American Journal of Applied Mathematics
2021; 9(6): 192-210
http://www.sciencepublishinggroup.com/j/ajam
doi: 10.11648/j.ajam.20210906.11
ISSN: 2330-0043 (Print); ISSN: 2330-006X (Online)

Using Full Cohort Information to Improve the Estimation Efficiency of Marginal Hazard Model for Multivariate Failure Times in Case-Cohort Studies

Hongtao Zhang[1], Haibo Zhou[2], David Couper[2], Jianwen Cai[2,*]

[1]Bristol Myers Squibb, Berkeley Heights, New Jersey, USA
[2]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA

Email address:

cai@bios.unc.edu (Jianwen Cai)
[*]Corresponding author

To cite this article:

Hongtao Zhang, Haibo Zhou, David Couper, Jianwen Cai. Using Full Cohort Information to Improve the Estimation Efficiency of Marginal Hazard Model for Multivariate Failure Times in Case-Cohort Studies. *American Journal of Applied Mathematics*. Vol. 9, No. 6, 2021, pp. 192-210. doi: 10.11648/j.ajam.20210906.11

Received: June 8, 2021; **Accepted:** December 2, 2021; **Published:** December 24, 2021

Abstract: The case-cohort design is widely used in large cohort studies when it is prohibitively costly to measure some exposures for all subjects in the full cohort, especially in studies where the disease rate is low. To investigate the effect of a risk factor on different diseases, multiple case-cohort studies using the same subcohort are usually conducted. To compare the effect of a risk factor on different types of diseases, times to different disease events need to be modeled simultaneously. Existing case-cohort estimators for multiple disease outcomes utilize only the relevant covariate information in cases and subcohort controls, though many covariates are measured for everyone in the full cohort. Intuitively, making full use of the relevant covariate information can improve efficiency. To this end, we consider a class of doubly-weighted estimators for both regular and generalized case-cohort studies with multiple disease outcomes. The asymptotic properties of the proposed estimators are derived and our simulation studies show that a gain in efficiency can be achieved with a properly chosen weight function. We apply the proposed method to re-analyze a data set from Atherosclerosis Risk in Communities (ARIC) study to showcase the gain in efficiency. Concluding remarks and future researches are also discussed.

Keywords: Case-cohort Study, Multiple Disease Outcomes, Survival Analysis

1. Introduction

The case-cohort design is widely used in large cohort studies when it is prohibitively costly to assemble exposure history for all subjects in the full cohort. First introduced by Prentice [1], the case-cohort design requires a random sample in the full cohort, named 'subcohort'. All subjects in the full cohort are followed until failure or censoring occurs, but complete exposure information is collected only for subjects who experienced failure and for those subjects selected into the subcohort. The case-cohort design is a special form of two-phase sampling design [2].

For data from case-cohort studies for a single disease outcome, many methods have been proposed under the Cox proportional hazard model framework. A pseudo-likelihood approach, which modified the partial likelihood by weighting the contributions of cases and subcohort controls differently, has been studied [1, 3]. Barlow [4] provided an easier alternative approach to computing the asymptotic variance. Chen and Lo [5] used a refined procedure to estimate the at-risk average to achieve efficiency gain. Reference [6] considered a stratified case-cohort design and used time-varying weights based on the at-risk process to improve the efficiency of the parameter estimates. Despite the advances in methods for univariate case-cohort designs, literature on the marginal models for case-cohort data with multiple disease outcomes is scarce. Kang and Cai [7] proposed a weighted estimating equation approach to fit a marginal proportional hazard model with multiple diseases. Kim and Cai [8] proposed a modified weight function in order to use the covariate information from subjects who have other disease(s).

Both of these methods use only the exposure information collected on cases and subjects in the subcohort. See also [9] for a comprehensive review.

In many studies, information about some exposures is available on all subjects in the full cohort, while information on other exposures that is costly to collect is available only for the cases and subjects in the subcohort. The former is referred to as the first-phase covariate data, and the latter as second-phase covariate data. For example, the Atherosclerosis Risk in Communities (ARIC) study is a large cohort study that involved 15,792 participants. An aim in one ancillary study to the ARIC study was to assess lipoprotein-associated phospholipase A_2 (Lp-PLA_2) as potential risk factor for atherosclerosis and its sequelae, so that physicians may consider using Lp-PLA_2 as a complementary risk factor beyond the traditional ones. Given the large cohort size and funding limitations, assaying Lp-PLA_2 for all the participants was cost-prohibitive. As an alternative, case-cohort studies were carried out: Lp-PLA_2 was obtained only for participants with an incident coronary heart disease (CHD) or stroke event, or in a selected subcohort of all participants [10, 11]. Lp-PLA_2 is thus the second-phase covariate (the exposure of primary interest) and the first-phase covariates are those collected on the full cohort, such as race, gender, lipid measurements, etc. To compare the effect of Lp-PLA_2 on incident stroke and CHD, the two disease outcomes need to be modeled simultaneously to properly account for their correlation. The methods proposed by [7] and [8] can be applied in this situation. However, only covariate information collected on the cases and subjects in the subcohort are used. It is desirable to use relevant covariate information collected on the full cohort to improve efficiency. For a single survival outcome, Kulich and Lin [12] proposed a doubly-weighted estimator that uses all available first-phase covariate data and postulated a regression model for second-phase covariate(s) on first-phase covariate(s), followed up by references [13, 14] in which calibration methods were used for weight estimation. However, such type of approach has not been explored for multiple diseases. In this paper, we aim to investigate a doubly-weighted approach to improve efficiency with multiple diseases with data from multiple traditional case-cohort studies. Furthermore, we also consider generalized case-cohort designs. Generalized case-cohort designs are usually conducted when the disease is not rare, but there are limited resources. Under such situation, instead of taking all the cases, a random sample of cases outside the subcohort is drawn [7, 15]. It will be of interest to examine the doubly-weighted approach for the generalized case-cohort studies.

2. Model and Estimation

2.1. Notations and Model Definition

Suppose that there are n independent subjects in the full cohort and K disease outcomes of interest. Consider independent vectors of potential failure times $T_i = (T_{i1}, \ldots, T_{iK})^T, i = 1, \ldots, n, k = 1, \ldots, K$. Similarly, we use $C_i = (C_{i1}, \ldots, C_{iK})^T$ to denote the potential right censoring time vectors. In practice, it is common to have $C_{i1} = \ldots C_{iK} = C_i$. The observed time is $X_{ik} = T_{ik} \wedge C_{ik}$. Let $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ denote the event indicator, $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)$ the counting process, and $Y_{ik}(t) = I(X_{ik} \geq t)$ the at-risk process for disease k of subject i, respectively. Let $Z_{ik}(t)$ be a $p \times 1$ potentially time-dependent covariate vector that can be decomposed into two components: a $p_1 \times 1$ vector of first-phase covariates $V_{ik}(t)$, and a $p_2 \times 1$ vector of second-phase time-independent covariates W_{ik}. The second-phase covariates are typically time-independent in practice, and they are usually measured at the baseline. The proposed method and the asymptotic results can easily be extended to time-dependent second-phase covariates setup. The potentially time-dependent first-phase covariates are assumed to be 'external' in the sense that they are not affected by the outcome processes [16]. We assemble all the covariates into a vector $Z_i = (Z_{i1}, \ldots, Z_{iK})^T$ and denote τ the study end time. Define the marginal hazard for disease outcome k

$$\lambda_{ik}(t|Z_{ik}(t)) = \lim_{\delta \to 0} \frac{1}{\delta} P\{t \leq T_{ik} < t + \delta | T_{ik} \geq t, Z_{ik}(t)\}.$$

Suppose that potential failure time T_{ik} arises from a Cox-type proportional marginal hazards model [17]

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

where $\lambda_{0k}(t)$ is an unspecified, baseline marginal hazard function and β_0 is a $p \times 1$ vector of fixed unknown regression parameters of interest. We note that disease-specific covariate effects can be accommodated by model (1) by redefining the regression coefficient vector and the covariate vector. For example, if we are interested in the disease-specific effect model:

$$\lambda_{ik}(t|Z_{ik}(t)) = Y_{ik}(t)\lambda_{0k}(t)\exp(\beta_k^T Z_{ik}(t)), \tag{2}$$

where $Z_{ik}(t)$ can be the same or different for different k and β_k denotes the disease-k-specific effect for covariate $Z_{ik}(t)$, $k = 1, ..., K$. The above model can be rewritten in the form of (1) by defining $\beta^* = (\beta_1^T, \ldots, \beta_K^T)^T$, a collection of all disease-specific parameters, and $Z_{ik}^*(t) = (0_{i1}^T, \ldots, Z_{ik}(t)^T, \ldots, 0_{iK}^T)^T$. Note that $\beta^{*T} Z_{ik}^*(t) = \beta_k^T Z_{ik}(t)$, therefore, the following model is equivalent to (2):

$$\lambda_{ik}(t|Z_{ik}^*(t)) = Y_{ik}(t)\lambda_{0k}(t)\exp(\beta^{*T} Z_{ik}^*(t)).$$

2.2. Estimation

If the data were complete, for $d = 0, 1, 2$, define $S_{k,F}^{(d)}(\beta, t) = n^{-1}\sum_{i=1}^n Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}$, with $a^{\otimes 0} = 1, a^{\otimes 1} = a, a^{\otimes 2} = aa^T$. The relative risk parameter β_0 can be estimated by solving the pseudo partial likelihood score equation

$$U_F(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\{Z_{ik}(t) - \bar{Z}_{k,F}(\beta,t)\}dN_{ik}(t) = 0, \tag{3}$$

Where $\bar{Z}_{k,F}(\beta,t) = S^{(1)}_{k,F}(\beta,t)/S^{(0)}_{k,F}(\beta,t)$. Under the case-cohort design, (3) cannot be calculated because covariate vector $Z_{ik}(t)$ is not fully observed for subjects that are neither in the subcohort nor among sampled cases. Instead, we consider a weighted version of the pseudo likelihood score function in which information from a completely observed subject represents multi-fold information for subjects who were not sampled to have their second-phase covariates measured.

Assume that we sample without replacement to obtain a subcohort of size $\tilde{n}$. Subcohort sampling is followed by the sampling of non-subcohort cases, that is, for disease k, we sample m_k subjects without replacement from cases that are outside the subcohort. Let ξ_i be an indicator of subcohort membership which equals 1 if subject i is sampled into the subcohort and 0 otherwise. Similarly, we define η_{ik} as the indicator for the ith subject outside the subcohort with the kth disease being selected into the sample. For any i, the subcohort sampling probability $\tilde{\alpha} = Pr(\xi_i = 1) = \tilde{n}/n$ and disease-specific case sampling probability $\tilde{q}_k = Pr(\eta_{ik} = 1|\Delta_{ik} = 1, \xi_i = 0) = m_k/(n_k - \tilde{n}_k)$, where n_k and $\tilde{n}_k$ denote the number of cases for the kth disease in the cohort and in the subcohort, respectively. With multiple disease outcomes of interest, the case-cohort samples are usually drawn separately: for each disease k, participants who have the event are sampled with probability $\tilde{q}_k$. This separate sampling may lead to subjects with more than one diseases to be included in multiple samples (overlapping subjects). Our estimation method can accommodate such situation.

A marginal proportional hazards model for case-cohort studies with multiple disease outcomes was first investigated in [7], which embedded the at-risk processes in estimating $\tilde{\alpha}$ and $\tilde{q}_k$. The motivation for using the doubly-weighted estimator arises from the intuition that one could incorporate additional information beyond the at-risk processes and hence obtain a more efficient estimator. Further, it is desirable to have the flexibility of weighting each covariate in (1) differently, which could lead to improved precision. We use the superscript/subscript 'KC' and 'DW' to indicate that the quantity, function or estimate is obtained from implementing the $\tilde{\beta}_{II}$ estimator in [7] and our doubly-weighted estimator, respectively.

Let

$$\tilde{w}_{ik}(t) = \Delta_{ik}\xi_i I_p + (1 - \Delta_{ik})\xi_i\hat{\alpha}_k(t)^{-1} + \Delta_{ik}(1 - \xi_i)\eta_{ik}\hat{q}_k(t)^{-1},$$

where

$$\hat{\alpha}_k(t) = \{\sum_{i=1}^{n}(1 - \Delta_{ik})A_{ik}(t)\}^{-1}\{\sum_{i=1}^{n}(1 - \Delta_{ik})\xi_i A_{ik}(t)\}, \tag{4}$$

and

$$\hat{q}_k(t) = \{\sum_{i=1}^{n}\Delta_{ik}(1 - \xi_i)B_{ik}(t)\}^{-1} \times \{\sum_{i=1}^{n}\Delta_{ik}(1 - \xi_i)\eta_{ik}B_{ik}(t)\}, \tag{5}$$

Where $A_{ik}(t)$ and $B_{ik}(t)$ denote diagonal matrices with p potentially different random processes on their respective diagonals. Each of the p covariates in model (1) can have its dedicated process in order to estimate the sampling probabilities more precisely. Define $S^{(d)}_{k,DW}(\beta,t) = n^{-1}\sum_{i=1}^{n}\tilde{w}_{ik}(t)Y_{ik}(t)Z_{ik}(t)^{\otimes d}e^{\beta^T Z_{ik}(t)}, d = 0, 1, 2,$ and the at-risk average process $\bar{Z}_{k,DW}(\beta,t) = \{S^{(0)}_{k,DW}(\beta,t)\}^{-1}\{S^{(1)}_{k,DW}(\beta,t)\}$. Following the idea of generalized estimating equation (GEE) approach for categorical and continuous outcome data, under independence working assumption, we propose to consider a doubly-weighted score equation:

$$U_{DW}(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau}\tilde{w}_{ik}(t) \times \{Z_{ik}(t) - \bar{Z}_{k,DW}(\beta,t)\}dN_{ik}(t) = 0. \tag{6}$$

Note that $U_{DW}(\beta)$ reduces to the score function in [12] when there is only one disease of interest. We will show that $\hat{\beta}_{DW}$ solving (6) is consistent and asymptotically follows a normal distribution with a sandwich type of variance. Because model (1) includes model (2) as a special case, these asymptotic properties extend to the estimates of the parameters in model (2).

Unlike other weighting schemes where weights and $S^{(0)}_k(\beta,t)$ are scalar functions, both $S^{(0)}_{k,DW}(\beta,t)$ and $\tilde{w}_{ik}(t)$ in the doubly-weighted estimating equation in (6) are $p \times p$ diagonal matrices. The second level weights, $A_{ik}(t)$ and $B_{ik}(t)$ in (4) and (5), are diagonal matrices with p potentially

different random processes on their respective diagonals. The Kang and Cai estimator [7] is a special case of the doubly-weighted estimator class by setting both $A_{ik}(t)$ and $B_{ik}(t)$ to $Y_{ik}(t) \cdot I_p$. The estimator in [8] also belongs to this class by setting $A_{ik}(t) = \{\prod_{j=1}^{K}(1-\Delta_{ij})\}Y_{ik}(t)$ and $B_{ik}(t)$ is not applicable in a traditional case-cohort study. Another choice of second level weight is similar to the 'optimal' weight proposed in [12]. It is a $p \times p$ diagonal matrix in the form of

$$A_{ik}(t) = diag[\{\hat{Z}_{ik}(t) - \bar{Z}_{k,KC}(\hat{\beta}_{KC}, t)\}exp\{\hat{\beta}_{KC}^T \hat{Z}_{ik}(t)\}Y_{ik}(t)], \tag{7}$$

Where $\hat{\beta}_{KC}$ and $\bar{Z}_{k,KC}(\hat{\beta}_{KC}, t)$ are parameter estimate and estimated at-risk average process obtained from implementing the Kang and Cai [7] method. $\hat{Z}_{ik} = (\hat{Z}_{ik,1}, \ldots, \hat{Z}_{ik,p})^T$ is a p-vector of covariates, where $\hat{Z}_{ik,p}$ is the observed value if the subject is in the case-cohort sample, otherwise it is estimated using first-phase covariate information. Calculating weight (7) requires another consistent and asymptotically normally distributed estimator, and in our case we used $\hat{\beta}_{KC}$. The calculation of $A_{ik}(t)$ in the form of (7) incorporates the first phase covariates outside the case-cohort sample to improve statistical efficiency. In this sense, specifying second-level weights $A_{ik}(t)$ (and $B_{ik}(t)$) in matrix form allows potential fuller use of available covariate information.

The doubly-weighted estimator $\hat{\beta}_{DW}$ can be obtained via a Newton-Raphson algorithm by iteratively solving (6) until the convergence criterion is met. Specifically, the estimator in the step $k+1$ is $\beta_{DW}^{(k+1)} = \beta_{DW}^{(k)} - D_{DW}(\beta_{DW}^{(k)})^{-1}U_{DW}(\beta_{DW}^{(k)})$, where $D_{DW}(\beta)$ is the derivative of $U_{DW}(\beta)$ with respect to β. Due to the matrix nature of $S_{k,DW}^{(0)}(\beta, t)$, special attention is needed to compute $D_{DW}(\beta)$. The explicit form of $D_{DW}(\beta)$ is given in Appendix 1 in the supplementary material.

We propose to use a Breslow-Aalen type estimator for the baseline cumulative hazard function $\Lambda_{0k}(t)$. The form of the estimator is the same as the one proposed in [7] with the estimator for β replaced by $\hat{\beta}_{DW}$. Specifically,

$$\hat{\Lambda}_{0k}(\hat{\beta}_{DW}, t) = \int_0^t \frac{\sum_{j=1}^n \rho_{jk}(u)dN_{jk}(u)}{nS_{k,KC}^{(0)}(\hat{\beta}_{DW}, u)},$$

where

$$\rho_{jk}(u) = \Delta_{ik}\xi_i + (1-\Delta_{ik})\xi_i\hat{\alpha}_k^{KC}(u)^{-1} + \Delta_{ik}(1-\xi_i)\eta_{ik}\hat{q}_k^{KC}(u)^{-1},$$

$$\alpha_k^{KC}(u) = \{\sum_{i=1}^n (1-\Delta_{ik})Y_{ik}(u)\}^{-1}\{\sum_{i=1}^n (1-\Delta_{ik})\xi_i Y_{ik}(u)\},$$

$$q_k^{KC}(u) = \{\sum_{i=1}^n \Delta_{ik}(1-\xi_i)Y_{ik}(u)\}^{-1} \times \{\sum_{i=1}^n \Delta_{ik}(1-\xi_i)\eta_{ik}Y_{ik}(u)\},$$

And $S_{k,KC}^{(0)}(\beta, u) = n^{-1}\sum_{i=1}^n \rho_{ik}(u)Y_{ik}(u)e^{\beta^T Z_{ik}(u)}$ are the scalar functions used in [7]. Based on the results in [7], this estimator is consistent and converges weakly to a zero mean Gaussian process if $\hat{\beta}_{DW}$ is a consistent estimator of β_0. We will establish the consistency of $\hat{\beta}_{DW}$ in the next section.

3. Asymptotic Properties of General Doubly Weighted Estimator

3.1. Asymptotic Results

We present the asymptotic properties of the doubly-weighted estimator. For $k = 1, \ldots, K$, define the following limiting quantities:

$$\begin{aligned}
s_k^{(d)}(\beta, t) &= E\{S_{k,F}^{(d)}(\beta, t)\}(d = 0, 1, 2), \\
\bar{z}_k(\beta, t) &= s_k^{(1)}(\beta, t)/s_k^{(0)}(\beta, t), \\
v_k(\beta, t) &= \frac{s_k^{(2)}(\beta, t)s_k^{(0)}(\beta, t) - s_k^{(1)}(\beta, t)^{\otimes 2}}{s_k^{(0)}(\beta, t)^2}, \\
G_k(\beta) &= \int_0^\tau v_k(\beta, t)s_k^{(0)}(\beta, t)d\Lambda_{0k}(t).
\end{aligned}$$

We assume the usual regularity conditions, as required in [18]:

Assumption 3.1. $(T_i, C_i, Z_i), i = 1, \ldots, n$ are independent and identically distributed.

Assumption 3.2. $pr\{Y_{ik}(t) = 1\} > 0$ for $t \in [0, \tau], i = 1, \ldots, n$ and $k = 1, \ldots, K$.

Assumption 3.3. $|Z_{ik}(0)| + \int_0^\tau |dZ_{ik}(t)| < D_z < \infty$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$ almost surely, where D_z is a constant.

Assumption 3.4. $G_k(\beta_0)$ is positive definite for $k = 1, \ldots, K$.

Assumption 3.5. (Finite interval) $\int_0^\tau \lambda_{0k}(t)dt < \infty$ for $k = 1, \ldots, K$.

Assumption 3.6. (Asymptotic stability) There exists a neighborhood $\mathcal{B}$ of β_0 such that

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \|S_{k,F}^{(d)}(\beta, t) - s_k^{(d)}(\beta, t)\| \to_p 0$$

for $d = 0, 1, 2$ and $k = 1, \ldots, K$.

Assumption 3.7. (Asymptotic regularity) For all $\beta \in \mathcal{B}$ and $k = 1, \ldots, K$: $s_k^{(1)}(\beta, t) = \frac{\partial}{\partial\beta} s_k^{(0)}(\beta, t)$, $s_k^{(2)}(\beta, t) = \frac{\partial^2}{\partial\beta\partial\beta^T} s_k^{(0)}(\beta, t)$ where $s_k^{(0)}(\cdot, t), s_k^{(1)}(\cdot, t), s_k^{(2)}(\cdot, t)$ are continuous functions of $\beta \in \mathcal{B}$, uniformly in $t \in [0, \tau]$ and $s_k^{(0)}(\cdot, t)$ is bounded away from 0 on $\mathcal{B} \times [0, \tau]$.

Assumption 3.8. (Lindeberg condition) There exists a $\delta > 0$ such that as $n \to \infty$

$$n^{-1/2} \sup_{i,k,t} \|Z_{ik}(t)\| Y_{ik}(t) I\{\beta_0^T Z_{ik}(t) > -\delta \|Z_{ik}(t)\|\} \to_p 0.$$

We also need the following conditions concerning case-cohort samples and second level weights:

Assumption 3.9. (Nontrivial subcohort and case sampling) As $n \to \infty$, $\tilde{\alpha}$ converges to a constant on $(0, 1]$; similarly, for $k = 1, \ldots, K$, $\tilde{q}_k$ converges to a constant on $(0, 1]$.

Assumption 3.10. For each component $Z_{ik,l}(t)$ of $Z_{ik}(t)$, $var \int_0^\tau |dV_{ik,l}(t)| < \infty$, where $V_{ik,l}(t) = Z_{ik,l}(t) exp\{\beta_0^T Z_{ik}(t)\}$. For each diagonal element $A_{ik,l}(t)$ of $A_{ik}(t)$, $var \int_0^\tau |dA_{ik,l}(t)| < \infty$. Diagonal elements of $B_{ik}(t)$ require a similar condition.

Assumption 3.11. $A_{ik}(t)$ is independent of ξ_i, and $B_{ik}(t)$ is independent of η_{ik}, for $k = 1, \ldots, K$.

Assumption 3.12. the absolute values of the diagonal elements of $\mu_k(t) \equiv E_k[(1 - \Delta_{1k})A_{1k}(t)]$ and $\theta_k(t) \equiv E_k[\Delta_{1k} B_{1k}(t)]$ are bounded away from 0 for all $t \in [0, \tau]$.

Assumption 3.12 is required in order to prove the asymptotic properties of $\hat{\alpha}_k(t)$ and $\hat{q}_k(t)$. As long as the elements on the diagonal of $A_{ik}(t)$ or $B_{ik}(t)$ are nonnegative (e.g., $Y_{ik}(t)$), this condition is trivial. However, this assumption may not hold if we use the weight function (7). We relax this condition in the next section. This will enable us to use arbitrary second level weights.

We present the asymptotic results here and provide the outline of the proof in Appendix 2 in the supplementary material. Define

$$\begin{aligned}
M_{ik}(t) &= N_{ik}(t) - \int_0^t Y_{ik}(u) e^{\beta_0^T Z_{ik}(u)} d\Lambda_{0k}(u), \\
\tilde{Z}_{ik}(\beta, t) &= Z_{ik}(t) - \bar{z}_k(\beta, t), \\
M_{\bar{z},ik}(\beta) &= \int_0^\tau \tilde{Z}_{ik}(\beta, t) dM_{ik}(t), \\
R_{ik}(\beta, t) &= Y_{ik}(t) \tilde{Z}_{ik}(\beta, t) e^{\beta^T Z_{ik}(u)}.
\end{aligned}$$

Asymptotic properties of $\hat{\beta}_{DW}$ are summarized in the following theorem:

Theorem 3.1. (Asymptotic properties of $\hat{\beta}_{DW}$)
Under conditions 3.1-3.12, $\hat{\beta}_{DW}$ solving the estimating equation $U_{DW}(\hat{\beta}_{DW}) = 0$ is a consistent estimator of β_0 and

$$\sqrt{n}(\hat{\beta}_{DW} - \beta_0) \to_d N(0, G(\beta_0)^{-1} \Sigma(\beta_0) G(\beta_0)^{-1}),$$

where $G(\beta) = \sum_k G_k(\beta)$ and

$$\Sigma(\beta_0) = Q(\beta_0) + \frac{1 - \tilde{\alpha}}{\tilde{\alpha}} V^I(\beta_0) + (1 - \tilde{\alpha}) \sum_k pr(\Delta_{1k} = 1) \frac{1 - \tilde{q}_k}{\tilde{q}_k} V_k^{II}(\beta_0), \tag{8}$$

where

$$\begin{aligned}
Q(\beta_0) &= E\left\{ \sum_k M_{\bar{z},1k}(\beta_0) \right\}^{\otimes 2}, \\
V^I(\beta_0) &= var\left\{ \sum_k (1 - \Delta_{1k}) \int_0^\tau \{R_{1k}(\beta_0, t) - \mu_k(t)^{-1} A_{1k}(t) E[(1 - \Delta_{1k}) R_{1k}(\beta_0, t)]\} d\Lambda_{0k}(t) \right\}, \\
V_k^{II}(\beta_0) &= var\left\{ M_{\bar{z},1k}(\beta_0) - \int_0^\tau \theta_k(t)^{-1} B_{1k}(t) \times E[\Delta_{1k} dM_{\bar{z},1k}(\beta_0, t)] \middle| \Delta_{1k} = 1, \xi_1 = 0 \right\}.
\end{aligned}$$

Throughout the manuscript, we assume that the subcohort of size $\tilde{n}$ is sampled without replacement. The key to prove the asymptotic results in this setting is to apply Lemma 1 in the supplementary material, which is specific for sampling without replacement scheme. The asymptotic variance of $\hat{\beta}_{DW}$ has three components: the variance of the full data, the variation due to subcohort sampling, and the variation due to further case sampling if a generalized case-cohort design is used. Unknown quantities can be estimated by substituting proper consistent estimators for their theoretical counterparts. See Appendix 2 in the supplementary material for details.

3.2. Generalization to Arbitrary Second Level Weight

In this section, we relax assumption 3.12, which will enable us to use arbitrary second level weights. For notational simplicity, we drop the subscript l by assuming $p = 1$. For $p \geq 2$, the operation is on each diagonal element of $A_{ik}(t)$ and $B_{ik}(t)$. We break down the second level weight by dynamic grouping based on the sign of $A_{ik}(t)$ and $B_{ik}(t)$. Specifically, denote $\gamma_{ik}^{+}(t) = I(A_{ik}(t) \geq 0)$, $\gamma_{ik}^{-}(t) = I(A_{ik}(t) < 0)$, and let $A_{ik}^{+}(t) = \gamma_{ik}^{+}(t)A_{ik}(t)$, $A_{ik}^{-}(t) = -\gamma_{ik}^{-}(t)A_{ik}(t)$. We then have an estimate of α using only the second level weights that are non-negative:

$$\hat{\alpha}_k^{+}(t) = \{\sum_i (1-\Delta_{ik})A_{ik}^{+}(t)\}^{-1}\{\sum_i \xi_i(1-\Delta_{ik})A_{ik}^{+}(t)\}.$$

$\hat{\alpha}_k^{-}(t)$ is defined similarly. For the second level weights $B_{ik}(t)$, we analogously define the quantities:

$$\begin{aligned}
&\zeta_{ik}^{+}(t) = I(B_{ik}(t) \geq 0), B_{ik}^{+}(t) = \zeta_{ik}^{+}(t)B_{ik}(t),\\
&\zeta_{ik}^{-}(t) = I(B_{ik}(t) < 0), B_{ik}^{-}(t) = -\zeta_{ik}^{-}(t)B_{ik}(t),\\
&\hat{q}_k^{+}(t) = \{\sum_{i=1}^{n} \Delta_{ik}(1-\xi_i)B_{ik}^{+}(t)\}^{-1} \times \{\sum_{i=1}^{n} \Delta_{ik}(1-\xi_i)\eta_{ik}B_{ik}^{+}(t)\},\\
&\hat{q}_k^{-}(t) = \{\sum_{i=1}^{n} \Delta_{ik}(1-\xi_i)B_{ik}^{-}(t)\}^{-1} \times \{\sum_{i=1}^{n} \Delta_{ik}(1-\xi_i)\eta_{ik}B_{ik}^{-}(t)\}.
\end{aligned}$$

Finally, the generalized weight function is

$$\tilde{w}_{ik}(t) = \Delta_{ik}\xi_i I_p + (1-\Delta_{ik})\xi_i \times [\gamma_{ik}^{+}(t)\hat{\alpha}_k^{+}(t)^{-1} + \gamma_{ik}^{-}(t)\hat{\alpha}_k^{-}(t)^{-1}] + \Delta_{ik}(1-\xi_i)\eta_{ik} \times [\zeta_{ik}^{+}(t)\hat{q}_k^{+}(t)^{-1} + \zeta_{ik}^{-}(t)\hat{q}_k^{-}(t)^{-1}].$$

The expressions of asymptotic variance also need to be modified to accommodate the grouping:

$$\begin{aligned}
V^I(\beta_0) = var\Bigg\{ \sum_k (1-\Delta_{1k}) \int_0^{\tau} \Bigg\{ & R_{1k}(\beta_0,t) - \gamma_{1k}^{+}(t)\mu_k^{+}(t)^{-1}A_{1k}^{+}(t)E^{+}[(1-\Delta_{1k})R_{1k}(\beta_0,t)] -\\
& \gamma_{1k}^{-}(t)\mu_k^{-}(t)^{-1}A_{1k}^{-}(t)E^{-}[(1-\Delta_{1k})R_{1k}(\beta_0,t)] \Bigg\} \times d\Lambda_{0k}(t) \Bigg\},
\end{aligned}$$

where $\mu_k^{+}(t) = E[(1-\Delta_{1k})A_{1k}(t)|A_{1k}(t) \geq 0]$ and $E^{+}[(1-\Delta_{1k})R_{1k}(\beta_0,t)] = E[(1-\Delta_{1k})R_{1k}(\beta_0,t)|A_{1k} \geq 0]$. $\mu_k^{-}(t)$ and $E^{-}[(1-\Delta_{1k})R_{1k}(\beta_0,t)]$ are analogously defined. Also,

$$\begin{aligned}
V_k^{II}(\beta_0) = var\Bigg\{ M_{\bar{z},1k}(\beta_0) - \Bigg\{ & \int_0^{\tau} \zeta_{1k}^{+}(t)\theta_k^{+}(t)^{-1}B_{1k}^{+}(t) \times E^{+}[dM_{\bar{z},1k}(\beta_0)|\Delta_{1k}=1] -\\
& \int_0^{\tau} \zeta_{1k}^{-}(t)\theta_k^{-}(t)^{-1}B_{1k}^{-}(t) \times E^{-}[dM_{\bar{z},1k}(\beta_0)|\Delta_{1k}=1] \Bigg\} \Bigg| \Delta_{1k}=1, \xi_1=0 \Bigg\}.
\end{aligned}$$

$\theta_k^{+}(t)$, $\theta_k^{-}(t)$, $E^{+}[dM_{\bar{z},1k}(\beta_0)|\Delta_{1k}=1]$ and $E^{-}[dM_{\bar{z},1k}(\beta_0)|\Delta_{1k}=1]$ are computed likewise. Due to the grouping, we need to split the sample to estimate the unknown quantities separately by the sign of the second level weight. Thus in general, a larger sample size is required to achieve satisfactory asymptotic properties.

3.3. Generalization to Stratified Sampling Design

Suppose that a cohort of size n can be partitioned into H mutually exclusive strata based on some first-phase covariates. We extend the method to stratified case-cohort studies, whereby sampling is conducted within each stratum with possibly different sampling probabilities. Specifically, let n_h denote the number of subjects in the hth stratum in the full cohort ($h = 1, \ldots, H$) and $n = n_1 + \cdots + n_H$. Let $p_h = n_h/n$. Within the hth stratum, we sample $\tilde{n}_h$ subcohort members via simple random sampling with probability being $\tilde{\alpha}_h = P(\xi_{hi} = 1) = \tilde{n}_h/n_h$. Total subcohort size $\tilde{n} = \tilde{n}_1 + \cdots + \tilde{n}_H$. Subsequently, for the kth disease outcome within the hth stratum, we sample m_{hk} cases outside the subcohort with probability $\tilde{q}_{hk} = m_{hk}/(n_{hk} - \tilde{n}_{hk})$, where n_{hk} and $\tilde{n}_{hk}$ are the numbers of subjects with the kth disease outcome in the hth stratum in the cohort and in the subcohort, respectively.

We consider the following model with the stratified sampling design,

$$\lambda_{hik}(t|Z_{hik}(t)) = Y_{hik}(t)\lambda_{0k}(t)e^{\beta_0^T Z_{hik}(t)}. \quad (9)$$

We use superscript/subscript 'ST' to denote the stratified version of quantities. The proposed estimator $\hat{\beta}_{DW}^{ST}$ solves the following estimating equation

$$U_{DW}^{ST}(\beta) = \sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{k=1}^{K}\int_0^\tau \tilde{w}_{hik}(t)\{Z_{hik}(t) - \bar{Z}_{k,DW}(\beta,t)\}dN_{hik}(t) = 0, \quad (10)$$

where $\tilde{w}_{hik}(t) = \Delta_{hik}\xi_{hi} + (1-\Delta_{hik})\xi_{hi}\hat{\alpha}_{hk}^{-1}(t) + \Delta_{hik}(1-\xi_{hi})\eta_{hik}\hat{q}_{hk}^{-1}(t)$. Estimating equation (10) utilizes weights that are estimated within each sampling stratum. The baseline cumulative hazard function $\Lambda_{0k}(t)$ is estimated by a Breslow-Aalen type estimator $\hat{\Lambda}_{0k}^{ST}(\hat{\beta}_{DW}^{ST}, t)$ where

$$\hat{\Lambda}_{0k}^{ST}(\beta,t) = \int_0^t \frac{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\rho_{hjk}(u)dN_{hjk}(u)}{n\sum_{h=1}^{H}\sum_{j=1}^{n_h}\rho_{hjk}(u)Y_{hjk}(u)e^{\beta^T Z_{hjk}(u)}},$$

where $\rho_{hjk}(u) = \Delta_{hik}\xi_{hi} + (1-\Delta_{hik})\xi_{hi}\hat{\alpha}_{hk}^{KC}(u)^{-1} + \Delta_{hik}(1-\xi_{hi})\eta_{hik}\hat{q}_{hk}^{KC}(u)^{-1}$ is the stratified version of the weight function used in [7].

Using arguments similar to those in Appendix 2 of the supplementary material, the asymptotic properties of $\hat{\beta}_{DW}^{ST}$ can be derived. It can be shown that $\sqrt{n}(\hat{\beta}_{DW}^{ST} - \beta_0)$ converges to a zero-mean normal distribution with variance function $G^{-1}(\beta_0)\Sigma^{ST}(\beta_0)G^{-1}(\beta_0)$ where

$$\Sigma^{ST}(\beta_0) = \sum_{h=1}^{H} p_h[Q_h(\beta_0) + \frac{1-\tilde{\alpha}_h}{\tilde{\alpha}_h}V_h^I(\beta_0) + (1-\tilde{\alpha}_h)\sum_{k=1}^{K} pr(\Delta_{1k}=1)\frac{1-\tilde{q}_{hk}}{\tilde{q}_{hk}}V_{hk}^{II}(\beta_0)],$$

$$Q_h(\beta_0) = E\Big\{\sum_k M_{\bar{z},h1k}(\beta_0)\Big\}^{\otimes 2},$$

$$V_h^I(\beta_0) = var\Big\{\sum_k(1-\Delta_{h1k})\int_0^\tau\{R_{h1k}(\beta_0,t) - \mu_{hk}(t)^{-1}A_{h1k}(t)E[(1-\Delta_{h1k})R_{h1k}(\beta_0,t)]\}\times d\Lambda_{0k}(t)\Big\},$$

$$V_{hk}^{II}(\beta_0) = var\Big\{M_{\bar{z},h1k}(\beta_0) - \Big\{\int_0^\tau \theta_{hk}(t)^{-1}B_{h1k}(t)\times E[dM_{\bar{z},h1k}(\beta_0)|\Delta_{h1k}=1]\Big\}\Big|\Delta_{h1k}=1,\xi_{h1}=0\Big\}.$$

4. Simulation Studies

We performed extensive simulation studies to examine the performance of the proposed doubly-weighted estimator with finite sample size. Suppose that a case-cohort study was conducted to investigate diseases 1 and 2 ($K = 2$). We considered the following set-up. There are three covariates of interest: Z_1 and Z_3 are two first-phase covariates where $Z_1 \sim N(0.3, 0.46^2)$ and $Z_3 \sim N(1, 0.5^2)$; Z_2 is the second-phase covariate for which values are available only for subcohort members and sampled cases. We assumed that Z_2 has a first-phase continuous surrogate $\tilde{Z}_2$ that follows a $N(0.5, 0.5^2)$ distribution. We introduced $Z_4 \sim N(0.5, \sigma_4^2)$ to represent the presence of auxiliary covariates. We set $Z_2 = \tilde{Z}_2 + Z_4 + \epsilon$ where $\epsilon \sim N(0, \sigma_\epsilon^2)$ and $\tilde{Z}_2, Z_4, \epsilon$ are mutually independent. Therefore, $\sigma^2 = \sigma_4^2 + \sigma_\epsilon^2$ controls the correlation between Z_2 and its first-phase surrogate $\tilde{Z}_2$. Specifically, $corr(Z_2, \tilde{Z}_2) = (2\sqrt{0.5^2+\sigma^2})^{-1}$.

We assumed that the marginal distribution of T_{ik} is exponential with failure rate $\lambda_{0k}e^{\beta_0^T Z_{ik}}$ where β_0 is the true regression parameter vector. Correlated failure time data were generated from the Clayton-Cuzick model [19], in which the joint survival function of $T_i = (T_{i1}, \ldots, T_{iK})^T$, denote by $S(t_{1i}, \ldots, t_{Ki}|Z_{1i}, \ldots, Z_{Ki})$, has the form

$$\left\{\sum_{k=1}^{K}\exp\left(\frac{\int_0^{t_{ik}}\lambda_{0k}(t)e^{\beta_0' Z_{ik}}dt}{\theta}\right) - (K-1)\right\}^{-\theta}.$$

The positive parameter θ measures the strength of the correlation among $(T_{i1}, \ldots, T_{iK})$. The relationship between θ and Kendall's τ_θ is $\tau_\theta = 1/(2\theta+1)$. The smaller θ, the larger Kendall's τ_θ, hence the stronger the correlation. With $K = 2$, the correlated pair of failure times (T_{i1}, T_{i2}) are generated

from independent uniform variates u_{i1}, u_{i2} via

$$T_{i2} = -\log(1-u_{i2})e^{-\beta_0' Z_{i2}},$$
$$T_{i1} = \theta \log\{(1-a) + a(1-u_{i1})^{-(1+\theta)^{-1}}\}e^{-\beta_0' Z_{i1}},$$

where $a = (1-u_{i2})^{-\theta^{-1}}$. The baseline hazard functions were set to 0.3 for disease 1 and 0.5 for disease 2 ($K = 2$). Right-censoring time $C_i = C_{i1} = C_{i2}$ was generated from the uniform distribution on $[0, r]$, hence the censoring percentage was controlled by the parameter r.

4.1. Traditional Case-cohort Design

We first examined the performance of doubly-weighted estimator under the traditional case-cohort design, using the second level weight $A_{ik}(t)$ in the form of (7). We simulated full study cohort samples of size $n = 3000$. We selected a subcohort of size 300 or 450 ($\tilde{\alpha} = 0.1$ or 0.15) and then collected all the cases outside the subcohort. Right-censoring parameter r was selected so that the event rates were roughly 4% and 7% for diseases 1 and 2, respectively. Values 0.05, 0.50, 10 were considered for parameter θ, corresponding to Kendall's τ_θ of 0.91, 0.50, 0.05, to represent strong to weak correlation between the two disease outcomes. Lastly, we set $\sigma_4^2 = 0.2$ and $\sigma_\epsilon^2 = 0.06$ so that $corr(Z_2, \tilde{Z}_2) = 0.7$.

In our simulation where $p = 3$, the first-phase covariates $\hat{Z}_{ik,1}$ and $\hat{Z}_{ik,3}$ were their respective observed values. For the subjects in the subcohort and the cases, the second-phase covariate $\hat{Z}_{ik,2}$ equaled the observed values, while for non-cases outside the subcohort $Z_{ik,2}$ was missing and $\hat{Z}_{ik,2}$ equaled the estimated value. We postulated a linear model to estimate the second-phase covariate $Z_{ik,2}$ for non-subcohort controls. Using the fully observed data on subcohort controls and cases, regressing Z_2 on its surrogate $\tilde{Z}_2$ yielded an R^2 around 0.5. If we incorporated the first-phase covariates Z_1, Z_3 and Z_4, the R^2 increased to 0.85. This mimicked the situation that auxiliary information was used to improve the capability of predicting missing Z_2. We then obtained $\hat{Z}_2$ for non-subcohort controls and implemented the doubly-weighted estimator $\hat{\beta}_{DW}$. For comparison purposes, we computed estimator II, denoted $\hat{\beta}_{KC}$, in [7]. The estimator based on the full cohort $\hat{\beta}_F$, which is not feasible in practice with case-cohort designs, was also obtained as a benchmark. Results presented were based on 2000 simulations for each setting.

Table 1. *Comparison of three estimators: case-cohort design with* $\beta_0 = (0.5, 0.0, 0.2)^T$.

			$\hat{\beta}_F$			$\hat{\beta}_{KC}$				$\hat{\beta}_{DW}$					
$\tilde{n}$	τ_θ	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	$RE_{DW\|KC}$	
300	0.91	β_1	0.500	0.329	0.325	0.95	0.515	0.463	0.450	0.94	0.495	0.329	0.331	0.94	1.98
		β_2	-0.002	0.095	0.095	0.95	-0.003	0.132	0.133	0.94	-0.002	0.104	0.099	0.93	1.61
		β_3	0.199	0.139	0.137	0.94	0.204	0.193	0.190	0.95	0.197	0.138	0.140	0.94	1.96
	0.50	β_1	0.500	0.278	0.277	0.95	0.513	0.421	0.413	0.95	0.495	0.279	0.284	0.94	2.28
		β_2	-0.001	0.083	0.081	0.95	-0.001	0.123	0.122	0.94	-0.001	0.091	0.084	0.93	1.83
		β_3	0.201	0.118	0.117	0.95	0.207	0.178	0.174	0.95	0.199	0.117	0.120	0.95	2.31
	0.05	β_1	0.504	0.257	0.263	0.95	0.517	0.406	0.403	0.95	0.499	0.258	0.270	0.95	2.48
		β_2	-0.002	0.078	0.077	0.95	-0.003	0.121	0.119	0.94	-0.002	0.087	0.080	0.92	1.93
		β_3	0.200	0.110	0.110	0.95	0.206	0.172	0.170	0.95	0.198	0.110	0.114	0.95	2.44
450	0.91	β_1	0.500	0.329	0.325	0.95	0.503	0.419	0.407	0.94	0.497	0.328	0.327	0.94	1.63
		β_2	-0.002	0.095	0.095	0.95	-0.003	0.121	0.120	0.95	-0.002	0.103	0.097	0.93	1.38
		β_3	0.199	0.139	0.137	0.94	0.203	0.174	0.172	0.95	0.197	0.138	0.138	0.94	1.59
	0.50	β_1	0.500	0.278	0.277	0.95	0.500	0.371	0.368	0.95	0.497	0.277	0.280	0.94	1.79
		β_2	-0.001	0.083	0.081	0.95	-0.002	0.110	0.108	0.95	-0.001	0.091	0.083	0.92	1.46
		β_3	0.201	0.118	0.117	0.95	0.205	0.157	0.155	0.95	0.200	0.117	0.118	0.95	1.80
	0.05	β_1	0.504	0.257	0.263	0.95	0.503	0.354	0.357	0.95	0.502	0.257	0.266	0.95	1.90
		β_2	-0.002	0.078	0.077	0.95	-0.003	0.108	0.105	0.94	-0.002	0.086	0.078	0.92	1.58
		β_3	0.200	0.110	0.110	0.95	0.204	0.151	0.150	0.95	0.199	0.110	0.112	0.95	1.88

NOTE: ESD, empirical standard deviation; ESE, estimated standard error; CR, estimated standard error coverage rate of the nominal 95% confidence intervals; $RE_{DW|KC} = ESD_{KC}^2/ESD_{DW}^2$, efficiency of $\hat{\beta}_{DW}$ relative to $\hat{\beta}_{KC}$. The full cohort contained 3000 subjects.

We considered two sets of values of true regression parameters $\beta_0 = (0.5, 0.0, 0.2)^T$ and $\beta_0 = (0.5, 1.2, 0.2)^T$. Results summarized in Tables 1 and 2 show that the doubly-weighted estimator was approximately unbiased. As the subcohort size $\tilde{n}$ increased, the average of the estimated standard error became closer to the empirical standard deviation and the 95% confidence interval had satisfactory coverage rate. More importantly, $\hat{\beta}_{DW}$ could be much more efficient than $\hat{\beta}_{KC}$. In comparable scenarios, our efficiency gain over $\hat{\beta}_{KC}$ ($RE \approx 3$) was higher than that of the Kulich-Lin estimator over BII estimator in [6] ($RE \approx 1.1$). It was observed that as the correlation between T_1 and T_2 decreased, the efficiency gain increased, which means that with smaller correlation, the decrease in the variance of $\hat{\beta}_{DW}$ was faster than that of $\hat{\beta}_{KC}$. A potential explanation is that when correlation decreases, the estimator $\hat{\beta}_{KC}$ is more precise, i.e., has smaller variance. Note that the second-level weight (7) involved $\hat{\beta}_{KC}$. Because $\hat{\beta}_{KC}$ is more precise when

the correlation between T_1 and T_2 decreases, this, in turn, improves the performance of $\hat{\beta}_{DW}$. The relative efficiency was smaller when the subcohort size increased, but a gain in efficiency was still noticeable.

Table 2. *Comparison of three estimators: case-cohort design with* $\beta_0 = (0.5, 1.2, 0.2)^T$.

			$\hat{\boldsymbol{\beta}}_{\boldsymbol{F}}$				$\hat{\boldsymbol{\beta}}_{\boldsymbol{KC}}$				$\hat{\boldsymbol{\beta}}_{\boldsymbol{DW}}$				
$\tilde{n}$	τ_θ	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	$\text{RE}_{DW\|KC}$	
300	0.91	β_1	0.496	0.309	0.311	0.95	0.507	0.515	0.492	0.94	0.482	0.315	0.334	0.95	2.67
		β_2	1.199	0.094	0.093	0.95	1.235	0.167	0.153	0.92	1.223	0.098	0.112	0.95	2.90
		β_3	0.201	0.133	0.131	0.95	0.208	0.216	0.207	0.94	0.195	0.136	0.140	0.95	2.52
	0.50	β_1	0.497	0.267	0.269	0.95	0.507	0.480	0.459	0.93	0.484	0.275	0.290	0.95	3.05
		β_2	1.199	0.084	0.082	0.95	1.235	0.159	0.144	0.91	1.226	0.089	0.101	0.95	3.19
		β_3	0.199	0.114	0.113	0.96	0.210	0.202	0.193	0.94	0.194	0.117	0.122	0.95	2.98
	0.05	β_1	0.501	0.250	0.248	0.95	0.512	0.457	0.442	0.94	0.488	0.258	0.269	0.95	3.14
		β_2	1.201	0.077	0.075	0.94	1.234	0.152	0.137	0.91	1.229	0.083	0.095	0.94	3.35
		β_3	0.201	0.106	0.104	0.95	0.209	0.197	0.186	0.94	0.196	0.110	0.113	0.95	3.21
450	0.91	β_1	0.496	0.309	0.311	0.95	0.504	0.448	0.436	0.95	0.485	0.312	0.324	0.96	2.06
		β_2	1.199	0.094	0.093	0.95	1.222	0.143	0.136	0.93	1.237	0.098	0.104	0.94	2.13
		β_3	0.201	0.133	0.131	0.95	0.207	0.188	0.183	0.94	0.197	0.134	0.136	0.95	1.97
	0.50	β_1	0.497	0.267	0.269	0.95	0.505	0.412	0.400	0.94	0.487	0.270	0.280	0.95	2.33
		β_2	1.199	0.084	0.082	0.95	1.222	0.135	0.126	0.93	1.238	0.088	0.093	0.93	2.35
		β_3	0.199	0.114	0.113	0.96	0.207	0.175	0.168	0.93	0.196	0.115	0.118	0.95	2.32
	0.05	β_1	0.501	0.250	0.248	0.95	0.508	0.391	0.383	0.94	0.491	0.253	0.259	0.95	2.39
		β_2	1.201	0.077	0.075	0.94	1.222	0.128	0.120	0.93	1.241	0.082	0.085	0.93	2.44
		β_3	0.201	0.106	0.104	0.95	0.208	0.170	0.161	0.93	0.197	0.108	0.109	0.95	2.48

NOTE: The full cohort contained 3000 subjects.

Table 3. *Comparison of three estimators: generalized case-cohort design with* $\beta_0 = (0.5, 0.0, 0.2)^T$.

			$\hat{\boldsymbol{\beta}}_{\boldsymbol{F}}$				$\hat{\boldsymbol{\beta}}_{\boldsymbol{KC}}$				$\hat{\boldsymbol{\beta}}_{\boldsymbol{DW}}$				
$\tilde{n}$	τ_θ	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	$\text{RE}_{DW\|KC}$	
400	0.91	β_1	0.505	0.149	0.147	0.95	0.525	0.343	0.336	0.96	0.489	0.254	0.270	0.97	1.82
		β_2	0.001	0.043	0.043	0.96	0.000	0.102	0.099	0.94	0.000	0.072	0.081	0.97	2.01
		β_3	0.199	0.059	0.062	0.96	0.201	0.143	0.142	0.96	0.201	0.106	0.114	0.97	1.82
	0.50	β_1	0.508	0.133	0.129	0.95	0.522	0.331	0.323	0.95	0.497	0.247	0.260	0.96	1.80
		β_2	0.000	0.037	0.038	0.96	-0.001	0.096	0.095	0.94	-0.001	0.069	0.077	0.98	1.94
		β_3	0.199	0.051	0.054	0.96	0.199	0.128	0.136	0.96	0.194	0.098	0.110	0.97	1.71
	0.05	β_1	0.513	0.115	0.113	0.95	0.532	0.322	0.311	0.94	0.517	0.233	0.267	0.97	1.91
		β_2	0.001	0.031	0.033	0.96	0.004	0.088	0.091	0.95	0.007	0.066	0.074	0.98	1.78
		β_3	0.198	0.046	0.047	0.96	0.203	0.125	0.131	0.96	0.203	0.100	0.106	0.97	1.56
600	0.91	β_1	0.505	0.149	0.147	0.95	0.515	0.286	0.273	0.94	0.516	0.220	0.219	0.96	1.69
		β_2	0.001	0.043	0.043	0.96	0.001	0.082	0.080	0.94	-0.001	0.058	0.062	0.97	2.00
		β_3	0.199	0.059	0.062	0.96	0.197	0.114	0.115	0.94	0.198	0.086	0.091	0.97	1.76
	0.50	β_1	0.508	0.133	0.129	0.95	0.505	0.272	0.259	0.94	0.504	0.189	0.198	0.97	2.07
		β_2	0.000	0.037	0.038	0.96	0.002	0.076	0.076	0.94	-0.002	0.056	0.059	0.96	1.84
		β_3	0.199	0.051	0.054	0.96	0.199	0.106	0.109	0.95	0.200	0.079	0.084	0.96	1.80
	0.05	β_1	0.513	0.115	0.113	0.95	0.512	0.258	0.246	0.94	0.522	0.178	0.188	0.96	2.10
		β_2	0.001	0.031	0.033	0.96	0.003	0.072	0.072	0.96	0.001	0.049	0.056	0.97	2.16
		β_3	0.198	0.046	0.047	0.96	0.195	0.111	0.104	0.92	0.195	0.075	0.079	0.95	2.19

NOTE: The full cohort contained 4000 subjects.

4.2. Generalized Case-cohort Design

We then examined the performance of the doubly-weighted estimator under a generalized case-cohort design with non-rare diseases. In practice, it is common to take a 'balanced' sample in which the numbers of cases and controls are roughly the same. Let the proportion with disease k be P_k. By simple algebra, we obtained that $\tilde{q}_k$, the case sampling proportion to achieve the expected case/control ratio R_k for disease k, is independent of full cohort size n:

$$\tilde{q}_k = \frac{[(1-P_k)R_k - P_k]\tilde{\alpha}}{P_k(1-\tilde{\alpha})}. \tag{11}$$

As was discussed in Section 2.2, generalized case-cohort samples are obtained separately for each disease outcome k

and overlapping subjects in the sample are allowed. Therefore, it is not needed to adjust for the disease-specific case sampling proportion $\tilde{q}_k$ to reflect that some subjects may have more than one disease outcomes of interest.

We considered the full cohort size of 4000. We then selected a subcohort of size 400 or 600 ($\tilde{\alpha} = 0.1$ or 0.15). The right-censoring parameter r was set to 0.25 so that the event rate was 19% for disease 1 and 28% for disease 2. Based on (11), the corresponding vectors of q_k to achieve roughly a 1:1 case/control ratio were $(0.36, 0.18)$ or $(0.58, 0.28)$, respectively.

We set both $A_{ik}(t)$ and $B_{ik}(t)$ to be in the same form as in (7). Results based on 2000 simulations are presented in Tables 3 and 4. Both estimators were generally unbiased. However, when the subcohort sampling proportion was below 0.1 (results not presented), the standard deviation of $\hat{\beta}_{DW}$ could not be estimated accurately and the efficiency gain was minimal. This phenomenon echoed our discussion in section 3.3 that the doubly-weighted estimator requires a larger sample size to obtain a stable variance estimator. On the other hand, $\hat{\beta}_{KC}$ yielded a good standard deviation estimator regardless of $\tilde{\alpha}$. We can see that the doubly-weighted estimator is more efficient than $\hat{\beta}_{KC}$, although the magnitude of the efficiency gain was not as large compared to the traditional case-cohort design. The correlation between the two diseases did not appear to affect the relative efficiency.

Table 4. *Comparison of three estimators: generalized case-cohort design with* $\beta_0 = (0.5, 1.2, 0.2)^T$.

			$\hat{\beta}_F$			$\hat{\beta}_{KC}$				$\hat{\beta}_{DW}$					
$\tilde{n}$	τ_θ	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	Mean	ESD	ESE	CR	$RE_{DW\|KC}$	
400	0.91	β_1	0.505	0.152	0.146	0.93	0.521	0.347	0.338	0.94	0.501	0.319	0.282	0.92	1.18
		β_2	1.200	0.048	0.046	0.94	1.225	0.100	0.101	0.95	1.183	0.085	0.088	0.94	1.38
		β_3	0.198	0.062	0.061	0.95	0.212	0.141	0.143	0.94	0.199	0.132	0.119	0.91	1.14
	0.50	β_1	0.503	0.134	0.131	0.94	0.534	0.328	0.327	0.95	0.536	0.308	0.272	0.91	1.13
		β_2	1.201	0.041	0.041	0.95	1.224	0.102	0.097	0.94	1.183	0.085	0.086	0.93	1.44
		β_3	0.200	0.055	0.055	0.95	0.202	0.137	0.138	0.95	0.192	0.128	0.113	0.92	1.15
	0.05	β_1	0.502	0.111	0.111	0.95	0.513	0.313	0.309	0.95	0.504	0.291	0.298	0.92	1.16
		β_2	1.200	0.036	0.036	0.95	1.220	0.096	0.091	0.94	1.185	0.084	0.084	0.94	1.31
		β_3	0.201	0.046	0.047	0.95	0.206	0.130	0.130	0.95	0.197	0.128	0.111	0.90	1.03
600	0.91	β_1	0.505	0.152	0.146	0.93	0.511	0.274	0.276	0.95	0.502	0.243	0.216	0.92	1.27
		β_2	1.200	0.048	0.046	0.94	1.209	0.086	0.082	0.93	1.182	0.071	0.071	0.93	1.47
		β_3	0.198	0.062	0.061	0.95	0.197	0.115	0.116	0.95	0.200	0.102	0.091	0.92	1.27
	0.50	β_1	0.503	0.134	0.131	0.94	0.508	0.268	0.264	0.95	0.496	0.235	0.205	0.91	1.30
		β_2	1.201	0.041	0.041	0.95	1.209	0.080	0.079	0.94	1.185	0.068	0.065	0.92	1.38
		β_3	0.200	0.055	0.055	0.95	0.204	0.117	0.111	0.94	0.201	0.095	0.087	0.92	1.52
	0.05	β_1	0.502	0.111	0.111	0.95	0.499	0.241	0.246	0.96	0.499	0.219	0.197	0.91	1.21
		β_2	1.200	0.036	0.036	0.95	1.204	0.076	0.073	0.93	1.184	0.062	0.061	0.94	1.50
		β_3	0.201	0.046	0.047	0.95	0.201	0.103	0.103	0.94	0.201	0.091	0.082	0.92	1.28

NOTE: The full cohort contained 4000 subjects.

5. Data Analysis

We applied the proposed procedures to a data set from the Atherosclerosis Risk in Communities (ARIC) study [10, 11]. The ARIC study is a large cohort study which enrolled 15,792 middle-aged men and women from four US communities. A baseline examination was conducted from 1987 to 1989, with 3 more examinations at roughly 3-year intervals through 1998. Participants were followed up for incident CHD, including CHD-related death, and ischemic incident stroke, a first definite or probable hospitalized stroke through 1998. It was of interest to examine whether lipoprotein-associated phospholipase, Lp-PLA_2, was associated with increased risk for incident CHD and ischemic stroke. After applying the exclusion criteria, a total of 12,363 subjects comprised the full cohort for this analysis. In order to preserve stored plasma samples and reduce cost, case-cohort studies were implemented, one for CHD and one for stroke. Specifically, a subcohort of participants was sampled. This subcohort together with those participants who have had CHD by 12/31/1998 constitute the case-cohort sample for the CHD case-cohort study. Similarly, the case-cohort study for stroke contains participants who are in the subcohort and those who have had stroke by 12/31/1998. For participants who are in either case-cohort study, their stored blood sample from visit 1 was thawed and measured for Lp-PLA_2.

Those who were still alive or disease-free by 12/31/1998 or lost to follow-up were treated as censored at 12/31/1998 or at the last contact, respectively. The subcohort was selected using stratified sampling based on gender, race (white versus black), and age group (below versus above 55). Table 5 shows participants' characteristics at visit 2 among different subgroups.

In this analysis, the two disease outcomes of interest were incident CHD and incident ischemic stroke. A total of 603 CHD cases and 183 ischemic incident stroke cases, along with 777 subcohort subjects, were included in the sample. As some participants had both disease outcomes, the total number of serum samples assayed was 1,470. The main exposure of interest was the tertile group indicators of Lp-PLA_2

(low/moderate/high Lp-PLA$_2$, with the reference level being the low Lp-PLA$_2$ group). Potential confounders included in the model were three first-phase stratum covariates, age at visit 2, gender and race, so that our model was comparable to model 1 in [10]. We used a disease-specific effect model for CHD and stroke to allow different effects for the same set of covariates. This resulted in a total of 10 regression parameters to be estimated.

We implemented the proposed doubly-weighted estimator $\hat{\beta}_{DW}$ with second level weight (7). To this end, we built a prediction model for the second-phase covariate Lp-PLA$_2$ (in mg/dL) among non-subcohort controls. The first-phase covariates used in the regression model for Lp-PLA$_2$ were race, gender, LDL-C, HDL-C and smoking status (never smoked/former smoker/current smoker). We assigned the tertile group indicators based on the predicted values. For comparison purposes, we calculated $\hat{\beta}_{KC}$, the estimator in [7]. We used the stratified version of estimating equation (10) and variance estimators to accommodate the stratified sampling nature of this ancillary study to the ARIC study. The coefficient estimates, standard errors and associated p-values are presented in table 6. There was fair agreement between the two methods in terms of point estimates. The findings matched those reported in univariate analyses [10, 11]. In terms of efficiency, $\hat{\beta}_{DW}$ outperformed $\hat{\beta}_{KC}$: despite a negligible (no more than 6%) increase in standard errors of 3 parameter estimates, $\hat{\beta}_{DW}$ yielded noticeably more efficient results elsewhere. The most noteworthy finding was for the high Lp-PLA$_2$ group: using the doubly-weighted estimator, we had strong evidence that it was significantly associated with elevated incident CHD risk (HR: 1.729, 95% CI: 1.092, 2.736), compared to the low Lp-PLA$_2$ group. On the other hand, $\hat{\beta}_{KC}$ deemed the effect non-significant (HR: 1.567, 95% CI: 0.846, 2.903). Other first-phase risk factors that were statistically associated with elevated risks were advancing age (CHD and stroke), white race (CHD) and male sex (stroke). Based on $\hat{\beta}_{DW}$, we performed a Wald test with 2 degrees of freedom to compare the corresponding coefficients for the Lp-PLA$_2$ group indicators between the two diseases. The p-value for the Wald test was 0.6580, suggesting the Lp-PLA$_2$ effects for the two diseases were not significantly different.

Table 5. *Baseline Characteristics of the ARIC Study.*

	CHD (n=604)	**Stroke (n = 183)**	**Subcohort (n = 777)**	**Full (n = 12,363)**
Age (SD), years	58.6 (5.44)	59.7 (5.54)	56.9 (5.57)	56.8 (5.70)
Male Sex, %	67.7	55.7	42.7	42.2
White Race, %	77.1	56.8	75.2	75.6
Lp-PLA$_2$ (SD), mg/L	0.427 (0.14)	0.451 (0.17)	0.378 (0.13)	N/A
Lp-PLA$_2$: Moderate †, %	31.5	22.4	33.9	N/A
Lp-PLA$_2$: High ‡, %	48.0	53.6	34.4	N/A

†Lp-PLA$_2$ between 0.310 and 0.422 mg/L
‡Lp-PLA$_2$ above 0.422 mg/L

Table 6. *Coefficient Estimates of Disease-Specific Effect Model.*

	$\hat{\boldsymbol{\beta}}_{\boldsymbol{DW}}$			$\hat{\boldsymbol{\beta}}_{\boldsymbol{KC}}$		
	Estimate	**Std Err**	**P-value**	**Estimate**	**Std Err**	**P-value**
Disease: CHD						
Age in years/10	0.5279	0.1076	$<$ .0001	0.4756	0.2020	0.0185
Male sex	1.0346	0.2411	$<$ .0001	0.9798	0.2730	0.0003
White race	-0.0904	0.2359	0.7016	-0.1692	0.2591	0.5137
Lp-PLA$_2$: Moderate	0.4135	0.3469	0.2333	0.2573	0.3296	0.4350
Lp-PLA$_2$: High	0.5474	0.2343	0.0195	0.4490	0.3146	0.1535
Disease: Stroke						
Age in years/10	0.9702	0.2297	$<$ .0001	1.0108	0.4175	0.0155
Male sex	0.6109	0.4134	0.1395		0.4328 0.4413	0.3267
White race	-0.9571	0.3936	0.0150	-1.2054	0.3906	0.0020
Lp-PLA$_2$: Moderate	-0.1162	0.6003	0.8465	-0.2028	0.5989	0.7349
Lp-PLA$_2$: High	0.4435	0.3697	0.2303	0.6961	0.4849	0.1511

6. Recommendations

(Generalized) case-cohort designs have been widely used in medical studies when it is prohibitively costly to measure some exposures for all subjects in the full cohort. In this research, we proposed a class of doubly-weighted estimators for multiple disease outcomes. With the choice of second level weights $A_{ik}(t)$ and $B_{ik}(t)$ being almost arbitrary, this class encompasses many estimators as special cases including that in [7]. We derived the doubly-weighted estimating equation and the asymptotic properties associated with the estimator solving it. We then implemented the method using a specific form of second level weight which utilized first-phase

covariate information collected. Simulation studies showed considerable efficiency gain over the estimator in [7] and the method was also applied on the data from ARIC study.

In univariate generalized case-cohort design, the power of the design could depend on various factors such as event rate, distribution of exposure/covariates, etc [15]. Reference [20] derived the optimal allocation for stratified case-cohort studies under various situations with only one disease outcome. It will be of interest to investigate the optimal allocation with multiple disease outcomes.

With multiple time-to-event outcomes, an alternative to the (generalized) case-cohort design is the nested case-control design, e.g., references [21, 22]. It will also be of interest to compare the performances of the two classes of designs. In terms of utilizing auxiliary information to improve efficiency, it may also be of interest, to explore whether the calibration approach [13, 14] can be generalized to the multivariate failure time case, where the variance-covariance matrix among the coefficients need to be properly handled. For the traditional case-cohort design, reference [23] proposed a class of updated estimators to improve the efficiency using auxiliary information. Extending their method to the generalized case-cohort setting could be another direction of future research.

When implementing the doubly-weighted estimator with second level weight (7), we need to build a prediction model for the unobserved second-phase covariates. In both simulation studies and the real data application, we chose to build the model using linear regression. Kernel regression and polynomial regression with carefully calibrated smoothing parameter can be explored, if flexible forms of the covariates are desired. Other choices of second level weights are possible. For example, in a similar fashion to [24], we can incorporate the Nadaraya-Watson kernel estimator in the second level weight.

Throughout this paper, we have assumed a Cox-type marginal proportional hazards model. Additive hazards models, which model risk differences, have often been used as an alternative to the proportional hazards model. For data arising from multiple case-cohort studies, reference [25] proposed a marginal additive hazards model based on a weighted estimating equation approach. They also considered the generalized case-cohort design. To improve efficiency, extending the proposed doubly-weighted approach to the marginal additive hazards model will allow us to make full use of first-phase covariate information, and thus may merit further investigation.

Appendix

Appendix 1. Explicit Form of $D_{DW}(\beta)$

Due to the matrix nature of $S^{(0)}_{k,DW}(\beta,t)$, special attention is required to compute $D_{DW}(\beta)$. Let $l,l'=1,\ldots,p$, we can explicitly express $\tilde{w}_{ik}(t)$ and $Z_{ik}(t)$:

$$\tilde{w}_{ik}(t)=diag\{\tilde{w}_{ik,1}(t),\tilde{w}_{ik,2}(t),\ldots,\tilde{w}_{ik,p}(t)\},\quad Z_{ik}(t)=[Z_{ik,1}(t),Z_{ik,2}(t),\ldots,Z_{ik,p}(t)]^T.$$

Define the scalar functions

$$S^{(0)}_{k,DW,l}(\beta,t)=n^{-1}\sum_{i=1}^{n}\tilde{w}_{ik,l}(t)Y_{ik}(t)exp\{\beta^T Z_{ik}(t)\},$$

$$S^{(1)}_{k,DW,ll'}(\beta,t)=n^{-1}\sum_{i=1}^{n}\tilde{w}_{ik,l}(t)Z_{ik,l'}(t)Y_{ik}(t)exp\{\beta^T Z_{ik}(t)\},$$

and

$$S^{(2)}_{k,DW,ll'}(\beta,t)=n^{-1}\sum_{i=1}^{n}\tilde{w}_{ik,l}(t)Z_{ik,l}(t)Z_{ik,l'}(t)Y_{ik}(t)exp\{\beta^T Z_{ik}(t)\}.$$

Let $V_{k,DW}(\beta,t)$ be the derivative of $-\bar{Z}_{k,DW}(\beta,t)$ with respect to β. We have

$$D_{DW}(\beta)=\frac{\partial U_{DW}(\beta)}{\partial\beta^T}=\sum_{k=1}^{K}\int_0^{\tau}\tilde{w}_{ik}(t)V_{k,DW}(\beta,t)d\sum_{i=1}^{n}N_{ik}(t),$$

where the lth row of $V_{k,DW}(\beta,t)$ has the form

$$S^{(0)}_{k,DW,l}(\beta,t)^{-2}\Big\{S^{(1)}_{k,DW,ll}(\beta,t)[S^{(1)}_{k,DW,l1}(\beta,t),\ldots,S^{(1)}_{k,DW,lp}(\beta,t)]-[S^{(2)}_{k,DW,l1}(\beta,t),\ldots,S^{(2)}_{k,DW,lp}(\beta,t)]S^{(0)}_{k,DW,l}(\beta,t)\Big\}.$$

Appendix 2. Proof of Theorem 1

The following two lemmas are important in deriving the asymptotic results and are applied repeatedly.

Lemma 6.1. Let $\xi = (\xi_1, ..., \xi_n)^T$ be a random vector containing $\tilde{n}$ ones and $n - \tilde{n}$ zeros, with each permutation equally likely. Let $B_i(t), i = 1, ..., n$ be independent and identically distributed 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 $B(t) = \{B_1(t), ..., B_n(t)\}^T$ 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 $l^\infty[0, \tau]$ 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.

This lemma is stated as Lemma A1 in [7]. Its proof involves the central limit theorem for finite population sampling from [26] and example 3.6.14 of [27]. A special case of this lemma is obtained by setting $\xi = J_n$ where J_n is an n-vector of ones.

We need the following results on the asymptotic properties of $\hat{\alpha}_k(t)$ and $\hat{q}_k(t)$. We present and prove Lemma B6.2, Lemma B6.4 and Theorem 1 assuming a single covariate in (1). With multiple covariates, $\hat{\alpha}_k(t)$, $\hat{q}_k(t)$ and $S^{(0)}_{k,DW}(\beta, t)$ are p-by-p diagonal matrices, and the arguments below pertain to each of the p processes on the diagonal.

Lemma 6.2.

$$n^{1/2}(\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1}) = \{\tilde{\alpha}\mu_k(t)\}^{-1} n^{-1/2} \sum_{i=1}^{n} (1 - \xi_i/\tilde{\alpha})(1 - \Delta_{ik}) A_{ik}(t) + o_p(1), \tag{12}$$

in which $\mu_k(t)$ is defined as $E[(1 - \Delta_{1k})A_{1k}(t)]$. Also, we have similar results for $\hat{q}_k(t)$:

$$n^{1/2}(\hat{q}_k(t)^{-1} - \tilde{q}_k^{-1}) = \{\tilde{q}_k(1 - \tilde{\alpha})\theta_k(t)\}^{-1} n^{-1/2} \sum_{i=1}^{n} (1 - \eta_{ik}/\tilde{q}_k)\Delta_{ik}(1 - \xi_i) B_{ik}(t) + o_p(1), \tag{13}$$

in which $\theta_k(t) = E[\Delta_{1k} B_{1k}(t)]$.

The detailed proof for equation (12), which utilizes assumptions 10 to 12, the special case of Lemma B6.1 and the functional delta method, can be found in [12]. Since the failure times for each disease are independently distributed, the proof of (12) can go through without modification. Equation (13) can be shown analogously.

We need another technical lemma from [28].

Lemma 6.3. Let $W(t)$ and $Z(t)$ be two sequences of bounded processes. Suppose that $W(t)$ is monotone and converges to $w(t)$ uniformly in t in probability and that $Z(t)$ converges weakly to a zero-mean process with continuous sample paths. Then

$$\int_0^t \{W(u) - w(u)\} dZ(u) \to 0, \quad \int_0^t Z(u) d\{W(u) - w(u)\} \to 0$$

uniformly in t in probability.

The next lemma states the uniform convergence of $\bar{Z}_{k,DW}(\beta, t)$, to the limit of its full cohort counterpart.

Lemma 6.4. (Convergence of the at-risk average process) For any k,

$$\sup_{\beta,t} \left\| \bar{Z}_{k,DW}(\beta, t) - \bar{z}_k(\beta, t) \right\| \to_p 0.$$

Proof We first show that $\sup_{\beta,t} \|S^{(d)}_{k,DW}(\beta, t) - S^{(d)}_{k,F}(\beta, t)\| \to_p 0$ uniformly in t and β for $d = 0, 1$. We start with

$$S^{(d)}_{k,DW}(\beta, t) - S^{(d)}_{k,F}(\beta, t) = n^{-1} \sum_i \{\tilde{w}_{ik}(t) - 1\} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)$$

Expanding the weight function $\tilde{w}_{ik}(t)$ and rearranging terms on the right-hand side (RHS), we get

$$\begin{aligned}
S^{(d)}_{k,DW}(\beta, t) - S^{(d)}_{k,F}(\beta, t) = {} & n^{-1} \sum_i (\frac{\xi_i}{\tilde{\alpha}} - 1) Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \\
& - n^{-1} \sum_i (\frac{\eta_{ik}}{\tilde{q}_k} - 1) \Delta_{ik} \xi_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \\
& - n^{-1} \sum_i (\frac{\xi_i}{\tilde{\alpha}} - 1) \Delta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \\
& + n^{-1} \sum_i (\frac{\eta_{ik}}{\tilde{q}_k} - 1) \Delta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \\
& + n^{-1} \sum_i (\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1})(1 - \Delta_{ik}) \xi_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \\
& + n^{-1} \sum_i (\hat{q}_k(t)^{-1} - \tilde{q}_k^{-1}) \Delta_{ik} (1 - \xi_i) \eta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t).
\end{aligned}$$

Taking the norm on both sides,

$$
\begin{aligned}
&\left\| S^{(d)}_{k,DW}(\beta,t) - S^{(d)}_{k,F}(\beta,t) \right\| \\
\leq &\left\| n^{-1}\sum_i (\frac{\xi_i}{\tilde{\alpha}} - 1) Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\| && (14)\\
+ &\left\| n^{-1}\sum_i (\frac{\eta_{ik}}{\tilde{q}_k} - 1) \Delta_{ik}\xi_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\| && (15)\\
+ &\left\| n^{-1}\sum_i (\frac{\xi_i}{\tilde{\alpha}} - 1) \Delta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\| && (16)\\
+ &\left\| n^{-1}\sum_i (\frac{\eta_{ik}}{\tilde{q}_k} - 1) \Delta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\| && (17)\\
+ &\left\| n^{-1}\sum_i (\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1})(1-\Delta_{ik})\xi_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\| && (18)\\
+ &\left\| n^{-1}\sum_i (\hat{q}_k(t)^{-1} - \tilde{q}_k^{-1}) \Delta_{ik}(1-\xi_i)\eta_{ik} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) \right\|. && (19)
\end{aligned}
$$

We now show each of the six terms converges to 0 in probability uniformly in β and t. (14) converges to 0 in probability uniformly in t by the special case of Lemma B6.1. Specifically,

$$\|n^{-1}\sum_i (\frac{\xi_i}{\tilde{\alpha}} - 1) Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)\| = \|n^{-1}\sum_i \frac{\xi_i}{\tilde{\alpha}} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t) - n^{-1}\sum_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)\|.$$

By iterated expectation argument conditioning on everything but ξ_i, it is clear that $E[\frac{\xi_i}{\tilde{\alpha}} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)] = E[Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)] = \mu_B(t)$. Also, by assumption, $\frac{\xi_i}{\tilde{\alpha}} Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)$ has finite variation on $[0,\tau]$. Hence, the aforementioned lemma guarantees the convergence of (14) to 0, uniformly in t and β. Through similar arguments, (15) - (17) converge to 0 in probability uniformly in t and β, respectively.

We then show that (18) converges to 0 in probability uniformly in t and β. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\|n^{-1}\textstyle\sum_i (\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1})(1-\Delta_{ik})\xi_i Z^d_{ik}(t) e^{\beta^T Z_{ik}(t)} Y_{ik}(t)\| \\
\leq \quad &\|\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1}\| \cdot n^{-1}\textstyle\sum_i (1-\Delta_{ik})\xi_i \|Z^d_{ik}(t)\| e^{\beta^T Z_{ik}(t)} Y_{ik}(t),
\end{aligned}
$$

The latter converges to 0 in probability, uniformly in t and β. This can be justified by noting $\hat{\alpha}_k(t)^{-1} - \tilde{\alpha}^{-1}$ converges to 0 in probability uniformly in t, in view of Lemma B6.2. Also by the Lindeberg condition, $n^{-1}\sum_i (1-\Delta_{ik})\xi_i \|Z^d_{ik}(t)\| e^{\beta^T Z_{ik}(t)} Y_{ik}(t)$ converges to a finite quantity. Likewise, (19) can be shown to converge to 0 in probability uniformly in t and β. Therefore, we have shown that $\sup_{\beta,t} \|S^{(d)}_{k,DW}(\beta,t) - S^{(d)}_{k,F}(\beta,t)\| \to_p 0$ uniformly in t and β, for $d = 0, 1$. This result, in combination with assumption 6, leads to the conclusion that $\sup_{\beta,t} \|S^{(d)}_{k,DW}(\beta,t) - s^{(d)}_k(\beta,t)\| \to_p 0$ uniformly in t and β.

To obtain the main result of the lemma, we start with

$$
\begin{aligned}
\sup_{\beta,t} \left\| \bar{Z}_{k,DW}(\beta,t) - \bar{z}_k(\beta,t) \right\| &= \sup_{\beta,t} \left\| \bar{Z}_{k,DW}(\beta,t) - \bar{Z}_{k,F}(\beta,t) + \bar{Z}_{k,F}(\beta,t) - \bar{z}_k(\beta,t) \right\| \\
&\leq \sup_{\beta,t} \left\| \bar{Z}_{k,DW}(\beta,t) - \bar{Z}_{k,F}(\beta,t) \right\| \\
&+ \sup_{\beta,t} \left\| \bar{Z}_{k,F}(\beta,t) - \bar{z}_k(\beta,t) \right\|
\end{aligned}
$$

Clearly, the second term on the RHS of the inequality converges to 0 in probability based on the full data result. The first term

can be written as:

$$\sup_{\beta,t}\left\|\frac{S^{(0)}_{k,F}(\beta,t)\{S^{(1)}_{k,DW}(\beta,t)-S^{(1)}_{k,F}(\beta,t)\}+S^{(1)}_{k,F}(\beta,t)\{S^{(0)}_{k,F}(\beta,t)-S^{(0)}_{k,DW}(\beta,t)\}}{S^{(0)}_{k,DW}(\beta,t)S^{(0)}_{k,F}(\beta,t)}\right\|$$
$$\leq\sup_{\beta,t}\left\|\frac{S^{(0)}_{k,F}(\beta,t)\{S^{(1)}_{k,DW}(\beta,t)-S^{(1)}_{k,F}(\beta,t)\}}{S^{(0)}_{k,DW}(\beta,t)S^{(0)}_{k,F}(\beta,t)}\right\|$$
$$+\sup_{\beta,t}\left\|\frac{S^{(1)}_{k,F}(\beta,t)\{S^{(0)}_{k,F}(\beta,t)-S^{(0)}_{k,DW}(\beta,t)\}}{S^{(0)}_{k,DW}(\beta,t)S^{(0)}_{k,F}(\beta,t)}\right\|.$$

Both terms converge to 0 in probability by assumption 6 and that $\sup_{\beta,t}\|S^{(d)}_{k,DW}(\beta,t)-S^{(d)}_{k,F}(\beta,t)\|\to_p 0$ uniformly in t and β, for $d=0,1$. This completes the proof. We are now in place to prove theorem 1. *Proof* The consistency of $\hat{\beta}_{DW}$ can be shown by the extension of [29]. Denote $n^{-1}U_{DW}(\beta)$ by $\tilde{U}_{DW}(\beta)$. $\hat{\beta}_{DW}$ is consistent if all four conditions below hold: (i) $\partial\tilde{U}_{DW}(\beta)/\partial\beta^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0; (ii) $\partial\tilde{U}_{DW}(\beta)/\partial\beta^T$ is negative definite with probability going to one as $n\to\infty$; (iii) $-\partial\tilde{U}_{DW}(\beta)/\partial\beta^T$ converges to $G(\beta_0)$ in probability uniformly for β in an open neighborhood of β_0; (iv) $\tilde{U}_{DW}(\beta)$ converges to 0 in probability.

We need to verify the four conditions to establish consistency. The form of $\partial\tilde{U}_{DW}(\beta)/\partial\beta^T$ was given in Appendix 1, hence (i) holds due to the continuity of each part. (ii) and (iii) are satisfied if we can show $\|-\partial\tilde{U}_{DW}(\beta)/\partial\beta^T-G(\beta)\|$ converges to 0 in probability uniformly in $\beta\in\mathcal{B}$ as $n\to\infty$. We make the decomposition

$$\begin{aligned}\left\|-\frac{\partial\tilde{U}_{DW}(\beta)}{\partial\beta^T}-G(\beta)\right\| &\leq \left\|\sum_{k=1}^{K}\int_0^\tau\{V_{k,DW}(\beta,t)-v_k(\beta,t)\}n^{-1}d\sum_{i=1}^{n}N_{ik}(t)\right\|\\ &+ \left\|\sum_{k=1}^{K}\int_0^\tau v_k(\beta,t)n^{-1}d\sum_{i=1}^{n}M_{ik}(t)\right\|\\ &+ \left\|\sum_{k=1}^{K}\int_0^\tau v_k(\beta,t)\{S^{(0)}_{k,DW}(\beta,t)-s^{(0)}_k(\beta,t)\}d\Lambda_{0k}(t)\right\|. \qquad (20)\end{aligned}$$

Each term on the RHS of (20) will be shown to converge to 0, uniformly in $\beta\in\mathcal{B}$. While proving Lemma B6.4, we showed that $\sup_{\beta,t}\|S^{(d)}_{k,DW}(\beta,t)-s^{(d)}_k(\beta,t)\|\to_p 0$ uniformly in t and β, for $d=0,1$. From the derivation in Appendix 1, it follows naturally that $V_{k,DW}(\beta,t)$ converges to $v_k(\beta,t)$ uniformly in t and β. By the Lenglart inequality, for any $\delta,\rho>0$, there exists n_0 such that for $n\geq n_0$,

$$P[n^{-1}\bar{N}_k(\tau)>c]\leq\frac{\delta}{c}+P[\int_0^\tau S^{(0)}_{k,DW}(\beta_0,t)\lambda_{0k}(t)dt>\delta].$$

By assumption 6, for $\delta>\int_0^\tau s^{(0)}_k(\beta_0,t)\lambda_{0k}(t)dt$, $P[\int_0^\tau S^{(0)}_{k,DW}(\beta_0,t)\lambda_{0k}(t)dt>\delta]\to 0$ as $n\to\infty$. Then $\lim_{c\uparrow\infty}\lim_{n\to\infty}P[n^{-1}\bar{N}_k(\tau)>c]=0$. Therefore, the first term on the RHS converges to 0 in probability uniformly in $\beta\in\mathcal{B}$ as $n\to\infty$.

For the second term, $n^{-1}\sum_{i=1}^{n}\int_0^\tau v_k(\beta,t)dM_{ik}(t)$ is a local square integrable martingale. The Lenglart inequality implies that, for any $\delta,\rho>0$, there exists n_0 such that for $n\geq n_0$,

$$P\left[\left\|n^{-1}\int_0^\tau\{v_k(\beta,t)\}_{ll'}d\bar{M}_k(t)\right\|>\rho\right]\leq\frac{\delta}{\rho^2}+P\left[n^{-1}\int_0^\tau\{v_k(\beta,t)\}^2_{ll'}S^{(0)}_{k,DW}(\beta_0,t)\lambda_{0k}(t)dt>\delta\right]$$

where the subscript ll' denotes the (l,l') element of the matrix. Assumptions 5-7 ensure that $P[n^{-1}\int_0^\tau\{v_k(\beta,t)\}^2_{ll'}S^{(0)}_{k,DW}(\beta_0,t)\lambda_{0k}(t)dt>\delta]$ converges to 0 in probability uniformly in $\beta\in\mathcal{B}$ for any δ. Then the second term on the RHS of (20) also converges to 0 in probability uniformly in $\beta\in\mathcal{B}$ as $n\to\infty$, since δ can be arbitrarily small.

Finally, by assumptions 4-6 and uniform convergence of $S^{(0)}_{k,DW}(\beta,t)$ to $s^{(0)}_k(\beta,t)$ in probability, the last term on the RHS of (20) converges to 0 uniformly in $\beta\in\mathcal{B}$ as $n\to\infty$. Therefore, the left-hand side of (20) converges to 0 uniformly in $\beta\in\mathcal{B}$ as $n\to\infty$. Then conditions (ii) and (iii) are satisfied.

Convergence of $\tilde{U}_{DW}(\beta)$ to zero in probability shows that (iv) is satisfied. Therefore, $\hat{\beta}_{DW}$ is a consistent estimator of β_0.

To establish the asymptotic normality of the doubly-weighted score process, we make the decomposition of $n^{-1/2}U_{DW}(\beta_0)$

$$\begin{aligned}
n^{-1/2}U_{DW}(\beta_0) &= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \tilde{w}_{ik}(t)\{Z_{ik}(t)-\bar{Z}_{k,DW}(\beta,t)\}dN_{ik}(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \tilde{w}_{ik}(t)\{Z_{ik}(t)-\bar{Z}_{k,DW}(\beta,t)\}dM_{ik}(t)\\
&= n^{-1/2}\sum_{k}\sum_{i}\int_0^\tau \{Z_{ik}(t)-\bar{z}_k(\beta_0,t)\}dM_{ik}(t) \qquad (21)\\
&+ n^{-1/2}\sum_{k}\sum_{i}\int_0^\tau \{\bar{z}_k(\beta_0,t)-\bar{Z}_{k,DW}(\beta,t)\}dM_{ik}(t) \qquad (22)\\
&+ n^{-1/2}\sum_{k}\sum_{i}\int_0^\tau (\tilde{w}_{ik}(t)-1)\{Z_{ik}(t)-\bar{z}_k(\beta_0,t)\}dM_{ik}(t) \qquad (23)\\
&+ n^{-1/2}\sum_{k}\sum_{i}\int_0^\tau (\tilde{w}_{ik}(t)-1)\{\bar{z}_k(\beta_0,t)-\bar{Z}_{k,DW}(\beta,t)\}dM_{ik}(t) \qquad (24)\\
&+ o_p(1).
\end{aligned}$$

Using the example in 2.11.16 of [27], the Kolmogorov-Centsov theorem, Lemma B6.2 and B6.3, (22) and (24) can be shown to converge to 0 in probability, uniformly in t. In [18], (21) was shown to converge to a zero mean normal distribution with covariance matrix $Q(\beta_0)$, where $Q(\beta_0) = E[\sum_{k=1}^{K}\int_0^\tau \tilde{Z}_{ik}(\beta,t)dM_{ik}(t)]^{\otimes 2}$.

We can further decompose (23) by expanding $\tilde{w}_{ik}(t)$:

$$\begin{aligned}
& n^{-1/2}\sum_{k}\sum_{i}\int_0^\tau (\tilde{w}_{ik}(t)-1)dM_{\bar{z},ik}(\beta_0,t) \qquad (25)\\
= \;& n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})\xi_i\int_0^\tau (\hat{\alpha}_k^{-1}(t)-\tilde{\alpha}^{-1})dM_{\bar{z},ik}(\beta_0,t) \qquad (26)\\
+ \;& n^{-1/2}\sum_{k}\sum_{i}\Delta_{ik}(1-\xi_i)\eta_{ik}\int_0^\tau (\hat{q}_k^{-1}(t)-\tilde{q}_k^{-1})dM_{\bar{z},ik}(\beta_0,t) \qquad (27)\\
+ \;& n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)M_{\bar{z},ik}(\beta_0) \qquad (28)\\
+ \;& n^{-1/2}\sum_{k}\sum_{i}\Delta_{ik}(1-\xi_i)(\eta_{ik}\tilde{q}_k^{-1}-1)M_{\bar{z},ik}(\beta_0). \qquad (29)
\end{aligned}$$

By (12), (26) is equal to

$$\begin{aligned}
& n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})\xi_i\times\int_0^\tau [\{\tilde{\alpha}\mu_k(t)\}^{-1}n^{-1}\sum_{j}(1-\xi_j\tilde{\alpha}^{-1})(1-\Delta_{jk})A_{jk}(t)]\tilde{Z}_{ik}(\beta_0,t)dM_{ik}(t)\\
= \;& -n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)\times\int_0^\tau \mu_k(t)^{-1}A_{ik}(t)\{n^{-1}\sum_{j}\xi_j\tilde{\alpha}^{-1}(1-\Delta_{jk})\tilde{Z}_{jk}(\beta_0,t)dM_{jk}(t)\}\\
= \;& n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)\times\int_0^\tau \mu_k(t)^{-1}A_{ik}(t)\{n^{-1}\sum_{j}\xi_j\tilde{\alpha}^{-1}(1-\Delta_{jk})\tilde{Z}_{jk}(\beta_0,t)Y_{jk}(t)e^{\beta_0^T Z_{jk}(t)}\}d\Lambda_{0k}(t).
\end{aligned}$$

The last equation is obtained by a martingale decomposition of $M_{jk}(t)$ and the fact that $(1-\Delta_{jk})dN_{jk}(t)=0$. Similarly, we have (28) equal to

$$\begin{aligned}
& -n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)\tilde{Z}_{ik}(\beta_0,t)Y_{ik}(t)e^{\beta_0^T Z_{ik}(t)}d\Lambda_{0k}(t)\\
= \;& -n^{-1/2}\sum_{k}\sum_{i}(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)R_{ik}(\beta_0,t)d\Lambda_{0k}(t).
\end{aligned}$$

The quantity $n^{-1}\sum_j \xi_j\tilde{\alpha}^{-1}(1-\Delta_{jk})\tilde{Z}_{jk}(\beta_0,t)Y_{jk}(t)e^{\beta_0^T Z_{jk}(t)}$ converge in probability to $E[(1-\Delta_{1k})R_{1k}(\beta_0,t)]$ uniformly in t, by the special case of Lemma B6.1. We can then combine (26) and (28) and can show that the combined term is asymptotically equivalent to

$$n^{-1/2}\sum_k\sum_i(1-\Delta_{ik})(\xi_i\tilde{\alpha}^{-1}-1)\int_0^\tau\{\mu_k(t)^{-1}A_{ik}(t)E[(1-\Delta_{1k})R_{1k}(\beta_0,t)]-R_{ik}(\beta_0,t)\}d\Lambda_{0k}(t). \tag{30}$$

Repeating the above procedure to combine (27) and (29), their summation is asymptotically equivalent to

$$n^{-1/2}\sum_k\sum_i\Delta_{ik}(1-\xi_i)(\eta_{ik}\tilde{q}_k^{-1}-1)[M_{\bar{z},ik}(\beta_0)-\int_0^\tau\theta_k(t)^{-1}B_{ik}(t)E[\Delta_{1k}dM_{\bar{z},k1}(\beta_0,t)]]. \tag{31}$$

By Lemma B6.1 and B6.2, both (30) and (31) can be shown to converge to a zero mean normal distribution.

By the law of total expectation, the three terms (21), (30) and (31) are pairwise uncorrelated, which implies independence under normality. Specifically, the covariances between (30) and (31), (21) and (31) are both 0 by conditioning on filtration $\mathcal{F}(\tau)$ and ξ. The covariance between (21) and (30) is 0 by conditioning on $\mathcal{F}(\tau)$. Therefore, $n^{-1/2}U_{DW}(\beta_0)$ is asymptotically normally distributed with mean zero and we can compute the contributions of (21), (30) and (31) to the asymptotic variance separately.

Following conditional arguments, the second component (30) has asymptotic variance $\frac{1-\tilde{\alpha}}{\tilde{\alpha}}V^I(\beta_0)$, in which

$$V^I(\beta_0)=var\Big\{\sum_{k=1}^K(1-\Delta_{1k})\int_0^\tau\{R_{1k}(\beta_0,t)-\mu_k^{-1}(t)A_{1k}(t)E[(1-\Delta_{1k})R_{1k}(\beta_0,t)]\}d\Lambda_{0k}(t)\Big\}.$$

Similarly, the asymptotic variance of (31) is $(1-\tilde{\alpha})\sum_{k=1}^K pr(\Delta_{1k}=1)\frac{1-\tilde{q}_k}{\tilde{q}_k}V_k^{II}(\beta_0)$, where

$$V_k^{II}(\beta_0)=var\Big\{M_{\bar{z},1k}(\beta_0)-\int_0^\tau\theta_k(t)^{-1}B_{1k}(t)E[\Delta_{1k}dM_{\bar{z},1k}(\beta_0,t)]|\Delta_{1k}=1,\xi_1=0\Big\}.$$

The desirable asymptotic distribution of $\hat{\beta}_{DW}$ then follows from the Taylor expansion of $U_{DW}(\hat{\beta}_{DW})$ around β_0 and Slutsky's theorem.

The quantities $G(\beta_0)$, $Q(\beta_0)$, $\frac{1-\tilde{\alpha}}{\tilde{\alpha}}V^I(\beta_0)$ and $(1-\tilde{\alpha})\sum_{k=1}^K pr(\Delta_{1k}=1)V_k^{II}(\beta_0)$ can be consistently estimated by $\hat{G}(\hat{\beta}_{DW})$, $\hat{Q}(\hat{\beta}_{DW})$, $\frac{1-\tilde{\alpha}}{\tilde{\alpha}}\hat{V}^I(\hat{\beta}_{DW})$ and $(1-\tilde{\alpha})\sum_k\hat{pr}(\Delta_{1k}=1)\hat{V}_k^{II}(\hat{\beta}_{DW})$, respectively, where

$$\hat{G}(\beta)=-n^{-1}D_{DW}(\beta),\ \ \hat{Q}(\beta)=n^{-1}\sum_{i=1}^n\frac{\xi_i}{\tilde{\alpha}}\Big[\sum_{k=1}^K\hat{M}_{\bar{z},ik}(\beta)\Big]^{\otimes 2},$$

where

$$\begin{aligned}\hat{M}_{\bar{z},ik}(\beta) &= \Delta_{ik}[Z_{ik}(X_{ik})-S_{k,DW}^{(0)}(\beta,X_{ik})^{-1}S_{k,DW}^{(1)}(\beta,X_{ik})]\\ &- n^{-1}\sum_{j=1}^n\frac{\Delta_{jk}Y_{ik}(X_{jk})e^{\beta^T Z_{ik}(X_{jk})}}{\hat{S}_{k,KC}^{(0)}(\beta,X_{jk})}\rho_{jk}(X_{jk})[Z_{ik}(X_{jk})-S_{k,DW}^{(0)}(\beta,X_{jk})^{-1}S_{k,DW}^{(1)}(\beta,X_{jk})],\end{aligned}$$

$$\begin{aligned}\hat{V}^I(\beta) &= n^{-1}\sum_{i=1}^n\frac{\xi_i}{\tilde{\alpha}}\Big[\sum_{k=1}^K(1-\Delta_{ik})\cdot n^{-1}\sum_{j=1}^n\frac{\Delta_{jk}}{\hat{S}_{k,KC}^{(0)}(\beta,X_{jk})}\cdot\rho_{jk}(X_{jk})\\ &\times\ \{\hat{R}_{ik}(\beta,X_{jk})-\hat{\mu}_k(X_{jk})^{-1}A_{ik}(X_{jk})\hat{E}[(1-\Delta_{1k})R_{1k}(\beta,X_{jk})]\}\Big]^{\otimes 2}\\ &- \Big[n^{-1}\sum_{i=1}^n\frac{\xi_i}{\tilde{\alpha}}\sum_{k=1}^K(1-\Delta_{ik})\cdot n^{-1}\sum_{j=1}^n\frac{\Delta_{jk}}{\hat{S}_{k,KC}^{(0)}(\beta,X_{jk})}\cdot\rho_{jk}(X_{jk})\\ &\times\ \{\hat{R}_{ik}(\beta,X_{jk})-\hat{\mu}_k(X_{jk})^{-1}A_{ik}(X_{jk})\hat{E}[(1-\Delta_{1k})R_{1k}(\beta,X_{jk})]\}\Big]^{\otimes 2},\end{aligned}$$

$$
\begin{aligned}
\hat{V}_k^{II}(\beta) &= (n_k-\tilde{n}_k)^{-1}\sum_{i=1}^{n}\frac{\eta_{ik}}{\tilde{q}_k}\Bigg[\hat{M}_{\bar{z},ik}(\beta)-(n_k-\tilde{n}_k)^{-1}\times \\
&\quad \sum_{j=1}^{n}\Delta_{jk}(1-\xi_j)\frac{\eta_{jk}}{\tilde{q}_k}\hat{\theta}_k(X_{jk})^{-1}B_{ik}(X_{jk})(Z_{jk}(X_{jk})-S_{k,DW}^{(0)}(\beta,X_{jk})^{-1}S_{k,DW}^{(1)}(\beta,X_{jk})) \\
&+ n^{-1}\sum_{j=1}^{n}\frac{\Delta_{jk}}{\hat{S}_{k,KC}^{(0)}(\beta,X_{jk})}\rho_{jk}(X_{jk})\hat{\theta}_k(X_{jk})^{-1}B_{ik}(X_{jk})\hat{E}[R_{1k}(\beta,X_{jk})|\Delta_{1k}=1,\xi_i=0]\Bigg]^{\otimes 2} \\
&- \Bigg[(n_k-\tilde{n}_k)^{-1}\sum_{i=1}^{n}\frac{\eta_{ik}}{\tilde{q}_k}\Bigg(\hat{M}_{\bar{z},ik}(\beta)-(n_k-\tilde{n}_k)^{-1}\times \\
&\quad \sum_{j=1}^{n}\Delta_{jk}(1-\xi_j)\frac{\eta_{jk}}{\tilde{q}_k}\hat{\theta}_k(X_{jk})^{-1}B_{ik}(X_{jk})(Z_{jk}(X_{jk})-S_{k,DW}^{(0)}(\beta,X_{jk})^{-1}S_{k,DW}^{(1)}(\beta,X_{jk})) \\
&+ n^{-1}\sum_{j=1}^{n}\frac{\Delta_{jk}}{\hat{S}_{k,KC}^{(0)}(\beta,X_{jk})}\rho_{jk}(X_{jk})\hat{\theta}_k(X_{jk})^{-1}B_{ik}(X_{jk})\hat{E}[R_{1k}(\beta,X_{jk})|\Delta_{1k}=1,\xi_i=0]\Bigg)\Bigg]^{\otimes 2}
\end{aligned}
$$

$$
\hat{R}_{ik}(\beta,t) = \{Z_{ik}(t)-S_{k,DW}^{(0)}(\beta,t)^{-1}S_{k,DW}^{(1)}(\beta,t)\}Y_{ik}(t)e^{\beta^T Z_{ik}(t)},
$$

$$
\hat{\mu}_k(t) = n^{-1}\sum_{i=1}^{n}(1-\Delta_{ik})A_{ik}(t),
$$

$$
\hat{E}[(1-\Delta_{1k})R_{1k}(\beta,t)] = n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}(1-\Delta_{ik})\hat{R}_{ik}(\beta,t),
$$

$$
\hat{\theta}_k(t) = n^{-1}\sum_{i=1}^{n}\Delta_{ik}B_{ik}(t),
$$

$$
\hat{E}[R_{1k}(\beta,t)|\Delta_{1k}=1,\xi_1=0] = (n_k-\tilde{n}_k)^{-1}\sum_{l=1}^{n}\Delta_{lk}(1-\xi_l)\frac{\eta_{lk}}{\tilde{q}_k}\hat{R}_{lk}(\beta,t).
$$

References

[1] R. L. Prentice, "A case-cohort design for epidemiologic cohort studies and disease prevention trials," *Biometrika*, vol. 73, no. 1, pp. 1–11, 1986.

[2] N. Breslow and J. Wellner, "Weighted Likelihood for Semiparametric Models and Two-phase Stratified Samples, with Application to Cox Regression," *Scand. J. Statist.*, vol. 34, no. 1, pp. 86–102, 2007.

[3] S. G. Self and R. L. Prentice, "Asymptotic Distribution Theory and Efficiency Results for Case-Cohort Studies," *The Annals of Statistics*, vol. 16, no. 1, pp. 64–81, 1988.

[4] W. E. Barlow, "Robust variance estimation for the case-cohort design.," *Biometrics*, vol. 50, no. 4, pp. 1064–72, 1994.

[5] K. Chen and S.-H. Lo, "Case-cohort and case-control analysis with Cox's model," *Biometrika*, vol. 86, no. 4, pp. 755–764, 1999.

[6] O. Borgan, B. Langholz, S. O. Samuelsen, L. Goldstein, and J. Pogoda, "Exposure stratified case-cohort designs," *Lifetime Data Analysis*, vol. 6, no. 1, pp. 39–58, 2000.

[7] S. Kang and J. Cai, "Marginal hazards model for case-cohort studies with multiple disease outcomes," *Biometrika*, vol. 96, no. 4, pp. 887–901, 2009.

[8] S. Kim, J. Cai, and W. Lu, "More efficient estimators for case-cohort studies," *Biometrika*, vol. 100, no. 3, p. 695, 2013.

[9] J. Ding, T.-S. Lu, J. Cai, and H. Zhou, "Recent progresses in outcome-dependent sampling with failure time data," *Lifetime data analysis*, vol. 23, no. 1, pp. 57–82, 2017.

[10] C. M. Ballantyne, R. C. Hoogeveen, H. Bang, J. Coresh, A. R. Folsom, G. Heiss, and A. R. Sharrett, "Lipoprotein-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*, vol. 109, no. 7, pp. 837–842, 2004.

[11] C. M. Ballantyne, R. C. Hoogeveen, H. Bang, J. Coresh, A. R. Folsom, L. E. Chambless, M. Myerson, K. K. Wu, A. R. Sharrett, and E. Boerwinkle, "Lipoprotein-associated phospholipase A2, high-sensitivity C-reactive protein, and risk for incident ischemic stroke in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study.," *Arch Intern Med*, vol. 165, pp. 2479–2484, 2005.

[12] M. Kulich and D. Y. Lin, "Improving the Efficiency of Relative-Risk Estimation in Case-Cohort Studies," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 832–844, 2004.

[13] N. E. Breslow, T. Lumley, C. M. Ballantyne, L. E. Chambless, and M. Kulich, "Improved Horvitz-Thompson Estimation of Model Parameters from Two-phase Stratified Samples: Applications in Epidemiology," *Statistics in Biosciences*, vol. 1, no. 1, pp. 32–49, 2009.

[14] N. E. Breslow, T. Lumley, C. M. Ballantyne, L. E. Chambless, and M. Kulich, "Using the whole cohort in the analysis of case-cohort data," *American Journal of Epidemiology*, vol. 169, no. 11, pp. 1398–1405, 2009.

[15] J. Cai and D. Zeng, "Power calculation for case-cohort studies with nonrare events," *Biometrics*, vol. 63, no. 4, pp. 1288–1295, 2007.

[16] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2002.

[17] J. Cai and R. L. Prentice, "Estimating equations for hazard ratio parameters based on correlated failure time data," *Biometrika*, vol. 82, no. 1, pp. 151–164, 1995.

[18] C. F. Spiekerman and D. Y. Lin, "Marginal Regression Models for Multivariate Failure Time Data," *Journal of the American Statistical Association*, vol. 93, no. 443, p. 1164, 1998.

[19] D. Clayton and J. Cuzick, "Multivariate Generalizations of the Proportional Hazards Model," *Journal of the Royal Statistical Society. Series A*, vol. 148, no. 2, pp. 82–117, 1985.

[20] W. Hu, J. Cai, and D. Zeng, "Sample size/power calculation for stratified case–cohort design," *Statistics in medicine*, vol. 33, no. 23, pp. 3973–3985, 2014.

[21] O. Saarela, S. Kulathinal, E. Arjas, and E. Läärä, "Nested case-control data utilized for multiple outcomes: a likelihood approach and alternatives," *Statistics in medicine*, vol. 27, no. 28, pp. 5991–6008, 2008.

[22] N. C. Støer and S. O. Samuelsen, "Comparison of estimators in nested case-control studies with multiple outcomes," *Lifetime data analysis*, vol. 18, no. 3, pp. 261–283, 2012.

[23] Y. Yan, H. Zhou, and J. Cai, "Improving efficiency of parameter estimation in case-cohort studies with multivariate failure time data," *Biometrics*, vol. 73, no. 3, pp. 1042–1052, 2017.

[24] L. Qi, C. Y. Wang, and R. L. Prentice, "Weighted Estimators for Proportional Hazards Regression With Missing Covariates," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1250–1263, 2005.

[25] S. Kang, J. Cai, and L. Chambless, "Marginal additive hazards model for case-cohort studies with multiple disease outcomes: an application to the Atherosclerosis Risk in Communities (ARIC) study," *Biostatistics*, vol. 14, no. 1, pp. 28–41, 2012.

[26] J. Hajek, "Limiting distributions in simple random sampling from a finite population," *Publications of the Mathematics Institute of the Hungarian Academy of Science*, vol. 5, no. 361, p. 74, 1960.

[27] A. W. Van der vaart and J. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics, Springer New York, 2012.

[28] D. Y. Lin, "On fitting Cox's proportional hazards models to survey data," *Biometrika*, vol. 87, no. 1, pp. 37–47, 2000.

[29] R. V. Foutz, "On the Unique Consistent Solution to the Likelihood Equations," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 147–148, 1977.