehat{\alpha}_{lk}(t)$ may also be considered.

The solution to $\widehat{U}(\beta) = 0_{p \times 1}$ is defined to be the estimator of the hazards regression parameter β_0 . We will denote the estimator which uses time-invariant weight functions as $\hat{\beta}_I$ and time-varying weight functions as $\hat{\beta}_{II}$. The corresponding pseudo-partial-likelihood functions will be denoted by $\widehat{U}_I(\beta)$ and $\widehat{U}_{II}(\beta)$, respectively.

2.4 Estimation of regression parameters under the generalized stratified case-cohort design

For the generalized stratified case-cohort design, the weight function needs to be modified to appropriately account for the sampling of cases outside the subcohort. Specifically, cases outside the subcohort who are sampled are weighted by $\widehat{q}_{lk}^{-1}(t)$ where $\widehat{q}_{lk}(t)$ denotes the number of sampled non-subcohort cases with the k th type of disease outcome in the l th stratum divided by the number of non-subcohort cases with the k th type of disease outcome in the l th stratum remaining in the risk set at time t . Then, the weight function $\omega_{lik}(t)$ has the following form:

$$\omega_{lik}(t) = (1 - \Delta_{lik}) \frac{\xi_{li}}{\alpha_{lk}(t)} + \Delta_{lik} \xi_{li} + \Delta_{lik} (1 - \xi_{li}) \frac{\eta_{lik}}{\widehat{q}_{lk}(t)} \text{ where } \widehat{q}_{lk}(t) = \frac{\sum_{i=1}^{n_l} \Delta_{lik} (1 - \xi_{li}) \eta_{lik} Y_{lik}(t)}{\sum_{i=1}^{n_l} \Delta_{lik} (1 - \xi_{li}) Y_{lik}(t)}.$$

Note that the proposed weight functions reduce to the ones for the original stratified case-cohort design if all cases outside the subcohort are sampled since $\widehat{q}_{lk} = 1$ for all k and l .

For the estimation of β_0 under the generalized stratified case-cohort design, the following weighted estimating functions with the weight function $\omega_{lik}(t)$ is considered:

$$\tilde{U}(\beta) = \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \int_0^\tau \omega_{lik}(u) \left\{ Z_{lik}(u) - \frac{\tilde{S}_k^{(1)}(\beta, u)}{\tilde{S}_k^{(0)}(\beta, u)} \right\} dN_{lik}(u), \quad (4)$$

where $\tilde{S}_k^{(d)}(\beta, t) = n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \omega_{lik}(t) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)}$ for $d = 0$ and 1 .

The solution to the equations $\tilde{U}(\beta) = 0_{p \times 1}$ is the estimator for the hazards regression parameter β_0 . We will denote the estimator which uses time-invariant weight functions as $\tilde{\beta}_I$ and time-varying weight functions as $\tilde{\beta}_{II}$, respectively. The corresponding weighted estimating functions are $\tilde{U}_I(\beta)$ and $\tilde{U}_{II}(\beta)$, respectively.

2.5 Estimation of the cumulative baseline hazard function

Let $\Lambda_{0k}(t)$ denote the cumulative baseline hazard function for the k th type of disease outcome at time t where $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u) du$. Then for the estimator of $\Lambda_{0k}(t)$, we consider the following Breslow-Aalen type estimators $\hat{\Lambda}_{0k}(\beta, t)$ for the original stratified case-cohort design and $\tilde{\Lambda}_{0k}(\beta, t)$ for the generalized stratified case-cohort design where

$$\hat{\Lambda}_{0k}(\beta, t) = \int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \rho_{lik}(u) dN_{lik}(u)}{n \hat{S}_k^{(0)}(\beta, u)} \text{ and} \quad (5)$$

$$\tilde{\Lambda}_{0k}(\beta, t) = \int_0^t \frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \omega_{lik}(u) dN_{lik}(u)}{n \tilde{S}_k^{(0)}(\beta, u)}, \quad (6)$$

respectively.

We, again, use the superscript I ($\hat{\Lambda}_{0k}^I(\cdot, t)$, $\tilde{\Lambda}_{0k}^I(\cdot, t)$) and II ($\hat{\Lambda}_{0k}^{II}(\cdot, t)$, $\tilde{\Lambda}_{0k}^{II}(\cdot, t)$) to denote the estimator using time-invariant and time-varying weight functions, respectively.

3 Asymptotic properties

We will focus on the asymptotic properties of the estimators under the generalized stratified case-cohort design, $\tilde{\beta}_I$, $\tilde{\beta}_{II}$, $\tilde{\Lambda}_{0k}^I(\cdot, t)$, and $\tilde{\Lambda}_{0k}^{II}(\cdot, t)$. This is because the estimators under the original stratified case-cohort studies, $\hat{\beta}_I$, $\hat{\beta}_{II}$, $\hat{\Lambda}_{0k}^I(\cdot, t)$, and $\hat{\Lambda}_{0k}^{II}(\cdot, t)$, are special cases of those under the generalized stratified case-cohort studies. Thus, their asymptotic properties can be directly drawn from those under the generalized case-cohort studies.

3.1 Conditions

In order to establish the consistency and asymptotic normality of the estimators for the generalized stratified case-cohort studies, the following sets of conditions are needed:

- (A) $(T_{li}, C_{li}, Z_{li}), i = 1, \dots, n_l$ 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) $\text{pr}\{Y_{lik}(\tau) > 0\} > 0$ for all $i = 1, \dots, n_l$, $k = 1, \dots, K$ and $l = 1, \dots, L$.
- (C) $|Z_{lik}(0)| + \int_0^\tau |dZ_{lik}(u)| < C_z < \infty$ almost surely for all $i = 1, \dots, n_l$, $k = 1, \dots, K$, and $l = 1, \dots, L$ where Z_{likj} is the j th component of Z_{lik} and C_z is some constant.
- (D) The matrix $A_k(\beta_0) = \int_0^\tau v_k(\beta_0, u) s_k^{(0)}(\beta_0, u) \lambda_{0k}(u) du$ is positive definite.
- (E) $\int_0^\tau \lambda_{0k}(u) du < \infty$, for all $k = 1, \dots, K$.
- (F) There exists a neighborhood $\mathcal{B}$ of β_0 that satisfies the following conditions, as $n \rightarrow \infty$: for all $k = 1, \dots, K$, and $d = 0, 1, 2$, $\beta \in \mathcal{B}$

$$\sup_{t \in [0, \tau]} \|S_k^{(d)}(\beta, t) - s_k^{(d)}(\beta, t)\| \xrightarrow{P} 0$$
where $E\{S_k^{(d)}(\beta, t)\} = s_k^{(d)}(\beta, t)$ 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 $\mathcal{B} \times [0, \tau]$.

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

- (G) As $n \rightarrow \infty$,
 - i. For all $l = 1, \dots, L$, $\tilde{\alpha}_l$ converges to a constant $\alpha_l \in (0, 1]$;
 - ii. For all $k = 1, \dots, K$, and $l = 1, \dots, L$, $\tilde{q}_{lk}$ converges to a constant q_{lk} in $(0, 1]$.
- (H) n_l/n converges to a constant $p_l \in [0, 1]$ for all $l = 1, \dots, L$ as $n \rightarrow \infty$.

Here and hereafter the norms for the vector a , matrix A , and function f are defined as the following:

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

3.2 Asymptotic properties of $\tilde{\beta}_I$ and $\tilde{\Lambda}_{0k}^I(\tilde{\beta}_I, t)$

The key component of the derivation of the asymptotic results involves a decomposition of the proposed estimating function into three asymptotically uncorrelated pieces and some negligible

terms. These three components represent, respectively, the whole cohort counterpart, one arising from sampling of a subcohort, and one arising from sampling of cases outside the subcohort.

Let us provide some lemmas which will be frequently used in proving the theorems.

Lemma 1—Let $W_n(t)$ and $G_n(t)$ be two sequences of bounded processes. For some constant τ , assume that the following conditions (a) – (c) hold where

- a. $\sup_{0 \leq t \leq \tau} \|W_n(t) - W(t)\| \xrightarrow{P} 0$ for some bounded process $W(t)$,
- b. $W_n(t)$ is monotone on $[0, \tau]$ and
- c. $G_n(t)$ converges to a zero-mean process with continuous sample paths. Under Conditions (a) – (c),

$$\sup_{0 \leq t \leq \tau} \left\| \int_0^t \{W_n(u) - W(u)\} dG_n(u) \right\| \xrightarrow{P} 0, \quad \sup_{0 \leq t \leq \tau} \left\| \int_0^t G_n(u) d\{W_n(u) - W(u)\} \right\| \xrightarrow{P} 0.$$

PROOF—Lemma 1 is given as a lemma in Lin (2000). Its proof follows from the strong embedding theorem (Shorack & Wellner, 1986, p47–48), lemma 1 of Lin et al. (2000) and the triangular argument of a norm.

Lemma 2 is an extension of the proposition in Kulich & Lin (2000).

Lemma 2—Let $\xi = (\xi_1, \dots, \xi_n)$ be a random vector containing $\tilde{n}$ ones and $n - \tilde{n}$ zeros, with each permutation equally likely. Let $B_i(t)$, $i = 1, \dots, n$, be i.i.d. real-valued random processes on $[0, \tau]$ with $E\{B_i(t)\} = \mu_B(t)$, $Var\{B_i(0)\} < \infty$ and $Var\{B_i(\tau)\} < \infty$. Let $B(t) = \{B_1(t), \dots, B_n(t)\}$ and ξ be independent. Suppose that almost all paths of $B_i(t)$ have finite variation. Then,

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

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

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

converges in probability to 0 uniformly in t .

PROOF—This lemma is an extension of the proposition in Kulich & Lin (2000). The proof of this lemma follows from Hájek (1960)'s central limit theorem for finite population sampling and Example 3.6.14 of van der Vaart & Wellner (1996). Specifically, suppose first that the $B_i(t)$'s have nondecreasing sample paths then the finite-dimensional convergence follows from Hájek (1960)'s central limit theorem for finite population sampling while the tightness follows from Example 3.6.14 of van der Vaart & Wellner (1996). In the general case, since almost every path $b(t)$ of $B(t)$ have finite variation, $b(t)$ can be written as $b_1^*(t) - b_2^*(t)$, where $b_1^*(t)$ and $b_2^*(t)$ are nonnegative, nondecreasing in t . Hence $B_i(t) = B_{i1}^*(t) - B_{i2}^*(t)$, where $B_{i1}^*(t)$ and $B_{i2}^*(t)$ are marginally tight since they meet the condition of Example 3.6.14 of van der Vaart & Wellner

(1996). This implies that they are jointly tight. The joint finite-dimensional convergence of the normalized $n^{-1/2} \sum_{i=1}^n \xi_i \{B_{i1}^*(t) - \mu_{B_{i1}^*}(t)\}$ and $n^{-1/2} \sum_{i=1}^n \xi_i \{B_{i2}^*(t) - \mu_{B_{i2}^*}(t)\}$ follows again from Hájek (1960)'s central limit theorem for finite population sampling. Therefore, $n^{-1/2} \sum_{i=1}^n \xi_i \{B_i(t) - \mu_{B_i}(t)\}$ converges weakly in $\ell^\infty[0, \tau]$ to zero mean Gaussian processes. It then follows that $n^{-1} \sum_{i=1}^n \xi_i \{B_i(t) - \mu_{B_i}(t)\}$ converges to 0 in probability uniformly in t .

Note that ξ_{li} is the subcohort membership indicator and η_{lik} is the sampling indicator for the i th subject with the k th disease within the l th stratum outside the subcohort where both the sampling of the subcohort and the cases outside the subcohort were conducted by simple random sampling without replacement. Thus, it is clear that our ξ_{li} 's and η_{lik} 's satisfy the conditions in lemma 2.

Theorem 1—Under Conditions (A) – (H), $\tilde{\beta}_I$ solving $\tilde{U}_I(\beta) = 0$ is a consistent estimator of β_0 . Also, $n^{1/2}(\tilde{\beta}_I - \beta_0)$ is asymptotically normally distributed with mean zero and with variance matrix $\Sigma_I(\beta_0)$ of the following form

$$A(\beta_0)^{-1} \sum_{l=1}^L p_l \left\{ Q_l(\beta_0) + \frac{1 - \alpha_l}{\alpha_l} V_l^{I,(1)}(\beta_0) + (1 - \alpha_l) \sum_{k=1}^K pr(\Delta_{lk}=1) \left(\frac{1 - q_{lk}}{q_{lk}} \right) V_{lk}^{I,(2)}(\beta_0) \right\} A(\beta_0)^{-1}$$

where

$$\begin{aligned} A(\beta) &= \sum_{k=1}^K A_k(\beta), \quad Q_l(\beta) = E_l \left\{ \sum_{k=1}^K M_{z,l1k}(\beta, \tau) \right\}^{\otimes 2}, \\ V_l^{I,(1)}(\beta) &= Var_l \left\{ \sum_{k=1}^K \int_0^\tau (1 - \Delta_{lk}) R_{lk}(\beta, u) d\Lambda_{0k}(u) \right\}, \\ V_{lk}^{I,(2)}(\beta) &= Var_l \{ M_{z,l1k}(\beta, \tau) | \Delta_{lk}=1, \xi_{l1}=0 \}, \quad R_{lk}(\beta, t) = Y_{lik}(t) \tilde{Z}_{lik}(\beta, t) e^{\beta^T Z_{lik}(t)}, \\ e_k(\beta, t) &= \frac{s_k^{(1)}(\beta, t)}{s_k^{(0)}(\beta, t)}, \quad v_k(\beta, t) = \frac{s_k^{(2)}(\beta, t) s_k^{(0)}(\beta, t) - s_k^{(1)}(\beta, t)^{\otimes 2}}{s_k^{(0)}(\beta, t)^2}, \\ A_k(\beta) &= \int_0^\tau v_k(\beta, u) s_k^{(0)}(\beta, u) \lambda_{0k}(u) du, \\ \tilde{Z}_{lik}(\beta, t) &= Z_{lik}(t) - e_k(\beta, t), \quad M_{z,l1k}(\beta, t) = \int_0^t \tilde{Z}_{lik}(\beta, u) dM_{lik}(\beta, u), \\ \text{and } M_{lik}(\beta, t) &= N_{lik}(t) - \int_0^t Y_{lik}(u) \lambda_{0k}(u) e^{\beta^T Z_{lik}(u)} du. \end{aligned}$$

Note that E_l , Var_l , and Cov_l denote the expectation, the variance and the covariance within the l th stratum, respectively.

We now study the asymptotic properties of $\tilde{\Lambda}_{0k}^I(\tilde{\beta}_I, t)$ ($k=1, \dots, K$). Let

$W^I(t) = n^{1/2} [\{ \tilde{\Lambda}_{01}^I(\tilde{\beta}_I, t) - \Lambda_{01}(t) \}, \dots, \{ \tilde{\Lambda}_{0K}^I(\tilde{\beta}_I, t) - \Lambda_{0K}(t) \}]^T$ and $\mathcal{W}^I(t) = \{ \mathcal{W}_1^I(t), \dots, \mathcal{W}_K^I(t) \}^T$ where $\mathcal{W}^I(t)$ is a zero-mean Gaussian process with the covariance function between $\mathcal{W}_j^I(t_1)$ and $\mathcal{W}_k^I(t_2)$ ($1 \leq j, k \leq K$ and $0 \leq t_1, t_2 \leq \tau$) is $\phi_{jk}^I(t_1, t_2)(\beta_0)$ where

$$\begin{aligned}
\varphi_{jk}^I(t_1, t_2)(\beta_0) &= \sum_{l=1}^L p_l \left(E_l \{v_{l1j}(\beta_0, t_1) v_{l1k}(\beta_0, t_2)\} + \frac{1-\alpha_l}{\alpha_l} E_l \{\psi_{l1j}^I(\beta_0, t_1) \psi_{l1k}^I(\beta_0, t_2)\} + (1-\alpha_l) \right. \\
&\quad \times \left[I(j=k) pr(\Delta_{l1k}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) \text{Cov}_l \left\{ \int_0^{t_1} \frac{dM_{l1k}(u)}{s_k^{(0)}(\beta_0, u)}, \int_0^{t_2} \frac{dM_{l1k}(u)}{s_k^{(0)}(\beta_0, u)} \middle| \Delta_{l1k}=1, \xi_{l1}=0 \right\} \right. \\
&\quad + pr(\Delta_{l1j}=1) \left(\frac{1-q_{lj}}{q_{lj}} \right) \text{Cov}_l \left\{ \int_0^{t_1} \frac{dM_{l1j}(u)}{s_j^{(0)}(\beta_0, u)}, r_k(\beta_0, t_2)^T A(\beta_0)^{-1} \mathbf{M}_{z,l1j}(\beta_0) \middle| \Delta_{l1j}=1, \xi_{l1}=0 \right\} \\
&\quad + pr(\Delta_{l1k}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) \text{Cov}_l \left\{ \int_0^{t_2} \frac{dM_{l1k}(u)}{s_k^{(0)}(\beta_0, u)}, r_j(\beta_0, t_1)^T A(\beta_0)^{-1} \mathbf{M}_{z,l1k}(\beta_0) \middle| \Delta_{l1k}=1, \xi_{l1}=0 \right\} \\
&\quad + \sum_{m=1}^K pr(\Delta_{l1m}=1) \left(\frac{1-q_{lm}}{q_{lm}} \right) r_j(\beta_0, t_1)^T A(\beta_0)^{-1} \text{Var}_l \{ \mathbf{M}_{z,l1m}(\beta_0) | \Delta_{l1m}=1, \xi_{l1}=0 \} \\
&\quad \times A(\beta_0)^{-1} r_k(\beta_0, t_2) \Big], v_{lik}(\beta, t) = r_k(\beta, t)^T A(\beta)^{-1} \sum_{m=1}^K \mathbf{M}_{z,lim}(\beta) + \int_0^t \{s_k^{(0)}(\beta, u)\}^{-1} dM_{lik}(u), \\
&\quad \psi_{lik}^I(\beta, t) = r_k(\beta, t)^T A(\beta)^{-1} \sum_{m=1}^K (1 - \Delta_{lim}) \int_0^t R_{lim}(\beta, u) d\Lambda_{0m}(u) \\
&\quad + (1 - \Delta_{lik}) \int_0^t \frac{Y_{lik}(u) e^{\beta^T Z_{lik}(u)}}{s_k^{(0)}(\beta, u)} d\Lambda_{0k}(u) \quad \text{and} \quad r_k(\beta, t) = - \int_0^t e_k(\beta, u) d\Lambda_{0k}(u).
\end{aligned}$$

Also, let $D[0, \tau]^K$ be a metric space consisting of right-continuous functions $f(t)$ with left-hand limits where $f(t) = \{f_1(t), \dots, f_K(t)\}^T$ and $f_k(t): [0, \tau] \rightarrow \mathcal{R}$. This metric space is equipped with a uniform metric $d_k(f, g) = \sup_{k,t \in [0, \tau]} \{|f_k(t) - g_k(t)|: 1 \leq k \leq K\}$ for $f, g \in D[0, \tau]^K$.

Theorem 2—Under Conditions (A) – (H), for each $k = 1, \dots, K$, $\tilde{\Lambda}_{0k}^I(\tilde{\beta}_l, t)$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0, \tau]$. Also, $W^I(t)$ converges weakly to $w^I(t)$ in $D[0, \tau]^K$.

Proofs for Theorems 1 and 2 can be derived from those for their time-varying counterparts which will be provided in the next subsection. More detailed explanation on this is deferred to Section 3.4

3.3 Asymptotic properties of $\tilde{\beta}_{ll}$ and $\tilde{\Lambda}_{0k}^I(\tilde{\beta}_{ll}, t)$

In order to establish the asymptotic properties of the proposed estimators with time-varying weight functions, we need the following lemma on the asymptotic properties of time-varying sampling probability estimators $\hat{\alpha}_{lk}(t)$ and $\hat{q}_{lk}(t)$.

Lemma 3—For all $l = 1, \dots, L$ and $k = 1, \dots, K$,

- i. $\hat{\alpha}_{lk}(t)$ and $\tilde{\alpha}_l$ converge to the same limit uniformly in t and

$$n^{1/2} \{ \hat{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1} \} = \frac{1}{p_l \tilde{\alpha}_l E_l \{ (1 - \Delta_{l1k}) Y_{l1k}(t) \}} n^{-1/2} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) (1 - \Delta_{lik}) Y_{lik}(t) \right\} + o_p(1).$$

- ii. $\hat{q}_{lk}(t)$ and $\tilde{q}_{lk}$ converge to the same limit uniformly in t and

$$n^{1/2} \{ \hat{q}_{lk}(t)^{-1} - \tilde{q}_{lk}^{-1} \} = \frac{1}{\tilde{q}_{lk} p_l (1 - \tilde{\alpha}_l) E_l \{ \Delta_{l1k} Y_{l1k}(t) \}} n^{-1/2} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\eta_{lik}}{\tilde{q}_{lk}} \right) \Delta_{lik} (1 - \xi_{li}) Y_{lik}(t) \right\} + o_p(1).$$

PROOF—For each l and k , it follows from the Taylor expansion of $\hat{\alpha}_{lk}(t)^{-1}$ around $\tilde{\alpha}_l$,

$$\begin{aligned} \widehat{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1} &= -\frac{1}{\alpha_*(t)^2} \{\widehat{\alpha}_{lk}(t) - \tilde{\alpha}_l\} \\ &= \frac{\tilde{\alpha}_l}{\alpha_*(t)^2} \cdot \frac{1}{\sum_{i=1}^{n_l} (1 - \Delta_{lik}) Y_{lik}(t)} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) (1 - \Delta_{lik}) Y_{lik}(t) \right\} \end{aligned}$$

where $\alpha_*(t)$ is on the line segment between $\widehat{\alpha}_{lk}(t)$ and $\tilde{\alpha}_l$. Then,

$$n^{1/2} \left\{ \widehat{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1} \right\} = \frac{\tilde{\alpha}_l}{\alpha_*(t)^2} \cdot \frac{n}{n_l} \cdot \frac{n_l}{\sum_{i=1}^{n_l} (1 - \Delta_{lik}) Y_{lik}(t)} n^{-1/2} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) (1 - \Delta_{lik}) Y_{lik}(t) \right\}$$

By Glivenko-Cantelli lemma, $n_l^{-1} \sum_{i=1}^{n_l} (1 - \Delta_{lik}) Y_{lik}(t)$ converges to $E_l \{(1 - \Delta_{lik}) Y_{lik}(t)\}$ in probability uniformly in t . In view of lemma 2, $n^{-1/2} \sum_{i=1}^{n_l} \left(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1 \right) (1 - \Delta_{lik}) Y_{lik}(t)$ converges to a zero-mean Gaussian process since $(1 - \Delta_{lik}) Y_{lik}(t)$ is bounded and monotone function in t . This implies $n^{-1} \sum_{i=1}^{n_l} \left(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1 \right) (1 - \Delta_{lik}) Y_{lik}(t)$ converges to 0 in probability uniformly in t and consequently, $\widehat{\alpha}_{lk}(t)$ and $\tilde{\alpha}_l$ converges to the same limit uniformly in t . This ensures $\alpha_*(t)$ also converges to the same limit as $\tilde{\alpha}_l$. Combining these results, it follows from Slutsky's theorem and Condition (H) that

$$\begin{aligned} n^{1/2} \left\{ \widehat{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1} \right\} &= \frac{1}{\tilde{\alpha}_l p_l E_l \{(1 - \Delta_{lik}) Y_{lik}(t)\}} n^{-1/2} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) (1 - \Delta_{lik}) Y_{lik}(t) \right\} \\ &+ \left[\frac{\tilde{\alpha}_l}{\alpha_*(t)^2} \cdot \frac{n}{n_l} \cdot \frac{n_l}{\sum_{i=1}^{n_l} (1 - \Delta_{lik}) Y_{lik}(t)} - \frac{1}{\tilde{\alpha}_l p_l E_l \{(1 - \Delta_{lik}) Y_{lik}(t)\}} \right] n^{-1/2} \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) (1 - \Delta_{lik}) Y_{lik}(t) \\ &= \frac{1}{\tilde{\alpha}_l p_l E_l \{(1 - \Delta_{lik}) Y_{lik}(t)\}} n^{-1/2} \left\{ \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) (1 - \Delta_{lik}) Y_{lik}(t) \right\} + o_p(1). \end{aligned}$$

(ii) can be shown via similar arguments.

Now, we state the asymptotic behavior of the regression parameter estimator $\tilde{\beta}_{II}$ in the following theorem:

Theorem 3—Under Conditions (A) – (H), $\tilde{\beta}_{II}$ solving $\tilde{U}_{II}(\beta) = 0$ is a consistent estimator of β_0 . Also, $n^{1/2}(\tilde{\beta}_{II} - \beta_0)$ is asymptotically normally distributed with mean zero and with variance matrix $\Sigma_{II}(\beta_0)$ of the following form

$$A(\beta_0)^{-1} \sum_{l=1}^L p_l \left\{ Q_l(\beta_0) + \frac{1 - \alpha_l}{\alpha_l} V_l^{II,(1)}(\beta_0) + (1 - \alpha_l) \sum_{k=1}^K p_r(\Delta_{lk}=1) \left(\frac{1 - q_{lk}}{q_{lk}} \right) V_{lk}^{II,(2)}(\beta_0) \right\} A(\beta_0)^{-1}$$

where

$$\begin{aligned} V_l^{II,(1)}(\beta) &= \text{Var}_l \left(\sum_{k=1}^K (1 - \Delta_{lk}) \int_0^\tau \left[R_{lk}(\beta, u) - \frac{Y_{lk}(u) E_l \{(1 - \Delta_{lk}) R_{lk}(\beta, u)\}}{E_l \{(1 - \Delta_{lk}) Y_{lk}(u)\}} \right] d\Lambda_{0k}(u) \right) \text{ and} \\ V_{lk}^{II,(2)}(\beta) &= \text{Var}_l \left[M_{\xi, lk}(\beta, \tau) - \int_0^\tau Y_{lk}(u) \frac{E_l \{dM_{\xi, lk}(\beta, u) | \Delta_{lk}=1, \xi_{lk}=0\}}{E_l \{Y_{lk}(u) | \Delta_{lk}=1\}} \Big| \Delta_{lk}=1, \xi_{lk}=0 \right]. \end{aligned}$$

PROOF—The proof for the consistency of $\tilde{\beta}_H$ is based on the application of the Inverse Function Theorem in Foutz (1977). One can show $\tilde{\beta}_H$ to be consistent for β_0 provided:

- i. $\frac{\partial \{n^{-1} \tilde{U}_H(\beta)\}}{\partial \beta^T}$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0 ,
- ii. $\frac{\partial \{n^{-1} \tilde{U}_H(\beta_0)\}}{\partial \beta^T}$ is negative definite with probability going to one as $n \rightarrow \infty$,
- iii. $\frac{\partial \{n^{-1} \tilde{U}_H(\beta)\}}{\partial \beta}$ converges to $A(\beta_0)$ in probability uniformly for β in an open neighborhood about β_0 ,
- iv. $n^{-1} \tilde{U}_H(\beta) \rightarrow 0$ in probability.

One can write

$$\frac{\partial \{n^{-1} \tilde{U}_H(\beta)\}}{\partial \beta^T} = -n^{-1} \sum_{k=1}^K \int_0^\tau \tilde{V}_k(\beta, u) d \left\{ \sum_{l=1}^L \sum_{i=1}^{n_l} \omega_{lik}^H(u) N_{lik}(u) \right\} \quad \text{where} \quad (7)$$

$$\tilde{V}_k(\beta, t) = \frac{\tilde{S}_{H,k}^{(2)}(\beta, t) \tilde{S}_{H,k}^{(0)}(\beta, t) - \tilde{S}_{H,k}^{(1)}(\beta, t)^2}{\tilde{S}_{H,k}^{(0)}(\beta, t)^2}$$

Then, (i) is clearly satisfied on the basis of (7) and Condition (F). Now, following Andersen & Gill (1982),

$$\begin{aligned} \left\| \left[-\frac{\partial \{n^{-1} \tilde{U}_H(\beta)\}}{\partial \beta^T} \right] - A(\beta) \right\| &\leq \left\| \sum_{k=1}^K \int_0^\tau \{ \tilde{V}_k(\beta, u) - v_k(\beta, u) \} n^{-1} d \left\{ \sum_{l=1}^L \sum_{i=1}^{n_l} N_{lik}(u) \right\} \right\| \\ &\quad + \left\| \sum_{k=1}^K \int_0^\tau \{ \tilde{V}_k(\beta, u) - v_k(\beta, u) \} d \left[n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \{ \omega_{lik}^H(u) - 1 \} N_{lik}(t) \right] \right\| \\ &\quad + \left\| \sum_{k=1}^K \int_0^\tau v_k(\beta, u) n^{-1} d \left\{ \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta, u) \right\} \right\| + \left\| \sum_{k=1}^K n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \{ \omega_{lik}^H(u) - 1 \} \int_0^\tau v_k(\beta, u) dM_{lik}(\beta, u) \right\| \\ &\quad + \left\| \sum_{k=1}^K \int_0^\tau v_k(\beta, u) \{ \tilde{S}_{H,k}^{(0)}(\beta, u) - s_k^{(0)}(\beta, u) \} \lambda_{0k}(u) du \right\| \end{aligned} \quad (8)$$

Each of the terms on the right side of the above inequality can be shown to converge to zero, uniformly in $\beta \in \mathcal{B}$. To show the first term on the right side of (8), we will first show that

$$\sup_{\substack{t \in [0, \tau] \\ \beta \in \mathcal{B}}} \|\tilde{V}_k(\beta, t) - v_k(\beta, t)\| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty \text{ for } k=1, \dots, K.$$

It suffices to show that $\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| \tilde{S}_{H,k}^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \right\| \xrightarrow{P} 0$ as $n \rightarrow \infty$ for $d=0, 1$ and 2 . One can write

$$\begin{aligned}
n^{1/2} \left\{ \tilde{S}_{ll,k}^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \right\} &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) (1 - \Delta_{lik}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \\
&\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{\eta_{lik}}{\bar{q}_k} - 1 \right) \Delta_{lik} (1 - \xi_{li}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \\
&\quad + \left\{ \hat{\alpha}_{lk}(t)^{-1} - \tilde{\alpha}_l^{-1} \right\} n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} (1 - \Delta_{lik}) \xi_{li} Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \\
&\quad + \left\{ \hat{q}_{lk}(t)^{-1} - \tilde{q}_{lk}^{-1} \right\} n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \Delta_{lik} (1 - \xi_{li}) \eta_{lik} Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)}.
\end{aligned} \tag{9}$$

Then by lemma 3,

$$\begin{aligned}
(9) &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) (1 - \Delta_{lik}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \\
&\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{\eta_{lik}}{\bar{q}_k} - 1 \right) \Delta_{lik} (1 - \xi_{li}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \\
&\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\alpha_l} \right) (1 - \Delta_{lik}) \frac{Y_{lik}(t)}{p_l E_l \{ (1 - \Delta_{lik}) Y_{lik}(t) \}} \\
&\quad \times \left\{ n^{-1} \sum_{j=1}^{n_l} (1 - \Delta_{ljk}) \frac{\xi_{lj}}{\alpha_l} Y_{ljk}(t) Z_{ljk}(t)^{\otimes d} e^{\beta^T Z_{ljk}(t)} \right\} \\
&\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(1 - \frac{\eta_{lik}}{\bar{q}_k} \right) (1 - \xi_{li}) \Delta_{lik} \frac{Y_{lik}(t)}{p_l E_l \{ \Delta_{lik} Y_{lik}(t) \}} \\
&\quad \times \left\{ n^{-1} \sum_{j=1}^{n_l} \Delta_{ljk} \frac{(1 - \xi_{lj})}{(1 - \alpha_l)} \frac{\eta_{ljk}}{\bar{q}_k} Y_{ljk}(t) Z_{ljk}(t)^{\otimes d} e^{\beta^T Z_{ljk}(t)} \right\} + o_p(1)
\end{aligned} \tag{10}$$

It follows from lemma 2 that, for $d = 0, 1$ and 2 , $n^{-1} \sum_{j=1}^{n_l} (1 - \Delta_{ljk}) \frac{\xi_{lj}}{\alpha_l} Y_{ljk}(t) Z_{ljk}(t)^{\otimes d} e^{\beta^T Z_{ljk}(t)}$ and $n^{-1} \sum_{j=1}^{n_l} \Delta_{ljk} \frac{(1 - \xi_{lj})}{(1 - \alpha_l)} \frac{\eta_{ljk}}{\bar{q}_k} Y_{ljk}(t) Z_{ljk}(t)^{\otimes d} e^{\beta^T Z_{ljk}(t)}$ converge to $p_l E_l \{ (1 - \Delta_{lik}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \}$ and $p_l p_r (\Delta_{lik} = 1) E_l \{ Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \mid \Delta_{lik} = 1, \xi_{li} = 0 \}$ in probability uniformly in t under Condition (G), respectively. Thus, from (10)

$$\begin{aligned}
&n^{1/2} \left\{ \tilde{S}_{ll,k}^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \right\} \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} (1 - \Delta_{lik}) \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) Y_{lik}(t) \\
&\quad \times \left[Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} - \frac{E_l \{ (1 - \Delta_{lik}) Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \}}{E_l \{ (1 - \Delta_{lik}) Y_{lik}(t) \}} \right] \\
&\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} (1 - \xi_{li}) \Delta_{lik} \left(\frac{\eta_{lik}}{\bar{q}_k} - 1 \right) Y_{lik}(t) \left[Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \right. \\
&\quad \left. - \frac{E_l \{ Y_{lik}(t) Z_{lik}(t)^{\otimes d} e^{\beta^T Z_{lik}(t)} \mid \Delta_{lik} = 1, \xi_{li} = 0 \}}{E_l \{ Y_{lik}(t) \mid \Delta_{lik} = 1 \}} \right] + o_p(1)
\end{aligned} \tag{11}$$

It then follows from lemma 2 that, for $d = 0, 1$ and 2 , $n^{1/2} \left\{ \tilde{S}_{ll,k}^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \right\}$ converges weakly to zero-mean Gaussian processes under Condition (G). Consequently, together with condition (F),

$$\sup_{\substack{t \in [0, \tau] \\ \beta \in \mathcal{B}}} \| \tilde{S}_{ll,k}^{(d)}(\beta, t) - S_k^{(d)}(\beta, t) \| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty \text{ for } d=0, 1, \text{ and } 2. \tag{12}$$

Since $s_k^{(0)}(\beta, t)$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$ by condition (F), it follows from the above convergence results that for $k = 1, \dots, K$, $\tilde{U}_k(\beta, t)$ converges to $v_k(\beta, t)$ in probability uniformly in t and β .

By combining these results with the Lengart inequality for $n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} N_{lik}(\tau)$ (Andersen & Gill, 1982, p1115), it follows that the first term on the right side of (8) converges to zero in probability, uniformly in $\beta \in \mathcal{B}$, as $n \rightarrow \infty$.

The second term and the fourth terms on the right side of (8) can be shown to converge to zero by applying lemma 2. The third term can be shown to converge to zero by the Lengart

inequality for $\sum_{k=1}^K \int_0^\tau v_k(\beta, t) n^{-1} d \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(t)$ (Andersen & Gill, 1982, p1115).

Conditions (D), (E) and (F) ensure the boundedness of $\sup_{t, \beta} \{v_k(\beta, t)\}_{jj'}$ and $\Lambda_{0k}(\tau)$ for $k = 1, \dots, K$ and $j, j' = 1, \dots, p$. Thus, together with the uniform convergence of $\tilde{S}_{ll,k}^{(0)}(\beta, t)$ to $s_k^{(0)}(\beta, t)$ in probability, the last term on the right side of (8) converges to zero in probability, uniformly in $\beta \in \mathcal{B}$ as $n \rightarrow \infty$. Hence,

$$-\frac{\partial \{n^{-1} \tilde{U}_{ll}(\beta)\}}{\partial \beta^T} \xrightarrow{p} A(\beta) \text{ as } n \rightarrow \infty \text{ uniformly in } \beta \in \mathcal{B}$$

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

For (iv), we will show that $n^{-1/2} \tilde{U}_{ll}(\beta_0)$ is asymptotically equivalent to

$$\begin{aligned} & n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K (1 - \Delta_{lik}) \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) \int_0^\tau [R_{lik}(\beta_0, u) - Y_{lik}(u) \\ & \quad \times \frac{E_l\{(1 - \Delta_{lik})R_{lik}(\beta_0, u)\}}{E_l\{(1 - \Delta_{lik})Y_{lik}(u)\}}] d\Lambda_{0k}(u) + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \Delta_{lik} (1 - \xi_{li}) \left(\frac{\eta_{lik}}{\bar{q}_{lik}} - 1 \right) \\ & \quad \times \left[\mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) - \int_0^\tau Y_{lik}(u) \frac{E_l\{d\mathbf{M}_{\tilde{z},lik}(\beta_0, u) | \Delta_{lik}=1, \xi_{li}=0\}}{E_l\{Y_{lik}(u) | \Delta_{lik}=1\}} \right]. \end{aligned} \quad (13)$$

Specifically, one can decompose $n^{-1/2} \tilde{U}_{ll}(\beta_0)$ into the following four parts:

$$\begin{aligned} & n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \int_0^\tau \left\{ e_k(\beta, u) - \frac{\tilde{S}_{ll,k}^{(1)}(\beta_0, u)}{\tilde{S}_{ll,k}^{(0)}(\beta_0, u)} \right\} dM_{lik}(\beta_0, u) \\ & \quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \int_0^\tau \{\omega_{lik}^{II}(u) - 1\} \tilde{Z}_{lik}(\beta_0, u) dM_{lik}(u) \\ & \quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \int_0^\tau \{\omega_{lik}^{II}(u) - 1\} \left\{ e_k(\beta, u) - \frac{\tilde{S}_{ll,k}^{(1)}(\beta_0, u)}{\tilde{S}_{ll,k}^{(0)}(\beta_0, u)} \right\} dM_{lik}(\beta_0, u) \end{aligned} \quad (14)$$

The second term on the right-hand side of (14) can be shown to converge to zero uniformly in

t . Note that, for fixed t , $n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta_0, t)$ is a sum of i.i.d. zero-mean random variables. Based on Conditions (C) and (E), $M_{lik}(\beta_0, t)$ is of bounded variation and therefore can be written as a difference of two monotone functions in t . It then follows from the example

of 2.11.16 of van der Vaart & Wellner (1996, p215) that $n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta_0, t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{W}_{Mk}(t)$. It can be shown that $E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}$

$(s)\}^4] \leq C\{\Lambda_{0k}(t) - \Lambda_{0k}(s)\}^2$ for some constant $C > 0$. Specifically, $E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^4] = 3(E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2])^2$ since $\mathcal{W}_{Mk}(t)$ is a zero-mean normal random variable for a fixed t . Then $E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2] = E\{\mathcal{W}_{Mk}(t)^2\} + E\{\mathcal{W}_{Mk}(s)^2\} - 2E\{\mathcal{W}_{Mk}(t)\mathcal{W}_{Mk}(s)\} = E\{\mathcal{W}_{Mk}(t)^2\} -$

$$\begin{aligned} & \{E\mathcal{W}_{Mk}(t)^2\} \\ &= E\{M_{lik}(\beta_0, t)^2\} \\ &= E\left\{\int_0^t Y_{lik}(u) e^{\beta_0^T Z_{lik}(u)} \lambda_{0k}(u) du\right\}, \\ & E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2] \\ &= E\left\{\int_s^t Y_{lik}(u) e^{\beta_0^T Z_{lik}(u)} \lambda_{0k}(u) du\right\} \leq e^{C_z} E\left\{\int_s^t \lambda_{0k}(u) du\right\} \end{aligned}$$

$E\{\mathcal{W}_{Mk}(s)^2\}$ for $s \leq t$. Since $\Lambda_{0k}(\cdot)$ is differentiable and $\lambda_0(\cdot)$ is bounded on $[0, \tau]$, there exists a constant M , such that $\Lambda_{0k}(t) - \Lambda_{0k}(s) \leq M(t - s)$ for $s \leq t$. Therefore,

$$E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2] \leq C_z^*(t - s) \text{ and}$$

$E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^4] \leq 3(E[\{\mathcal{W}_{Mk}(t) - \mathcal{W}_{Mk}(s)\}^2])^2 \leq \tilde{C}_z^*(t - s)^2$ for some constant $\tilde{C}_z^*$. Then, by the Kolmogorov-Centsov Theorem (Karatzas & Shreve, 1988, p53), $\mathcal{W}_{Mk}(t)$ has continuous sample paths. In addition, since $\tilde{S}_{lik}^{(1)}(\beta_0, t)/\tilde{S}_{lik}^{(0)}(\beta_0, t)$ is of bounded variation based on (12) and Conditions (C) and (F), we can write $\tilde{S}_{lik}^{(1)}(\beta_0, t)/\tilde{S}_{lik}^{(0)}(\beta_0, t) = Z_{k1}^*(t) - Z_{k2}^*(t)$ where both $Z_{k1}^*(t)$ and $Z_{k2}^*(t)$ are nonnegative, monotone in t and bounded. Therefore, $\tilde{S}_{lik}^{(1)}(\beta_0, t)/\tilde{S}_{lik}^{(0)}(\beta_0, t)$ is a sum of two monotone functions. Hence, it follows from lemma 1 that the second term on the right-hand side of (14) converges to 0 uniformly in t .

By similar arguments, the fourth term on the right-hand side of (14) can be shown to converge to 0 uniformly in t .

The third term on the right-hand side of (14) can be further decomposed as

$$\begin{aligned} & n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K (1 - \Delta_{lik}) \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) \\ & + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K (1 - \Delta_{lik}) \xi_{li} \int_0^\tau \left\{ \hat{\alpha}_{lk}^{-1}(u) - \tilde{\alpha}_l^{-1} \right\} \tilde{Z}_{lik}(\beta_0, u) dM_{lik}(\beta_0, u) \\ & + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \Delta_{lik} (1 - \xi_{li}) \left(\frac{\eta_{lik}}{\tilde{q}_k} - 1 \right) \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) \\ & + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \Delta_{lik} (1 - \xi_{li}) \eta_{lik} \int_0^\tau \left\{ \hat{q}_{lk}^{-1}(u) - \tilde{q}_{lk}^{-1} \right\} \tilde{Z}_{lik}(\beta_0, u) dM_{lik}(\beta_0, u) \end{aligned} \quad (15)$$

The second term on the right side of (15) is asymptotically equivalent to

$$n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K (1 - \Delta_{lik}) \left(1 - \frac{\xi_{li}}{\alpha_l} \right) \int_0^\tau Y_{lik}(u) \frac{E_l\{(1 - \Delta_{lik})R_{lik}(\beta_0, u)\}}{E_l\{(1 - \Delta_{lik})Y_{lik}(u)\}} d\Lambda_{0k}(u) \quad (16)$$

by (i) in lemma 3 and applying lemma 2. Likewise, by (ii) in lemma 3 and applying lemma 2, the last term on the right side of (15) is asymptotically equivalent to

$$n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \Delta_{lik} (1 - \xi_{li}) \left(1 - \frac{\eta_{lik}}{\tilde{q}_k} \right) \int_0^\tau Y_{lik}(u) \frac{E_l\{d\mathbf{M}_{\tilde{z},lik}(\beta_0, u) | \Delta_{lik}=1, \xi_{li}=0\}}{E_l\{Y_{lik}(u) | \Delta_{lik}=1\}} \quad (17)$$

By combining (16) and (17), (15) is asymptotically equivalent to

$$\begin{aligned}
& n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K (1 - \Delta_{lik}) \left(\frac{\xi_{li}}{\alpha_l} - 1 \right) \int_0^\tau \left[R_{lik}(\beta_0, u) - Y_{lik}(u) \frac{E_l\{(1 - \Delta_{lik})R_{lik}(\beta_0, u)\}}{E_l\{(1 - \Delta_{lik})Y_{lik}(u)\}} \right] d\Lambda_{0k}(u) \\
& + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \Delta_{lik} (1 - \xi_i) \left(\frac{\eta_{lik}}{q_{lk}} - 1 \right) \\
& \times \left[\mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) - \int_0^\tau Y_{lik}(u) \frac{E_l\{d\mathbf{M}_{\tilde{z},lik}(\beta_0, u) | \Delta_{lik}=1, \xi_{li}=0\}}{E_l\{Y_{lik}(u) | \Delta_{lik}=1\}} \right]
\end{aligned} \tag{18}$$

Combining the above results, we have shown that $n^{-1/2} \tilde{U}_H(\beta_0)$ is asymptotically equivalent to (13). Under the regularity conditions, the first term on the right-hand side of (13) is

asymptotically zero-mean normal with covariance matrix $\sum_{l=1}^L p_l Q_l(\beta_0)$ where

$Q_l(\beta_0) = E\{\sum_{k=1}^K \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau)\}^{\otimes 2}$ by Spiekerman & Lin (1998). The second and the third terms on the right-hand side of (13) can be shown to be asymptotically zero-mean normal with

covariance matrix $\sum_{l=1}^L p_l \frac{1-\alpha_l}{\alpha_l} V_l^{II,(1)}(\beta_0)$ and $\sum_{l=1}^L p_l (1 - \alpha_l) \sum_{k=1}^K pr(\Delta_{lik}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) V_{lk}^{II,(2)}(\beta_0)$ by lemma 2, respectively. It follows from conditional expectation arguments that these three terms are mutually independent. Therefore, $n^{-1/2} \tilde{U}_H(\beta_0)$ is asymptotically normally distributed with mean zero and with finite variance

$$\sum_{l=1}^L p_l \left\{ Q_l(\beta_0) + \frac{1 - \alpha_l}{\alpha_l} V_l^{II,(1)}(\beta_0) + (1 - \alpha_l) \sum_{k=1}^K pr(\Delta_{lik}=1) \left(\frac{1 - q_{lk}}{q_{lk}} \right) V_{lk}^{II,(2)}(\beta_0) \right\}.$$

Hence $n^{-1} \tilde{U}_H(\beta)$ converges to zero in probability. Thus, (iv) is satisfied.

By (i),(ii),(iii) and (iv), it follows that there exists a unique sequence $\tilde{\beta}_H$ s.t. $n^{-1} \tilde{U}_H(\tilde{\beta}_H) = 0$ with probability converging to one as $n \rightarrow \infty$ and with $\tilde{\beta}_H$ converging in probability to β_0 by Theorem 2 (Foutz, 1977).

The asymptotic normality of $\tilde{\beta}_H$ follows from the consistency of $\tilde{\beta}_H$ and a Taylor series expansion of $\tilde{U}_H(\beta)$. This completes the proof.

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

Theorem 4—Under Conditions (A) – (H), for each $k = 1, \dots, K$, $\tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_H, t)$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0, \tau]$. Also,

$W^{II}(t) = n^{1/2} [\{\tilde{\Lambda}_{01}^{II}(\tilde{\beta}_H, t) - \Lambda_{01}(t)\}, \dots, \{\tilde{\Lambda}_{0K}^{II}(\tilde{\beta}_H, t) - \Lambda_{0K}(t)\}]^T$ converges weakly to a zero-mean Gaussian process $W^{II}(t)$ in $D[0, \tau]^K$ where $\mathcal{W}^{II}(t) = \{\mathcal{W}_1^{II}(t), \dots, \mathcal{W}_K^{II}(t)\}^T$. The covariance function between $\mathcal{W}_j^{II}(t_1)$ and $\mathcal{W}_k^{II}(t_2)$ is

$$\begin{aligned}
\varphi_{jk}^{II}(t_1, t_2)(\beta_0) = & \sum_{l=1}^L p_l \left(E_l \{ \nu_{lj}(\beta_0, t_1) \nu_{lk}(\beta_0, t_2) \} + \frac{1-\alpha_l}{\alpha_l} E_l \{ \psi_{lj}^{II}(\beta_0, t_1) \psi_{lk}^{II}(\beta_0, t_2) \} + (1-\alpha_l) \right. \\
& \times \left[I(j=k) pr(\Delta_{lk}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) Cov_l \left\{ \zeta_{lk}^{(1)}(\beta_0, t_1), \zeta_{lk}^{(1)}(\beta_0, t_2) | \Delta_{lk}=1, \xi_{l1}=0 \right\} \right. \\
& + pr(\Delta_{lj}=1) \left(\frac{1-q_{lj}}{q_{lj}} \right) Cov_l \left\{ \zeta_{lj}^{(1)}(\beta_0, t_1), r_k(\beta_0, t_2)^T A(\beta_0)^{-1} \zeta_{lj}^{(2)}(\beta_0, t_2) | \Delta_{lj}=1, \xi_{l1}=0 \right\} \\
& + pr(\Delta_{lk}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) Cov_l \left\{ \zeta_{lk}^{(1)}(\beta_0, t_2), r_j(\beta_0, t_1)^T A(\beta_0)^{-1} \zeta_{lk}^{(2)}(\beta_0, t_1) | \Delta_{lk}=1, \xi_{l1}=0 \right\} \\
& \left. + \sum_{m=1}^K pr(\Delta_{lm}=1) \left(\frac{1-q_{lm}}{q_{lm}} \right) \right. \\
& \left. \times r_j(\beta_0, t_1)^T A(\beta_0)^{-1} Cov_l \left\{ \zeta_{lm}^{(2)}(\beta_0, t_1), \zeta_{lm}^{(2)}(\beta_0, t_2) | \Delta_{lm}=1, \xi_{l1}=0 \right\} A(\beta_0)^{-1} r_k(\beta_0, t_2) \right] \Bigg).
\end{aligned}$$

where

$$\begin{aligned}
\zeta_{lik}^{(1)}(\beta, t) = & \int_0^t \frac{1}{s_k^{(0)}(\beta, u)} \left[dM_{lik}(\beta, u) - Y_{lik}(u) \frac{E_l \{ dM_{lik}(\beta, u) | \Delta_{lk}=1, \xi_{l1}=0 \}}{E_l \{ Y_{lik}(u) | \Delta_{lk}=1 \}} \right] \text{ and} \\
\zeta_{lik}^{(2)}(\beta, t) = & \mathbf{M}_{\tilde{z}, lik}(\beta, \tau) - \int_0^t Y_{lik}(u) \frac{E_l \{ d\mathbf{M}_{\tilde{z}, lik}(\beta, u) | \Delta_{lk}=1, \xi_{l1}=0 \}}{E_l \{ Y_{lik}(u) | \Delta_{lk}=1 \}}.
\end{aligned}$$

PROOF—One can make the following decomposition

$$\begin{aligned}
& n^{1/2} \{ \tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_{II}, t) - \Lambda_{0k}(t) \} \\
& = n^{1/2} \int_0^t \left\{ \frac{1}{n\tilde{s}_{II,k}^{(0)}(\tilde{\beta}_{II}, u)} - \frac{1}{n\tilde{s}_{II,k}^{(0)}(\beta_0, u)} \right\} d \left\{ \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta, u) \right\} \\
& + n^{1/2} \int_0^t \left\{ \frac{1}{n\tilde{s}_{II,k}^{(0)}(\tilde{\beta}_{II}, u)} - \frac{1}{n\tilde{s}_{II,k}^{(0)}(\beta_0, u)} \right\} d \left[\sum_{l=1}^L \sum_{i=1}^{n_l} \{ \omega_{lik}^{II}(u) - 1 \} M_{lik}(\beta, u) \right] \\
& + n^{1/2} \int_0^t \left\{ \frac{1}{\tilde{s}_{II,k}^{(0)}(\tilde{\beta}_{II}, u)} - \frac{1}{\tilde{s}_{II,k}^{(0)}(\beta_0, u)} \right\} \tilde{s}_{II,k}^{(0)}(\beta_0, u) d\Lambda_{0k}(u) + \int_0^t \frac{1}{\tilde{s}_{II,k}^{(0)}(\beta_0, u)} d \left\{ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta, u) \right\} \\
& + \int_0^t \frac{1}{\tilde{s}_{II,k}^{(0)}(\beta_0, u)} d \left[n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \{ \omega_{lik}^{II}(u) - 1 \} M_{lik}(\beta, u) \right]
\end{aligned} \tag{19}$$

By the Taylor expansion of $\tilde{s}_{II,k}^{(0)}(\tilde{\beta}_{II}, u)^{-1}$ around β_0 , the first term on the right-hand side of (19) is equivalent to

$$\int_0^t \left\{ -\frac{\tilde{s}_{II,k}^{(1)}(\beta^*, u)^T}{\tilde{s}_{II,k}^{(0)}(\beta^*, u)^2} \right\} (\tilde{\beta}_{II} - \beta_0) d \left\{ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta_0, u) \right\} \tag{20}$$

where β^* is on the line segment between $\tilde{\beta}_{II}$ and β_0 . Then, as $n \rightarrow \infty$, (20) converges to 0 uniformly in t in probability by lemma 1 since $\tilde{s}_{II,k}^{(1)}(\beta, u)/\tilde{s}_{II,k}^{(0)}(\beta, u)$ is of bounded variation, $\tilde{\beta}_{II}$ is consistent for β_0 , and $n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta_0, u)$ converges weakly to a zero-mean Gaussian process with continuous sample path. The second term can be shown to converge to 0 uniformly in t in probability by similar arguments.

Again, it follows from the Taylor expansion of $\tilde{s}_{II,k}^{(0)}(\tilde{\beta}_{II}, u)^{-1}$ around β_0 , the uniform convergence of $\tilde{s}_{II,k}^{(1)}(\beta, u)$ and $\tilde{s}_{II,k}^{(0)}(\beta, u)$, the consistency of $\tilde{\beta}_{II}$ for β_0 and the boundedness of $\Lambda_{0k}(t)$ on $[0, \tau]$ that the third term is asymptotically equivalent to

$$n^{1/2} r_k(\beta_0, t)^T (\tilde{\beta}_{II} - \beta_0).$$

The fourth term can be shown to be asymptotically equivalent to

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

by lemma 1 since $\tilde{S}_{II,k}^{(0)}(\beta, u)$ converges to $s_k^{(0)}(\beta, u)$ uniformly in t and

$n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} M_{lik}(\beta_0, u)$ converges weakly to a zero-mean Gaussian process with continuous sample path. For the last term on the right-hand side of (19), it follows from lemma 3, and the uniform convergence of $\tilde{S}_{II,k}^{(0)}(\beta_0, t)^{-1}$ to $s_k^{(0)}(\beta_0, t)^{-1}$, where $s_k^{(0)}(\beta_0, t)$ is bounded away from 0 that

$$\begin{aligned} & \int_0^t \frac{1}{\tilde{S}_{II,k}^{(0)}(\beta_0, u)} d \left[n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \{ \omega_{lik}^{II}(u) - 1 \} M_{lik}(\beta_0, u) \right] \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left[\left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) (1 - \Delta_{lik}) \int_0^t \frac{1}{\tilde{S}_{II,k}^{(0)}(\beta_0, u)} Y_{lik}(u) e^{\beta_0^T Z_{lik}(u)} d\Lambda_{0k}(u) \right. \\ & \quad - (1 - \Delta_{lik}) \int_0^t \frac{1}{\tilde{S}_{II,k}^{(0)}(\beta_0, u)} \{ \hat{\alpha}_{lk}(u)^{-1} - \tilde{\alpha}_l^{-1} \} \xi_{li} Y_{lik}(u) e^{\beta_0^T Z_{lik}(u)} d\Lambda_{0k}(u) \\ & \quad + \Delta_{lik} (1 - \xi_{li}) \left(\frac{\eta_{lik}}{\tilde{q}_k} - 1 \right) \int_0^t \frac{1}{\tilde{S}_{II,k}^{(0)}(\beta_0, u)} dM_{lik}(\beta_0, u) \\ & \quad \left. + \Delta_{lik} (1 - \xi_{li}) \int_0^t \frac{1}{\tilde{S}_{II,k}^{(0)}(\beta_0, u)} \{ \hat{q}_k(u)^{-1} - \tilde{q}_k^{-1} \} \eta_{lik} dM_{lik}(\beta_0, u) \right] \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left[\left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) (1 - \Delta_{lik}) \int_0^t Y_{lik}(u) \right. \\ & \quad \times \left[e^{\beta_0^T Z_{lik}(u)} - \frac{E_l \{ (1 - \Delta_{lik}) Y_{lik}(u) e^{\beta_0^T Z_{lik}(u)} \}}{E_l \{ (1 - \Delta_{lik}) Y_{lik}(u) \}} \right] \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta_0, u)} \\ & \quad \left. - \Delta_{lik} (1 - \xi_{li}) \left(1 - \frac{\eta_{lik}}{\tilde{q}_k} \right) \int_0^t \frac{1}{s_k^{(0)}(\beta_0, u)} \left[dM_{lik}(u) - Y_{lik}(u) \frac{E_l \{ dM_{lik}(u) | \Delta_{lik}=1, \xi_{li}=0 \}}{E_l \{ Y_{lik}(u) | \Delta_{lik}=1 \}} \right] \right] \end{aligned} \quad (21)$$

Now by combining the above results and using the asymptotic expansion of $n^{1/2}(\tilde{\beta}_{II} - \beta_0)$ where

$$\begin{aligned} n^{1/2} (\tilde{\beta}_{II} - \beta_0) &= A(\beta_0)^{-1} n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \sum_{k=1}^K \{ \mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) \\ & \quad + (1 - \Delta_{lik}) \left(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1 \right) \int_0^\tau \left[R_{lik}(\beta_0, u) - Y_{lik}(u) \frac{E_l \{ (1 - \Delta_{lik}) R_{lik}(\beta_0, u) \}}{E_l \{ (1 - \Delta_{lik}) Y_{lik}(u) \}} \right] d\Lambda_{0k}(u) \\ & \quad + \Delta_{lik} (1 - \xi_{li}) \left(\frac{\eta_{lik}}{\tilde{q}_k} - 1 \right) \\ & \quad \times \left[\mathbf{M}_{\tilde{z},lik}(\beta_0, \tau) - \int_0^\tau Y_{lik}(u) \frac{E_l \{ d\mathbf{M}_{\tilde{z},lik}(\beta_0, u) | \Delta_{lik}=1, \xi_{li}=0 \}}{E_l \{ Y_{lik}(u) | \Delta_{lik}=1 \}} \right] \} + o_p(1), \end{aligned}$$

we have

$$n^{1/2} \{ \tilde{\Lambda}_{0k}^{II}(\tilde{\beta}_{II}, t) - \Lambda_{0k}(t) \} = n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} v_{lik}(\beta_0, t) + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \psi_{lik}^{II}(\beta_0, t) + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} v_{lik}^*(\beta_0, t) + o_p(1)$$

where

$$\begin{aligned}
 v_{lik}(\beta, t) &= r_k(\beta, t)^T A(\beta)^{-1} \sum_{m=1}^K \mathbf{M}_{z,lim}(\beta, \tau) + \int_0^t \{s_k^{(0)}(\beta, u)\}^{-1} dM_{lik}(\beta, u), \\
 \psi_{lik}^{II}(\beta, t) &= r_k(\beta, t)^T A(\beta)^{-1} \sum_{m=1}^K (1 - \Delta_{lim}) \int_0^t [R_{lim}(\beta, u) \\
 &\quad - \frac{Y_{lim}(u) E_l \{(1 - \Delta_{lim}) R_{lim}(\beta, u)\}}{E_l \{(1 - \Delta_{lim}) Y_{lim}(u)\}}] d\Lambda_{0m}(u) + (1 - \Delta_{lik}) \int_0^t Y_{lik}(u) \left[e^{\beta^T Z_{lik}(u)} \right. \\
 &\quad \left. - \frac{E_l \{(1 - \Delta_{lik}) Y_{lik}(u) e^{\beta^T Z_{lik}(u)}\}}{E_l \{(1 - \Delta_{lik}) Y_{lik}(u)\}} \right] \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\beta, u)} \text{ and} \\
 v_{lik}^*(\beta, t) &= r_k(\beta, t)^T A(\beta)^{-1} \sum_{m=1}^K \Delta_{lim} (1 - \xi_i) \left(\frac{\eta_{lim}}{q_{lim}} - 1 \right) \zeta_{lim}^{(2)}(\beta, t) + \Delta_{lik} (1 - \xi_{li}) \left(\frac{\eta_{lik}}{q_{lik}} - 1 \right) \zeta_{lik}^{(1)}(\beta, t), \\
 \zeta_{lik}^{(1)}(\beta, t) &= \int_0^t \frac{1}{s_k^{(0)}(\beta, u)} \left[dM_{lik}(\beta, u) - Y_{lik}(u) \frac{E_l \{dM_{lik}(\beta, u) | \Delta_{lik}=1, \xi_{li}=0\}}{E_l \{Y_{lik}(u) | \Delta_{lik}=1\}} \right] \text{ and} \\
 \zeta_{lik}^{(2)}(\beta, t) &= \mathbf{M}_{z,lik}(\beta, \tau) - \int_0^t Y_{lik}(u) \frac{E_l \{d\mathbf{M}_{z,lik}(\beta, u) | \Delta_{lik}=1, \xi_{li}=0\}}{E_l \{Y_{lik}(u) | \Delta_{lik}=1\}}
 \end{aligned}$$

Now, let $W^{(1)}(t) = \{W_1^{(1)}(t), \dots, W_K^{(1)}(t)\}^T$ where

$$W_k^{(1)}(t) = n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} v_{lik}(\beta_0, t), \quad W^{(2)}(t) = \{W_1^{(2)}(t), \dots, W_K^{(2)}(t)\}^T \text{ where}$$

$$W_k^{(2)}(t) = n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\alpha_l}\right) \psi_{lik}^{II}(\beta_0, t), \text{ and } W^{(3)}(t) = \{W_1^{(3)}(t), \dots, W_K^{(3)}(t)\}^T \text{ where}$$

$W_k^{(3)}(t) = n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} v_{lik}^*(\beta_0, t)$ for $k = 1, \dots, K$. Then, $W^{(1)}(t)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}^{(1)}(t) = \{\mathcal{W}_1^{(1)}(t), \dots, \mathcal{W}_K^{(1)}(t)\}^T$ in $D[0, \tau]^K$ where the covariance

function between $\mathcal{W}_j^{(1)}(t_1)$ and $\mathcal{W}_k^{(1)}(t_2)$ is $\sum_{l=1}^L p_l E_l \{v_{l1j}(\beta_0, t_1) v_{l1k}(\beta_0, t_2)\}$ by Spiekerman & Lin (1998, Thm.2). $W^{(2)}(t)$ also can be shown to converge weakly to a zero-mean Gaussian

process $\mathcal{W}^{(2)}(t) = \{\mathcal{W}_1^{(2)}(t), \dots, \mathcal{W}_K^{(2)}(t)\}^T$. For any finite number of time points $(t_1, \dots, t_D)$, the finite dimensional distribution of $W^{(2)}(t)$ is asymptotically the same as those of $\mathcal{W}^{(2)}(t)$ by lemma 2 and Cramer-Wold device. Since the space $D[0, \tau]^K$ is equipped with the uniform metric, it suffices to show the marginal tightness of $W_k^{(2)}(t)$ for each k . The marginal tightness follows directly by applying lemma 2 to $W_k^{(2)}(t)$. Thus, $W^{(2)}(t)$ converges weakly to a zero-mean

Gaussian process where the covariance function between $\mathcal{W}_j^{(2)}(t_1)$ and $\mathcal{W}_k^{(2)}(t_2)$ is

$\sum_{l=1}^L p_l \frac{1-\alpha_l}{\alpha_l} E_l \{\psi_{l1j}^{II}(\beta_0, t_1) \psi_{l1k}^{II}(\beta_0, t_2)\}$. The weak convergence $W^{(3)}(t)$ to a zero-mean Gaussian process $\mathcal{W}^{(3)}(t)$ follows from the similar arguments with the covariance function between $\mathcal{W}_j^{(2)}(t_1)$ and $\mathcal{W}_k^{(3)}(t_3)$ being

$$\begin{aligned}
 &\sum_{l=1}^L p_l (1 - \alpha_l) \left[I(j=k) pr(\Delta_{lik}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) Cov_l \left\{ \zeta_{lik}^{(1)}(\beta_0, t_1), \zeta_{lik}^{(1)}(\beta_0, t_2) | \Delta_{lik}=1, \xi_{li}=0 \right\} \right. \\
 &\quad + pr(\Delta_{l1j}=1) \left(\frac{1-q_{lj}}{q_{lj}} \right) Cov_l \left\{ \zeta_{l1j}^{(1)}(\beta_0, t_1), r_k(\beta_0, t_2)^T A(\beta_0)^{-1} \zeta_{l1j}^{(2)}(\beta_0, t_2) | \Delta_{l1j}=1, \xi_{l1}=0 \right\} \\
 &\quad + pr(\Delta_{lik}=1) \left(\frac{1-q_{lk}}{q_{lk}} \right) Cov_l \left\{ \zeta_{lik}^{(1)}(\beta_0, t_2), r_j(\beta_0, t_1)^T A(\beta_0)^{-1} \zeta_{lik}^{(2)}(\beta_0, t_1) | \Delta_{lik}=1, \xi_{li}=0 \right\} \\
 &\quad \left. + \sum_{m=1}^K pr(\Delta_{lim}=1) \left(\frac{1-q_{lm}}{q_{lm}} \right) \right. \\
 &\quad \times \left. r_k(\beta_0, t_1)^T A(\beta_0)^{-1} Cov_l \left\{ \zeta_{lim}^{(2)}(\beta_0, t_1), \zeta_{lim}^{(2)}(\beta_0, t_2) | \Delta_{lim}=1, \xi_{li}=0 \right\} A(\beta_0)^{-1} r_j(\beta_0, t_2) \right].
 \end{aligned}$$

It follows from the conditional expectation argument that these three terms are mutually independent. Therefore, $W^H(t) = W^{(1)}(t) + W^{(2)}(t) + W^{(3)}(t)$ converges weakly to a zero-mean Gaussian process $W^H(t) = \mathcal{W}^{(1)}(t) + \mathcal{W}^{(2)}(t) + \mathcal{W}^{(3)}(t)$ where the covariance function between $\mathcal{W}_j^H(t_1)$ and $\mathcal{W}_k^H(t_2)$ is $\varphi_{jk}^H(t_1, t_2)(\beta_0)$. This completes the proofs.

3.4 Proofs of Theorems 1 and 2

Proofs for Theorems 1 and 2 basically follow the same steps used for those for Theorems 3 and 4. However, the steps involving the asymptotic expansions of $\hat{\alpha}_{lk}(t)^{-1}$ and $\hat{q}_{lk}(t)^{-1}$ around α_l and $\tilde{q}_{lk}$ (lemma 3) can now be omitted. Specifically, the third and fourth terms in (9) and (10), and the second and fourth terms in (15) and (21) vanish.

4 Discussion

In this paper, we considered fitting marginal hazards models for failure time data with multiple disease outcomes from two types of stratified case-cohort study designs: the original and the generalized stratified case-cohort designs. In either design, subcohort members are sampled via stratified random sampling with possibly different sampling proportions within each stratum where the strata were constructed based on the information available for the entire cohort members. After the selection of the subcohort members, we sample all the remaining cases outside the subcohort under the original stratified case-cohort design whereas we are allowed to select a part of cases outside the subcohort via stratified random sampling under the generalized stratified case-cohort design. For estimation, we proposed weighted estimating equation approach for regression parameters and Breslow-Aalen type estimator for cumulative baseline hazards functions. We also provided a detailed proofs for deriving the asymptotic properties of the proposed estimators. The proposed estimators were shown to have desirable asymptotic properties such as consistency and asymptotic normality.

One modification to the generalized case-cohort study design in the current paper might be worth considering. Instead of sampling cases outside the sub-cohort within each stratum separately, one might want to sample cases outside the subcohort from the whole cohort regardless of their strata. Our proposed methods can be easily adapted to this design simply by redefining the strata, i.e., defining cases outside the subcohort as a separate single stratum.

Acknowledgments

This research is partly supported by National Institutes of Health NHLBI Grant R01 HL-57444.

References

- Andersen P, Gill R. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*. 1982; 10:1100–1120.
- Barlow W. Robust variance estimation for the case-cohort design. *Biometrics*. 1994; 50:1064–1072. [PubMed: 7786988]
- Borgan O, Langholz B, Samuelsen SO, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. *Lifetime Data Analysis*. 2000; 6:39–58. [PubMed: 10763560]
- Breslow NE, Wellner JA. Weighted likelihood for semipara-metric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics*. 2007; 34:86–102.
- Chen K, Lo S. Case-cohort and case-control analysis with cox's model. *Biometrika*. 1999; 86:755–764.
- Chen K. Generalized case-cohort sampling. *Journal of the Royal Statistical Society, Series B*. 2001; 63:791–809.
- Cullen KJ. Mass health examinations in the Busselton population, 1966 to 1970. *Australian Journal of Medicine*. 1972; 2:714–718.

- Foutz RV. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*. 1977; 72:147–148.
- Hájek J. Limiting distributions in simple random sampling from a finite population. *Pub Math Inst Hungar Acad Sci*. 1960; 5:361–374.
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1951; 47:663–685.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2. New York: Wiley, John & Sons; 2002.
- Kang S, Cai J. Marginal hazards model for case-cohort studies with multiple disease outcomes. *Biometrika*. 2009; 96:887–901.
- Karatzas, I.; Shreve, SE. *Brownian Motion and Stochastic Calculus*. 2. New York: Springer-Verlag; 1988.
- Kulich M, Lin DY. Additive hazards regression for case-cohort studies. *Biometrika*. 2000; 87:73–87.
- Kulich M, Lin DY. Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association*. 2004; 99:832–844.
- Langholz B, Thomas D. Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *American Journal of Epidemiology*. 1990; 131:169–176. [PubMed: 2403467]
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B*. 2000; 62:711–730.
- Lin DY. On fitting cox's proportional hazards models to survey data. *Biometrika*. 2000; 87:37–47.
- Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
- Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies. *Annals of Statistics*. 1988; 16:64–81.
- Shorack, GR.; Wellner, JA. *Empirical Processes with Applications to Statistics*. New York: Wiley; 1986.
- Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*. 1998; 93:1164–1175.
- Thisted, R. *Elements of Statistical Computing*. New York: Chapman & Hall; 1988.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. New York: Springer-Verlag; 1996.
- Wacholder S, Gail M, Pee D. Efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics*. 1991; 47:63–76. [PubMed: 2049514]
- Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*. 1989; 84:1065–1073.