A proportional hazards model for time-to-event data with epidemiological bias

Qiaozhen ZHANG[1], Hongsheng DAI[2]*, Bo FU[3,4]

[1] *LPMC and Institute of Statistics, Nankai University, Tianjin 300071, China*

[2] *Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester,CO4 3SQ, UK*

[3] *Administrative Data Research Centre for England, University College London, 222 Euston Road, London NW1 2DA, UK*

[4] *Institute of Child Health, University College London, 30 Guilford Street, London WC1 N 1EH, UK*

Abstract

In hepatitis C virus (HCV) epidemiological studies, the estimation of progression to cirrhosis and prognostic effects of associated risk factors is of particular importance when projecting national disease burden. However, the progression estimates obtained from conventional methods could be distorted due to a referral bias [11]. In recent years, several approaches have been developed to handle this epidemiological bias in analyzing time-to-event data. This paper proposes a new estimation approach for this problem under a semiparametric proportional hazards framework. The new method uses a martingale approach based on the mean rate function, rather than the traditional hazard rate function, and develops an iterative algorithm to estimate the Cox regression parameter and baseline hazard rate simultaneously. The consistency and asymptotic properties of the proposed estimators are derived theoretically and evaluated via simulation studies. The new

*Corresponding author. *Email address:* hdaia@essex.ac.uk (Hongsheng Dai).

method is also applied to a real HCV cohort study.

Keywords: Censoring; Martingale; Proportional hazards model; Referral bias; Truncation.

1 Introduction

In hepatitis C virus (HCV) epidemiological studies, unbiased estimation of progression from HCV infection to a particular outcome event of interest, such as liver cirrhosis, is of considerable importance when projecting national HCV disease burden [2, 7, 11, 17, 20]. However widely varying estimates of progression rate have been reported, mainly due to the referral bias in hepatitis C epidemiology [8, 9, 10, 11, 27, 29]. Clinically the patients with more rapid disease progression are preferentially referred to specialist clinics at later stages of disease [11]. If so, the conventional analysis based on liver clinic cohorts will lead to a biased estimate of the progression rate among the HCV patient community [9, 11]. [10] proposed a pseudo score weighting approach to correct the estimation biases and recover the parameters. However, its performance depends on strong assumptions about referral patterns and a sound estimate of the selection probability. It may be difficult to justify the assumptions about referral patterns in practice.

To reduce the uncertainty involved in the assumptions, [6] modeled such referral bias under a survival analysis framework with truncation. Consider the data set studied in [11], where 387 HCV-infected individuals were recruited from Edinburgh Royal Infirmary's liver clinic by the end of 1999. An individual's information is only available if he or she was referred to the clinic cohort before the end of study recruitment, that is, the observed data is subject to a univariate truncation $R \leqslant L$ [6]. Here R is the period from infection to referral to the liver clinic and L is the time from infection to the end of recruitment. Our main interest

is the incubation period (time T) from HCV infection to development of cirrhosis, which is subject to right censoring at C, the time from infection to last diagnosis (such as an invasive biopsy test) follow-up. The event times (R, T) are correlated, because referral is increasingly likely the closer a patient is to developing cirrhosis [11]. [6] proposed a new approach by transforming the epidemiological bias problem to a nonparametric bivariate survival analysis modelling with the presences of both censoring and truncation and proposed new estimators for the bivariate distribution function based on the idea of a polar coordinate transformation [5]. To further study how the progression risk (hazard rate) of the cirrhosis event is affected by risk factors, [29] proposed an accelerated failure time model to obtain a robust debiasing estimate for the effect of covariates by modelling the dependency of outcome event time on referral time. The proposed method in [29], however, cannot deal with time-dependent covariates, which may be a constraint for its application since the incubation time from HCV infection to cirrhosis usually lasts for a quite long period (more than 10 years) and time-dependent covariates are often available as potential predictors over follow-up. More recently, [27] developed a maximum likelihood estimating approach for this referral bias problem using parametric Weibull regression models.

In this paper, we will develop a new estimation approach under Cox proportional hazards framework for modelling time-to-event data with referral bias, where time-dependent covariates are also allowed. For the pair of time-to-event variables (R, T) discussed above, we consider the following survival model for the hazard function of T, the cirrhosis time,

$$\Lambda(dt) \;=\; \Lambda_0(dt) \cdot \exp\left[\boldsymbol{W}^\top(t_-)\boldsymbol{\beta}\right], \tag{1}$$

where $\boldsymbol{W}(t)$ is a possibly time-dependent p-dimensional covariate vector, t_- means

the left limit, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top$ and Λ_0 is a completely unspecified continuous baseline cumulative hazard rate function. Note that if one is interested in the cross-ratio function estimation, the method proposed in [14] is applicable for such censored and truncated data. We here focus on the estimation for the above hazard rate function for T, which is the main interest in this application [29]. For model (1), if the cirrhosis time T itself is subject to both censoring and truncation, we can use the methods proposed in [12, 22, 23] and [25]. For the analysis of length-biased data, which may appear in observational studies where the observed samples are not randomly selected from the population of interest but with probability proportional to their length [24], recent research includes [3, 15, 16, 18, 19, 28]. [4] considered nonparametric analysis of bivariate left-truncated competing risks data. However, all these methods are not applicable in our study because of a different censoring and truncation framework.

The main challenge in our study comes from the correlations between T and R (referral is increasingly likely the closer a patient is to developing cirrhosis) and between C and L (the last diagnosis time for cirrhosis is usually close to the end of study recruitment). For analysing bivariate time-to-event data, some recent works include the frailty models in [30] for censored data and the new copula models under interval sampling in [32, 33]. In this paper, we model the correlation between R and T and the correlation between C and L through a non-parametric approach, instead of using parametric frailty or copula methods. We will propose a new likelihood estimation approach, where the parameter $\boldsymbol{\beta}$ and Λ_0 will be estimated simultaneously via an iteration algorithm. We will show that such an iterative algorithm will give a consistent estimate.

This paper is organized as follows. Section 2 introduces the notations and the likelihood function and develops the estimation algorithm. The large sample

properties of the estimates are also given in this section. Monte Carlo simulation studies and a real data analysis are provided in Section 3 and Section 4, respectively. Section 5 gives a discussion for further research work.

2 Methodology

2.1 Preliminaries

We follow the notations in the previous section. The cirrhosis time T with hazard rate function (1) is of our main interest. Throughout this paper, we assume that (L, C) is independent of the bivariate event times (R, T) and assume that $\boldsymbol{W}_i(t)$ is independent of (L, C), similar to [29]. Such an assumption is reasonable in this particular application, since the truncation time L (end of recruitment) is determined independently before the study and the censoring time C (last follow-up time) is usually a certain period after L. Therefore both L and C are not related to the individual information R, T and $\boldsymbol{W}$. Such an assumption is also necessary in the theoretical aspects. This is because if (L, C) and $\boldsymbol{W}$ are correlated then we would need to model their relations, say via a bivariate proportional hazard models for (L, C) and $\boldsymbol{W}$. However, this will be very challenging since (L, C) are under both truncation and censoring (see the discussion in Section 5). We will focus on model (1), the proportional hazard model for the univariate variable T, but use the information R to remove the selection bias (truncation bias).

We denote $X = \min\{T, C\}$ and $\delta = I(T \leq C)$. Subjects for the whole HCV-infected population are denoted by $[R, L, X, \delta, \boldsymbol{W}(t)]$. Since only patients with $R \leq L$ can be observed, we denote the ith observed data as $X_i = \min\{T_i, C_i\}$, $\delta_i = I(T_i \leq C_i)$, R_i, L_i and $\boldsymbol{W}_i(t)$. In our study, we also assume that $\Pr(R \leq T) = 1$, which means that a patient will be referred to hospital before he or she develops a

severe cirrhosis event [29]. Note that throughout this paper, we use letters without subscript i, such as X or L, representing an object in the whole population and letters with a subscript label, such as X_i or L_i, representing observations from the study sample (subject to truncation).

Define $\mathcal{F}_t^i = \sigma\{\boldsymbol{W}_i(u), u \leq t\}$ and $N_i(dt) = I(X_i \in dt, \delta_i = 1)$, the counting process of observed failures for the ith individual, and $H_i(t) = I(L_i \geq t)$. We use the short notation $\cdot \in dt$ for $\cdot \in [t, t+dt)$. Denote $\mathcal{F}_t$ as the filtration for a subject in the (unbiased) population and $G(t,s) = \Pr(L > t, C > s)$ as the joint survival function for (L, C). The truncation probability is denoted by $\gamma = \Pr(R \leq L)$. We assume that subjects with different covariates have the same truncation probability γ. Although this is a strong assumption, it is reasonable in this hepatitis cohort study. This is because HCV infected patients usually have no symptoms and referral to hospital may be due to a health check by chance which implies no causal relation between R and $\boldsymbol{W}$. Note that the method proposed in this paper can be easily extended to the more general framework, where the truncation probability depends on the covariate information $\mathcal{F}_\infty$, i.e., $\gamma_{\mathcal{F}} = \Pr(R \leq L|\mathcal{F}_\infty)$, since the truncation probabilities can be estimated via a simple univariate model.

Under model (1), we can show that (see Appendix A)

$$\frac{\mathrm{E}\{H_i(t)N_i(dt)|\mathcal{F}_{t-}^i\}}{G(t_-, t_-)} = \frac{1}{\gamma}\Lambda_i(dt)e^{-\Lambda_i(t_-)}, \tag{2}$$

where $\Lambda_i(t) = \int_0^t \exp[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}]\Lambda_0(ds)$. Define

$$A_i(dt|\boldsymbol{\beta}, \Lambda_0) = \gamma^{-1}G(t_-, t_-)\Lambda_i(dt)e^{-\Lambda_i(t_-)}, \;\; A_i(t|\boldsymbol{\beta}, \Lambda_0) = \int_0^t A_i(ds|\boldsymbol{\beta}, \Lambda_0),$$
$$\widetilde{N}_i(dt) = H_i(t)N_i(dt), \;\; \widetilde{N}_i(t) = \int_0^t \widetilde{N}_i(ds). \tag{3}$$

In the right-hand side of equation (2) the term $\Lambda_i(dt)e^{-\Lambda_i(t_-)}$ is actually the probability that T_i occurs at the interval $(t, t+dt]$. We call it the *mean rate function* here, rather than probability density function, to emphasize its link with the hazard rate function.

Note that model (2) involves the observed data (X_i, δ_i, L_i) (i.e., the process $\widetilde{N}_i(t)$) and $\boldsymbol{W}_i$ (i.e., $\mathcal{F}_t^i$). If G and γ are given, model (2) does not involve R_i. This is because our main interest is to study the marginal hazard rate of T and its relation with covariate $\boldsymbol{W}$. The referral time R is mainly treated as a confounding factor (not having causal effects on T) and it leads to the selection bias due to truncation. Therefore the information of R is used only when we estimate G and γ to remove the truncation bias in the analysis.

We view $\{d\widetilde{N}_i(t), 0 \le t < \infty, i = 1, \ldots, n\}$ as the observed data. If the joint survival function G and the truncation probability γ are known, then the 0-1 valued Bernoulli random variable $d\widetilde{N}_i(t)$ has probability of being 1 as $\mathrm{E}\{\widetilde{N}_i(dt)|\mathcal{F}_{t_-}^i\} = A_i(dt|\boldsymbol{\beta}, \Lambda_0)$. Following [1] we can write the likelihood for $(\boldsymbol{\beta}, \Lambda_0)$ given (G, γ) as

$$\begin{aligned} & \mathcal{L}_n := \mathcal{L}_n(\boldsymbol{\beta}, \Lambda_0; G, \gamma) \\ = & \prod_{0 \le s < \infty} \left\{ \prod_i \left[A_i(ds|\boldsymbol{\beta}, \Lambda_0)\right]^{\widetilde{N}_i(ds)} \cdot \left[1 - \sum_i A_i(ds|\boldsymbol{\beta}, \Lambda_0)\right]^{1-\sum_i \widetilde{N}_i(ds)} \right\}, \end{aligned}$$

which can be further written as, using the product integration theory in [13],

$$
\begin{aligned}
& \mathcal{L}_n := \mathcal{L}_n(\boldsymbol{\beta}, \Lambda_0; G, \gamma) \qquad (4)\\
= & \prod_{0\le s<\infty} \left\{ \prod_i \left[A_i(ds|\boldsymbol{\beta}, \Lambda_0)\right]^{\widetilde{N}_i(ds)} \right\} \exp\left[-\sum_{i=1}^{n} \int_0^\infty A_i(ds|\boldsymbol{\beta}, \Lambda_0) \right] \\
= & \prod_{0\le s<\infty} \left\{ \left[\frac{G(s_-, s_-)}{\gamma} \Lambda_0(ds) \right]^{\sum_i \widetilde{N}_i(ds)} \right\} \prod_{0\le s<\infty} \left\{ \prod_i \left[\exp(\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-)) \right]^{\widetilde{N}_i(ds)} \right\} \\
& \cdot \exp\left\{ -\int_0^\infty \left[\sum_{i=1}^{n} \exp(\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-)) \right] \frac{G(s_-, s_-)}{\gamma} \Lambda_0(ds) \right\}.
\end{aligned}
$$

Equation (4) is the full likelihood for $\boldsymbol{\beta}$ and Λ_0, if G and γ are treated as known. This likelihood function does not involve R_i, since the information of R_i is not needed if G and γ are given. We will demonstrate in later sections that R_i will only be used when we estimate G and γ.

2.2 A formula for Λ_0

In this subsection, we provide a very important formula for Λ_0, which will be used to motivate the new estimation method. In practice, to estimate Λ_0 we can use a piece-wise right-continuous function, which only has positive jumps when $\sum_i \widetilde{N}_i(s)$ has jumps. This is reasonable and necessary since from (4) we know that $\Lambda_0(ds)$ is only identifiable when $\sum_i \widetilde{N}_i(ds)$ is not 0. Therefore we can treat Λ_0 as a function with only a discrete number of parameters (the sizes of its jumps) to be estimated. Given any fixed $\boldsymbol{\beta}, \gamma, G$, the estimating equations for $\Lambda_0(ds)$ is $\frac{\partial \ln \mathcal{L}_n}{\partial \Lambda_0(ds)} = 0$. Using the result, $\frac{\partial \Lambda_i(u)}{\partial \Lambda_0(ds)} = \exp\left[\boldsymbol{W}_i(s_-)^\top \boldsymbol{\beta}\right]$ for $s \le u$, and from the

estimating equations $\frac{\partial \ln \mathcal{L}_n}{\partial \Lambda_0(ds)} = 0$, we further have

$$\begin{aligned}
\Lambda_0(ds) &= \frac{n^{-1}\sum_i \widetilde{N}_i(ds)}{\boldsymbol{Q}(\boldsymbol{\beta}, \Lambda_0, s) + \boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \Lambda_0, s)\frac{G(s_-,s_-)}{\gamma}}, \qquad (5) \\
\boldsymbol{Q}(\boldsymbol{\beta}, \Lambda_0, s) &= n^{-1}\int_s^\infty \sum_{i=1}^n \exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}\right]\widetilde{N}_i(du) - \int_s^\infty \boldsymbol{R}(\boldsymbol{\beta}, \Lambda_0, s, u)\frac{G(u_-,u_-)}{\gamma}\Lambda_0(du). \\
&= n^{-1}\sum_{i=1}^n \int_s^\infty \exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}\right]\left[\widetilde{N}_i(du) - A_i(du|\boldsymbol{\beta}, \Lambda_0)\right],
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \Lambda_0, s) &= n^{-1}\sum_{i=1}^n \exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-)\right], \\
\boldsymbol{R}(\boldsymbol{\beta}, \Lambda_0, s, u) &= n^{-1}\sum_{i=1}^n \exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta})\right]\exp\left[\boldsymbol{W}_i^\top(u_-)\boldsymbol{\beta} - \Lambda_i(u_-)\right].
\end{aligned}$$

Given $\boldsymbol{\beta}, \gamma, G$, the above equation (5) for Λ_0 can be solved numerically using recursive algorithms. Such iterative algorithm, however, is not computationally efficient. Therefore, for simplicity, we consider the following approximation formula for $\Lambda_0(ds)$,

$$\Lambda_0(ds) \approx \frac{n^{-1}\sum_i \widetilde{N}_i(ds)}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \Lambda_0, s)\frac{G(s-,s-)}{\gamma}}, \qquad (6)$$

since $\mathrm{E}\left\{\left[\tilde{N}_i(du) - A_i(du|\boldsymbol{\beta}, \Lambda_0)\right]|\mathcal{F}_{u_-}^i\right\} = 0$ and further $\boldsymbol{Q}(\boldsymbol{\beta}, \Lambda_0, s)$ converges to 0.

2.3 Estimation for $\boldsymbol{\beta}$ and Λ_0 and large sample properties.

Equation (6) motivates us to consider the following simple working-likelihood function. Replace Λ_0 in (4) with (6), the likelihood function $\mathcal{L}_n$ becomes

$$\tilde{\mathcal{L}}_n(\boldsymbol{\beta}, \Lambda_0) \;=\; \prod_{0\le s<\infty} \left\{ \prod_i \left[\frac{\exp\left(\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-)\right)}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \Lambda_0, s)} \right]^{\widetilde{N}_i(ds)} \right\}.$$

This implies that we can consider the following log-likelihood function

$$\ln \tilde{\mathcal{L}}_n(\boldsymbol{\beta}, \Lambda_0) = n^{-1} \sum_i \int_0^\infty \left\{ \left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-) \right] - \ln \left[\boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \Lambda_0, s) \right] \right\} \widetilde{N}_i(ds).$$

Then we have the following theorem. Proof is provided in Appendix B.

Theorem 2.1. *The estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\Lambda}_0$, which maximize $\log \tilde{\mathcal{L}}_n(\boldsymbol{\beta}, \Lambda_0)$, are consistent, i.e., they converge to the true parameter values, $\boldsymbol{\beta}^*$ and Λ_0^*, in probability.*

In practice, it is computationally inefficient to calculate $\tilde{\boldsymbol{\beta}}$ and $\tilde{\Lambda}_0$ by directly maximizing $\tilde{\mathcal{L}}_n$. Instead, we can find the estimates for $\boldsymbol{\beta}$ and Λ_0 via iteration algorithms. Given Λ_0, the estimate for $\boldsymbol{\beta}$ can be found by solving the following estimating equations (by taking partial derivatives for $\log \tilde{\mathcal{L}}_n$ over $\boldsymbol{\beta}$),

$$\boldsymbol{U}_n(\boldsymbol{\beta}, \Lambda_0) = n^{-1} \sum_{i=1}^n \int_0^\infty \left[\boldsymbol{W}_i(s_-) - \boldsymbol{\Lambda}_i^{(1)}(s_-) - \mathrm{E}^{(1)}(\boldsymbol{\beta}, \Lambda_0, s) \right] \widetilde{N}_i(ds) = \mathbf{0}, \quad (7)$$

where $\boldsymbol{\Lambda}_i^{(1)}(s) = \frac{\partial \Lambda_i(s)}{\partial \boldsymbol{\beta}} = \int_0^s \boldsymbol{W}_i(u_-) \exp\left[\boldsymbol{W}_i^\top(u_-)\boldsymbol{\beta}\right] \Lambda_0(du)$ and

$$\begin{aligned}
\mathrm{E}^{(1)}(\boldsymbol{\beta}, \Lambda_0, s) &= \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \Lambda_0, s)}{S^{(0)}(\boldsymbol{\beta}, \Lambda_0, s)}, \\
\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \Lambda_0, s) &= \frac{\partial S^{(0)}(\boldsymbol{\beta}, \Lambda_0, s)}{\partial \boldsymbol{\beta}} \\
&= n^{-1} \sum_{i=1}^{n} \left[\boldsymbol{W}_i(s_-) - \boldsymbol{\Lambda}_i^{(1)}(s_-)\right] \exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta} - \Lambda_i(s_-)\right].
\end{aligned}$$

Given $\boldsymbol{\beta}$, for simplicity and motivated by equation (6) we can find an estimate $\hat{\Lambda}_0$ by solving the following recursive formula

$$\hat{\Lambda}_0(ds) = \frac{n^{-1} \sum_i \widetilde{N}_i(ds)}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}, \hat{\Lambda}_0, s) \frac{\hat{G}(s-, s-)}{\hat{\gamma}}}. \tag{8}$$

Here $\hat{G}$ and $\hat{\gamma}$ are consistent estimates for the true parameters G^* and γ^*. Therefore we consider the following Algorithm 1 in practice.

Algorithm 1 Algorithm for estimating Λ_0 and $\boldsymbol{\beta}$.

Given consistent estimates $\hat{G}$ and $\hat{\gamma}$, with starting point $\hat{\boldsymbol{\beta}}^{[0]}, \hat{\Lambda}_0^{[0]}$
repeat
  Given $\hat{\boldsymbol{\beta}}^{[m]}$, calculate $\hat{\Lambda}_0^{[m+1]}$ by solving (8)
  Given $\hat{\Lambda}_0^{[m+1]}$, solve equation (7) to get $\hat{\boldsymbol{\beta}}^{[m+1]}$
until $\hat{\Lambda}_0^{[m]}$ and $\hat{\boldsymbol{\beta}}^{[m]}$ converge to some $\hat{\Lambda}_0$ and $\hat{\boldsymbol{\beta}}$
Output $\hat{\Lambda}_0$ and $\hat{\boldsymbol{\beta}}$ as the estimate

The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0$, obtained from Algorithm 1, are asymptotically equivalent to solve equations (7) and (8) simultaneously. Such estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0$ are equivalent to $\tilde{\boldsymbol{\beta}}$ and $\tilde{\Lambda}_0$ (given in Theorem 2.1) for large enough n, since the pair $(\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_0)$ asymptotically satisfies equations (7) and (8) as well. Therefore we have the following proposition.

Proposition 2.1. *Suppose that $\hat{G}$ and $\hat{\gamma}$ are consistent to the true parameters G^* and γ^*. The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0$, obtained from Algorithm 1, are consistent, i.e.,*

they converge to the true parameter values, $\boldsymbol{\beta}^*$ and Λ_0^*, in probability.

Note that consistent estimates $\hat{G}$ and $\hat{\gamma}$ can be obtained via the method in [6].

Theorem 2.2. *The asymptotic distribution for $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$ is given by*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*)).$$

The formula for $\boldsymbol{\Psi}(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*) = \lim_{n\to\infty} \boldsymbol{\Psi}_n(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*)$ is given by equation (13).

The asymptotic distribution for $\sqrt{n}\left[\hat{\Lambda}_0(t) - \Lambda_0^*(t)\right]$ is given by

$$\sqrt{n}\left[\hat{\Lambda}_0(t) - \Lambda_0^*(t)\right] \xrightarrow{d} \mathcal{N}(0, \sigma_\Lambda^2(t)),$$

for some function $\sigma_\Lambda^2(t)$.

An estimate for $\boldsymbol{\Psi}(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*)$ is given by $\boldsymbol{\Psi}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, \hat{G}, \hat{\gamma})$ and an estimate for $\sigma_\Lambda^2(t)$ is given by (details in the supplementary file)

$$\begin{aligned}
\hat{\sigma}_\Lambda^2(t) \;=\; & \left[\int_0^t \frac{\boldsymbol{S}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)}{S^{(0)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)} \hat{\Lambda}_0(ds)\right]^\top \boldsymbol{\Psi}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, \hat{G}, \hat{\gamma}) \left[\int_0^t \frac{\boldsymbol{S}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)}{S^{(0)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)} \hat{\Lambda}_0(ds)\right] \\
& + \int_0^t \left[\frac{1}{S^{(0)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)\hat{G}(s_-, s_-)\hat{\gamma}^{-1}}\right] \hat{\Lambda}_0(ds) \qquad (9) \\
& + 2\left[\int_0^t \frac{\boldsymbol{S}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)}{S^{(0)}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, s)} \hat{\Lambda}_0(ds)\right]^\top \boldsymbol{I}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0)^{-1} \boldsymbol{\Phi}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0, \hat{G}, \hat{\gamma}),
\end{aligned}$$

where

$$\begin{aligned}
& \boldsymbol{\Phi}_n(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*) \\
= \; & n^{-1} \sum_{i=1}^n \int_0^\infty \left[n^{-1} \sum_i \int_s^\infty \boldsymbol{J}_i(\boldsymbol{\beta}^*, \Lambda_0^*, s, u) \widetilde{N}_i(du)\right] \frac{1}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}^*, \Lambda_0^*, s) G^*(s_-, s_-) \gamma^{*-1}} \Lambda_0^*(ds),
\end{aligned}$$

with $\boldsymbol{J}_i$ and $\boldsymbol{I}_n$ given by equations (11) and (12) respectively.

3 Monte Carlo simulations

In this section, we evaluate the finite sample performance of the proposed methods via Monte Carlo simulations. We choose a small sample size $n = 100$ and a larger sample size $n = 300$. All simulation results are based on 500 replicated simulations. Truncation times L and censoring times C are generated respectively from $C = av_1 + bv_2$ and $L = cv_1 + dv_2 + h \cdot U[0, 1]$, where v_1 and v_2 are standard exponentially random variables, and we can adjust the censoring/truncation probabilities via changing the values of (a, b, c, d, h). We considered two simulation scenarios, corresponding to different correlation levels of T and R.

3.1 Scenario 1

We consider a time varying covariate $W_1(t) = W_1 + 0.1t$ and time-independent covariate W_2 and W_3, where $W_1 \sim U(0, 1)$ and $W_2 \sim$ Bernoulli(0.5), $W_3 \sim$ Bernoulli(0.5). The baseline hazard rate function is $\Lambda_0^*(dt) = \lambda_0^*(t)dt = \exp(\alpha t)dt$, with $\alpha = 0.3$. The survival times R and T are generated as follows: $R = e\xi$ and $T = \frac{1}{(\alpha+\beta_1^*)} \log[(\alpha + \beta_1^*)\xi \exp(-\boldsymbol{W}^\top \boldsymbol{\beta}^*) + 1]$, where ξ follows standard exponential distribution. Note that such T and R are correlated and such T has hazard rate function $\Lambda^*(dt) = \exp(\alpha t)dt \exp(W_1(t)\beta_1^* + W_2\beta_2^* + W_3\beta_3^*)$. The constant e, reflecting the correlation level between R and T, is chosen as $e = 0.8$. This gives a 0.9-correlation between T and R and such a strong correlation is very similar to the real data analysis provided in Section 4. We choose $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \beta_3^*)^\top = (0.2, 0.4, 0.3)^\top$.

The simulation results are presented in Table 1 ($n = 300$) and Table 2 ($n = 100$). In the tables, the empirical estimates, empirical standard deviation ($\hat{s}_\beta$),

and the average standard deviation estimates ($\hat{\sigma}_\beta$) for the proposed estimates of $\hat{\boldsymbol{\beta}}$ are reported. The tables also present the coverage probabilities of 95% confidence intervals constructed based on the normal approximation. As shown in Table 1, the estimate $\hat{\boldsymbol{\beta}}$ based on the new method are virtually unbiased and that the variance estimates agree with the empirical variances quite well. The coverage probabilities also show that the variance estimates for $\hat{\boldsymbol{\beta}}$ work well. Table 1 also provides the results based on the simple Cox regression analysis without considering truncation, which gives severely biased estimates. All the estimates in the rows (i), based on the new method, have smaller bias than the results in the rows (ii), based on standard Cox model analysis. The standard error estimates of the new method are larger than the results based on the standard Cox models. This is because the new method uses $\hat{G}$, the joint survival function estimate for (L, C), to adjust the truncation bias and this introduces extra variation to the estimates. When we decrease from $n = 300$ to $n = 100$, the results (provided in Table 2) have similar patterns, i.e., the estimates of the proposed method dominate the estimates based on simple Cox model results.

Figure 1 displays the estimate of the baseline hazard rate function, along with their 95 percent point-wise confidence bands. The light-colored (yellow) solid line is the estimated hazard rate function, which match almost exactly the dark-colored (red) solid line (the true hazard rate function), except for the part of the tail. The green dotted line is the mean curve of the estimated 95% confidence interval for the baseline hazard rate, which is very close to the Monte Carlo estimate of the 95% confidence interval for the baseline hazard rate. These results indicate good performance of the proposed method with moderate sample size.

Note that, the coverage probability of the standard Cox's regression model in Table 1 (sample size $n = 300$) is much smaller than that in Table 2 (sample size

$n = 100$). This is reasonable since the standard Cox's regression analysis actually uses the biased data set (without considering truncation) and therefore gives a biased estimate. Therefore, the larger the sample size, the more significant bias of the estimate (the smaller of the coverage probability).

Table 1: Summary of simulation results for Scenario 1, $n = 300$; (i) results based on the new method; (ii) results based on standard Cox model without considering truncation.

$n = 300$			Cens.%= 15				Cens.%= 40			
	$\boldsymbol{\beta}$		$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*	$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*
$\gamma = 0.4$	0.2	(i)	0.228	0.152	0.169	0.96	0.223	0.155	0.161	0.96
		(ii)	0.277	0.122	0.126	0.91	0.262	0.137	0.144	0.93
	0.4	(i)	0.453	0.164	0.173	0.96	0.388	0.167	0.167	0.97
		(ii)	0.571	0.126	0.129	0.75	0.554	0.144	0.145	0.82
	0.3	(i)	0.318	0.162	0.175	0.96	0.288	0.164	0.178	0.96
		(ii)	0.428	0.121	0.127	0.85	0.410	0.134	0.144	0.91
$\gamma = 0.6$	0.2	(i)	0.204	0.154	0.157	0.96	0.190	0.176	0.179	0.95
		(ii)	0.270	0.121	0.126	0.93	0.240	0.160	0.153	0.94
	0.4	(i)	0.433	0.171	0.168	0.95	0.388	0.185	0.188	0.93
		(ii)	0.536	0.128	0.128	0.83	0.474	0.151	0.154	0.92
	0.3	(i)	0.324	0.163	0.161	0.95	0.291	0.168	0.179	0.95
		(ii)	0.394	0.123	0.127	0.90	0.347	0.151	0.153	0.94

3.2 Scenario 2

In the second simulation scenario, we consider the same time varying covariate $W_1(t)$, time-independent covariate W_2 and W_3 and the same baseline hazard rate function, as those in the previous section. The survival times R and T are generated as follows, under a different correlation structure from that in Scenario 1: $T = \frac{1}{(\alpha+\beta^*)} \ln[(\alpha + \beta_1^*)\xi \exp(-\boldsymbol{W}^\top \boldsymbol{\beta}^*) + 1]$, where ξ follows the standard exponential distribution and R is drawn from a gamma$(3\xi, 5)$. The correlation of R and T is about 0.55, which represents a much weaker correlation between R and T than that in Scenario 1. We choose the same true parameter values as that in

Table 2: Summary of simulation results for Scenario 1, $n = 100$; (i) results based on the new method; (ii) results based on standard Cox model without considering truncation.

$n = 100$			Cens.%= 15				Cens.%= 40			
	$\boldsymbol{\beta}$		$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*	$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*
$\gamma = 0.4$	0.2	(i)	0.233	0.275	0.289	0.96	0.218	0.282	0.289	0.97
		(ii)	0.285	0.228	0.222	0.94	0.262	0.253	0.251	0.93
	0.4	(i)	0.433	0.280	0.285	0.96	0.387	0.291	0.298	0.96
		(ii)	0.559	0.237	0.228	0.89	0.548	0.270	0.258	0.90
	0.3	(i)	0.320	0.293	0.308	0.96	0.291	0.284	0.281	0.96
		(ii)	0.418	0.232	0.226	0.92	0.397	0.269	0.257	0.92
$\gamma = 0.6$	0.2	(i)	0.233	0.261	0.268	0.96	0.183	0.283	0.293	0.96
		(ii)	0.274	0.228	0.221	0.93	0.248	0.269	0.266	0.95
	0.4	(i)	0.450	0.272	0.285	0.96	0.389	0.302	0.312	0.95
		(ii)	0.544	0.239	0.227	0.91	0.493	0.282	0.273	0.95
	0.3	(i)	0.334	0.271	0.287	0.96	0.279	0.304	0.314	0.96
		(ii)	0.399	0.234	0.226	0.91	0.347	0.285	0.272	0.93

Scenario 1. The simulation results are presented in Table 3 and Table 4.

As what we discovered in Scenario 1, all the estimates in the rows (i), based on the new method, have smaller bias than the results in the rows (ii), based on standard Cox model analysis. However, in Scenario 2, the bias of the estimates based on the standard Cox model is not as severe as those in Scenario 1. For example in Table 1, with 40% censoring and 40% truncation, the estimated values for the standard Cox regression analysis $\hat{\boldsymbol{\beta}} = (0.262, 0.554, 0.410)^\top$ are severely biased comparing to the true value $(0.2, 0.4, 0.3)^\top$. On the other hand, in Table 3, with 40% censoring and 40% truncation, the standard Cox regression estimates $\hat{\boldsymbol{\beta}} = (0.227, 0.464, 0.350)^\top$ have much less bias although they are still not good enough. The reason for this is that the selection bias of T comes from the truncation of R. If R and T are highly correlated, the truncation on R will lead to severely biased sample for T; if the correlation of R and T is not high, the truncation on R will have less impact on T. On the contrary, the new method can deal

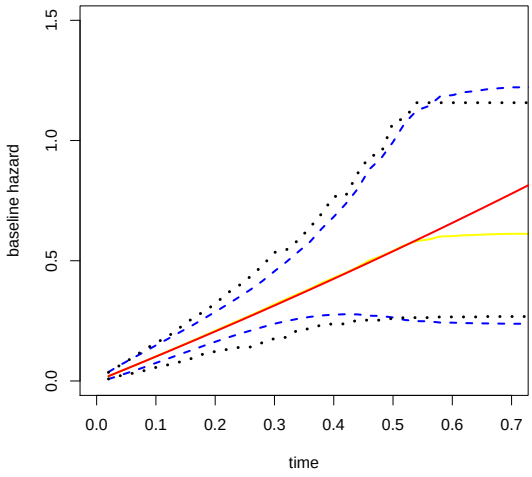

Figure 1: Baseline hazard estimation. True baseline function: the dark (red color) solid curve; the estimated baseline function: the light (yellow color) solid curve; the mean of estimated 95% confidence intervals: the dotted curves; the 95% confidence interval for the baseline hazard rate estimates: the dashed curves

with the truncation on R and it give more consistent results for both scenarios with either small correlation or large correlation of R an T.

4 A real example

We apply the proposed method to the Edinburgh hepatitis C data previously studied in [11]. The aim of our study was to determine how the progression risk (hazard rate) of the cirrhosis event is affected by the three risk factors: age at infection, HIV co-infection (yes:1 or no:0) and heavy alcohol consumption (yes:1 or no:0). The cirrhosis event T may be censored and the correlated referral event is right-truncated.

In an earlier paper, we proposed an accelerated failure time model for censored survival data under referral bias for this application [29], where Schoenfeld residual analysis [21] indicated that the proportional hazard assumption is likely to

Table 3: Summary of simulation results for Scenario 2, $n = 300$; (i) results based on the new method; (ii) results based on standard Cox model without considering truncation.

$n = 300$			Cens.%= 15				Cens.%= 40			
	$\boldsymbol{\beta}$		$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*	$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*
$\gamma = 0.4$	0.2	(i)	0.209	0.133	0.141	0.95	0.217	0.165	0.169	0.95
		(ii)	0.233	0.121	0.125	0.93	0.227	0.148	0.144	0.95
	0.4	(i)	0.398	0.160	0.167	0.96	0.397	0.167	0.173	0.95
		(ii)	0.467	0.133	0.127	0.90	0.464	0.149	0.145	0.93
	0.3	(i)	0.298	0.146	0.148	0.96	0.298	0.163	0.169	0.96
		(ii)	0.356	0.131	0.126	0.93	0.350	0.149	0.145	0.93
$\gamma = 0.6$	0.2	(i)	0.211	0.164	0.161	0.95	0.214	0.165	0.172	0.95
		(ii)	0.236	0.126	0.127	0.93	0.224	0.151	0.154	0.95
	0.4	(i)	0.407	0.170	0.178	0.95	0.401	0.173	0.175	0.95
		(ii)	0.462	0.122	0.129	0.94	0.449	0.148	0.155	0.94
	0.3	(i)	0.314	0.179	0.181	0.95	0.306	0.176	0.188	0.96
		(ii)	0.356	0.2129	0.128	0.92	0.348	0.160	0.154	0.93

Table 4: Summary of simulation results for Scenario 2, $n = 100$; (i) results based on the new method; (ii) results based on standard Cox model without considering truncation.

$n = 100$			Cens.%= 15				Cens.%= 40			
	$\boldsymbol{\beta}$		$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*	$\hat{\boldsymbol{\beta}}$	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	*CP.*
$\gamma = 0.4$	0.2	(i)	0.204	0.234	0.246	0.96	0.187	0.275	0.279	0.95
		(ii)	0.238	0.230	0.221	0.96	0.242	0.256	0.253	0.94
	0.4	(i)	0.369	0.258	0.263	0.96	0.389	0.288	0.283	0.96
		(ii)	0.465	0.244	0.225	0.93	0.467	0.275	0.258	0.94
	0.3	(i)	0.278	0.249	0.241	0.96	0.284	0.285	0.279	0.96
		(ii)	0.362	0.228	0.224	0.94	0.359	0.263	0.258	0.94
$\gamma = 0.6$	0.2	(i)	0.217	0.244	0.257	0.96	0.173	0.300	0.319	0.95
		(ii)	0.247	0.240	0.224	0.94	0.246	0.283	0.269	0.95
	0.4	(i)	0.410	0.250	0.256	0.95	0.370	0.312	0.316	0.94
		(ii)	0.463	0.236	0.229	0.94	0.460	0.290	0.275	0.95
	0.3	(i)	0.307	0.251	0.269	0.95	0.285	0.297	0.305	0.94
		(ii)	0.336	0.237	0.228	0.93	0.326	0.284	0.275	0.95

be violated for heavy alcohol consumption; also see Figure 2. We therefore fit the proposed proportional hazards model (1) to the data with $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$, corresponding to the covariates (age W_1, HIV coinfection W_2, heavy alcohol consumption W_3, and the interaction of W_3 and time), in order to guarantee the validation of the proportional hazards assumption for 'heavy alcohol consumption'. This proportional hazards model with a time-dependent covariate (interaction term) allows that the effect that heavy alcohol consumption may change over time. When the model includes the interaction term, the proportional hazards assumption becomes reasonable. This can be seen from the Schoenfeld residual plots presented in Figure 3, which implied that both β_3 (for heavy alcohol consumption) and β_4 (for the interaction of alcohol with time) are almost constant over time.

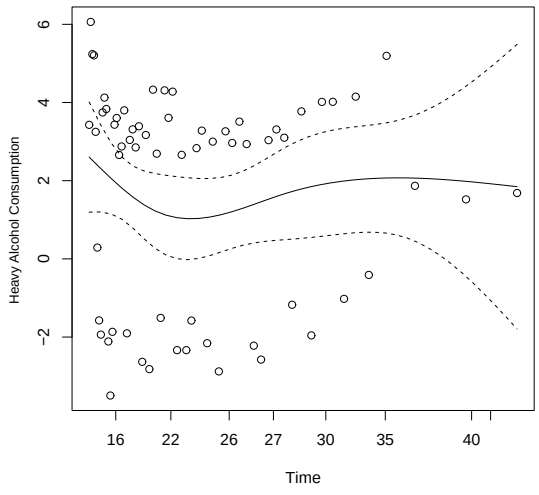

Figure 2: Schoenfeld residual plot for alcohol, based on analysis without interaction term.

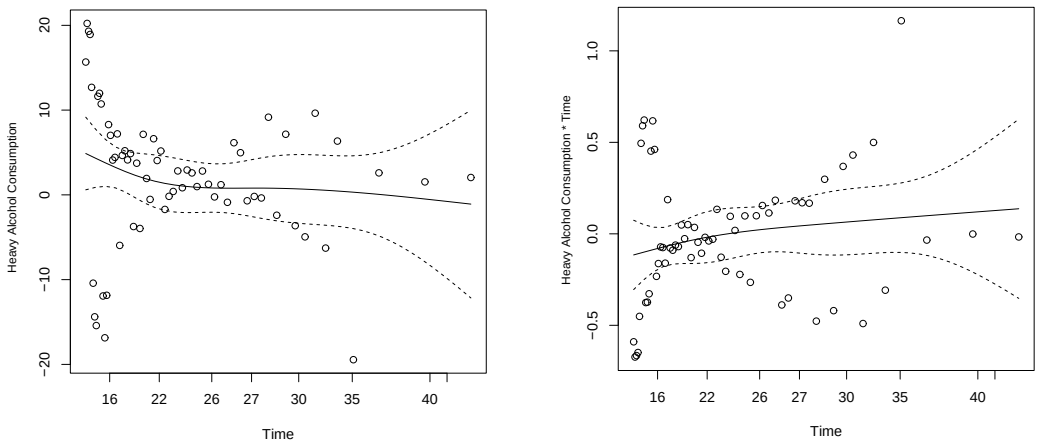

Figure 3: Results based on the model with interaction term. Left plot: Schoenfeld residual plot for alcohol; right plot: Schoenfeld residual plot for the interaction of alcohol and time from infection.

Table 5 summarizes the estimates of regression parameters obtained from our method. The results from the truncated model, where the referral bias is considered, show that age at infection (β_1), HIV co-infection (β_2) and heavy alcohol in-take (β_3) are significantly identified as risk factors associated with more rapid disease progression to cirrhosis. The interaction effect of heavy alcohol assumption and time has a negative estimate -0.043, but it is not significant. It implied that the effect of heavy alcohol consumption is slightly decreasing during the disease progression although the change is not significant.

Table 5: Summary of Data analysis

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Parameter	$\hat{\boldsymbol{\beta}}$	0.074	1.646	3.400	-0.043
	se	0.014	0.839	1.609	0.072

5 Discussion

Under the proportional hazards model framework, we presented a likelihood procedure for the estimation of the coefficients for time-varying covariates and the baseline hazard function under referral bias. The challenge in our study is the referral bias, which is due to a random right-truncation on a time event R which is highly correlated to the time event T of main interest. The proposed method deals with the truncation bias via a nonparametric approach. The asymptotic results are developed and our simulation studies indicate that the proposed method has good statistical properties.

We applied the new methods to a real HCV cohort study where traditional analysis without taking into account referral bias often gave biased estimate for the effects of risk factors (see [10, 11]). Such referral bias often happens in chronic disease epidemiology cohort studies, where entry into the cohort is dependent on a subject's progression to an event of interest and patients are more likely to be referred to specialist clinics at later stages of disease (see [11]).

This paper only focused on dealing with the analysis for the main survival events. In such long-term epidemiological studies, repeated outcome measurements for each subject may also be available and of interest. It would be interesting to extend the method developed here to a joint modelling framework for survival events and longitudinal measurements. In addition, because chronic disease epidemiological studies usually last a long period of time, certain time-independent variables may have time-varying effects. It is also of interest to develop new methods to deal with time-varying coefficients [26]. These are left to future work. Another possible future research work is to extend the proposed methodology to handle other types of sampling bias, for example the length-biased data [24].

Another possible future research work is to consider (R, T) and (L, C) are conditional independent given $\boldsymbol{W}$, for appropriate application problems. Because that (R, T) and (L, C) are truncated or censored by each other, we need to model the relation of (R, T) and $\boldsymbol{W}$ and the relation of (L, C) and $\boldsymbol{W}$ simultaneously. When using the inverse probability weighted method to remove the truncation bias, the joint distribution G of (L, C) will depend on $\boldsymbol{W}$. This will raise further challenges in the iteration algorithms. In addition, the truncation probability will depends on $\boldsymbol{W}$, i.e., $\gamma = \Pr(R \leq L|\boldsymbol{W})$. This truncation probability will depends on how (R, T) and (L, C) are modelled with respect to $\boldsymbol{W}$. New methods need to be develop to estimate such truncation probabilities. However, if (L, C) is independent of $\boldsymbol{W}$, the truncation probability can be estimated based on the formula, $\gamma = \Pr(R \leq L|\boldsymbol{W}) = \int_0^\infty \Pr(R \leq s|\boldsymbol{W})F_L(ds)$, where $F_L(s) = \Pr(L \leq s)$, and the proposed method in this paper can be easily extended to such cases.

A Proof of equation (2)

Proof. Because of the assumption $\Pr(R \leq T) = 1$, we have that $\{T \in dt, L \geq t\}$ implies $R \leq L$. Then it follows that

$$\begin{aligned}
\frac{\mathrm{E}\left\{H_i(t)N_i(dt)|\mathcal{F}_{t-}^i\right\}}{G(t_-, t_-)} &= \frac{\Pr(T_i \in dt, C_i \geq T_i, L_i \geq t|\mathcal{F}_{t-}^i)}{G(t_-, t_-)} \\
&= \frac{\Pr(T \in dt, C \geq t, L \geq t|R \leq L, \mathcal{F}_{t-}^i)}{G(t_-, t_-)} \\
&= \frac{\Pr(T \in dt, C \geq t, L \geq t|\mathcal{F}_{t-}^i)}{\Pr(R \leq L)G(t_-, t_-)}.
\end{aligned}$$

Since we denote $\gamma = \Pr(R \leq L)$, we have

$$\frac{\mathrm{E}\left\{H_i(t)N_i(dt)|\mathcal{F}_{t-}^i\right\}}{G(t_-, t_-)} = \frac{\Pr(T \in dt, C \geq t, L \geq t|\mathcal{F}_{t-}^i)}{\gamma G(t_-, t_-)}.$$

From the independency assumption of $(R, T, \boldsymbol{W})$ and (L, C), we have

$$\frac{\mathrm{E}\left\{H_i(t)N_i(dt)|\mathcal{F}^i_{t_-}\right\}}{G(t_-, t_-)} = \frac{1}{\gamma}\Pr(T \in dt|\mathcal{F}^i_{t_-}) = \frac{1}{\gamma}\Lambda_i(dt)e^{-\Lambda_i(t_-)}.$$

□

B Proof of Theorem 2.1

Before proceeding to the proof of Theorem 2.1, we introduce the following lemma.

Lemma B.1. *For any positive values $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ we have*

$$n^{-1}\sum_{i=1}^{n} \ln\left(\frac{x_i}{y_i}\right) \cdot y_i \leq \ln\left(\frac{\bar{x}}{\bar{y}}\right)\bar{y}.$$

Proof. It is straightforward to prove the lemma, if we view $y_i/\sum_j y_j, i = 1, \ldots, n$ as a probability distribution for the values $z_i = x_i/y_i, i = 1, \ldots, n$. This is because

$$\mathrm{E}\left\{\ln(z)\right\} = \sum_{i=1}^{n} \ln\left(\frac{x_i}{y_i}\right) \cdot \frac{y_i}{\sum_i y_i} \leq \log\left[\mathrm{E}\{z\}\right] = \ln\left(\sum_i \frac{x_i}{y_i}\frac{y_i}{\sum_i y_i}\right) = \ln\left(\frac{\bar{x}}{\bar{y}}\right),$$

which implies that the lemma is true. □

Now we prove Theorem 2.1. Following the notations in previous sections, we denote Λ_0^*, $\boldsymbol{\beta}^*$, G^* and γ^* as the true parameter values. Denote $\Lambda_i^*(s) =$

$\int_0^s e^{\boldsymbol{W}_i^\top(u_-)\boldsymbol{\beta}^*}\Lambda_0^*(du)$. Define

$$\begin{aligned}\mathcal{Z}_n(\boldsymbol{\beta},\Lambda_0,t) &= n^{-1}\sum_{i=1}^n\int_0^t \boldsymbol{W}_i^\top(s_-)(\boldsymbol{\beta}-\boldsymbol{\beta}^*)A_i(ds|\boldsymbol{\beta}^*,\Lambda_0^*)\\ &\quad -n^{-1}\sum_{i=1}^n\int_0^t\left[\Lambda_i(s_-)-\Lambda_i^*(s_-)\right]A_i(ds|\boldsymbol{\beta}^*,\Lambda_0^*)\\ &\quad -n^{-1}\int_0^t\ln\left[\frac{S^{(0)}(\boldsymbol{\beta},\Lambda_0,s)}{S^{(0)}(\boldsymbol{\beta}^*,\Lambda_0^*,s)}\right]\sum_i A_i(ds|\boldsymbol{\beta}^*,\Lambda_0^*).\end{aligned}$$

We can rewrite the above formula as

$$\begin{aligned}&\mathcal{Z}_n(\boldsymbol{\beta},\Lambda_0,t)\\ =\;& n^{-1}\sum_{i=1}^n\int_0^t\ln\left\{\frac{\exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}-\Lambda_i(s_-)\right]}{\exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^*-\Lambda_i^*(s_-)\right]}\right\}\exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^*-\Lambda_i^*(s_-)\right]\frac{G^*(s_-,s_-)}{\gamma^*}\Lambda_0^*(ds)\\ &-\int_0^t\ln\left[\frac{n^{-1}\sum_{i=1}^n e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}-\Lambda_i(s_-)}}{n^{-1}\sum_{i=1}^n e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^*-\Lambda_i^*(s_-)}}\right]\left[n^{-1}\sum_{i=1}^n e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^*-\Lambda_i^*(s_-)}\right]\frac{G^*(s_-,s_-)}{\gamma^*}\Lambda_0^*(ds).\end{aligned}$$

Then with Lemma B.1, we have that $\mathcal{Z}_n(\boldsymbol{\beta},\Lambda_0,ds)\le 0$ for any $(\boldsymbol{\beta},\Lambda_0)$. On the other hand it is obvious that $\mathcal{Z}_n(\boldsymbol{\beta},\Lambda_0,\infty)$ and $\mathcal{Z}(\boldsymbol{\beta},\Lambda_0,\infty)=\lim_n\mathcal{Z}_n(\boldsymbol{\beta},\Lambda_0,\infty)$ reach the maximum value 0, at $\boldsymbol{\beta}^*,\Lambda_0^*$.

Define

$$\begin{aligned}&\mathcal{X}_n(\boldsymbol{\beta},\Lambda_0,t)\\ =\;& n^{-1}\sum_{i=1}^n\int_0^t\boldsymbol{W}_i^\top(s_-)(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\widetilde{N}_i(ds)-n^{-1}\sum_{i=1}^n\int_0^t\left[\Lambda_i(s_-)-\Lambda_i^*(s_-)\right]\widetilde{N}_i(ds)\\ &-n^{-1}\int_0^t\ln\left[\frac{S^{(0)}(\boldsymbol{\beta},\Lambda_0,s)}{S^{(0)}(\boldsymbol{\beta}^*,\Lambda_0^*,s)}\right]\sum_i\widetilde{N}_i(ds),\end{aligned}$$

which has the same maximum point as $\ln\tilde{\mathcal{L}}_n(\boldsymbol{\beta},\Lambda_0)$, since they are only different in a constant. Clearly $\lim(\mathcal{X}_n-\mathcal{Z}_n)=0$ since $\mathcal{X}_n-\mathcal{Z}_n$ is a martingale, and further $\lim\mathcal{X}_n=\mathcal{Z}$.

Therefore $\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_0$, the maximum point for $\log \tilde{\mathcal{L}}_n(\boldsymbol{\beta}, \Lambda_0)$ (or the maximum point for $\mathcal{X}_n(\boldsymbol{\beta}, \Lambda_0, \infty)$), converges to the true parameter value $\boldsymbol{\beta}^*$ and Λ_0^*, the maximum point of $\mathcal{Z}$, under the mild assumption that $(\boldsymbol{\beta}^*, \Lambda_0^*)$ is the unique maximum point for $\mathcal{Z}$.

C Proof of Theorem 2.2

First we define, for $u \geq s$,

$$\begin{aligned}
& \boldsymbol{F}(\boldsymbol{\beta}, \Lambda_0, s, u) := \frac{\partial \, \mathrm{E}^{(1)}(\boldsymbol{\beta}, \Lambda_0, u)}{\partial \Lambda_0(ds)} \\
= \; & \frac{n^{-1} \sum_i \left\{ -\boldsymbol{W}_i(s_-) e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}} - \left[\boldsymbol{W}_i^*(u_-) - \boldsymbol{\Lambda}_i^{(1)}(u_-) \right] e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}} \right\} e^{\boldsymbol{W}_i^\top(u_-)\boldsymbol{\beta} - \Lambda_i(u_-)}}{S^{(0)}(\boldsymbol{\beta}, \Lambda_0, u)} \\
& + \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \Lambda_0, u) n^{-1} \sum_{i=1}^n e^{\boldsymbol{W}_i^\top(u_-)\boldsymbol{\beta} - \Lambda_i(u_-)} e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}}}{S^{(0)}(\boldsymbol{\beta}, \Lambda_0, u)^2}.
\end{aligned}$$

By the first-order Taylor expansion of the score function $\boldsymbol{U}_n(\boldsymbol{\beta}, \Lambda_0)$ around the true value $\boldsymbol{\beta}^*, \Lambda_0^*(ds)$, we have

$$\begin{aligned}
& \sqrt{n} \boldsymbol{U}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0) \\
\overset{d}{\approx} \; & \sqrt{n} \boldsymbol{U}_n(\boldsymbol{\beta}^*, \Lambda_0^*) + \sqrt{n} \boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + \sqrt{n} \int_0^\infty \frac{\partial \boldsymbol{U}_n(\boldsymbol{\beta}^*, \Lambda_0^*)}{\partial \Lambda_0^*(ds)} \left[\hat{\Lambda}_0(ds) - \Lambda_0^*(ds) \right] \\
= \; & \sqrt{n} \boldsymbol{U}_n(\boldsymbol{\beta}^*, \Lambda_0^*) + \sqrt{n} \boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \\
& - n^{-1/2} \sum_i \int_0^\infty \int_s^\infty \boldsymbol{J}_i(\boldsymbol{\beta}^*, \Lambda_0^*, s, u) \widetilde{N}_i(du) \left[\hat{\Lambda}_0(ds) - \Lambda_0^*(ds) \right], \qquad (10)
\end{aligned}$$

where

$$\boldsymbol{J}_i(\boldsymbol{\beta}^*, \Lambda_0^*, s, u) = \boldsymbol{W}_i(s_-) e^{\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^*} + \boldsymbol{F}(\boldsymbol{\beta}^*, \Lambda_0^*, s, u), \qquad (11)$$

and

$$
\begin{aligned}
\boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*) &= -n^{-1}\sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{\Lambda}_i^{(2)*}(s) + \mathrm{E}^{(2)}(\boldsymbol{\beta}^*, \Lambda_0^*, s)\right]\widetilde{N}_i(ds), \qquad (12)\\
\boldsymbol{\Lambda}_i^{(2)*}(s) &= \int_0^s \boldsymbol{W}_i(u_-)^{\otimes 2} e^{\boldsymbol{W}_i^\top(u-)\boldsymbol{\beta}^*}\Lambda_0^*(du),\\
\mathrm{E}^{(2)}(\boldsymbol{\beta}^*, \Lambda_0^*, s) &= \frac{\boldsymbol{S}^{(2)}(\boldsymbol{\beta}^*, \boldsymbol{\Lambda}_0^*, s)}{S^{(0)}(\boldsymbol{\beta}^*, \boldsymbol{\Lambda}_0^*, s)} - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}^*, \boldsymbol{\Lambda}_0^*, s)^{\otimes 2}}{S^{(0)}(\boldsymbol{\beta}^*, \boldsymbol{\Lambda}_0^*, s)^2},\\
\boldsymbol{S}^{(2)}(\boldsymbol{\beta}^*, \boldsymbol{\Lambda}_0^*, s) &= n^{-1}\sum_{i=1}^{n}\left\{-\boldsymbol{\Lambda}_i^{(2)*}(s) + \left[\boldsymbol{W}_i(s_-) - \boldsymbol{\Lambda}_i^{(1)*}(s)\right]^{\otimes 2}\right\}\exp\left[\boldsymbol{W}_i^\top(s_-)\boldsymbol{\beta}^* - \Lambda_i^*(s)\right].
\end{aligned}
$$

The notation $\boldsymbol{a}\otimes\boldsymbol{b}$ denotes the outer product of the two column vectors $\boldsymbol{a}, \boldsymbol{b}$.

Then according to $\boldsymbol{U}_n(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0) = \boldsymbol{0}$ and (8), we have

$$
\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\\
\overset{d}{\approx}\ & \boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*)^{-1}\left\{-\sqrt{n}\boldsymbol{U}_n(\boldsymbol{\beta}^*, \Lambda_0^*) + n^{-1/2}\sum_i\int_0^{\infty}\int_s^{\infty}\boldsymbol{J}_i(\boldsymbol{\beta}^*, \Lambda_0^*, s, u)\widetilde{N}_i(du)\left[\hat{\Lambda}_0(ds) - \Lambda_0^*(ds)\right]\right\}\\
=\ & -n^{-1/2}\boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*)^{-1}\sum_i\left\{\int_0^{\infty}\left[\boldsymbol{W}_i(s_-) - \boldsymbol{\Lambda}_i^{(1)*}(s) - \mathrm{E}^{(1)}(\boldsymbol{\beta}^*, \Lambda_0^*, s)\right]M_i(ds)\right.\\
& \left. - \int_0^{\infty}\left[n^{-1}\sum_k\int_s^{\infty}\boldsymbol{J}_k(\boldsymbol{\beta}^*, \Lambda_0^*, s, u)\widetilde{N}_k(du)\right]\frac{1}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}^*, \Lambda_0^*, s)G^*/\gamma^*}M_i(ds)\right\} + o_p(1),
\end{aligned}
$$

where $M_i(ds) = \tilde{N}_i(ds) - A_i(ds|\boldsymbol{\beta}^*, \Lambda_0^*)$.

Following the martingale theory for counting processes, we similarly have that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$ is asymptotically normal with mean $\boldsymbol{0}$ and variance-covariance matrix

$\boldsymbol{\Psi}(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*) = \lim_{n\to\infty} \boldsymbol{\Psi}_n(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*)$, where

$$
\begin{aligned}
& \boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*)\boldsymbol{\Psi}_n(\boldsymbol{\beta}^*, \Lambda_0^*, G^*, \gamma^*)\boldsymbol{I}_n(\boldsymbol{\beta}^*, \Lambda_0^*) \qquad (13)\\
= \; & n^{-1}\sum_{i=1}^{n}\int_0^\infty \left[\boldsymbol{W}_i(s_-) - \boldsymbol{\Lambda}_i^{(1)*}(s) - \mathrm{E}^{(1)}(\boldsymbol{\beta}^*, \Lambda_0^*, s)\right]^{\otimes 2} A_i(ds|\boldsymbol{\beta}^*, \Lambda_0^*)\\
& + \int_0^\infty \left[n^{-1}\sum_k \int_s^\infty \boldsymbol{J}_k(\boldsymbol{\beta}^*, \Lambda_0^*, s, u)\widetilde{N}_k(du)\right]^{\otimes 2} \frac{\Lambda_0^*(ds)}{\boldsymbol{S}^{(0)}(\boldsymbol{\beta}^*, \Lambda_0^*, s)G^*/\gamma^*}\\
& - \left\{\int_0^\infty \left[n^{-1}\sum_k \int_s^\infty \boldsymbol{J}_k(\boldsymbol{\beta}^*, \Lambda_0^*, s, u)\widetilde{N}_k(du)\right]\Lambda_0^*(ds)\right\}^2.
\end{aligned}
$$

The asymptotic normality can be proved similarly for $\hat{\Lambda}_0$, based on the equation (10). □

ACKNOWLEDGMENTS

We are grateful to the associate editor and two referees for their insightful comments and constructive suggestions which have essentially improved the paper. Bo Fu was partially supported by the UK Medical Research Council research grant (MR/M025152/1) and a key international collaboration project (71420107023) funded by National Natural Science Foundation of China.

References

[1] P. K. Andersen, O. Borgan, R. D. Gill, et al., Statistical models based on counting processes, Springer-Verlag, New York Inc., 1993

[2] S. M. Bird, D. J. Goldberg, S. J. Hutchinson, Projecting severe sequelae of injection-related hepatitis C virus epidemic in the UK. Part 2: Preliminary UK estimates of prevalent injection-related hepatitis C carriers, and derivation

of progression rates to liver cirrhosis by gender and age at hepatitis C virus infection, J. Epidemiol. Biostat. 6 (2001) 267–277.

[3] K. C. G. Chan, Y. Q. Chen, C. Z. Di, Proportional mean residual life model for right-censored length-biased data, Biometrika 99 (2012) 995–1000.

[4] Y. Cheng, P. S. Shen, Z. Zhang, et al., Nonparametric association analysis of bivariate left-truncated competing risks data, Biom. J. 58 (2015) 635–651.

[5] H. Dai, Y. Bao, An inverse probability weighted estimator for the bivariate distribution function under right censoring, Statist. Probab. Lett. 79 (2009) 1789–1797.

[6] H. Dai, B. Fu, A polar coordinate transformation for estimating bivariate survival functions with randomly censored and truncated data, J. Statist. Plann. Inference 142 (2012) 248–262.

[7] S. Deuffic-Burban, J. B. Wong, A. J.Valleron, et al., Comparing the public health burden of chronic hepatitis C and HIV infection in France, J. Hepatol. 40 (2004) 319–326.

[8] G. J. Dore, A. J. Freeman, M. Law, et al., Is severe liver disease a common outcome for people with chronic hepatitis C, J. Gastroenterol. Hepatol. 17 (2002) 423–430.

[9] A. J. Freeman, G. J. Dore, M. G. Law, et al., Estimating progression to cirrhosis in chronic hepatitis C virus infection, Hepatology 34 (2001) 809–816.

[10] B. Fu, B. D. Tom, S. M. Bird, Re-weighted inference about hepatitis C virus-infected communities when analysing diagnosed patients referred to liver clinics, Stat. Methods Med. Res. 18 (2009) 303–320.

[11] B. Fu, B. D. Tom, T. Delahooke, et al. Event-biased referral can distort

estimation of hepatitis C virus progression rate to cirrhosis, and of prognostic influences, J. Clin. Epidemiol. 60 (2007) 1140–1148.

[12] R. B. Geskus, Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring, Biometrics 67 (2011) 39–49.

[13] R. D. Gill, S. Johansen, A survey of product-integration with a view toward application in survival analysis, Ann. Statist. 18 (1990) 1501–1555.

[14] T. Hu, X. Lin, B. Nan, Cross-ratio estimation for bivariate failure times with left truncation, Lifetime Data Anal. 20 (2014) 23–37.

[15] C. Y. Huang, J. Qin, Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia, J. Amer. Statist. Assoc. 107 (2012) 946–957.

[16] C. Y. Huang, J. Qin, D. A. Follmann, A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling, Biometrika 99 (2012) 199–210.

[17] M. G. Law, G. J. Dore, N. Bath, et al., Modelling hepatitis C virus incidence, prevalence and long-term sequelae in Australia, 2001, Int. J. Epidemiol. 32 (2003) 717–724.

[18] J. Qin, J. Ning, H. Liu, et al., Maximum likelihood estimations and EM algorithms with length-biased data, J. Amer. Statist. Assoc. 106 (2011) 1434–1449.

[19] J. Qin, Y. Shen, Statistical methods for analyzing right-censored length-biased data under Cox model, Biometrics 66 (2010) 382–392.

[20] J. A. Salomon, M. C. Weinstein, J. K. Hammitt, et al., Empirically calibrated

model of hepatitis C virus infection in the United States, Amer. J. Epidemiol. 156 (2002) 761–773.

[21] D. Schoenfeld, Partial residuals for the proportional hazards regression model, Biometrika 69 (1982) 239–241.

[22] H. Shen, R. J. Cook, Regression with incomplete covariates and left-truncated time-to-event data, Statist. Med. 32 (2013) 1004–1015.

[23] P. S. Shen, Additive hazards model with truncated and doubly censored data, J. Appl. Statist. 40 (2013) 1520–1532.

[24] Y. Shen, J. Ning, J. Qin, Analyzing length-biased data with semiparametric transformation and accelerated failure time models, J. Amer. Statist. Assoc. 104 (2009) 1192–1202.

[25] Y. R. Su, J. L. Wang, Modeling left-truncated and right-censored survival data with longitudinal covariates, Ann. Statist. 40 (2012) 1465–1488.

[26] L. Tian, D. Zucker, L. J. Wei, On the Cox model with time-varying regression coefficients, J. Amer. Statist. Assoc. 100 (2005) 172–183.

[27] B. D. Tom, V. T. Farewell, S. M. Bird, Maximum likelihood and pseudo score approaches for parametric time-to-event analysis with informative entry times, Ann. Appl. Statist. 8 (2014) 726–746.

[28] W. Y. Tsai, Pseudo-partial likelihood for proportional hazards models with biased-sampling data, Biometrika 96 (2009) 601–615.

[29] H. Wang, H. Dai, B. Fu, Accelerated failure time models for censored survival data under referral bias, Biostatistics 14 (2013) 313–326.

[30] C. C. Wen, Y. H. Chen, A frailty model approach for regression analysis of bivariate interval-censored survival data, Statist. Sinica 23 (2013) 383–408.

[31] D. Zeng, J. Cai, A semiparametric additive rate model for recurrent events with an informative terminal event, Biometrika 97 (2010) 699–712.

[32] H. Zhu, M. C. Wang, Analysing bivariate survival data with interval sampling and application to cancer epidemiology, Biometrika 99 (2012) 345–361.

[33] H. Zhu, M. C. Wang, Nonparametric inference on bivariate survival data with interval sampling: association estimation and testing, Biometrika 101 (2014) 519–533.