JOURNAL OF APPLIED STATISTICS
2021, VOL. 48, NO. 5, 846–865
https://doi.org/10.1080/02664763.2020.1752633

Taylor & Francis
Taylor & Francis Group

Regression analysis of case-cohort studies in the presence of dependent interval censoring

Mingyue Du[a], Qingning Zhou[b], Shishun Zhao[a] and Jianguo Sun[c]

[a]Center for Applied Statistical Research and College of Mathematics, Jilin University, Changchun, People's Republic of China; [b]Department of Mathematics and Statistics, The University of North Carolina at Charlotte, Charlotte, NC, USA; [c]Department of Statistics, University of Missouri, Columbia, MO, USA

ABSTRACT

The case-cohort design is widely used as a means of reducing the cost in large cohort studies, especially when the disease rate is low and covariate measurements may be expensive, and has been discussed by many authors. In this paper, we discuss regression analysis of case-cohort studies that produce interval-censored failure time with dependent censoring, a situation for which there does not seem to exist an established approach. For inference, a sieve inverse probability weighting estimation procedure is developed with the use of Bernstein polynomials to approximate the unknown baseline cumulative hazard functions. The proposed estimators are shown to be consistent and the asymptotic normality of the resulting regression parameter estimators is established. A simulation study is conducted to assess the finite sample properties of the proposed approach and indicates that it works well in practical situations. The proposed method is applied to an HIV/AIDS case-cohort study that motivated this investigation.

ARTICLE HISTORY
Received 6 May 2019
Accepted 1 April 2020

KEYWORDS
Case-cohort design; dependent interval censoring; inverse probability weighting; proportional hazards model

1. Introduction

The case-cohort design is widely used as a means of reducing the cost in large cohort studies, especially when the disease rate is low and covariate measurements may be expensive (Prentice [27]; Scheike and Martinussen [30], Self and Prentice [32]). For the situation, instead of collecting the covariate information on all study subjects, it collects the covariate information only on the subjects whose failures are observed and on a subsample of the remaining subjects. Among others, one area where the design is often used is epidemiological cohort studies in which the outcomes of interest are times to failure events such as AIDS, cancer, heart disease and HIV infection. For such studies, in addition to the incomplete nature on covariate information, another feature is that the observations are usually interval-censored rather than right-censored due to the periodic follow-up nature of the study (Sun [34]).

By interval-censored data, we usually mean that the failure time of interest is known or observed only to belong to an interval instead of being observed exactly. It is easy to see that

CONTACT Jianguo Sun sunj@missouri.edu

interval-censored data include right-censored data as a special case. Furthermore, sometimes one may also face informative censoring, meaning that the failure time of interest and the censoring mechanism are correlated (Huang and Wolfe [13]; Wang et al. [37]). An example of informatively interval-censored data may arise in a periodic follow-up study of certain disease where study subjects may not follow the pre-specified visit schedules and instead pay clinical visits according to their disease status or how they feel with respect their treatments. Among others, Huang and Wolfe [13] and Sun [33] discussed the issue and pointed out that in the presence of informative censoring, the analysis that ignores it may result in biased or misleading results or conclusions. More discussion on informatively interval-censored data can be found in Sun [34].

One real study that motivated this investigation is the HVTN 505 Trial to assess the efficacy of a DNA prime-recombinant adenovirus type 5 boost (DNA/rAd5) vaccine to prevent human immunodeficiency virus type 1 (HIV-1) infection (Fong et al. [8]; Hammer et al. [10]; Janes et al. [14]). It is well-known that HIV-1 infection is deadly as it causes AIDS for which there is no cure and thus it is important and essential to develop a safe and effective vaccine for the prevention of the infection. The original study consists of 2504 men or transgender women who had sex with men were examined periodically, thus yielding only interval-censored data on the time to HIV-1 infection. For each subject, the information on four demographic covariates, age, race, BMI and behavioural risk, was collected, and in addition, for a subgroup of HIV infection cases and non-cases, a number of T cell response biomarkers and anti-body response biomarkers were also measured. One goal of the study is to determine or identify the important or relevant covariates or biomarkers for HIV-1 infection.

Many authors have discussed the analysis of case-cohort studies but most of the existing methods are for right-censored failure time data. For example, some of the early work on this was given by Prentice [27] and Self and Prentice [32], who proposed some pseudolikelihood approaches based on the modification of the commonly used partial likelihood method under the proportional hazards model. By following them, Chen and Lo [3] proposed an estimating equation approach that yields more efficient estimators than the pseudolikelihood estimator proposed in Prentice [27], and Chen [2] developed an estimating equation approach that applies to a class of cohort sampling designs, including the case-cohort design with the key estimating function constructed by a sample reuse method via local averaging. Also Marti and Chavance [25] and Keogh and White [18] proposed some multiple imputation methods and in particular, the latter method extended the former by considering more complex imputation models that include time and interaction or nonlinear terms. In addition, Kang and Cai [17] and Kim et al. [19] developed weighted estimating equation approaches for case-cohort studies with multiple disease outcomes, where the latter method improved the efficiency upon the former by utilizing more information in constructing the weights.

Interval-censored failure time data naturally occur in many areas, especially in the studies with periodic follow-ups, and a great deal of literature has been developed for their analysis (Chen et al. [5]; Finkelstein [7]; Sun [34]; Zhou et al. [40]). In particular, Sun [34] and Bogaerts et al. [1] provided comprehensive reviews of the existing literature on interval-censored data. Although there also exist some methods for either informatively interval-censored data or the interval-censored arising from case-cohort studies, there does not seem to exist an established procedure for informatively interval-censored data

arising from case-cohort studies. In particular, for the analysis of informatively interval-censored data, two types of approaches are commonly used and they are the frailty model approach and the copula model approach. For example, Zhang et al. (2005, 2007) and Wang et al. [36,38] gave some frailty model estimation procedures, while Ma et al. [23,24] and Zhao et al. (2015) proposed some copula model methods. For the analysis of the interval-censored data arising from case-cohort studies, Gilbert et al. [9] presented a midpoint imputation procedure and Li and Nan [20] considered a special case of interval-censored data, current status data, where the failure time of interest is either left- or right-censored (Jewell and van der Laan [15]). Also Zhou et al. [41] proposed a likelihood-based approach. However, all of the three methods above assume that the interval censoring mechanism is non-informative or independent of the failure time of interest. As discussed by many authors and above, the informative censoring is a serious and difficult issue and the use of the methods that do not take it into account can yield biased or misleading results and conclusions (Huang and Wolfe [13]; Ma et al. [23]). In the following, we will develop a frailty model approach, a generalization of the method proposed in Zhou et al. [41], for the analysis of the case-cohort studies yielding interval-censored data with informative censoring.

The remainder of the paper is organized as follows. We will begin in Section 2 with introducing some notation and models to be used throughout the paper and in particular, we will present joint frailty models for the failure time of interest and the underlying censoring mechanism. To estimate regression parameters, a sieve inverse probability weighting estimation procedure is then presented in Section 3 and in the method, Bernstein polynomials are employed to approximate unknown functions. Furthermore, we establish the consistency and asymptotic normality of the resulting estimators of regression parameters and provide a weighted bootstrap procedure for variance estimation. Section 4 presents some results obtained from an extensive simulation study conducted to assess the finite sample properties of the proposed methodology and they suggest that the method works well in practical situations. In Section 5, we apply the proposed method to the HIV/AIDS study described above and Section 6 gives some discussion and concluding remarks.

2. Notation and models

Consider a failure time study that consists of n independent subjects. For subject i, let T_i denote the failure time of interest and suppose that there exists a p-dimensional vector of covariates denoted by Z_i that may affect T_i, $i = 1, \ldots, n$. Also for subject i, suppose that there exist two examination times denoted by U_i and V_i with $U_i \leq V_i$ and one only observes $\Delta_{1i} = I(T_i \leq U_i)$ and $\Delta_{2i} = I(U_i < T_i \leq V_i)$, indicating if the failure time T_i is left-censored and interval-censored, respectively. Note that here U_i and V_i are random variables and assumed to be observed and they together with Δ_{1i} and Δ_{2i} give the observed interval-censored data on the T_i's (Sun [34]; Zhou et al. [41]).

For the case-cohort studies, as mentioned above, the information on covariates is available only for the subjects who either have experienced the failure event of interest or with $\Delta_{1i} = 1$ or $\Delta_{2i} = 1$ or are from the sub-cohort that is a random sample of the entire cohort. Define $\xi_i = 1$ if the covariate Z_i is available or observed and 0 otherwise, $i = 1, \ldots, n$. For the selection of the subcohort, by following Zhou et al. [40] and others, we will consider the independent Bernoulli sampling with the selection probability $q \in (0, 1)$. Then under

the assumption above, the probability that the covariate Z_i is observed is given by

$$Pr(\xi_i = 1) = \pi_q(\Delta_{1i}, \Delta_{2i}) = \Delta_{1i} + \Delta_{2i} + (1 - \Delta_{1i} - \Delta_{2i})q,$$

$i = 1, \ldots, n$, and the observed data have the form

$$O^{\xi} = \left\{ O_i^{\xi} = (U_i, V_i, \Delta_{1i}, \Delta_{2i}, \xi_i, \xi_i Z_i);\ i = 1, \ldots, n \right\}.$$

In contrast, if all covariates were observed, the full cohort data would be

$$O' = \{O_i = (U_i, V_i, \Delta_{1i}, \Delta_{2i}, Z_i);\ i = 1, \ldots, n\}.$$

To describe the covariate effects and dependent interval censoring, define $W_i = V_i - U_i$, $i = 1, \ldots, n$. By following Ma et al. [23], we will focus on the situation where the dependent censoring can be characterized by the correlation between the T_i's and W_i's. As mentioned in Ma et al. [23], one example where this may be the case is follow-up studies where some study subjects may tend to pay more or less clinical visits than the scheduled ones. More comments on this will be given below. For the covariate effects, we assume that there exists a latent variable b_i with mean one and known distribution but unknown variance η and given Z_i and b_i, the hazard functions of T_i and W_i have the forms

$$\lambda_i^{(T)}(t|Z_i, b_i) = \lambda_t(t) \exp(\beta_t' Z_i) b_i, \tag{1}$$

and

$$\lambda_i^{(W)}(t|Z_i, b_i) = \lambda_w(t) \exp(\beta_w' Z_i) b_i\,, \tag{2}$$

respectively. In the above, $\lambda_t(t)$ and $\lambda_w(t)$ are unknown baseline hazard functions and β_t and β_w are $p \times 1$ vectors of unknown regression parameters. Also it will be assumed that given Z_i and b_i, W_i is independent of U_i and T_i and W_i are independent. In other words, the correlation between T_i and W_i is measured by the parameter η. More comments on this are given below.

Define $\Delta_i = (\Delta_{1i}, \Delta_{2i})$ and $\theta = (\beta_t, \beta_w, \Lambda_t, \Lambda_w, \eta)$, where $\Lambda_t(t) = \int_0^t \lambda_t(u)du$ and $\Lambda_w(t) = \int_0^t \lambda_w(u)du$. Assume that b_i is independent of (U_i, Z_i) and the joint distribution of (U_i, Z_i) does not involve the parameters of interest. To motivate the proposed estimation procedure, note that conditional on (W_i, U_i, Z_i, b_i), the likelihood of the observation from subject i has the form

$$\begin{aligned} L_{\Delta_i|W_i,U_i,b_i}(\theta) = &\left[1 - \exp\{-\Lambda_t(U_i)\exp(\beta_t' Z_i) b_i\}\right]^{\Delta_{1i}} \left[\exp\{-\Lambda_t(U_i)\exp(\beta_t' Z_i) b_i\}\right. \\ &\left. - \exp\{-\Lambda_t(V_i)\exp(\beta_t' Z_i) b_i\}\right]^{\Delta_{2i}} \left[\exp\{-\Lambda_t(V_i)\exp(\beta_t' Z_i) b_i\}\right]^{1-\Delta_{1i}-\Delta_{2i}} \end{aligned}$$

Also note that conditional on (Z_i, b_i), the likelihood of the observation on W_i is given by

$$L_{W_i|b_i} = \left\{\lambda_w(W_i)\exp\{\beta_w' Z_i\} b_i \exp\{-\Lambda_w(W_i)\exp(\beta_w' Z_i) b_i\}\right\}^{\Psi_i}.$$

where $\Psi_i = I(W_i < \infty)$. This motivates the following inverse probability weighted log-likelihood function

$$l_{O^\xi}(\theta) = \sum_{i=1}^{n} l_i(\theta; O_i^\xi) = \sum_{i=1}^{n} p_i l_i(\theta; O_i)$$
$$= \sum_{i=1}^{n} p_i \log \left\{ \int L_{\Delta_i|W_i,U_i,b_i}(\theta) L_{W_i|b_i}(\theta) f(b_i;\eta)\, \mathrm{d}b_i \right\} \tag{3}$$

for estimation of θ, where $f(b_i;\eta)$ denotes the the density function of the b_i's and

$$p_i = \frac{\xi_i}{\pi_q(\Delta_{1i}, \Delta_{2i})} = \frac{\xi_i}{\Delta_{1i} + \Delta_{2i} + (1 - \Delta_{1i} - \Delta_{2i})q}.$$

If f is the gamma distribution, the function $l_{O^\xi}(\theta)$ has a closed form as

$$\begin{aligned} l_{O^\xi}(\theta) = \sum_{i=1}^{n} p_i \log \Big\{ & (\lambda_w \exp(\beta_w' Z_i))^{\Psi_i} \Big[(1 + (\eta \Lambda_w(W_i) \exp(\beta_w' Z_i))\Psi_i)^{-\eta^{-1}-\Psi_i} \\ & - (1 + \eta \Lambda_t(U_i) \exp(\beta_t' Z_i) + (\eta \Lambda_w(W_i) \exp(\beta_w' Z_i))\Psi_i)^{-\eta^{-1}-\Psi_i} \Big]^{\Delta_{1i}} \\ & \times \Big[(1 + \eta \Lambda_t(U_i) \exp(\beta_t' Z_i) + (\eta \Lambda_w(W_i) \exp(\beta_w' Z_i))\Psi_i)^{-\eta^{-1}-\Psi_i} \\ & - (1 + \eta \Lambda_t(U_i + W_i) \exp(\beta_t' Z_i) + (\eta \Lambda_w(W_i) \exp(\beta_w' Z_i))\Psi_i)^{-\eta^{-1}-\Psi_i} \Big]^{\Delta_{2i}} \\ & \times \Big[(1 + \eta \Lambda_t(U_i + W_i) \exp(\beta_t' Z_i) + (\eta \Lambda_w(W_i) \exp(\beta_w' Z_i))\Psi_i)^{-\eta^{-1}-\Psi_i} \Big]^{1-\Delta_{1i}-\Delta_{2i}} \Big\}. \end{aligned} \tag{4}$$

In the next section, for estimation of θ, we will discuss the maximization of the inverse probability weighted log-likelihood function $l_{O^\xi}(\theta)$.

3. Sieve inverse probability weighting estimation

Define the parameter space of θ

$$\Theta = \{\theta = (\beta_t, \beta_w, \eta, \psi) : \psi = (\Lambda_t(t), \Lambda_w(t))\} = \mathbf{B} \otimes M^1 \otimes M^2,$$

where $\mathbf{B} = \{(\beta_t, \beta_w, \eta) \in R^{2p} \times R^+, \| \beta_t \| + \| \beta_w \| + \| \eta \| \leq M\}$ with M being a positive constant and M^j denotes the collection of all bounded and continuous nondecreasing, nonnegative functions over the interval $[\sigma_j, \tau_j], j = 1, 2$. In practice, $[\sigma_1, \tau_1]$ is usually taken to be the range of the U_i's and V_i's and $[\sigma_2, \tau_2]$ the range of the W_i's. More comments on this are given below. For the maximization of the inverse probability weighted log-likelihood function $l_{O^\xi}(\theta)$, it is easy to see that this would not be straightforward since $l_{O^\xi}(\theta)$ involves unknown functions $\Lambda_t(t)$ and $\Lambda_w(t)$. To deal with this and by following Ma et al. [24], Zhou et al. [40] and others, we propose first to approximate the two functions by Bernstein polynomials.

More specifically, define the sieve space

$$\Theta_n = \{\theta_n = (\beta_t, \beta_w, \eta, \psi_n) : \psi_n = (\Lambda_{tn}(t),\ \Lambda_{wn}(t))\} = \mathbf{B} \otimes M_n^1 \otimes M_n^2.$$

with

$$M_n^1 = \left\{ \Lambda_{tn} : \Lambda_{tn}(t) = \sum_{k=0}^{m} \phi_{k1} B_k(t, m, \sigma_1, \tau_1), \right.$$
$$\left. \phi_{m1} \geq \cdots \geq \phi_{11} \geq \phi_{01} \geq 0,\ \sum_{k=0}^{m} |\phi_{k1}| \leq M_n \right\},$$

and

$$M_n^2 = \left\{ \Lambda_{wn} : \Lambda_{wn}(t) = \sum_{k=0}^{m} \phi_{k2} B_k(w, m, \sigma_2, \tau_2), \right.$$
$$\left. \phi_{m2} \geq \cdots \geq \phi_{12} \geq \phi_{02} \geq 0,\ \sum_{k=0}^{m} |\phi_{k2}| \leq M_n \right\}.$$

In the above,

$$B_k(t, m, \sigma_1, \tau_1) = C_m^k \left(\frac{t - \sigma_1}{\tau_1 - \sigma_1} \right)^k \left(1 - \frac{t - \sigma_1}{\tau_1 - \sigma_1} \right)^{m-k},$$

and

$$B_k(w, m, \sigma_2, \tau_2) = C_m^k \left(\frac{w - \sigma_2}{\tau_2 - \sigma_2} \right)^k \left(1 - \frac{w - \sigma_2}{\tau_2 - \sigma_2} \right)^{m-k},$$

$k = 0, \ldots, m$, which Bernstein polynomials of degree $m = o(n^\nu)$ for some $\nu \in (0, 1)$. Note that some restrictions are needed above on the parameters since $\Lambda_t(t)$ and $\Lambda_w(w)$ are nonnegative and nondecreasing functions. However, this can be easily removed by some reparameterization. For example, one can reparameterize the parameters $\{\phi_{0j}, \ldots, \phi_{mj}\}$ by the cumulative sums of the parameters $\{\exp(\phi_{0j}^*), \ldots, \exp(\phi_{mj}^*)\}, j = 1, 2$.

Let $\hat{\theta}_n = (\hat{\beta}_{tn}, \hat{\beta}_{wn}, \hat{\eta}_n, \hat{\Lambda}_{tn}, \hat{\Lambda}_{wn})$ denote the estimator of θ given by the value of θ that maximizes the inverse probability weighted log-likelihood function $l_{O^\xi}(\theta)$ over the sieve space Θ_n. Also let $\theta_0 = (\beta_{t0}, \beta_{w0}, \eta_0, \Lambda_{t0}, \Lambda_{w0})$ denote the true value of θ, $\hat{\vartheta}_n = (\hat{\beta}_{tn}, \hat{\beta}_{wn}, \hat{\eta}_n)$, $\vartheta_0 = (\beta_{t0}, \beta_{w0}, \eta_0)$, and for any $\theta^1 = (\beta_t^1, \beta_w^1, \eta^1, \Lambda_t^1, \Lambda_w^1)$ and $\theta^2 = (\beta_t^2, \beta_w^2, \eta^2, \Lambda_t^2, \Lambda_w^2)$ in the parameter space Θ, define the distance

$$d(\theta^1, \theta^2) = \left\{ \|\beta_t^1 - \beta_t^2\|^2 + \|\beta_w^1 - \beta_w^2\|^2 + \|\eta^1 - \eta^2\|^2 \right.$$
$$\left. + \|\Lambda_t^1 - \Lambda_t^2\|_2^2 + \|\Lambda_w^1 - \Lambda_w^2\|_2^2 \right\}^{1/2}.$$

Here $\|v\|$ denotes the Euclidean norm for a vector v, $\|\Lambda_t^1 - \Lambda_t^2\|_2^2 = \int [(\Lambda_t^1(u) - \Lambda_t^2(u))^2 + \psi(\Lambda_t^1(u + w) - \Lambda_t^2(u + w))^2] dG(u, w)$, and $\|\Lambda_w^1 - \Lambda_w^2\|_2^2 = \int \psi [\Lambda_w^1(w) - \Lambda_w^2(w)]^2 dG(u, w)$ with $G(u, w)$ denoting the joint distribution function of U and W. The following two theorems establish the asymptotic properties of $\hat{\theta}_n$.

Theorem 3.1: *Suppose that the regularity conditions* (C1)–(C4) *given in the Appendix hold. Then as* $n \to \infty$, *we have that* $d(\hat{\theta}_n, \theta_0) \to 0$ *almost surely and* $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\min\{(1-\nu)/2, \nu r/2\}})$, *where* $\nu \in (0,1)$ *is defined in* $m = o(n^{\nu})$ *and* r *in the regularity condition* (C3).

Theorem 3.2: *Suppose that the regularity conditions* (C1)–(C5) *given in the Appendix hold. Then as* $n \to \infty$ *and if* $\nu > 1/2r$, *we have that*

$$n^{1/2}(\hat{\vartheta}_n - \vartheta_0) = I^{-1}(\vartheta_0)\, n^{-1/2} \sum_{i=1}^{n} p_i l^*(\vartheta_0, O_i) + o_p(1) \to N(0, \Sigma)$$

in distribution, where

$$\Sigma = I^{-1}(\vartheta_0) + I^{-1}(\vartheta_0) E\left\{ \frac{1 - \pi_q(\Delta_1, \Delta_2)}{\pi_q(\Delta_1, \Delta_2)} \{l^*(\vartheta_0, O)\}^{\otimes 2} \right\} I^{-1}(\vartheta_0)$$

with $v^{\otimes 2} = vv'$ *for a vector* v *and* $I(\vartheta)$ *and* $l^*(\vartheta, O)$, *given in the Appendix, denoting the information matrix and efficient score for* $\vartheta = (\beta_t, \beta_w, \eta)$ *based on the complete data.*

The proof of the results given above is sketched in the Appendix. For the determination of the proposed estimator $\hat{\theta}_n$, different methods can be used and in the numerical studies below, the Matlab function fmincon is used. Also for the determination of $\hat{\theta}_n$, one needs to choose or specify the degree m of Bernstein polynomials, which controls the smoothness of the approximation. For this, one common approach is to perform the grid search by considering different values of m and choosing the one that minimizes

$$AIC = -2\, l_{O^{\xi}}(\hat{\theta}_n) + 2\,(2p + 2m + 3)$$

based on the AIC criterion. Note that instead of this, one may employ other criteria such as the BIC criterion and the numerical results indicate that they give similar performance. Also note that in the approximation of Λ_t and Λ_w, we used the same degree m and in practice, different m could be used too.

For inference about $\vartheta_0 = (\beta_{t0}, \beta_{w0}, \eta_0)$, of course, one needs to estimate the covariance matrix of $\hat{\vartheta}_n = (\hat{\beta}_{tn}, \hat{\beta}_{wn}, \hat{\eta}_n)$. For this, a natural way would be to derive a consistent estimator of Σ. On the other hand, one could see from the Appendix that Σ involves the information matrix $I(\vartheta_0)$ and the efficient score $l^*(\vartheta_0, O)$ and both of them do not have closed forms. Thus, it would be difficult to derive a consistent estimator and instead we propose to employ the weighted bootstraps procedure discussed in Ma and Kosorok [22], which is easy to implement and seems to work well in the numerical studies described below. Specifically, let $\{u_1, \ldots, u_n\}$ denote n independent realizations of a bounded positive random variable u satisfying $E(u) = 1$ and $var(u) = \epsilon_0 < \infty$ and define the new weights $p_i' = u_i\, p_i$, $i = 1, \ldots, n$. Also let $\hat{\vartheta}_n'$ denote the estimator of ϑ proposed above with replacing the p_i's by the p_i''s. Then if we repeat this B times, one can estimate the covariance matrix of $\hat{\vartheta}_n$ by the sample covariance matrix of the $\hat{\vartheta}_n'$'s. By following Ma and Kosorok [22], it can be shown that this weighted bootstrap variance estimator is consistent.

4. A simulation study

In this section, we report some results obtained from a simulation study conducted to evaluate the finite sample performance of the inverse probability weighted estimation procedure proposed in the previous sections. In the study, it was assumed that the covariate Z followed the Bernoulli distribution with the success probability of 0.5 and to generate the subcohort, as mentioned above, we considered the independent Bernoulli sampling with the selection probability being 0.1. For the proportion of the observed failure events or the event rate, we studied several cases including $p_e = 0.05$, 0.1 and 0.2. To generate interval-censored data, we first generated the U_i's from the uniform distribution over $(0, a)$ with a being a positive constant and the latent variable b_i's. Then the T_i's and W_i's were generated based on models (2.1) and (2.2) with $\lambda_t = 0.2t$, $0.1t$ or $4t/9$, $\lambda_w = 12t$ and the V_i's were defined as $V_i = U_i + W_i$ for all i. The results given below are based on the full cohort size $n = 1000$ or 2000 with 1000 replications.

Table 1 presents the results obtained on the proposed estimators $\hat{\beta}_{tn}$, $\hat{\beta}_{wn}$ and $\hat{\eta}_n$ with $n = 1000$, the true values of the parameters being $\beta_{t0} = \beta_{w0} = 0$, 0.2 or 0.5 and $\eta_0 = 0.8$, and the b_i's following the gamma distribution. The results include the estimated bias (Bias) given by the average of the proposed estimates minus the true value, the sample standard error (SSE), the average of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). Here we took the degree of Bernstein polynomials being $m = 3$ and the weighted bootstrap sample size $B = 100$ for variance estimation. Also for the variance estimation, we generated the random sample $\{u_1, \ldots, u_n\}$ repeatedly from the exponential distribution. Table 2 gives the estimation results obtained under the same set-up as above except $n = 2000$. One can see from the two tables that the results indicate that the proposed estimator seems to be unbiased and the weighted bootstrap variance estimation procedure seems to work well. Also they indicate that the normal approximation to the distribution of the proposed estimator appears to be reasonable. In addition, as expected, the estimation results became better when the percentage of the observed failure events or the full cohort size increased. We also considered other set-ups including different values for m and B and obtained similar results.

In the proposed estimation procedure, it has been assumed that the distribution of the latent variables b_i's is known up to a variance parameter. Hence in practice, one question of interest may be the robustness of the estimation procedure with respect to the distribution. To investigate this, we repeated the simulation study above giving the results in Table 1 with $p_e = 0.1$ except that we generated the b_i's from the log-normal distribution instead of the gamma distribution but assumed that they followed the gamma distribution. Table 3 presents the results obtained on the proposed estimators $\hat{\beta}_{tn}$ and $\hat{\beta}_{wn}$, including the Bias, the SSE, the ESE and the 95% empirical CP. As before, they suggest that the proposed methodology seems to work well or the estimators $\hat{\beta}_{tn}$ and $\hat{\beta}_{wn}$ appear to be robust with respect to the distribution of the latent variables.

For the problem discussed here, instead of the inverse probability weighting method proposed above, there exist two commonly used naive approaches that estimate regression parameters by using the regular likelihood approaches. One is to base the estimation only on the selected sub-cohort and the other is to base the estimation on a simple random sample that has the same size as the case-cohort sample. Let $\hat{\beta}_{t_{sub}}$ and $\hat{\beta}_{t_{srs}}$ denote the estimators of β_t given by the two naive methods above, respectively, and here we only

Table 1. Estimation of regression parameters with $n = 1000$.

p_e	Parameter	Bias	SSE	ESE	CP
5%	$\beta_t = 0$	−0.0107	0.3920	0.3903	0.9510
	$\beta_w = 0$	−0.0026	0.3180	0.3247	0.9460
	$\eta = 0.8$	−0.0034	0.2778	0.2863	0.9330
	$\beta_t = 0.2$	0.0074	0.4045	0.3958	0.9450
	$\beta_w = 0.2$	0.0163	0.3259	0.3273	0.9480
	$\eta = 0.8$	0.0180	0.2722	0.2852	0.9520
	$\beta_t = 0.5$	0.0272	0.4070	0.4032	0.9490
	$\beta_w = 0.5$	0.0203	0.3469	0.3330	0.9380
	$\eta = 0.8$	0.0200	0.2789	0.2813	0.9470
10%	$\beta_t = 0$	−0.0163	0.3428	0.3377	0.9330
	$\beta_w = 0$	−0.0047	0.2976	0.3106	0.9540
	$\eta = 0.8$	0.0347	0.2475	0.2544	0.9420
	$\beta_t = 0.2$	−0.0003	0.3413	0.3394	0.9530
	$\beta_w = 0.2$	0.0158	0.3127	0.3096	0.9400
	$\eta = 0.8$	0.0347	0.2509	0.2493	0.9410
	$\beta_t = 0.5$	0.0117	0.3447	0.3438	0.9470
	$\beta_w = 0.5$	0.0291	0.3146	0.3130	0.9410
	$\eta = 0.8$	0.0386	0.2500	0.2440	0.9370
20%	$\beta_t = 0$	−0.0067	0.3058	0.3022	0.9480
	$\beta_w = 0$	0.0074	0.2766	0.2740	0.9400
	$\eta = 0.8$	0.0071	0.2202	0.2159	0.9240
	$\beta_t = 0.2$	−0.0022	0.3066	0.3027	0.9410
	$\beta_w = 0.2$	0.0134	0.2757	0.2756	0.9480
	$\eta = 0.8$	0.0132	0.2178	0.2150	0.9340
	$\beta_t = 0.5$	0.0021	0.3089	0.3058	0.9410
	$\beta_w = 0.5$	0.0223	0.2786	0.2791	0.9440
	$\eta = 0.8$	0.0165	0.2220	0.2155	0.9230

focus on the estimation of β_t. Table 4 gives the estimation results given by the proposed method and the two naive approaches under the set-up similar to that for Table 2 with $p_e = 0.1$. Note that here for comparison, we also considered the approach given by Zhou et al. [40], which treated the observation process to be independent of the failure time of interest or ignored the correlation between the failure time and the observation process. The resulting estimator of β_t is denoted by $\hat{\beta}_{t_{in}}$ in the table. One can see from Table 4 that the proposed estimate clearly gave better performance than the two naive estimates and one would get biased results if ignoring the correlation between the failure time of interest and the observation process.

As pointed out by a reviewer and motivated by the real data discussed below, we also repeated the study that gave the results in Table 1 with $p_e = 0.05$ in which we generated the subcohort in the same way as before but only from none-case subjects instead of all subjects as above. In other words, the goal here is to assess the performance of the proposed approach for case–control studies. The obtained estimation results are presented in Table 5 and one can see that they are similar to those given in Table 1. In other words, it seems that the proposed estimation approach seems to give good performance for and can be applied to case–control studies too.

5. An application

In this section, we will apply the methodology proposed in the previous sections to the HVTN 505 Trial discussed above. It is a randomized, multiple-sites clinical trial of men or

Table 2. Estimation of regression parameters with $n = 2000$.

p_e	Parameter	Bias	SSE	ESE	CP
5%	$\beta_t = 0$	−0.0106	0.2718	0.2707	0.9480
	$\beta_w = 0$	0.0007	0.2255	0.2270	0.9470
	$\eta = 0.8$	0.0117	0.1945	0.2094	0.9620
	$\beta_t = 0.2$	0.0080	0.2772	0.2751	0.9480
	$\beta_w = 0.2$	0.0112	0.2299	0.2289	0.9510
	$\eta = 0.8$	0.0089	0.1872	0.2085	0.9730
	$\beta_t = 0.5$	0.0170	0.2813	0.2781	0.9440
	$\beta_w = 0.5$	0.0052	0.2299	0.2313	0.9510
	$\eta = 0.8$	0.0205	0.1877	0.2015	0.9660
10%	$\beta_t = 0$	−0.0046	0.2345	0.2344	0.9440
	$\beta_w = 0$	0.0081	0.2097	0.2148	0.9500
	$\eta = 0.8$	0.0346	0.1722	0.1847	0.9660
	$\beta_t = 0.2$	0.0058	0.2404	0.2371	0.9400
	$\beta_w = 0.2$	0.0060	0.2130	0.2172	0.9580
	$\eta = 0.8$	0.0371	0.1827	0.1830	0.9410
	$\beta_t = 0.5$	0.0083	0.2371	0.2407	0.9540
	$\beta_w = 0.5$	0.0151	0.2208	0.2193	0.9450
	$\eta = 0.8$	0.0372	0.1705	0.1830	0.9570
20%	$\beta_t = 0$	−0.0023	0.2083	0.2106	0.9500
	$\beta_w = 0$	−0.0096	0.1897	0.1928	0.9550
	$\eta = 0.8$	0.0060	0.1609	0.1671	0.9520
	$\beta_t = 0.2$	0.0011	0.2114	0.2122	0.9600
	$\beta_w = 0.2$	0.0099	0.1918	0.1938	0.9540
	$\eta = 0.8$	0.0075	0.1575	0.1645	0.9560
	$\beta_t = 0.5$	−0.0034	0.2197	0.2137	0.9370
	$\beta_w = 0.5$	0.0040	0.1953	0.1958	0.9430
	$\eta = 0.8$	0.0036	0.1597	0.1628	0.9430

Table 3. Estimation of regression parameters with $n = 1000$, $p_e = 0.1$ and misspecified frailty distribution.

Parameter	Bias	SSE	ESE	CP
$\beta_t = 0$	0.0040	0.3192	0.3221	0.9500
$\beta_w = 0$	−0.0017	0.2682	0.2600	0.9470
$\beta_t = 0.2$	0.0025	0.3218	0.3238	0.9500
$\beta_w = 0.2$	0.0014	0.2711	0.2617	0.9410
$\beta_t = 0.5$	−0.0035	0.3356	0.3289	0.9410
$\beta_w = 0.5$	0.0090	0.2728	0.2678	0.9510

Table 4. Comparison of the proposed and naive estimators for β_t with $n = 2000$ and $p_e = 0.1$.

Parameter		Bias	SSE	ESE	CP
$\beta_t = 0.8$	$\hat{\beta}_t$	-0.0129	0.2544	0.2546	0.9550
	$\hat{\beta}_{t_{sub}}$	−0.0379	0.5391	0.5308	0.9470
	$\hat{\beta}_{t_{srs}}$	−0.0031	0.3677	0.3697	0.9600
	$\hat{\beta}_{t_{in}}$	−0.1840	0.2142	0.2161	0.8610
$\beta_t = 1$	$\hat{\beta}_t$	−0.0027	0.2534	0.2595	0.9540
	$\hat{\beta}_{t_{sub}}$	0.0151	0.5655	0.5635	0.9560
	$\hat{\beta}_{t_{srs}}$	−0.0249	0.3959	0.3853	0.9390
	$\hat{\beta}_{t_{in}}$	−0.2282	0.2117	0.2206	0.8040

transgender women who had sex with men for assessing the efficacy of the DNA/rAd5 vaccine for HIV-1 infection (Fong et al. [8]; Hammer et al. [10]; Janes et al. [14]). As mentioned above, the original study consists of the subjects randomly assigned to receive either the DNA/rAd5 vaccine or placebo, and in the following, we will focus

Table 5. Estimation of regression parameters for case–control studies with $n = 1000$ and $p_e = 0.05$.

Parameter	Bias	SSE	ESE	CP
$\beta_t = 0$	-0.0048	0.4015	0.4086	0.9560
$\beta_w = 0$	0.0060	0.3196	0.3283	0.9570
$\eta = 0.8$	0.0022	0.3048	0.3189	0.9390
$\beta_t = 0.2$	−0.0177	0.3930	0.4108	0.9570
$\beta_w = 0.2$	−0.0158	0.3176	0.3290	0.9590
$\eta = 0.8$	0.0143	0.3076	0.3226	0.9550
$\beta_t = 0.5$	−0.0015	0.3960	0.4003	0.9570
$\beta_w = 0.5$	0.0055	0.3287	0.3354	0.9610
$\eta = 0.8$	0.0250	0.3266	0.3179	0.9450

only on the 1253 subjects in the vaccine group. It is well-known that HIV-1 infection is deadly as it causes AIDS for which there is no cure and thus it is important and essential to develop a safe and effective vaccine for the prevention of the infection. For each subject, four demographic covariates were observed and they are age, race, BMI and behavioural risk. In addition, to assess their relationship with the HIV infection, a number of T cell response biomarkers and antibody response biomarkers were measured for a cohort of 150 subjects consisting of all HIV infection cases (25) and other 125 randomly selected subjects among the vaccine recipients. The failure time of interest here is the time to true HIV-1 infection and for which, only interval-censored data are available.

In all previous analyses, the authors simplified the observed data into right-censored data and also did not consider the possibility of informative censoring (Fong et al. [8]; Hammer et al. [10]; Janes et al. [14]). They identified the T cell response biomarker Env CD8+ polyfunctionality score and the antibody response biomarker IgG.Cconenv03140CF.avi that may have significant effects on the HIV infection time. For simplicity, below we will refer these two biomarkers as to Env CD8 Score and IgG, respectively. For the analysis below, by following Fong et al. [8] and Janes et al. [14], we will focus on the cohort of 150 vaccine recipients, which can be treated as a case–control design with the full cohort being all subjects in the vaccine group, and investigate the relationship between the HIV infection time and the four demographic covariates plus the two biomarkers.

Table 6 presents the estimation results given by the application of the methodology proposed in the previous sections to the HVTN 505 Trial, including the estimated covariate effects $\hat{\beta}_{tn}$ and $\hat{\beta}_{wn}$, the estimated standard errors (ESE) and the p-values for testing the covariate effect being zero. Here for the degree of Bernstein polynomials, we tried several values, including $m = 2, 3, 4, 5, 6$ and 7, and the results above were obtained based on $m = 3$, which gave the smallest AIC defined above, and $B = 500$. One can see from Table 6 that the proposed estimation procedure suggests that among the six covariates considered here, two demographic covariates, race and behavioural risk, seem to be correlated with the HIV infection time and the two biomarkers also appear to have significant prognostic effects on the development of HIV infection. On the other hand, the age and BMI did not seem to have any effects on the HIV infection. In addition, the race and behavioural risk appear to have significant effects on the observation process too.

Table 6. Estimated covariate effects for the HVTN 505 Trial.

	Proposed method					
Covariate	$\hat{\beta}_{tn}$	SSE	*p-value*	$\hat{\beta}_{wn}$	SSE	*p-value*
age	-0.2116	0.2523	0.4018	0.0287	0.3174	0.9279
race	−0.7962	0.4676	0.0886	1.7492	0.6204	0.0048
BMI	−0.1560	0.3020	0.6055	0.1813	0.3621	0.6166
behavioural risk	1.1079	0.5677	0.0510	2.2763	0.6781	0.0008
Env CD8 Score	−0.9575	0.2286	0.0000	0.2661	0.4628	0.5652
IgG	−0.5085	0.1610	0.0016	0.2744	0.1611	0.0886
η	0.0030	2.0820	0.9989			
	Method given in Zhou et al. [40]					
Covariate	$\hat{\beta}_{t_{in}}$	*SSE*	*p-value*			
age	-0.2114	0.2580	0.4125			
race	−0.7985	0.4996	0.1100			
BMI	−0.1561	0.2832	0.5814			
behavioural risk	1.1086	0.7482	0.1385			
Env CD8 Score	−0.9574	0.2846	0.0008			
IgG	−0.5089	0.1620	0.0017			

For comparison, we also applied the method given in Zhou et al. [40], which assumed that the HIV infection time and the observation process were independent, to the data and included the estimated covariate effects, which are denoted by $\hat{\beta}_{t_{in}}$, in the table along with the estimated standard errors and the p-values. One can see from the table that one difference between the results given by the two methods is on the estimation of the effect of the behavioural risk factor, which did not see to have any effect on the development of HIV infection based on the method given in Zhou et al. [40]. One explanation for this may be due to the fact that the method given in Zhou et al. [40] ignored the existence of informative censoring.

6. Discussion and concluding remarks

This paper discussed the analysis of case-cohort studies that yield informatively interval-censored failure time data arising from the proportional hazards model. As discussed above, a great deal of literature has been developed for the analysis of case-cohort studies that give right-censored data. In practice, however, the observed information on the failure time is more likely and naturally given in the form of interval-censored data, which is especially the case for longitudinal or periodic follow-up studies. One major difference between right-censored data and interval-censored data is that the latter has a much more complex structure than the former, which makes the analysis of the latter much more difficult. Although a large amount of literature has also been established for the analysis of either interval-censored data or case-cohort studies, there is no method available for the informative censoring situation discussed above. As pointed out before and seen in Section 5, informative censoring often occurs naturally and for the situation, the analysis that ignores it could result in biased or misleading results and conclusions.

As discussed in Sections 4 and 5, a type of studies that is similar to case-cohort studies is the case–control study and the key difference between the two is the generation of the

subcohort. With the case-cohort design, the subcohort is sampled from all study subjects, while the case–control design samples the subcohort only from the subjects who do not experience the failure event of interest during the follow-up. It is apparent that the data structures under the two designs are different but on the other hand, the simulation study suggested that the proposed estimation approach seems to be valid too for the case–control design. A possible explanation for this is that the resulting data may carry similar information about the model and the regression parameters of interest given the low percentage of the event rate.

In practice, interval-censored data may be given in different forms (Sun [34]). For example, instead of the form discussed here, one may have case K or mixed interval-censored data (Wang et al. [37]). Note that for the analysis, one can still apply the proposed estimation procedure to these situations by expressing the data using the format described here. However, the derivation or establishment of the asymptotic properties may be different and one may need some other assumptions similar to those described in Huang [11] and Wang et al. [37]. In the previous sections, the focus has been on the informative censoring that can be characterized by models (2.1) and (2.2) or through latent variables. More specifically, it has been assumed that the magnitude of the informative censoring can be measured by the parameter η. It is apparent that as with most of frailty model approaches, a natural question would be if one can test $\eta = 0$. Unfortunately it does not seem to exist an established procedure for it in the literature. Another related question is the possibility of performing the goodness-of-fit tests on models (2.1) and (2.2). For this, if $\eta = 0$, one may apply the test procedures given in Ren and He [28] and McKeague and Utikal [26], respectively, to test them separately. However, it would be difficult or not straightforward to generalize either of them to the situation discussed here.

As mentioned above, to deal with the informative censoring, another commonly used method is the copula model approach, which directly models the joint distribution of the failure time of interest and censoring variables (Sun [34]). For example, Cui et al. [6] and Ma et al. [24] developed two such methods for regression analysis of current status data with informative censoring, a special case of interval-censored data where each subject is observed only once. Among others, Ma et al. [23] proposed a copula model approach for regression analysis of general interval-censored data. An advantage of the copula model approach is that it allows one to work or model the marginal distribution and the association parameter separately but it has the limitation that one needs to assume that the underlying copula function is known.

It is well-known that although the proportional hazards model is one of the most commonly used models for regression analysis of failure time data, sometimes one may prefer a different model or a different model may fit the data or describe the problem of interest better (Kalbfleisch and Prentice [16]). For example, the additive hazards model is usually preferred if the excess risk is of interest and one may want to consider the linear transformation model if the model flexibility is more important. Some literature has been developed for these and other models for regression analysis of general interval-censored data or the analysis of case-cohort studies that yield right-censored data. However, there does not seem to exist an established estimation procedure for the problem discussed here under other models. In other words, it would be useful to generalize the proposed method to the situation under the additive hazards or linear transformation model.

Acknowledgments

The authors wish to thank the Editor-in-Chief, the Associate Editor and three reviewers for their many critical and constructive comments and suggestions that greatly improved the paper. Also the authors want to thank Dr Peter Gibert for providing the HIV example data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was partially supported by the National Science Foundation of USA grant DMS-1916170, the National Natural Science Foundation of China grant 11671168, the Science and Technology Developing Plan of Jilin Province of China grant 20170101061JC, and the National Institute of Allergy and Infectious Disease of USA grant 1 R56 AI140953-01.

References

[1] K. Bogaerts,A. Komarek, and E Lesaffre, *Survival Analysis with Interval-Censored Data: A Practical Approach with Examples in R, SAS, and BUGS*, CRC Press, 2017.
[2] K. Chen, *Generalized case cohort sampling*, J. R. Statist. Soc. B 63 (2001), pp. 791–809.
[3] K. Chen and S.H. Lo, *Case-cohort and case-control analysis with Cox's model*, Biometrika 86 (1999), pp. 755–764.
[4] X. Chen, Y. Fan, and V. Tsyrennikov, *Efficient estimation of semiparametric multivariate copula models*, J. Am. Stat. Assoc. 101 (2006), pp. 1228–1240.
[5] D.G. Chen, J. Sun, and K Peace, *Interval-Censored Time-to-Event Data: Methods and Applications*, CRC Press, 2012.
[6] Q. Cui, H. Zhao, and J. Sun, *A new copula model-based method for regression analysis of dependent current status data*, Stat. Interface. 11 (2018), pp. 463–471.
[7] D.M. Finkelstein, *A proportional hazards model for interval-censored failure time data*, Biometrics 42 (1986), pp. 845–854.
[8] Y. Fong, X. Shen, V.C. Ashley, A. Deal, K.E. Seaton, C. Yu, S.P.. Grant, G. Ferrari, R.T. Bailer, R.A. Koup, D. Montefiori, B.F.. Haynes, M. Sarzotti-Kelsoe, B.S. Graham, L.N. Carpp, S.M. Hammer, M. Sobieszczyk, S. Karuna, E. Swann, E. DeJesus, M. Mulligan, I. Frank, S. Buchbinder, R.M. Novak, M.J. McElrath, S. Kalams, M. Keefer, N.A. Frahm, H.E. Janes, P.B. Gilbert, and G.D. Tomaras, *Modification of the association between T-Cell immune responses and human immunodeficiency virus type 1 infection risk by vaccine-induced antibody responses in the HVTN 505 trial*, J. Infect. Dis. 217 (2018), pp. 1280–1288.
[9] P.B. Gilbert, M.L. Peterson, D. Follmann, M.G. Hudgens, D.P. Francis, M. Gurwith, W.L. Heyward, D.V. Jobes, V. Popovic, S.G. Self, F. Sinangil, D. Burke, and P.W. Berman, *Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial*, J. Infect. Dis. 191 (2005), pp. 666–677.
[10] S.M. Hammer, M.E. Sobieszczyk, H. Janes, M.J. Mulligan, S.T. Karuna, D. Grove, B.A. Koblin, S.P. Buchbinder, M.C. Keefer, G.D. Tomaras, N. Frahm, J. Hural, C. Anude, B.S. Graham, M.E. Enama, E. Adams, E. DeJesus, R.M. Novak, I. Frank, C. Bentley, S. Ramirez, R. Fu, R.A. Koup, J.R. Mascola, G.J. Nabel, D.C. Montefiori, J. Kublin, M.J. McElrath, L. Corey, and P.B. Gilbert, HVTN 505 Study Team, *Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine*, N. Engl. J. Med. 369 (2013), pp. 2083–2092.
[11] J. Huang, *Asymptotic properties of nonparametric estimation based on partly interval-censored data*, Stat. Sin. 9 (1999), pp. 501–519.
[12] J. Huang and A.J. Rossini, *Sieve estimation for the proportional-odds failure-time regression model with interval censoring*, J. Am. Stat. Assoc. 92 (1997), pp. 960–967.

[13] X. Huang and R. Wolfe, *A frailty model for informative censoring*, Biometrics 58 (2002), pp. 510–520.
[14] H.E. Janes, K.W. Cohen, N. Frahm, S.C. De Rosa, B. Sanchez, J. Hural, C.A. Magaret, S. Karuna, C. Bentley, R. Gottardo, G. Finak, D. Grove, M. Shen, B.S. Graham, R.A. Koup, M.J. Mulligan, B. Koblin, S.P. Buchbinder, M.C. Keefer, E. Adams, C. Anude, L. Corey, M. Sobieszczyk, S.M. Hammer, P.B. Gilbert, and M.J. McElrath, *Higher T-cell responses induced by DNA/rAd5 HIV-1 preventive vaccine are associated with lower HIV-1 infection risk in an efficacy trial*, J. Infect. Dis. 215 (2017), pp. 1376–1385.
[15] N.P. Jewell and M. van der Laan, *Current status data: review, recent development and open problems*, Adv. Survival Anal. 35 (2004), pp. 625–643.
[16] J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, New York, 2002.
[17] S. Kang and J. Cai, *Marginal hazards model for case-cohort studies with multiple disease outcomes*, Biometrika 96 (2009), pp. 887–901.
[18] R.H. Keogh and I.R. White, *Using full-cohort data in nested case-control and case-cohort studies by multiple imputation*, Stat. Med. 32 (2013), pp. 4021–4043.
[19] S. Kim, J. Cai, and W. Lu, *More efficient estimators for case-cohort studies*, Biometrika 100 (2013), pp. 695–708.
[20] Z. Li and B. Nan, *Relative risk regression for current status data in case-cohort studies*, Can. J. Stat. 39 (2011), pp. 557–577.
[21] G.G. Lorentz, *Bernstein Polynomials*, Chelsea Publishing Co, New York, 1986.
[22] S. Ma and M.R. Kosorok, *Robust semiparametric M-estimation and the weighted bootstrap*, J. Multivar. Anal. 96 (2005), pp. 190–217.
[23] L. Ma, T. Hu, and J. Sun, *Cox regression analysis of dependent interval-censored failure time data*, Comput. Stat. Data Anal. 103 (2016), pp. 79–90.
[24] L. Ma, T. Hu, and J. Sun, *Sieve maximum likelihood regression analysis of dependent current status data*, Biometrika 102 (2015), pp. 731–738.
[25] H. Marti and M. Chavance, *Multiple imputation analysis of case-cohort studies*, Stat. Med. 30 (2011), pp. 1595–1607.
[26] I.W. McKeague and K.J. Utikal, *Goodness-of-fit tests for additive hazards and proportional hazards models*, Scand. J. Stat. 18 (1991), pp. 177–195.
[27] R.L. Prentice, *A case-cohort design for epidemiologic cohort studies and disease prevention trials*, Biometrika 73 (1986), pp. 1–11.
[28] J.J. Ren and B. He, *Estimation and goodness-of-fit for the Cox model with various types of censored data*, J. Stat. Plan. Inference. 141 (2011), pp. 961–971.
[29] T. Saegusa and J.A. Wellner, *Weighted likelihood estimation under two-phase sampling*, Ann. Stat. 41 (2013), pp. 269–295.
[30] T.H. Scheike and T. Martinussen, *Maximum likelihood estimation for Cox's regression model under case-cohort sampling*, Scand. J. Stat. 31 (2004), pp. 283–293.
[31] X. Shen, *On methods of sieves and penalization*, Ann. Stat. 25 (1997), pp. 2555–2591.
[32] X. Shen and W.H. Wong, *Convergence rate of sieve estimates*, Ann. Stat. 22 (1994), pp. 580–615.
[33] J. Sun, *A nonparametric test for current status data with unequal censoring*, J. R. Stat. Soc. B 61 (1999), pp. 243–250.
[34] J Sun, *The Statistical Analysis of Interval-censored Failure Time Data*, New York, Springer, 2006.
[35] A.W. van der Vaart and J.A Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York, 1996.
[36] P. Wang, H. Zhao, and J. Sun, *Regression analysis of case K interval-censored failure time data in the presence of informative censoring*, Biometrics 72 (2016), pp. 1103–1112.
[37] L.M. Wang, C.S McMahan, M.G Hudgens, and Z.P. Qureshi, *A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data*, Biometrics 72 (2016), pp. 222–231.
[38] S. Wang, C. Wang, P. Wang, and J. Sun, *Semiparametric analysis of the additive hazards model with informatively interval-censored failure time data*, Comput. Stat. Data Anal. 125 (2018), pp. 1–9.

[39] Y. Zhang, L. Hua, and J. Huang, *A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data*, Scand. J. Stat. 37 (2010), pp. 338–354.
[40] Q. Zhou, T. Hu, and J. Sun, *A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data*, J. Am. Stat. Assoc. 112 (2017a), pp. 664–672.
[41] Q. Zhou, H. Zhou, and J. Cai, *Case-cohort studies with interval-censored failure time data*, Biometrika 104 (2017b), pp. 17–29.

Appendix Proofs of the asymptotic properties of $\hat{\theta}_n$

In this appendix, we will sketch the proof of the asymptotic properties of the proposed estimator $\hat{\theta}_n$. Let τ denote the length of study. Then a single observation can be written as

$$O^{\xi} = \{U,\ \Psi W,\ \Psi = I(W < \tau - U),\ \Delta_1 = I(T \le U),\ \Delta_2 = I(U < T \le U + \Psi W),\ \xi Z,\ \xi\}.$$

To establish the asymptotic properties, we need the following regularity conditions, which are commonly used in the studies of interval-censored data and usually satisfied in practice (Huang and Rossini [12]; Zhang et al. [39]; Ma et al. [23]; Zhou et al. [40]).

(C1) The distribution of the covariate Z has a bounded support in R^p and is not concentrated on any proper subspace of R^p.
(C2) The true parameters $(\beta_{t0}, \beta_{w0}, \eta_0)$ lie in the interior of a compact set $\mathbf{B}$ in $R^{2p} \times R^{+}$.
(C3) The first derivative of $\Lambda_{t0}(\cdot)$ and $\Lambda_{w0}(\cdot)$, denoted by $\Lambda_{t0}^{(1)}(\cdot)$ and $\Lambda_{w0}^{(1)}(\cdot)$, is Holder continuous with exponent $\gamma \in (0, 1]$. That is, there exists a constant $K > 0$ such that $|\Lambda_{t0}^{(1)}(t_1) - \Lambda_{t0}^{(1)}(t_2)| \le K|t_1 - t_2|^{\gamma}$ for all $t_1, t_2 \in [\sigma, \tau]$, where $0 < \sigma < \tau < \infty$. Let $r = 1 + \gamma$.
(C4) There exists a constant $K > 0$ such that $Pl(\theta, O^{\xi}) - Pl(\theta_0, O^{\xi}) \le -Kd(\theta, \theta_0)^2$ for every θ in a neighbourhood of θ_0, where $l(\theta, O^{\xi})$ is the weighted log-likelihood function based on a single observation O^{ξ}.
(C5) The matrix $E(\{l^*(\vartheta_0, O)\}^{\otimes 2})$ is finite and positive definite, where $v^{\otimes 2} = vv'$ for a vector v, and $l^*(\vartheta, O)$ is the efficient score for $\vartheta = (\beta_t, \beta_w, \eta)$ based on the complete observation $O = \{U, \Psi W, \Psi, \Delta_1, \Delta_2, Z\}$ and will be given in the proof of Theorem 2.

For the proof, we will mainly employ the empirical process theory and some nonparametric techniques. Let $Pf = \int f(y)dP$ denote the expectation of $f(Y)$ under the probability measure P, and $P_n f = n^{-1}\sum_{i=1}^{n} f(Y_i)$, the expectation of $f(Y)$ under the empirical measure P_n. Define the covering number of the class $\mathcal{L}_n = \{l(\theta, O^{\xi}) : \theta \in \Theta_n\}$, where $l(\theta, O^{\xi})$ is the weighted log-likelihood function based on a single observation O^{ξ}. Also for any $\epsilon > 0$, define the covering number $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ as the smallest positive integer κ for which there exists $\{\theta^{(1)}, \ldots, \theta^{(\kappa)}\}$ such that

$$\min_{j \in \{1,\ldots,\kappa\}} \frac{1}{n}\sum_{i=1}^{n} \left| l(\theta, O_i^{\xi}) - l(\theta^{(j)}, O_i^{\xi}) \right| < \epsilon$$

for all $\theta \in \Theta_n$, where $\{O_1^{\xi}, \ldots, O_n^{\xi}\}$ represent the observed data and for $j = 1, \ldots, \kappa$, $\theta^{(j)} = (\beta_t^{(j)}, \beta_w^{(j)}, \eta^{(j)}, \Lambda_t^{(j)}, \Lambda_w^{(j)}) \in \Theta_n$. If no such κ exists, define $N(\epsilon, \mathcal{L}_n, L_1(P_n)) = \infty$. Also for the proof, we need the following two lemmas, whose proofs are similar to those for Lemmas 1 & 2 in Zhou et al. [40] and thus omitted.

Lemma A.1: *Assume that the regularity conditions* (C1)–(C3) *given above hold. Then we have that the covering number of the class* $\mathcal{L}_n = \{l(\theta, O^{\xi}) : \theta \in \Theta_n\}$ *satisfies*

$$N\Big(\epsilon, \mathcal{L}_n, L_1(P_n)\Big) \le KM_n^{2(m+1)}\epsilon^{-(2p+2m+3)}$$

for a constant K, where $m = o(n^{\nu})$ *with* $\nu \in (0, 1)$ *is the degree of Bernstein polynomials, and* $M_n = O(n^a)$ *with* $a > 0$ *controls the size of the sieve space* Θ_n.

Lemma A.2: *Assume that the regularity conditions* (C1)–(C3) *given above hold. Then we have that*

$$\sup_{\theta\in\Theta_n}\left|P_n l(\theta,O^\xi)-Pl(\theta,O^\xi)\right|\to 0$$

almost surely.

Proof of Theorem 3.1: We first prove the strong consistency of $\hat{\theta}_n$. Let $l(\theta,O^\xi)$ denote the weighted log-likelihood function based on a given single observation O^ξ and consider the class of functions $\mathcal{L}_n=\{l(\theta,O^\xi):\theta\in\Theta_n\}$. By Lemma A.1, the covering number of $\mathcal{L}_n$ satisfies

$$N\Big(\epsilon,\mathcal{L}_n,L_1(P_n)\Big)\le KM_n^{2(m+1)}\epsilon^{-(2p+2m+3)}.$$

Furthermore, by Lemma A.2, we have

$$\sup_{\theta\in\Theta_n}\left|P_n l(\theta,O^\xi)-Pl(\theta,O^\xi)\right|\to 0\quad\text{almost surely.}\tag{A1}$$

Note that $E(p|O)=1$, then $Pl(\theta,O^\xi)=P\{pl(\theta,O)\}=Pl(\theta,O)$ and θ_0 maximizes $Pl(\theta,O^\xi)$. Let $M(\theta,O^\xi)=-l(\theta,O^\xi)$, and define $K_\epsilon=\{\theta:d(\theta,\theta_0)\ge\epsilon,\theta\in\Theta_n\}$ for $\epsilon>0$ and

$$\zeta_{1n}=\sup_{\theta\in\Theta_n}|P_nM(\theta,O^\xi)-PM(\theta,O^\xi)|,\ \zeta_{2n}=P_nM(\theta_0,O^\xi)-PM(\theta_0,O^\xi).$$

Then

$$\inf_{K_\epsilon}PM(\theta,O^\xi)=\inf_{K_\epsilon}\Big\{PM(\theta,O^\xi)-P_nM(\theta,O^\xi)+P_nM(\theta,O^\xi)\Big\}\le\zeta_{1n}+\inf_{K_\epsilon}P_nM(\theta,O^\xi).\tag{A2}$$

If $\hat{\theta}_n\in K_\epsilon$, then we have

$$\inf_{K_\epsilon}P_nM(\theta,O^\xi)=P_nM(\hat{\theta}_n,O^\xi)\le P_nM(\theta_0,O^\xi)=\zeta_{2n}+PM(\theta_0,O^\xi).\tag{A3}$$

Define $\delta_\epsilon=\inf_{K_\epsilon}PM(\theta,O^\xi)-PM(\theta_0,O^\xi)$. Under Condition (C4), we have $\delta_\epsilon>0$. It follows from A2 and A3 that

$$\inf_{K_\epsilon}PM(\theta,O^\xi)\le\zeta_{1n}+\zeta_{2n}+PM(\theta_0,O^\xi)=\zeta_n+PM(\theta_0,O^\xi)$$

with $\zeta_n=\zeta_{1n}+\zeta_{2n}$, and hence $\zeta_n\ge\delta_\epsilon$.This gives $\{\hat{\theta}_n\in K_\epsilon\}\subseteq\{\zeta_n\ge\delta_\epsilon\}$, and by A1 and the strong law of large numbers, we have both $\zeta_{1n}\to 0$ and $\zeta_{2n}\to 0$ almost surely. Therefore, $\cup_{k=1}^\infty\cap_{n=k}^\infty\{\hat{\theta}_n\in K_\epsilon\}\subseteq\cup_{k=1}^\infty\cap_{n=k}^\infty\{\zeta_n\ge\delta_\epsilon\}$, which proves that $d(\hat{\theta}_n,\theta_0)\to 0$ almost surely.

Now we will show the convergence rate of $\hat{\theta}_n$ by using Theorem 3.4.1 of van der Vaart and Wellner [35]. Below we use $\tilde{K}$ to denote a universal positive constant which may differ from place to place. First note from Theorem 1.6.2 of Lorentz [21] that there exists a Bernstein polynomial Λ_{tn0} and Λ_{wn0} such that $\|\Lambda_{tn0}-\Lambda_{t0}\|_\infty=O(m^{-r/2})$ and $\|\Lambda_{wn0}-\Lambda_{w0}\|_\infty=O(m^{-r/2})$. Define $\theta_{n0}=(\beta_{t0},\beta_{w0},\eta_0,\Lambda_{tn0},\Lambda_{wn0})$. Then we have $d(\theta_{n0},\theta_0)=O(n^{-r\nu/2})$. For any $\rho>0$, define the class of functions $\mathcal{F}_\rho=\{l(\theta,O^\xi)-l(\theta_{n0},O^\xi):\theta\in\Theta_n,\ \rho/2<d(\theta,\theta_{n0})\le\rho\}$ for a given single observation O^ξ. One can easily show that $P(l(\theta_0,O^\xi)-l(\theta_{n0},O^\xi))\le\tilde{K}d(\theta_0,\theta_{n0})\le\tilde{K}n^{-r\nu/2}$. From Condition (C4), for large n, we have

$$\begin{aligned}P(l(\theta,O^\xi)-l(\theta_{n0},O^\xi))&=P(l(\theta,O^\xi)-l(\theta_0,O^\xi))+P(l(\theta_0,O^\xi)-l(\theta_{n0},O^\xi))\\&\le-\tilde{K}\rho^2+\tilde{K}n^{-r\nu/2}=-\tilde{K}\rho^2,\end{aligned}$$

for any $l(\theta,O^\xi)-l(\theta_{n0},O^\xi)\in\mathcal{F}_\rho$.

Following the calculations in Shen and Wong [32](p. 597), we can establish that for $0<\varepsilon<\rho$, $\log N_{[]}(\varepsilon,\mathcal{F}_\rho,L_2(P))\le\tilde{K}N\log(\rho/\varepsilon)$ with $N=2(m+1)$. Moreover, some algebraic manipulations yield that $P(l(\theta,O^\xi)-l(\theta_{n0},O^\xi))^2\le\tilde{K}\rho^2$ for any $l(\theta,O^\xi)-l(\theta_{n0},O^\xi)\in\mathcal{F}_\rho$. Under Conditions

(C1)–(C3), it is easy to see that $\mathcal{F}_\rho$ is uniformly bounded. Therefore, by Lemma 3.4.2 of van der Vaart and Wellner [35], we obtain

$$E_P\|n^{1/2}(P_n - P)\|_{\mathcal{F}_\rho} \le \tilde{K} J_{[]}\{\rho, \mathcal{F}_\rho, L_2(P)\}\left[1 + \frac{J_{[]}\{\rho, \mathcal{F}_\rho, L_2(P)\}}{\rho^2 n^{1/2}}\right]$$

where $J_{[]}\{\rho, \mathcal{F}_\rho, L_2(P)\} = \int_0^\rho [1 + \log N_{[]}\{\varepsilon, \mathcal{F}_\rho, L_2(P)\}]^{1/2} d\varepsilon \le \tilde{K} N^{1/2}\rho$. This yields $\phi_n(\rho) = N^{1/2}\rho + N/n^{1/2}$. It is easy to see that $\phi_n(\rho)/\rho$ is decreasing in ρ, and $r_n^2\phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} \le \tilde{K} n^{1/2}$, where $r_n = N^{-1/2}n^{1/2} = n^{(1-\nu)/2}$.

Finally note that $P_n\{l(\hat{\theta}_n, O^\xi) - l(\theta_{n0}, O^\xi)\} \ge 0$ and $d(\hat{\theta}_n, \theta_{n0}) \le d(\hat{\theta}_n, \theta_0) + d(\theta_0, \theta_{n0}) \to 0$ in probability. Thus by applying Theorem 3.4.1 of van der Vaart and Wellner [35], we have $n^{(1-\nu)/2} d(\hat{\theta}_n, \theta_{n0}) = O_p(1)$. This together with $d(\theta_{n0}, \theta_0) = O(n^{-r\nu/2})$ yields that $d(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu/2})$ and the proof is completed. ■

Proof of Theorem 3.2: Now we will prove the asymptotic normality of $\hat{\vartheta}_n = (\hat{\beta}_{tn}, \hat{\beta}_{wn}, \hat{\eta}_n)$. First we will establish the asymptotic normality for the estimator based on the complete observation $O = \{U, \Psi W, \Psi, \Delta_1, \Delta_2, Z\}$. With a little abuse of notation, we still denote the complete-data estimator as $\hat{\vartheta}_n$.

Let V denote the linear span of $\Theta - \theta_0$ and define the Fisher inner product for $v, \tilde{v} \in V$ as $< v, \tilde{v} >= P\{\dot{l}(\theta_0, O)[v]\dot{l}(\theta_0, O)[\tilde{v}]\}$ and the Fisher norm for $v \in V$ as $\|v\|^2 =< v, v >$, where

$$\dot{l}(\theta_0, O)[v] = \frac{\mathrm{d}l(\theta_0 + sv, O)}{\mathrm{d}s}\Big|_{s=0}$$

denotes the first order directional derivative of $l(\theta, O)$ at the direction $v \in V$ (evaluated at θ_0). Also let $\bar{V}$ be the closed linear span of V under the Fisher norm. Then $(\bar{V}, \|\cdot\|)$ is a Hilbert space. Furthermore, for a vector of $(2p + 1)$ dimension $b = (b_1', b_2', b_3)'$ with $\|b\| \le 1$ and any $v \in V$, define a smooth functional of θ as $h(\theta) = b_1'\beta_1 + b_2'\beta_2 + b_3\eta$ and

$$\dot{h}(\theta_0)[v] = \frac{\mathrm{d}h(\theta_0 + sv)}{\mathrm{d}s}\Big|_{s=0}$$

whenever the right hand-side limit is well defined. Then by the Riesz representation theorem, there exists $v^* \in \bar{V}$ such that $\dot{h}(\theta_0)[v] =< v, v^* >$ for all $v \in \bar{V}$ and $\|v^*\| = \|\dot{h}(\theta_0)\|$. Also note that $h(\theta) - h(\theta_0) = \dot{h}(\theta_0)[\theta - \theta_0]$. It thus follows from the Cramér-Wold device that to prove the asymptotic normality for $\hat{\vartheta}_n$, i.e. $n^{1/2}(\hat{\vartheta}_n - \vartheta_0) \to N(0, I^{-1}(\vartheta_0))$ in distribution, it suffices to show that

$$n^{1/2} < \hat{\theta}_n - \theta_0, v^* > \to_d N(0, b' I^{-1}(\vartheta_0) b) \tag{A4}$$

since $b'(\hat{\vartheta}_n - \vartheta_0) = h(\hat{\theta}_n) - h(\theta_0) = \dot{h}(\theta_0)[\hat{\theta}_n - \theta_0] =< \hat{\theta}_n - \theta_0, v^* >$. In fact, A4 holds since one can show that $n^{1/2} < \hat{\theta}_n - \theta_0, v^* > \to_d N(0, \|v^*\|^2)$ and $\|v^*\|^2 = b' I^{-1}(\vartheta_0) b$.

We first prove that $n^{1/2} < \hat{\theta}_n - \theta_0, v^* > \to_d N(0, \|v^*\|^2)$. Let $\delta_n = n^{-min\{(1-\nu)/2, r\nu/2\}}$ denote the rate of convergence obtained in Theorem 3.1, and for any $\theta \in \Theta$ such that $d(\theta, \theta_0) \le \delta_n$, define the first order directional derivative of $l(\theta, O)$ at the direction $v \in V$ as

$$\dot{l}(\theta, O)[v] = \frac{\mathrm{d}l(\theta + sv, O)}{\mathrm{d}s}\Big|_{s=0}$$

and the second-order directional derivative at the directions $v, \tilde{v} \in V$ as

$$\ddot{l}(\theta, O)[v, \tilde{v}] = \frac{\mathrm{d}^2 l(\theta + sv + \tilde{s}\tilde{v}, O)}{\mathrm{d}\tilde{s}\mathrm{d}s}\Big|_{s=0}\Big|_{\tilde{s}=0} = \frac{\mathrm{d}\dot{l}(\theta + \tilde{s}\tilde{v}, O)[\tilde{v}]}{\mathrm{d}\tilde{s}}\Big|_{\tilde{s}=0}.$$

Note that by Condition (C3) and Theorem 1.6.2 of Lorentz [21], there exists $\Pi_n v^* \in \Theta_n - \theta_0$ such that $\|\Pi_n v^* - v^*\| = O(n^{-vr/2})$. Furthermore, under the assumption $\nu > 1/2r$, we have $\delta_n\|\Pi_n v^* -$

$v^*\| = o(n^{-1/2})$. Define $r[\theta-\theta_0, O] = l(\theta, O) - l(\theta_0, O) - \dot{l}(\theta_0, O)[\theta-\theta_0]$ and let ε_n be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. Then by the definition of $\hat{\theta}_n$, we have

$$\begin{aligned}
0 &\le P_n[l(\hat{\theta}_n, O) - l(\hat{\theta}_n \pm \varepsilon_n \Pi_n v^*, O)] \\
&= \mp\varepsilon_n P_n \dot{l}(\theta_0, O)[\Pi_n v^*] + (P_n - P)\Big\{r[\hat{\theta}_n - \theta_0, O] - r[\hat{\theta}_n \pm \varepsilon_n \Pi_n v^* - \theta_0, O]\Big\} \\
&\quad + P\Big\{r[\hat{\theta}_n - \theta_0, O] - r[\hat{\theta}_n \pm \varepsilon_n \Pi_n v^* - \theta_0, O]\Big\} \\
&= \mp\varepsilon_n P_n \dot{l}(\theta_0, O)[v^*] \mp \varepsilon_n P_n \dot{l}(\theta_0, O)[\Pi_n v^* - v^*] + (P_n - P)\Big\{r[\hat{\theta}_n - \theta_0, O] \\
&\quad - r[\hat{\theta}_n \pm \varepsilon_n \Pi_n v^* - \theta_0, O]\Big\} + P\Big\{r[\hat{\theta}_n - \theta_0, O] - r[\hat{\theta}_n \pm \varepsilon_n \Pi_n v^* - \theta_0, O]\Big\} \\
&= \mp\varepsilon_n P_n \dot{l}(\theta_0, O)[v^*] \mp I_1 + I_2 + I_3.
\end{aligned}$$

We will investigate the asymptotic behaviour of I_1, I_2 and I_3. For I_1, it follows from Conditions (C1)–(C3), Chebyshev inequality and $\|\Pi_n v^* - v^*\| = o(1)$ that $I_1 = \varepsilon_n \times o_p(n^{-1/2})$. For I_2, by the mean value theorem, we obtain that

$$\begin{aligned}
I_2 &= (P_n - P)\Big\{l(\hat{\theta}_n, O) - l(\hat{\theta}_n \pm \varepsilon_n \Pi_n v^*, O) \pm \varepsilon_n \dot{l}(\theta_0, O)[\Pi_n v^*]\Big\} \\
&= \mp\varepsilon_n (P_n - P)\Big\{(\dot{l}(\tilde{\theta}, O) - \dot{l}(\theta_0, O))[\Pi_n v^*]\Big\},
\end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}_n$ and $\hat{\theta}_n \pm \varepsilon_n \Pi_n v^*$. By Theorem 2.8.3 of van der Vaart and Wellner [35], we know that $\{\dot{l}(\theta, O)[\Pi_n v^*] : \|\theta - \theta_0\| \le \delta_n\}$ is Donsker class. Therefore, by Theorem 2.11.23 of van der Vaart and Wellner [35], we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$. For I_3, note that

$$\begin{aligned}
P(r[\theta-\theta_0, O]) &= P\{l(\theta, O) - l(\theta_0, O) - \dot{l}(\theta_0, O)[\theta-\theta_0]\} \\
&= 2^{-1} P\{\ddot{l}(\tilde{\theta}, O)[\theta-\theta_0, \theta-\theta_0] - \ddot{l}(\theta_0, O)[\theta-\theta_0, \theta-\theta_0]\} \\
&\quad + 2^{-1} P\{\ddot{l}(\theta_0, O)[\theta-\theta_0, \theta-\theta_0]\} \\
&= 2^{-1} P\{\ddot{l}(\theta_0, O)[\theta-\theta_0, \theta-\theta_0]\} + \varepsilon_n \times o_p(n^{-1/2}),
\end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and θ and the last equation follows from Taylor expansion and Conditions (C1)–(C3). Therefore,

$$\begin{aligned}
I_3 &= -2^{-1}\{\|\hat{\theta}_n - \theta_0\|^2 - \|\hat{\theta}_n \pm \varepsilon_n \Pi_n v^* - \theta_0\|^2\} + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta}_n - \theta_0, \Pi_n v^* > + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta}_n - \theta_0, v^* > + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta}_n - \theta_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}),
\end{aligned}$$

where the last equality holds due to the facts $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$, Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \to \|v^*\|^2$. Combining the above facts, together with $P\dot{l}(\theta_0, O)[v^*] = 0$, we can establish that

$$\begin{aligned}
0 &\le P_n\{l(\hat{\theta}_n, O) - l(\hat{\theta}_n \pm \varepsilon_n \Pi_n v^*, O)\} \\
&= \mp\varepsilon_n P_n \dot{l}(\theta_0, O)[v^*] \pm \varepsilon_n < \hat{\theta}_n - \theta_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \mp\varepsilon_n (P_n - P)\{\dot{l}(\theta_0, O)[v^*]\} \pm \varepsilon_n < \hat{\theta}_n - \theta_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}).
\end{aligned}$$

Therefore, we obtain $\mp n^{1/2}(P_n - P)\{\dot{l}(\theta_0, O)[v^*]\} \pm n^{1/2} < \hat{\theta}_n - \theta_0, v^* > + o_p(1) \ge 0$ and then $n^{1/2} < \hat{\theta}_n - \theta_0, v^* >= n^{1/2}(P_n - P)\{\dot{l}(\theta_0, O)[v^*]\} + o_p(1) \to_d N(0, \|v^*\|^2)$ by the central limit theorem and $\|v^*\|^2 = \|\dot{l}(\theta_0, O)[v^*]\|^2$.

Next we will prove that $\|v^*\|^2 = b'I^{-1}(\vartheta_0)b$. For each component ϑ_q, $q = 1, 2, \ldots, (2p+1)$, we denote by $\psi_q^* = (b_{1q}^*, b_{2q}^*)$ the value of $\psi_q = (b_{1q}, b_{2q})$ minimizing

$$E\Big\{l_\vartheta \cdot e_q - l_{b_1}[b_{1q}] - l_{b_2}[b_{2q}]\Big\}^2,$$

where l_ϑ is the score function for ϑ, l_{b_j} is the score operator for Λ_j, $j = 1, 2$, and e_q is a $(2p+1)$-dimensional vector of zeros except the q-th element equal to 1.

Define the q-th element of $l^*(\vartheta, O)$ as $l_\vartheta \cdot e_q - l_{b_1}[b_{1q}^*] - l_{b_2}[b_{2q}^*]$, $q = 1, \ldots, (2p+1)$, and $I(\vartheta)$ as $E(\{l^*(\vartheta, O)\}^{\otimes 2})$. By Condition (C5), the matrix $I(\vartheta_0)$ is positive definite. Furthermore, by following similar calculations in Chen et al. [4](sec. 3.2), we obtain

$$\|v^*\|^2 = \|\dot{h}(\theta_0)\|^2 = \sup_{v \in \bar{V}: \|v\| > 0} \frac{|\dot{h}(\theta_0)[v]|^2}{\|v\|^2} = b'\Big[E\big(\{l^*(\vartheta_0, O)\}^{\otimes 2}\big)\Big]^{-1} b = b'I^{-1}(\vartheta_0)b\,.$$

Thus, we have shown that $n^{1/2}(\hat{\vartheta}_n - \vartheta_0) \to N(0, I^{-1}(\vartheta_0))$ in distribution for the estimator $\hat{\vartheta}_n$ based on the complete data.

Now consider the estimator $\hat{\vartheta}_n$ based only on the case-cohort data. Note that the weight $p = \xi/\pi_q(\Delta_1, \Delta_2)$ is bounded and does not depend on θ, and $E\{p|O\} = 1$. By Theorem 3.2 of Saegusa and Wellner [29], we have

$$n^{1/2}(\hat{\vartheta}_n - \vartheta_0) = I^{-1}(\vartheta_0)\, n^{-1/2} \sum_{i=1}^{n} p_i l^*(\vartheta_0, O_i) + o_p(1),$$

where $I(\vartheta)$ and $l^*(\vartheta, O)$, defined above, are the information and efficient score for ϑ based on the complete data. Note that

$$\begin{aligned}
\mathrm{var}\{pl^*(\vartheta_0, O)\} &= \mathrm{var}\big\{E\{pl^*(\vartheta_0, O)|O\}\big\} + E\big\{\mathrm{var}\{pl^*(\vartheta_0, O)|O\}\big\} \\
&= \mathrm{var}\{l^*(\vartheta_0, O)\} + E\left\{\mathrm{var}(\xi|O)\frac{\{l^*(\vartheta_0, O)\}^{\otimes 2}}{\pi_q^2(\Delta_1, \Delta_2)}\right\} \\
&= I(\vartheta_0) + E\left\{\frac{1 - \pi_q(\Delta_1, \Delta_2)}{\pi_q(\Delta_1, \Delta_2)}\{l^*(\vartheta_0, O)\}^{\otimes 2}\right\}.
\end{aligned}$$

Thus, we have

$$n^{1/2}(\hat{\vartheta}_n - \vartheta_0) \to N(0, \Sigma)$$

in distribution, where

$$\Sigma = I^{-1}(\vartheta_0) + I^{-1}(\vartheta_0)E\left\{\frac{1 - \pi_q(\Delta_1, \Delta_2)}{\pi_q(\Delta_1, \Delta_2)}\{l^*(\vartheta_0, O)\}^{\otimes 2}\right\} I^{-1}(\vartheta_0).$$

■