Electronic Journal of Statistics
Vol. 17 (2023) 1807–1847
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2133

Uniform confidence bands for hazard functions from censored prevalent cohort survival data*†

Ali Shariati‡

School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia
e-mail: ali.shariati@mq.edu.au

Hassan Doosti

School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia
e-mail: hassan.doosti@mq.edu.au

Vahid Fakoor

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: fakoor@um.ac.ir

Masoud Asgharian

Department of Mathematics and Statistics, McGill University, Montreal, Canada
e-mail: masoud.asgharian2@mcgill.ca

Abstract: Prevalent cohort studies are commonly conducted in many areas of research when incident cohort studies are deemed infeasible due to logistic or other constraints. While such studies are cost effective, it is known that survival data collected on prevalent cases do not form a representative sample from the target population. When the incidence (e.g. onset of disease) arise from a stationary Poisson process, it allows developing a more efficient methodology. While the stationarity assumption holds in many applications, to the best of our knowledge, the problem of establishing uniform confidence bands using data arisen in such settings has not been addressed in the current literature. We devise a method for obtaining uniform confidence bands for the cumulative hazard and the survival function built on their nonparametric maximum likelihood estimators (NPMLEs). To attain

*The authors would like to thank Professor Christina Wolfson for providing access to the CSHA data. Ali Shariati's research was supported by iMQRES No. 20191178 (Macquarie University funded scholarship). A part of this research was done during Ali Shariati's visit to McGill University. Masoud Asgharian's research was supported by NSERC RGPIN-2018-05618. The CSHA was supported by the Seniors Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada (project 6606-3954-MC[S]). The progression of dementia project within the CSHA was supported by Pfizer Canada through the Health Activity Program of the Medical Research Council of Canada and the Pharmaceutical Manufacturers Association of Canada; by the NHRDP (project 6603-1417-302[R]); by Bayer; and by the British Columbia Health Research Foundation (projects 38 [93-2] and 34 [96-1]).

†This article was first posted with the incomplete captions of the Figures 3–8. The captions were corrected on 11 December 2023.

‡Corresponding author.

this objective, we first present results on uniform strong consistency, weak convergence and asymptotic efficiency of the NPMLE of the cumulative hazard function. Given the intractable forms of the limiting processes in this case, the idea is to numerically approximate the functionals of the asymptotic processes of the normalized NPMLEs. Our simulation studies reveal the efficiency of the estimators for finite samples. The proposed procedures are illustrated using a set of real data on patients with dementia from the Canadian Study of Health and Aging.

Keywords and phrases: Asymptotic efficiency, uniform Confidence band, hazard function, informative censoring, length-bias, NPMLE, prevalent cohort, uniform strong consistency, weak convergence, dementia.

Received August 2022.

1. Introduction

Survival data typically comprise an initiating event, say onset of a disease, and a terminating event, say death due to the disease. The classical setting of survival analysis requires incident cases, i.e. subjects who have not experienced the initiating event before being recruited for follow-up. A viable alternative to follow-up studies on incident cases when logistics or other constraints render such studies impractical is conducting studies on prevalent cases, i.e. subjects who have experienced the initiating event before being recruited. Although prevalent cohort studies are typically cost effective, compared to incident cohort studies, it is well known that survival data collected on prevalent cases do not form a representative sample of the target population. Moreover, when the terminating event is subject to right censoring, it is known that such censoring is informative. As such statistical inference from prevalent cohort survival data should take into account the bias induced by response-selective sampling mechanism and the informative censoring. Failure to address the structural sampling bias leads to considerable overestimation of the survivor function, and the functionals of the survivor function, e.g. [42] for survival with dementia.

Prevalent survival data are typically left truncated by the gap time between the onset of the initiating event and the recruitment time. Left truncated right censored data have been extensively studied in survival analysis, attesting to the frequent use of the prevalent cohort survey design to investigate survival from onset of a disease [39, 40, 17]. The conditional approach to analysis of right censored prevalent cohort data, pioneered by [20] and [31], conditions away the observed truncation times (e.g., [32, 41, 40, 31, 19, 3, 39]). It is known that the conditional approach while robust is not most efficient when the left-truncation distribution is fully, or partially, known ([39]). There are many real instances where making assumptions about the form of the truncation distribution is supported by the data and the context of the application ([7, 26]). A case that has received particular attention in many areas of application ([7]) is the so-called stationarity where the initiating events, say the onset of a disease for different subjects, are assumed to be generated from a stationary point process. Such assumption is commonly referred to as the *stationarity* assumption ([9]). Under

the stationarity assumption the collected survival data are called length-biased ([1]). To account for the information about the left-truncation distribution and informative censoring, the unconditional approach was proposed and shown to be asymptotically most efficient (e.g., [35, 36, 16, 39, 6, 9, 23]). The validity of the stationarity assumption has been frequently verified in cohort surveys in the current literature which makes studies on unconditional approach more persuasive. Examples of length-biased data include epidemiological cohort studies, such as research on different chronic diseases and the early detection models [44]. Further instances found in literature contain cancer prevention trials, such as screening for breast cancer [43], ages at death for residents of a retirement centre [18], a survey on elders with dementia [42]. While prevalent cohort studies are the main motivation behind this paper, the statistical tools we propose are also useful for analysis of length-biased data arising from other different sampling procedures. Sampling procedures subject to length-bias have been reported in a variety of different studies in the literature, such as investigation into unemployment duration [13] (see [7, 2, 21] for a list of references).

A subtle but pivotal distinction missed by several authors and frequently in practice is that the Kaplan–Meier estimator is not a consistent estimator of the length-biased (prevalent-case) survival function due to informative censoring, and hence cannot be transformed to find a consistent estimator of the unbiased (incident-case) survival function. When there is no censoring, the classical empirical distribution is the unconditional NPMLE of the length-biased distribution function. However, in the presence of right censoring, the NPMLE proposed by [37], called hereafter *Vardi estimator*, for the multiplicative censoring model may be applied to estimate the length-biased survival function. The application of Vardi's NPMLE for the length-biased distribution from prevalent cohort cases was discussed by [6] and [8]. The asymptotic behavior of Vardi's NPMLE under multiplicative censoring was established by [38]. Similar results under the prospective prevalent cohort setting were derived by [8].

Many studies have been carried out on hazard functions in the presence of left truncation and right censoring (see [14] for a list of references), but all those works applicable to the sampling setting we study here were based on the conditional likelihood approach. Our approach here, however, centres around studying the unconditional NPMLE of the cumulative hazard function as a Vardi integral, i.e. integrals of appropriate integrands with respect to Vardi estimator, as the integrator. Besides, the other key parameter we study is the unbiased survival function as another Vardi integral. [6] proposed the unconditional NPMLE of the unbiased survival function and investigated its limiting behaviour. Furthermore, the higher efficiency of the survival function estimator for small sample sizes in comparison to a conditional estimator was exhibited. However, the intractable form of the limiting stochastic process given in [8] for the Vardi's NPMLE under prevalent cohort setting renders derivations of analytical forms of asymptotic processes of functionals of the Vardi's estimator practically impossible (e.g. [6]).

In this paper, we take a different approach. To derive uniform confidence bands for the cumulative hazard function of the target population of interest using prevalent cohort survival data, we first study the asymptotic behaviour

of the NPMLE of the cumulative hazard function. We prove the uniform strong consistency with the appropriate rate of convergence under assumptions that are weaker than those used in [6]. To obtain results over the whole domain of the distribution function, we directly derive the weak convergence of the NPMLE of the hazard function as a functional of Vardi's estimator. We then propose a computational method to surmount the obstacle associated with the intractable forms of the functionals of the asymptotic processes of the normalized NPMLEs. This method is applied to compute quantiles of the limiting distributions of the supremum of the stochastic integrals corresponding to the cumulative hazard and the survival function, needed to derive uniform confidence bands.

Even though we only present our results for censored prevalent cohort survival data under stationarity assumption, we further discuss (Subsection 3.1) that our methodology can be modified for other scenarios if the intensity of the incidence process is known. Our results are also valid under the multiplicative censoring model (Subsection 2.3), which unifies several important problems, such as an estimation problem in renewal processes, estimation under a decreasing density constraint, and deconvolution of an exponential random variable [37].

The rest of this article is organised as follows. In Section 2, we discuss data structure, definitions and assumptions required throughout the paper. Main theoretical results are discussed in Section 3. We then derive asymptotic efficiency of the mentioned NPMLEs in Section 4. The general framework to derive uniform confidence bands for the cumulative hazard function and survival function is discussed in Section 5. Section 6 provides a computational algorithm to approximate the limiting distributions. The following section reports results on simulation studies conducted to illustrate and inspect the performance of the NPMLEs and the corresponding confidence bands for finite sample sizes. Section 9 summarises the results inferred by applying the proposed procedures for the data on Canadian elderly population with dementia collected through the Canadian Study of Health and Aging (CSHA). Proofs of the theoretical statements are presented in Section 10.

2. Preliminaries

2.1. Data setup

Let X' denote the unbiased continuous lifetime (or failure time) of interest. In cross-sectional sampling, the truncation variable T' associated with X' indicates the time duration between the initiating and recruitment time. Therefore, there is a pair of random variables (X', T') for each subject in the population. It is often reasonable to assume that X' and T' are independent. Let $F_{X'}(x) := P(X' \leq x)$ and $f_{X'}$, defined on $\mathbb{R}^+ := [0, \infty)$, denote the respective cumulative distribution and probability density functions of the lifetime variable X'. In prevalent cohort studies, cases that have experienced the initiating event but are yet to encounter the terminating event are only observable. Thus, those subjects with $X' \geq T'$ are identifiable whereas the rest of cases are left truncated (see Figure 1). Denote the observation satisfying this condition $(X' \geq T')$ by X.

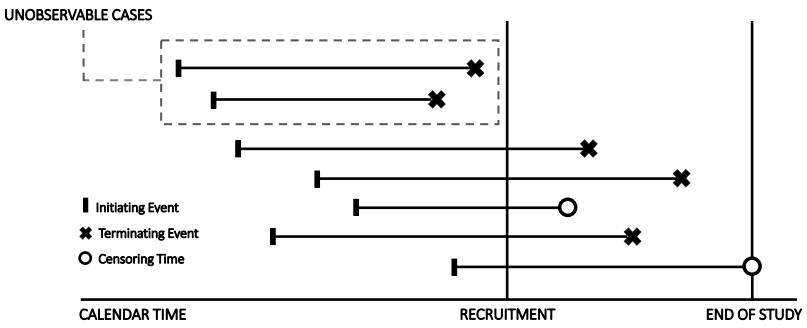

FIG 1. *Cross-sectional Sampling with Follow-up.*

Suppose that the stationarity assumption for the incidence rate holds, that implies the truncation variable follows a uniform distribution on some interval $[0, \tau']$, where $\tau' > \tau$. It is then obtained for $G(x) := P(X \leq x)$ that

$$G(x) = F_{X'|X'>T'}(x) = \mu_{X'}^{-1} \int_0^x sF_{X'}(ds), \quad x \geq 0, \tag{1}$$

where $\mu_{X'} = E(X') < \infty$ is the expected value corresponding to variable X'. Define g as the density function of the observed survival time corresponding to G. Then, G and g represent the so-called length-biased distribution function and length-biased probability density function.

For any observed case in the cross-sectional sampling with follow-up, we have the triple $(A, R \wedge C, \delta)$ related to the lifetime X in the presence of right-censoring, where A, R and C, are the current age (also known as the backward recurrence time), the residual lifetime (also forward recurrence time) and the residual censoring time, respectively. Also, δ is the censoring indicator variable defined by $\delta := I(R \leq C)$. It is worth noting that the residual censoring time is the remaining time from the recruitment until being censored. Moreover, it is a reasonable assumption in most real situations to consider C is independent of (A, R). Since our sample is consists of both censored and uncensored observations, denote the observed subject X by Y if it is uncensored and by Z once it is censored, and consequently $Y = A + R$ and $Z = A + C$. Bear in mind that, although C is independent of (A, R), the complete lifetime observed Y and the censoring time observed Z are not independent and $Cov(Y, Z) > 0$, which implies that the censoring mechanism is informative in a prevalent cohort study with follow-up [4]. In other words, subjects should survive until the recruitment to be observable in the sampling procedure under study, and then the remaining of their lifetimes (R) may be randomly censored (see Figure 2). Therefore, the Kaplan-Meier estimator does not result in a NPMLE for the length-biased distribution here. Alternatively, [37] derived a NPMLE for the common lifespan distribution function of the n independent and identically distributed stationary renewal processes. It is proven that prevalent cohort cases under the stationarity assumption and observations of n independent and identically distributed

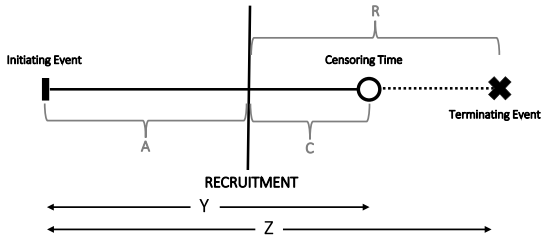

FIG 2. *Informative Censoring Structure.*

stationary renewal processes started a long time ago have the same NPMLE. The NPMLE of G is introduced in the next section.

To complete the preliminary definitions in this section, we need to present distributions of Y, Z and the recurrence times here. The joint distribution of A and R has the following representation [37, 15],

$$f_{A,R}(a,r) = \frac{f_{X'}(a+r)}{\mu_{X'}}.$$

Let p indicate the probability of censoring for each recruited subject during the period of follow-up, $p := P(\delta = 1)$. For the remaining censoring time C, define $F_C(c) := P(C \leq c)$, then $S_C(c) = 1 - F_C(c)$. Denote the respective letters $\mathcal{C}$ and $\mathcal{U}$ for censored and uncensored subjects. Define

$$G^{\mathcal{U}}(y) := P(A + R \leq y | \delta = 1) = \frac{1}{p} \int_0^y \frac{g(s)}{s} \int_0^s S_C(r) dr ds,$$

and

$$G^{\mathcal{C}}(z) := P(A + C \leq z | \delta = 0) = \frac{1}{1-p} \int_0^z F_C(s) \int_s^\infty u^{-1} dG(u) ds.$$

Suppose $g^{\mathcal{U}}$ and $g^{\mathcal{C}}$ are the respective probability density functions corresponding to $G^{\mathcal{U}}$ and $G^{\mathcal{C}}$. In addition, let f_r indicate the probability density function of backward/forward recurrence time [11]. Then, f_r has representation:

$$f_r(x) = \frac{1}{\mu_{X'}} \left(1 - F_{X'}(x)\right) = \int_x^\infty s^{-1} G(ds). \tag{2}$$

2.2. Overview: Definitions and NPMLE

Let n indicate the number of subjects observed at the recruitment time. Our sample includes the independent vectors $(A_i, R_i \wedge C_i, \delta_i)$ for $i = 1, 2, \ldots, n$. In terms of Y and Z, the sample consists of the random variables $Y_1, \ldots, Y_{N_1}$ and $Z_1, \ldots, Z_{N_2}$, where the number of uncensored subjects (N_1) and censored cases (N_2) are random, that is a disparity between the model under study in this article and n independent renewal process and multiplicative censoring [8]. Let

n_1 and n_2 be the values realized for the respective random variables N_1 and N_2. Let $(a_i, r_i \wedge c_i, \delta_i)$ indicate the values realized for the vector $(A_i, R_i \wedge C_i, \delta_i)$ for $i = 1, \ldots, n$. Denote $y_1, \ldots, y_{n_1}$ and $z_1, \ldots, z_{n_2}$ for the uncensored and censored realized values. Thus, the full likelihood function is as follows,

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \left(f_{A,R}(A = a_i, R = r_i)\right)^{\delta_i} \left(P(A = a_i, R > c_i)\right)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left(\frac{f_{X'}(y_i)}{\mu_{X'}}\right)^{\delta_i} \left(\int_{c_i}^{\infty} f_{A,R}(a_i, r)dr\right)^{1-\delta_i} \\
&\propto \prod_{i=1}^{n} \left(G(y_i)\right)^{\delta_i} \left(\int_{z_i}^{\infty} s^{-1}G(ds)\right)^{1-\delta_i}. \qquad (3)
\end{aligned}$$

Let $0 < t_1 < t_2 < \cdots < t_k$ denote the distinct sample values realized for the sample $\{(y_1, \ldots, y_{n_1}) \cup (z_1, \ldots, z_{n_2})\}$. Consequently, $k \leq n_1 + n_2 = n$ due to possibility of multiplicities in data sets in real applications. Let $\hat{G}$ denote Vardi's estimator which maximises the likelihood $\mathcal{L}$. [37] derived this estimator for the multiplicative censoring model, but its validity for the underlying model was noted by [8]. Given $\hat{G}$, the NPMLE of $F_{X'}$ has representation

$$\hat{F}(x) = \hat{\mu}_{X'} \int_0^x t^{-1}\hat{G}(dt). \qquad (4)$$

where

$$\hat{\mu}_{X'}^{-1} = \int_0^{\infty} x^{-1}\hat{G}(dx).$$

In the rest of this article, we conduct a study on the unbiased cumulative hazard function and the unbiased survival function from censored prevalent cohort survival data. We consider these quantities as the functionals of G. For this objective, it is necessary to introduce the process to which the empirical process $\mathcal{B}_n := \sqrt{n}(\hat{G} - G)$ converges asymptotically. Let $\mathcal{B}$ indicate the limiting process of $\mathcal{B}_n$. We also need to define the spaces on which $\mathcal{B}_n$ and $\mathcal{B}$ act. Let $\mathcal{D}_0[a, \infty]$ denote the space of all Cadlag functions $u(\cdot)$ on $[a, \infty]$ vanishing at a. The uniform norm $\|\cdot\|_{[a,b]}$ is defined by $\|u\|_{[a,b]} := \sup_{a \leq x \leq b}|u(x)|$. The space $\mathcal{D}_0[0, b]$ endowed with the uniform topology, the topology induced by the uniform norm, is a Banach space. This implies that the space of bounded linear operators on $\mathcal{D}_0[0, \infty]$ denoted by $\mathcal{L}(\mathcal{D}_0[0, \infty], \mathcal{D}_0[0, \infty])$ is a Banach algebra. The other fact about $\mathcal{D}_0[0, \infty]$ that we need is that cadlag functions have countably many jumps [24]. This assures that cadlag functions are Riemann integrable on bounded intervals. The stochastic process $\mathcal{B}$ is then defined by

$$\mathcal{B} := \mathcal{F}^{-1}(V), \qquad (5)$$

where $\mathcal{F}^{-1}$ is the inverse of $\mathcal{F}$, and $\mathcal{F} : \mathcal{D}_0[0, \infty] \to \mathcal{D}_0[0, \infty]$ is an invertible linear operator with representation

$$\mathcal{F} := \mathcal{G}_1 + \mathcal{G}_2$$

such that

$$\mathcal{G}_1(u)(x) := p\int_{0<y\leq x}\frac{g^{\mathcal{U}}(y)}{g(y)}u(dy)$$

$$\mathcal{G}_2(u)(x) := (1-p)\int_{0<z\leq x} z\left(\int_{z\leq s}\frac{u(s)}{s^2}ds\right)d\left[\left(\frac{f_r(x)}{f_r(z)}-1\right)\frac{g^{\mathcal{C}}(z)}{f_r(z)}\right],$$

and $u\in\mathcal{D}_0[0,b]$. In addition, V is defined as follows

$$\begin{aligned} V(x) \quad := \quad & p^{1/2}B_1\left(G^{\mathcal{U}}(x)\right)+(1-p)^{1/2}f_r(x)\int_{0<t\leq x}B_2\left(G^{\mathcal{C}}(t)\right)d\frac{1}{f_r(t)} \\ & +\left(\frac{p}{1-p}\right)^{1/2}\left(G^{\mathcal{U}}(x)-G(x)\right)Z, \end{aligned} \tag{6}$$

where B_1 and B_2 are independent Brownian bridges, and the random variable $Z\sim N(0,1)$ is independent from B_1 and B_2.

2.3. Assumptions

The condition defined by set $\mathcal{J}$ below does not actually impose any restriction to our results as long as $F_C(0)=0$. However, we are considering the possibility of the situation when we cannot follow-up on a proportion of cases recruited right after the recruitment with probability $P(C=0)=F_C(0)>0$. The information on such cases only comes from the current age A. Furthermore, since A and R are identically distributed under stationarity, $F_C(0)>0$ can also be considered as the share of uncensored observations with missing onset times. Besides, this condition was initially introduced by [8] and used to show that their results for $\hat{G}$ under prevalent cohort setup generalize those by [38] under multiplicative censoring. Our methodology inherits this generalization and can thus be used to derive uniform confidence bands under multiplicative censoring setup as well.

Let denote $\tau := \inf\{x : G(x)=1\}$, assume that $\tau<\infty$ and identify $\beta = F_C(0)\geq 0$. Define

$$\mathcal{J} := \left\{x\leq\tau : \left(\frac{2}{\alpha(x)}-\frac{1}{1-\beta}\right)\beta<1\right\},$$

where $\alpha(x) := \frac{1}{x}\int_0^x S_C(s)ds = pg^{\mathcal{U}}(x)/g(x)$. It is of note that $\mathcal{J}=[0,\tau]$ if $\beta=0$. Bear also in mind that $\alpha(x)$ is a decreasing function with $\lim_{x\to 0}\alpha(x)=1-\beta$. Thus, a necessary condition in order to $x\in\mathcal{J}$ is that $\beta<1/2$. Moreover, it is apparent that $\alpha(x)\geq S_C(x)$ and as a consequence a sufficient condition $x\in\mathcal{J}$ is that $F_C(x)<1/2$. For the interpretation of a condition similar to $\mathcal{J}$ see [6], although the alternative condition used by [6] is more restrictive than $\mathcal{J}$ here. The condition imposed by $x\in\mathcal{J}$ is not restrictive because β, the mass of the residual censoring distribution at 0, would be very small in practice. If $\beta=0.01$ for example, then a sufficient condition for $x\in\mathcal{J}$ is that $F_C(x)\leq 0.98$.

Theoretical results presented in this article have been derived under one of the assumptions below. While the assumption (A1) was initially used by [5], the assumption (A2) was employed by [6] first.

Assumption (A1). A sequence of real numbers γ_{n_1,n_2} meets the assumption (A1) if

$$\sum_{n_1,n_2} G\left(\gamma_{n_1,n_2}\right) < \infty,$$

where the summation is taken over the subsample sizes n_1 and n_2 taken jointly to infinity so that $\hat{p} := \sum_{i=1}^{n} \delta_i/n \to p$.

To sidestep problems concerning a singularity at the origin, we select a sequence of positive real numbers γ_{n_1,n_2} fulfilling (A1) and estimates $\hat{G}$ assigning zero probability below γ_{n_1,n_2}. Thus, the existence of such distribution functions $\hat{G}$ is of critical importance. Claim 1 below reveals such estimates $\hat{G}$ exist. But we first need the score equation corresponding to the likelihood function (3) derived in (2.1) of [5].

Claim 1. *Let assumption (A1) hold. Then, for all sufficiently large n_1 and n_2, the score equation has a solution $\hat{G}$ such that $\hat{G}(x) = 0$ for all $x < \gamma_{n_1,n_2}$.*

Claim 1 here coincides with Claim 1 in [5] and, therefore, its proof is omitted. Claim 1 indicates that the sequence γ_{n_1,n_2} helps us to control the behavior of $\hat{G}$ for large n_1 and n_2 to avoid problems related to a singularity at the origin. It is not tricky to find γ_{n_1,n_2} since X is continuous and so $G(0) = 0$. For instance, an obvious selection of γ_{n_1,n_2} satisfying the aforementioned criteria is $\gamma_{n_1,n_2} = G^{-1}(1/n^{\nu})$ where $\nu > 1$ and G^{-1} is the inverse function of G. However, our results in Section 3 indicates that γ_{n_1,n_2} ought to satisfy the following added constraint.

$$\gamma_{n_1,n_2}^{-1} \sqrt{\frac{\log\log n}{n}} \longrightarrow 0, \tag{7}$$

which means γ_{n_1,n_2} cannot go to zero as fast as $(\log\log n/n)^{1/2}$ goes. In the above example (i.e. $\gamma_{n_1,n_2} = G^{-1}(1/n^{\nu})$), the choice of ν and the behaviour of G are hence very essential in order for γ_{n_1,n_2} to satisfies (7). Hence, the condition (7) limits our results to those distribution functions G falling to zero fast enough. For example, it can be shown that such a sequence of γ_{n_1,n_2} exists when G is a member of the Gamma family of distribution functions with shape parameter $k > 2$. That in return means assumption (A1) holds for the target populations $F_{X'}$ in the Gamma family of distribution functions with $k > 1$. Similarly, it is observed that (A1) is met for Weibull family of distribution functions with shape parameter $k > 1$ as the target distribution functions.

Assumption (A2). We say the length-biased distribution G satisfies assumption $(A2)$ if there exists a $\gamma > 0$ such that for all $x < \gamma$ we have $G(x) = 0$.

Assumption (A2) states that we can be assured of survival of subjects for some mere length of time after the initiating event, which is a reasonable assumption in examples of practical importance. If we can assume such $\gamma > 0$ exists, we can simply choose $\gamma_{n_1,n_2} = \gamma/2$, and consequently assumption (A1) is not required.

Hence, assumption $(A1)$ is less strict than (A2). For example, any distribution function truncated between $[0, \gamma]$, for arbitrary γ, satisfies assumption (A2).

3. Asymptotic study

In this section, we introduce the NPMLEs of the hazard function and the survival function. We study the asymptotic behaviour of the NPMLE of the hazard function here. The asymptotic properties of the NPMLE of the survival function were first studied by [6]. Having reviewed the results given in [6], we prove the uniform strong consistency of the estimator of the survival function under a weaker assumption, which extends the consistency for distributions with domains defined over $[0, \tau]$.

3.1. Cumulative hazard function

Let define Λ as the cumulative hazard function of the unbiased population of interest as follows

$$\Lambda(x) := \int_0^x \frac{F_{X'}(dt)}{1 - F_{X'}(t)} = \int_0^x \frac{G\,(dt)}{t \int_t^\infty s^{-1} G(dt)} \tag{8}$$

where $x \in [0, \tau)$. After that, the NPMLE of Λ has representation

$$\hat{\Lambda}(x) := \int_0^x \frac{\hat{F}_{X'}(dt)}{1 - \hat{F}_{X'}(t)} = \int_0^x \frac{\hat{G}\,(dt)}{t \int_t^\infty s^{-1} \hat{G}(dt)},$$

where $x \in [0, t_k)$ and t_k is the maximum of observations. The uniform strong consistency of $\hat{\Lambda}$ has been studied in the next theorem.

Theorem 1. *Suppose that assumption* $(A1)$ *is satisfied. For any* $b \in \mathcal{J}$*, it is derived that, as* $n \to \infty$*,*

$$\|\hat{\Lambda} - \Lambda\|_{[0,b]} \xrightarrow{a.s.} 0.$$

Additionally, it is obtained for any $b \in \mathcal{J}$ *and* $b \leq \tau - \epsilon$ *that, if assumption (A1) is satisfied, as* $n \to \infty$*,*

$$\|\hat{\Lambda} - \Lambda\|_{[0,b]} \quad = \quad O\left(\frac{F_{X'}(\gamma_{n_1,n_2})}{1 - F_{X'}(\gamma_{n_1,n_2})} + \gamma_{n_1,n_2}^{-1} \sqrt{\frac{\log\log n}{n}} \right) \qquad a.s.$$

and if assumption (A2) holds, then, as $n \to \infty$*,*

$$\|\hat{\Lambda} - \Lambda\|_{[0,b]} = O\left(\sqrt{\frac{\log\log n}{n}} \right) \qquad a.s.$$

Proof. See Proofs. □

Define $U_n^1 := \sqrt{n}\left(\hat{\Lambda} - \Lambda\right)$. In the following theorem, asymptotic weak convergence of U_n is presented.

Theorem 2. *Under assumption (A2), it is proven for any $b \in \mathcal{J}$ that, as $n \to \infty$,*

$$U_n^1 \xrightarrow{\mathcal{W}} U^1 \quad in \quad \mathcal{D}_0[0,b],$$

where U^1 denotes a mean zero Gaussian process with covariance function

$$Cov\left(U^1(x_1), U^1(x_2)\right) = \int_0^\infty \int_0^\infty Cov\left(\mathcal{B}(t), \mathcal{B}(s)\right) \mathcal{I}_{x_1}^1(dt)\mathcal{I}_{x_2}^1(ds)$$

and $\mathcal{B}$ is defined in (5). In addition, U^1 has representation

$$U^1(x) := \int_0^\infty \mathcal{I}_x^1(t)\mathcal{B}(dt),$$

where

$$\mathcal{I}_x^1(t) := \mathcal{K}_x(t,t) - \int_0^t \mathcal{K}_x(t,s)\Lambda(ds)$$

and

$$\mathcal{K}_x(t,s) := \left\{ \frac{\mu_{X'}}{t\left(1 - F_{X'}(s)\right)} \right\} I_{[0,x]}(s).$$

Proof. See Proofs. □

Although our results are here driven under the stationarity assumption, one can mimic the proofs presented to obtain asymptotic results and confidence bands under other scenarios as long as the intensity of the incidence process is known [1, 9]. This, however, requires two separate steps. First, a new master equation should be used in [8], see Section 8. In addition, it entails modification of the representations (14-15) in Proofs and, thus, some subsequent steps. The same can be done in the case of survival function in the next section.

3.2. Survival function

Let define $S := 1 - F_{X'}$ as the survival function of the target population of interest. Following that, the NPMLE of the S for any $x > 0$ is given by

$$\hat{S}(x) = 1 - \hat{F}(x) = \hat{\mu} \int_x^\infty t^{-1}\hat{G}\left(dt\right),$$

where $\hat{F}$ is presented in (4). In a study on the unconditional NPMLE of the unbiased survival function, [6] showed that under assumption (A2), as $n \to \infty$,

$$\|\hat{S} - S\|_{[0,b]} = O\left(\sqrt{\frac{\log\log n}{n}}\right) \qquad a.s. \tag{9}$$

Additionally, let U^2 denote a mean zero Gaussian process with representation

$$U^2(x) := \int_0^\infty \mathcal{I}_x^2(t)\mathcal{B}(dt)$$

and covariance function

$$Cov\left(U^2(x_1), U^2(x_2)\right) = \int_0^\infty \int_0^\infty Cov\left(\mathcal{B}(t), \mathcal{B}(s)\right) \mathcal{I}_{x_1}^2(dt)\mathcal{I}_{x_2}^2(ds),$$

where $\mathcal{B}$ is given in (5) and

$$\mathcal{I}_x^2(t) := \mu_{X'} \left\{ \frac{I_{[x,\infty]}(t) - S(x)}{t} \right\}.$$

By defining $U_n^2 := \sqrt{n}\left(\hat{S} - S\right)$, [6] also derived that, under assumption $(A2)$,

$$U_n^2 \xrightarrow{\mathcal{W}} U^2 \quad \text{in} \quad \mathcal{D}_0[0, b], \tag{10}$$

as $n \to \infty$, for any $b \in \mathcal{J}$.

We first study the uniform strong consistency of the unbiased survival function under assumption (A1) in the below theorem. By extending the uniform strong consistency of the NPMLE of S for distributions satisfying assumption (A1), the theorem relaxes the restriction (A2), under which (9) was derived.

Theorem 3. *Suppose that assumption* $(A1)$ *holds. Then, it is derived for any* $b \in \mathcal{J}$ *that, as* $n \to \infty$,

$$\|\hat{S} - S\|_{[0,b]} = O\left(F_{X'}(\gamma_{n_1,n_2}) + \gamma_{n_1,n_2}^{-1} \sqrt{\frac{\log\log n}{n}} \right) \qquad a.s.$$

Proof. See Proofs. □

Although the weak convergence for the empirical process U_n^2 was proved by [6], it could not be used in practice to obtain confidence bands since $\mathcal{B}$ does not have an explicit and tractable form. We propose two statistics in Section 5 that asymptotically does not rely on an unknown parameter. Based on these statistics, we derive confidence bands for the cumulative hazard function and the survival function. To surmount the obstacle associated with the intractable forms of the proposed statistics, we derive a computational method to obtain the behaviour of the stochastic integrals U^i, for $i = 1$ and 2, and thus constructing the confidence bands in practice. However, we first study asymptotic efficiency of the proposed NPMLEs in the following section.

4. Asymptotic efficiency

From this point onwards, let denote θ^1 and θ^2 for the respective statistical functionals Λ and S (i.e. $\theta^1 \equiv \Lambda$ and $\theta^2 \equiv S$). Similarly, let define $\hat{\theta}^1 \equiv \hat{\Lambda}$ and

$\hat{\theta}^2 \equiv \hat{S}$. Given the representations (1) and (8), it is easy to find the form of functionals ϕ^i such that $\phi^i(G) = \theta^i$, and hence $\phi^i(\hat{G}) = \hat{\theta}^i$, for $i = 1$ and 2. Having considered the asymptotic efficiency of $\hat{G}$ given in [8], the probing question we try to address in this section is that whether the asymptotic efficiency of $\hat{G}$ is carried over into the efficiency of $\phi^i(\hat{G})$ as well. The following theorem prove the asymptotic efficiency of $\hat{\theta}^i$ in estimating θ^i, for $i = 1, 2$.

Theorem 4. *Let $G : [0, \tau] \to [0, 1]$ be the distribution function of the random variable X, and η denote a constant such that $1 - G(\eta) \geq \epsilon > 0$ for some $\epsilon > 0$. Under assumption (A2):*

i) The NPMLE $\hat{\theta}^1$ is asymptotically efficient estimator of θ^1 over $[0, \eta]$.
ii) The NPMLE $\hat{\theta}^2$ is asymptotically efficient estimator of θ^2 over $[0, \tau]$.

Proof. See Proofs. □

5. Uniform confidence bands

The theoretical results presented in Section 3 may be applied to establish uniform confidence bands for the statistical functionals under study. For this purpose, we need to propose a nonparametric statistics whose asymptotic distribution does not rely on any unknown parameter. For the Gaussian processes U^i, define $(\sigma_n^i(x))^2$ as uniform consistent estimators of $(\sigma^i(x))^2 := Cov(U^i(x), U^i(x))$. Inspired by a common method used for likelihood-based confidence bands [45, 47, 29], let define

$$\mathcal{S}_n^i(x, \theta^i) := \frac{U_n^i(x)}{\sigma_n^i(x)}, \tag{11}$$

where $i = 1$ and 2. Then, the stochastic processes $\mathcal{S}_n^i$ are nonparametric statistics that can be applied to derive confidence bands for θ^i. For this objective, we need to present the following theorem here.

Theorem 5. *Under assumption (A2), it is derived that, as $n \to \infty$,*

$$\mathcal{S}_n^i(\cdot, \theta^i) \xrightarrow{\mathcal{W}} \mathcal{S}^i(\cdot, \theta^i) \quad in \quad \mathcal{D}_0[0, b],$$

where $b \in \mathcal{J}$ and

$$\mathcal{S}^i(x, \theta^i) := \frac{U^i(x)}{\sigma^i(x)}.$$

Proof. See Proofs. □

Considering Theorem 5, we can obtain through the continuous mapping theorem that, as $n \to \infty$,

$$\sup_{x \in [0,b]} \left| \mathcal{S}_n^i(x, \theta^i) \right| \xrightarrow{\mathcal{D}} \sup_{x \in [0,b]} \left| \mathcal{S}^i(x, \theta^i) \right|, \tag{12}$$

where $\xrightarrow{\mathcal{D}}$ indicates convergence in distribution. Hence, an asymptotic $100(1-\alpha)\%$ confidence band for the statistical functional θ^i is inferred as follows:

$$\mathcal{C}_{\theta^i} = \left\{ \bar{\theta}^i(x) : \left| \mathcal{S}_n^i(x, \bar{\theta}^i(x)) \right| \leq q_\alpha^i, \quad x \in [0,b] \right\},$$

where q_α^i is the upper α-quantile of the distribution of $\sup_{x\in[0,b]} \left|\mathcal{S}^i(x,\theta^i)\right|$. It is then concluded that, as $n \to \infty$,

$$P\left(\theta^i \in \mathcal{C}_{\theta^i}\right) \longrightarrow 1-\alpha.$$

In Section 8, we discuss in details how to draw paths from the stochastic process $\mathcal{B}$. The behaviour of $\mathcal{B}$ can be employed to obtain the distribution of $\sup_{x\in[0,b]} \left|\mathcal{S}^i(x,\theta^i)\right|$, and thus reaching q_α^i for different values of α and corresponding $i = 1, 2$. Furthermore, we define two uniform consistent estimators for σ^i, indicated by σ_n^i and σ_n^{i*}, in Simulation.

The following Corollary presents the limiting distribution of $\mathcal{S}^i$, for $i = 1, 2$, at any fixed point.

Corollary 1. *For any $x_0 \in \mathcal{J}$, we have, as $n \to \infty$,*

$$\mathcal{S}_n^i(x_0, \theta^i(x_0)) \xrightarrow{\mathcal{D}} Z,$$

where the random variable Z follows the standard normal distribution.

Proof. See Proofs. □

Using Corollary 1, an asymptotic $100(1-\alpha)\%$ confidence interval for the mean residual life function at $x_0 \in \mathcal{J}$ is given by

$$\mathcal{C}_{\theta^i}(x_0) = \left\{ \bar{\theta}^i(x_0) : \left| \mathcal{S}_n^1(x_0, \bar{\theta}^i(x_0)) \right| \leq z_{\alpha/2} \right\},$$

where $z_{\alpha/2}$ is the upper $(\alpha/2)$-quantile of the normal standard distribution. It follows from Theorem 1 that, as $n \to \infty$,

$$P\left(\theta^i(x_0) \in \mathcal{C}_{\theta^i}(x_0)\right) \longrightarrow 1-\alpha.$$

6. Approximating the distribution of $\sup \mathcal{S}^i$

Given the intractable form of the stochastic processes $\mathcal{S}^i$, $i = 1, 2$, it is tricky to obtain the distributions of $\|\mathcal{S}^i\|_{[0,b]} := \sup_{x\in[0,b]} |\mathcal{S}_n^i(x,\theta^i)|$ analytically. Indeed, although the linear operator $\mathcal{F}$ is invertible, the inverse operator $\mathcal{F}^{-1}$ does not have a closed form, and consequently deriving an explicit expression for $\mathcal{B}$ defined in (5) is not possible. However, we here propose a computational method that can be applied to approximate the distribution of $\|\mathcal{S}^i\|_{[0,b]}$. We attain this objective in three phases. Initially, we discuss the problem of simulating paths from the process $\mathcal{B} = \mathcal{F}^{-1}(V)$ by means of a numerical method. After that, the generated paths of $\mathcal{B}$ are used to obtain paths of the stochastic integrals U^i, $i = 1, 2$. Finally, we propose an algorithm that can be employed to approximate $\mathcal{S}^i$ and then the distributions of $\|\mathcal{S}^i\|_{[0,b]}$.

6.1. Paths of the stochastic process $\mathcal{B}$

The important question that we seek to address in this subsection is how to computationally draw paths from the stochastic process $\mathcal{B}$ defined in (5). To reach this objective, the first step is to draw paths from V given in (6). For this purpose, assume a target population $F_{X'}$ and consider the corresponding length-biased distribution G defined in (1) along with an arbitrary censoring distribution F_C. We compute the values of $V(x)$ and thus $\mathcal{B}(x) = \mathcal{F}^{-1}(V(x))$ for $x \in \{0, x_1, \ldots, x_m, \tau\}$, where m denotes the number of grids considered. Without loss of generality, suppose that $0 < x_1 < \ldots < x_m < \tau$, and set $x_0 = 0$, $x_{m+1} = \tau$. By simulating a standard normal variable Z and Brownian bridges B_1 and B_2 defined in (6) at the respective points $G^{\mathcal{U}}(x_l)$ and $G^{\mathcal{C}}(x_l)$, we can obtain the value of $V(x_l)$ for $l = 0, 1, \ldots, m+1$. Given the values of the sample path $V(x_l)$ for $l = 0, \ldots, m+1$, the reset of the procedure to obtain $\mathcal{B}(x_l) = \mathcal{F}^{-1}V(x_l)$ is to consider the nonlinear equations $\mathcal{F}\left(u(x_l)\right) = V(x_l)$ for $l = 0, \ldots, m+1$ and deal with the linear operator $\mathcal{F}$ as a function of unknown variables $u(x_0), \ldots, u(x_{m+1})$, finding its inverse at points $V(x_0), \ldots, V(x_{m+1})$. For this aim, an appropriate numerical method is required that could be used to solve the system of $m+1$ nonlinear equations.

Figures 3 compares the behaviour of V and the corresponding $\mathcal{B}$ for two selectively different stochastic paths under two levels of censoring, 15% and 40%. Having assumed Gamma distribution with the respective parameters shape and rate equal to 4 and 2 as the target population, we obtained the corresponding functions for G, g and f_r. Given the share of censoring 15% ($p = 1 - 0.15$), we derived distribution functions $G^{\mathcal{C}}$ and $G^{\mathcal{U}}$. To generate V defined in (6), we simulated two independent Brownian bridges B_1 and B_2 along with a random variable Z from the standard normal distribution. We plugged all these quantities in (6) to complete generation of V. The corresponding $\mathcal{B} = \mathcal{F}^{-1}(V)$ was calculated via the procedure explained in the above paragraph. The respective paths of V and $\mathcal{B}$ are plotted in the upper-left diagram in Figure 3. Then, we changed the level of censoring to 40% ($p = 1 - 0.40$) and obtained the corresponding V and $\mathcal{B}$ exhibited in the upper-right diagram in Figure 3. The procedures were repeated to simulate different paths from V, shown in lower-left and -right diagrams in Figure 3. Comparing all plots, it was seen that V and $\mathcal{B}$ followed almost similar trends, although $\mathcal{B}$ had larger scale in comparison to V. There were minimal disparities between V and $\mathcal{B}$ under 15% censoring. However, the differences between two stochastic processes were revealed to increase markedly as the percentage of censoring rose to 40%. The fluctuations in V and $\mathcal{B}$ have slightly increased by increasing the proportions of censoring.

6.2. Asymptotic quantiles

Having generated paths from the stochastic process $\mathcal{B}$, we need to draw the corresponding paths from the stochastic integrals $\mathcal{S}$. The stochastic integrals U^i

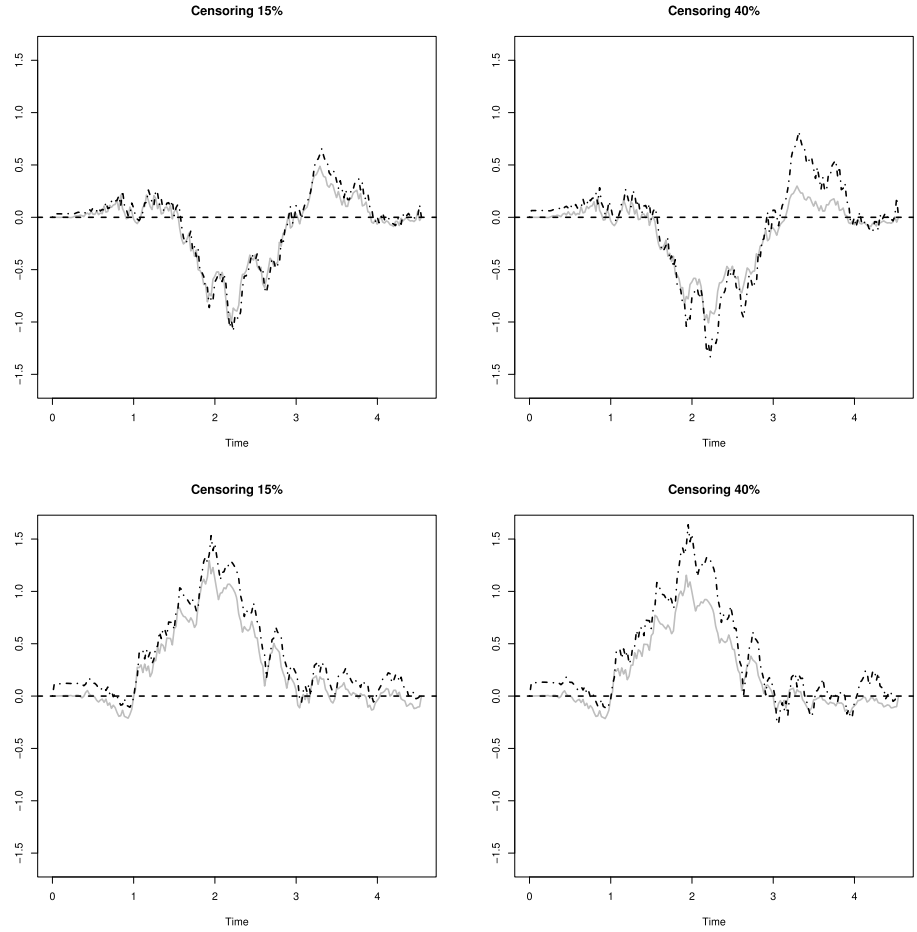

FIG 3. *Paths of the stochastic processes* V —— *and the corresponding paths of* $\mathcal{B}$ _.._..

are Gaussian processes that may be approximated by means of the equation

$$U^i(x) = \lim_{m\to\infty} \tilde{U}^i_m, \qquad (i = 1, 2),$$

where

$$\tilde{U}^i_m(x) := \sum_{l=1}^{m+1} \mathcal{I}^i_x(x_l) \left[\mathcal{B}(x_l) - \mathcal{B}(x_{l-1})\right]$$

and $\{x_l : l = 0, 1, \ldots, m+1\}$ is the partition of $[0, \tau]$ defined earlier in this section. Moreover, the limit is taken in quadratic mean over all the partitions such that, as $m \to \infty$,

$$\max_{1\le l\le m+1} (x_l - x_{l-1}) \longrightarrow 0.$$

The limiting distributions of $\sup \mathcal{S}$ could closely be approximated through the following procedure. First, we need to draw N paths from the stochastic process $\mathcal{B}$, and employ them to obtain the corresponding paths from the stochastic integral U^i. Suppose that $\{\tilde{U}^i_{m,j}(x); x \in [0,\tau]\}_{j=1}^N$ indicates N sample paths simulated from the stochastic integrals U^i, where m is the number of grids used to approximate U^i. For each $1 \leq l \leq m$, let $\tilde{\sigma}^i_N(x_l)$ denote the sample standard deviation of $\tilde{U}^i_{m,1}(x_l), \tilde{U}^i_{m,2}(x_l), \ldots, \tilde{U}^i_{m,N}(x_l)$. Define

$$\tilde{\mathcal{S}}^i_{N,m,j} := \max_{1\leq l\leq m} \frac{\tilde{U}^i_{m,j}(x_l)}{\tilde{\sigma}^i_N(x_l)}, \qquad (i = 1, 2),$$

where $j = 1, \ldots, N$. Afterwards, define $\tilde{q}^i_{N,m,\alpha}$ as the empirical upper α-quantile of $\tilde{\mathcal{S}}^i_{N,m,1}, \tilde{\mathcal{S}}^i_{N,m,2}, \ldots, \tilde{\mathcal{S}}^i_{N,m,N}$. Following [25], it is then derived that

$$\lim_{N\to\infty} \lim_{m\to\infty} P\left(\mathcal{S}^i_n(\cdot,\cdot) \leq \tilde{q}^i_{N,m,\alpha}\right) = P\left(\mathcal{S}^i_n(\cdot,\cdot) \leq q^i_\alpha\right).$$

Hence, $\tilde{q}^i_{N,m,\alpha}$ provides a close approximation of q^i_α for sufficiently large numbers of generated paths N and grids m. By choosing values $N = 1000$ and $m = 100$ to approximate the distribution of $\|\mathcal{S}^i\|_{[0,\tau]}$, our simulations indicated that the values of $\tilde{q}^i_{N,m,\alpha}$ remains unchanged against changes in underlying distribution function and level of censoring. The computational time for $N = 1000$ and $m = 200$ in a basic computer (CPU: Intel Core i7-7700) using *nleqslv* package in R for solving systems of m equations and parallel programming was about 5 minutes to reach the distributions of $\|\mathcal{S}^1\|_{[0,\tau]}$ and $\|\mathcal{S}^2\|_{[0,\tau]}$, simultaneously.

7. Variance function approximation

In order to estimate confidence bands $\mathcal{C}_{\theta^i}$, we also need to derive σ^i_n used in (11) for $i = 1, 2$. The right side of (12) is distribution free, and therefore can be approximated via the procedure discussed in Section 6 based on any parametric distribution. Given the intractable forms of U^i for $i = 1, 2$, it is not feasible to derive explicit forms for σ^i, and thus the procedure in Subsection 6.2 can be used which necessitates generating paths from U^i. However, we cannot directly apply such a procedure for σ^i_n because it requires unknown parameters of the population of interest. However, if there are stochastic processes $\hat{U}^i$ that do not rely on any unknown parameter and almost surely uniformly converges to U^i, the variance of $\hat{U}^i$ may be applied in practice instead. The objective of this section is to introduce such $\hat{U}^i$. For this aim, we need to find a process that almost surely uniformly converges to $\mathcal{B} = \mathcal{F}^{-1}(V)$ as for the first step.

Suppose that $\hat{f}_r$ is the estimator of f_r reached by plugging in the Vardi's estimator in (2). Let $\hat{G}^{\mathcal{U}}$ and $\hat{G}^{\mathcal{C}}$ denote the classical empirical distributions of $G^{\mathcal{U}}$ and $G^{\mathcal{C}}$ based on separate samples $Y_1, \ldots, Y_{N_1}$ and $Z_1, \ldots, Z_{N_2}$. Recall $\hat{p}$ defined in Section 2.3. Let define $\hat{V}$ as follow,

$$\hat{V}(x) := \hat{p}^{1/2} B_1\left(\hat{G}^{\mathcal{U}}(x)\right) + (1-\hat{p})^{1/2}\, \hat{f}_r(x) \int_{0<t\leq x} B_2\left(\hat{G}^{\mathcal{C}}(t)\right) d\frac{1}{\hat{f}_r(t)}$$

$$+\left(\frac{\hat{p}}{1-\hat{p}}\right)^{1/2}\left(\hat{G}^{\mathcal{U}}(x)-\hat{G}(x)\right)Z,$$

where the random normal variate Z and the Brownian bridges B_1 and B_2 are defined in (6). The below theorem is then followed.

Theorem 6. *For any $b \in \mathcal{J}$, it is proved that, as $n \to \infty$,*

$$\|\hat{V}-V\|_{[0,b]} \xrightarrow{a.s.} 0.$$

Proof. See Proofs. □

To propose a process that almost surely converges to $\mathcal{B}$, the following estimators should be defined. Let $\hat{g}^{\mathcal{U}}$ and $\hat{g}^{\mathcal{C}}$ denote the kernel density estimators of $g^{\mathcal{U}}$ and $g^{\mathcal{C}}$ as introduced by [27]. The weak and strong uniform consistency of $\hat{g}^{\mathcal{U}}$ and $\hat{g}^{\mathcal{C}}$ were studied in [22, 28]. Define the kernel density estimator of g by

$$\hat{g}(x) := \frac{1}{h}\int_0^{\infty} K\left(\frac{x-t}{h}\right)\hat{G}(dt), \tag{13}$$

where K is some kernel function, h is bandwidth and $\hat{G}$ is the Vardi's estimator. The asymptotic behavior of $\hat{g}$ under multiplicative censoring was studied by [5]. The following theorem proves the uniformly strong consistency of $\hat{g}$ for data collected in prevalent cohort studies under stationarity of incidence processes.

Theorem 7. *Suppose that h is a sequence of positive bandwidths tending to 0 as $n \to \infty$. Let K be a continuous function supported on $(-1,1)$ with $\int K(s)ds = 1$ and total variation $V_K < \infty$. Then, as $n \to \infty$,*

$$\|\hat{g}-g\|_{[0,\tau]} \xrightarrow{a.s.} 0.$$

Proof. See Proofs. □

Let $\hat{\beta}$ be the proportion of cases censored right after the recruitment including the subjects that are not followed as discussed in Subsection 2.3. Define

$$\hat{\mathcal{G}}_1(u)(x) := \int_{0<y\le x}\left((1-\hat{\beta})I_{[0,t_1)}(y)+\hat{p}\frac{\hat{g}^{\mathcal{U}}(y)}{\hat{g}(y)}I_{[t_1,t_k]}(y)\right)u(dy),$$

$$\hat{\mathcal{G}}_2(u)(x) := (1-\hat{p})\int_{0<z\le x} z\left(\int_{z\le t}\frac{u(t)}{t^2}dt\right)d\left[\left(\frac{\hat{f}_r(x)}{\hat{f}_r(z)}-1\right)\frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)}\right].$$

We can then define the following linear operator, $\hat{\mathcal{F}} : \mathcal{D}_0[0,\infty] \to \mathcal{D}_0[0,\infty]$,

$$\hat{\mathcal{F}} := \hat{\mathcal{G}}_1 + \hat{\mathcal{G}}_2.$$

$\hat{\mathcal{F}}$ and $\hat{V}$ can be employed to propose a process that almost surely uniformly converges to $\mathcal{B}$. In the next two theorems, we assume $\hat{g}$ used in $\hat{\mathcal{G}}^1$ is a uniformly consistent estimator of g, and thus the conditions of Theorem 7 hold.

Theorem 8. *Let define* $\hat{\mathcal{B}} := \hat{\mathcal{F}}^{-1}(\hat{V})$. *For any* $b \in \mathcal{J}$ *and sufficiently large* n,

$$\|\hat{\mathcal{B}} - \mathcal{B}\|_{[0,b]} \xrightarrow{a.s.} 0.$$

Proof. See Proofs. □

Let $\hat{\mathcal{I}}_x^i$ denote the estimator of $\mathcal{I}_x^i$ obtained by substituting $\hat{F}$, $\hat{\Lambda}$, $\hat{S}$ and $\hat{\mu}_{X'}$ for F, Λ, S and $\mu_{X'}$ in $\mathcal{I}_x^i$, for $i = 1, 2$. A process almost surely uniformly converging to U^i is given by

$$\hat{U}^i(x) := \int_0^\infty \hat{\mathcal{I}}_x^i(t)\hat{\mathcal{B}}(dt).$$

Theorem 9. *For any* $b \in \mathcal{J}$ *and sufficiently large* n, *it is derived under assumption (A1) that*

$$\|\hat{U}^i - U^i\|_{[0,b]} \xrightarrow{a.s.} 0.$$

Given Theorem 8, the proof of Theorem 9 is straightforward and so is omitted. Let denote $(\hat{\sigma}^i(x))^2 := Cov(\hat{U}^i(x), \hat{U}^i(x))$. Since processes U^i are mean zero Gaussian processes, it can be shown that

$$\|(\hat{\sigma}^i)^2 - (\sigma^i)^2\|_{[0,\tau]} \leq \int \left\|(\hat{U}^i)^2(\cdot,\omega) - (U^i)^2(\cdot,\omega)\right\|_{[0,\tau]} dP(\omega),$$

and hence, it is derived by means of the dominated convergence theorem and Theorem 9 that $(\hat{\sigma}^i)^2$ almost surely uniformly converges to $(\sigma^i)^2$.

However, given the intractable forms of the stochastic integrals $\hat{U}^i$, we can adopt the procedure given in Subsection 6.1 to simulate paths from the stochastic process $\hat{\mathcal{B}}$. Having generated paths from the process $\hat{\mathcal{B}}$, the stochastic integrals $\hat{U}^i$ can be approximated similar to the procedure given in Subsection 6.2. Let $\{\hat{U}^i_{m,j}(x); x \in [0,\tau]\}_{j=1}^N$ denote N paths simulated from the stochastic integrals $\hat{U}^i$, where m is the number of grids employed to approximate $\hat{U}^i$. For $1 \leq l \leq m$, let $\hat{\sigma}^i_N(x_l)$ denote the sample standard deviation of $\hat{U}^i_{m,1}(x_l), \ldots, \hat{U}^i_{m,N}(x_l)$. Then, $\hat{\sigma}^i_N(x_l)$ almost surely uniformly converges to σ^i.

8. Simulation

In this section, we conduct a comprehensive simulation study to illustrate the performance of NPMLEs and confidence bands and determine the empirical coverage probabilities of the proposed confidence bands $(\mathcal{C}_{\theta^i}, i = 1, 2)$ for small sample sizes. To simulate censored survival data emanating from a prevalent cohort under stationarity assumption of incidence, we first generated independent and identically distributed pairs of (X_i', T_i'), $i = 1, \ldots, n'$, where the failure times X_i' were simulated form the target population. The uniform distribution $U(0, \tau')$ was considered for truncation times T_i' to guarantee the stationarity assumption. In each scenario, the value of the parameter τ' was chosen such that $\tau' > \max_{1 \leq i \leq n'}(X_i')$. In addition, the value of n' in each repetition was selected large enough $(n' > n)$ to make sure that n subjects were observed. Only

n random cases from those pairs of (X_i', T_i') satisfying the condition $X_i' > T_i'$ were kept in the cohort, forming our prevalent cohort observations (X_i, T_i). The censoring variables C_i, $i = 1, \ldots, n$, were generated from a uniform distribution $U(0, \tau'')$, where τ'' in each scenario was considered such that the desired censoring percentage was reached on average for large number of repetitions, ensuring $\hat{p} \to p$ almost surely. Having imposed censoring, we considered $T_i + \min(R_i, C_i)$ for $i = 1, \ldots, n$ as our observations, where $R_i = X_i - T_i$.

To approximate the variance functions, we considered generating $N = 300$ paths from $\hat{U}^i$ to compute $\hat{\sigma}_N^i$ for $i = 1, 2$. The computational time to simultaneously reach $\hat{\sigma}_N^1$ and $\hat{\sigma}_N^2$ was just above 90 seconds for a personal computer (CPU: Intel Core i7-7700), which dived to a sixth after using parallel programming in R. In order to estimate $\hat{g}$, $\hat{g}^{\mathcal{U}}$ and $\hat{g}^{\mathcal{C}}$, the *Epanechnikov* kernel function were used. The Silverman's *rule-of-thumb* was used to obtain bandwidths h for $\hat{g}^{\mathcal{U}}$ and $\hat{g}^{\mathcal{C}}$. As for the bandwidth of $\hat{g}$ defined in (13), we have developed a cross-sectional procedure in an unpublished manuscript that may be used for censored prevalent cohort survival data. However, having approximated the optimal bandwidths one time, we considered a fixed value for h for each target population. One can also use the optimal bandwidth introduced in [5] for $\hat{g}$.

One of the objectives of simulation studies was to graphically evaluate the general adequacy of the estimators and elucidate the potential impact of increasing the censoring rate on their performance. For this purpose, two families of distribution functions, namely Weibull and Gamma, were considered with different choices of shape parameter k and scale parameter λ as the target populations of interest. To illustrate how the procedures perform, we considered the sample size $n = 400$, nominal level of confidence 95% and two levels of censoring 15% and 40% corresponding to $p = 1 - 0.15 = 0.85$ and $p = 0.6$, respectively. Bear in mind that the number of censored objects in each iteration was random although $\hat{p} \to p$ almost surely.

Figures 4 and 5 illustrate how the procedures performed for the target populations Weibull with parameters $k = 3$ and $\lambda = 0.5$ (Weibull $(3, .5)$) and Gamma with parameter $k = 4$ and $\lambda = 3$ (Gamma $(4, 3)$). Two diagrams at the top exhibit results for the cumulative hazard function while the two figures at the bottom show that for the survival function. Figures 4 and 5 also compare the performance of procedures under 15% censoring (two diagrams on the left) with that under 40% censoring (two figures on the right). The diagrams consisted in 100 iterations. The true cumulative hazard and survival curves were plotted in solid black lines. The average NPMLEs out of 100 repetitions were plotted with bullet points. All the 100 confidence bands estimated in each repetition of each setting were plotted in gray solid lines, shaping the gray clouds around the true curves. In each scenario, the average of upper-bound and lower-bound of the confidence bands out of 100 iterations were calculated separately and drawn using black dash-dotted lines. We found the NPMLEs proposed perform generally very well. As seen in the figures, the average of estimates of θ^1 and θ^2 for both target distributions located on the true curves perfectly under both levels of censoring. Although narrower and more precise, $\mathcal{C}_{\theta^1}$ for scenarios with 15% censoring touched the true hazard functions at the end in a few cases. All the ob-

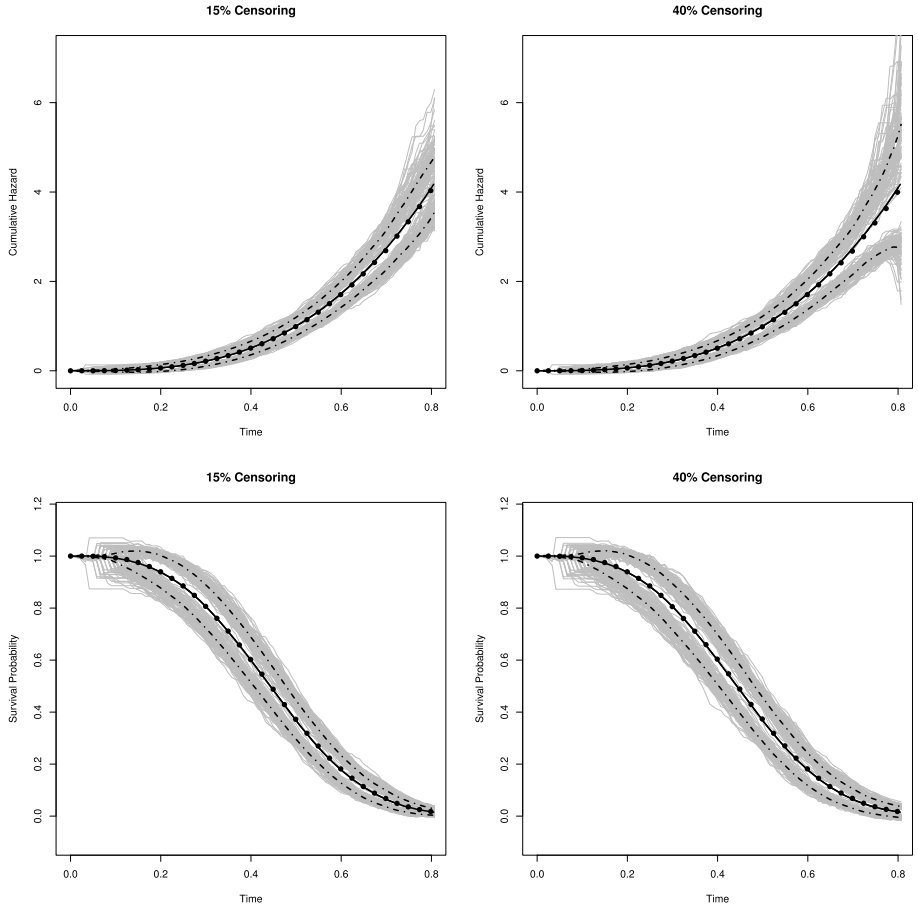

FIG 4. *Estimation for the cumulative hazard function and survival function for target population* $W(3, 0.5)$ *after adjustment for length-bis: True curves* ——, *average NPMLEs* •, *and confidence bands* ——, *average confidence bands* – – –.

served NPMLEs $\hat{\theta}^i$ were smooth, that have not been shown in the figures due to space limitations, although there were more fluctuations under 40% censoring.

According to Figures 4 and 5, the confidence bands for the cumulative hazard function ($\mathcal{C}_{\theta^1}$) became wider steadily by increasing in time. In contrast, despite an initial and marked surge, the widths of bands for the survival functions ($\mathcal{C}_{\theta^2}$) remained almost constant as the time increased before gradually narrowed at the end. In all sampling scenarios, the upper and lower limits of the confidence bands $\mathcal{C}_{\theta^1}$ and $\mathcal{C}_{\theta^2}$ touched the true curve in the left tails. It was observed that the proportions of the proposed confidence bands covering completely the hazard functions θ^1 in 15% censoring setting were between $94-97\%$, while those for survival functions θ^2 were slightly less by roughly 1% in three out of four scenarios. All confidence bands were very smooth and tight. However, compared

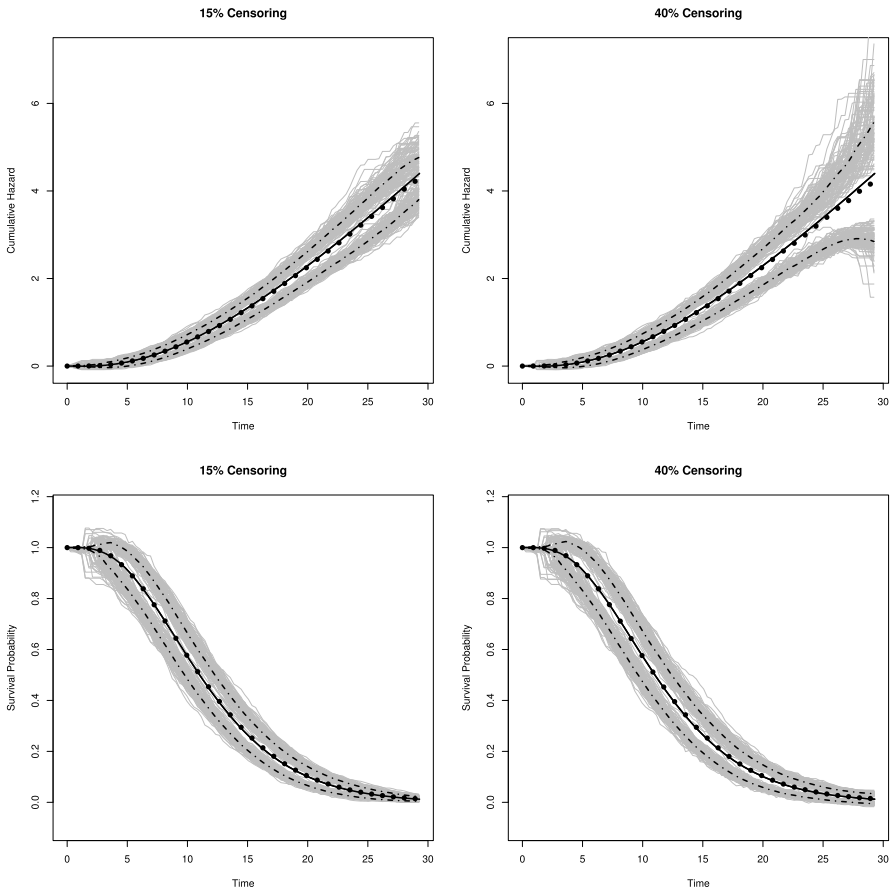

FIG 5. *Estimation for the cumulative hazard function and survival function for target population* $G(4,3)$ *after adjustment for length-bis: True curves* ——, *average NPMLEs* •, *and confidence bands* ——, *average confidence bands* – · – · –.

to 15% censoring scenarios, confidence bands became wider and fluctuated more in scenarios with 40% censoring.

The other aim of simulation studies was to inspect the performance of the confidence bands proposed for small sample sizes. For this objective, we considered sample sizes $n = 400$, 600 and 800 from two target populations $W(2,3)$ and $G(3,0.5)$. We also imposed three separate levels of censoring of 15%, 30% and 40% on data, representing minor, moderate and severe censoring. Also, the nominal level of confidence 95% was contemplated. The performance of the confidence bands were evaluated based on 1000 iterations of each case. The coverage probabilities were calculated as the proportion of confidence bands covering the real curve out of 1000 repetitions. It is worth mentioning that the real set of CSHA data that we analyse in Section 9 includes 826 cases with two types of dementia with censoring proportion just under 30%. Additionally, [10] reported

that CSHA data could be modelled by Weibull distribution function. Accordingly, we have considered simulation scenarios that are properly aligned with our real data structure.

In addition to $\tilde{\sigma}_N^i$ proposed in Section 7, we introduce a bootstrap method as an alternative method to estimate σ^i. By having q_α^i $(i = 1, 2)$, all we need for $\mathcal{C}_{\theta^i}$ is $\hat{\theta}^i$ and the bootstrap estimator given bellow. Therefore, one may find the bootstrap estimator easy to apply as the bootstrap method facilitates the procedure for estimating variance functions. Let $(A_j^*, R_j^* \wedge C_j^*, \delta_j^*)$, for $j = 1, \ldots, n$, be independent and identically distributed observations resampled from the original and fixed sample $(A_1, R_1 \wedge C_1, \delta_1), \ldots, (A_n, R_n \wedge C_n, \delta_n)$. Let $\hat{G}^*$ and $\hat{\theta}^{i*}$ denote the Vardi's estimator and the NPMLE of θ^i based on the bootstrap sample $(A_j^*, R_j^* \wedge C_j^*, \delta_j^*)$, where $j = 1, \ldots, n$ and $i = 1, 2$. Define $U_n^{i*} := \sqrt{n}(\hat{\theta}^{i*} - \hat{\theta}^i)$. By replicating the above resampling procedure B times, σ_n^{i*} is defined as the sample standard deviation of $U_{1n}^{i*}, \ldots U_{Bn}^{i*}$. Under assumption (A2), we have the uniform integrability property and therefore σ_n^{i*} is a uniform strong consistent estimator of σ^i. For the rest of this section, we considered $B = 500$ in our resampling procedures.

As seen in Figures 4 and 5, $\mathcal{C}_{\theta^i}$ tended to touch θ^i in the left tails for $i = 1, 2$. An obvious reason is that the confidence bands cannot cover the true parameters beyond the range of data, i.e. (t_1, t_k). It is also not surprising if $\hat{\sigma}_N^i$ and σ_n^{i*} cannot perform well in the tails. Estimation is even less efficient for small quantiles under length-bias owing to under-representation of small values of failure times, whereas estimation for large quantiles is noticeably more efficient due to over-representation of large values of survival times under length-biased sampling, which thus helps to avoid the estimation problems in the right tail. Therefore, we ignored any negligible intersection between θ^1 and $\mathcal{C}_{\theta^1}$ within the margin $[0, p_{05}]$ for sample size $n = 400$ in our computations to obtain the coverage probabilities, where p_{05} is the 5th-lower percentile of the target distribution. This problem gradually disappears as the sample size increases since $t_1 \to 0$ almost surely. To show that, the margins that we considered correspondingly for $n = 600$ and 800 were narrowed gradually to about $[0, p_{04}]$ and $[0, p_{03}]$, respectively. Our numerical results also indicated that $\mathcal{C}_{\theta^1}$ performs slightly better than $\mathcal{C}_{\theta^2}$ in the left tail. To compute the coverage probabilities of $\mathcal{C}_{\theta^2}$, the corresponding edges that we should consider were wider by a narrow margin of 1%, approximately.

Table 1 summarises the empirical coverage percentages observed for the proposed confidence bands $\mathcal{C}_{\theta^i}$ in covering the true parameters θ^i, for $i = 1, 2$. Generally, it was seen that both confidence bands performed well for all sample sizes using $\hat{\sigma}_N^i$. Considering $\mathcal{C}_{\theta^1}$ using $\hat{\sigma}_N^1$, the share of bands truly covered the hazard function accounted for $93.81 - 94.91\%$ for Weiblull distribution, while that for Gamma distribution were merely different with a range of $93.71 - 95.21\%$. The coverage probability of $\mathcal{C}_{\theta^1}$ grew minimally when the censoring level increased, except for Weibull distribution with $n = 400$. In comparison to $\hat{\sigma}_N^1$, the coverage probabilities of $\mathcal{C}_{\theta^1}$ using σ_n^{1*} were distinctly lower under 15% and 30% censoring when $n = 400$, but the performance of $\mathcal{C}_{\theta^1}$ improved gradually as the sample size increased, ranging from 91.02 to 94.31%. Contrastingly, the coverage prob-

TABLE 1
Empirical Coverage Probabilities of Confidence Bands (in %).

			W(2, 3)		**G(3, 0.5)**	
θ^i	**Size**	**Censoring**	σ_n^{i*}	$\hat{\sigma}_N^i$	σ_n^{i*}	$\hat{\sigma}_N^i$
$i=1$	400	0.15	91.82	94.31	91.92	93.71
		0.30	91.02	93.61	92.42	94.81
		0.40	90.62	94.31	91.32	94.51
	600	0.15	94.21	93.81	93.81	93.81
		0.30	93.21	94.41	94.31	94.61
		0.40	92.42	94.71	93.11	95.21
	800	0.15	93.51	94.31	93.01	94.11
		0.30	93.42	94.61	92.51	94.31
		0.40	91.42	94.91	91.62	94.51
$i=2$	400	0.15	91.42	93.21	92.31	93.21
		0.30	91.42	94.41	92.22	93.31
		0.40	90.12	94.51	91.52	94.51
	600	0.15	92.51	94.11	94.91	93.51
		0.30	92.81	94.61	92.92	94.81
		0.40	92.02	95.51	92.81	95.01
	800	0.15	93.01	94.11	95.11	94.31
		0.30	91.72	94.81	93.61	95.71
		0.40	91.02	95.71	91.92	95.01

ability remained relatively unchanged under 40% censoring, forming noticeably lower coverage probabilities between $90.12 - 92.81\%$.

Turning to $\mathcal{C}_{\theta^2}$ using $\hat{\sigma}_N^2$, the percentages of the confidence bands covering the true survival function accounted for 93.21 to 95.71. There was not any meaningful difference in the performance of $\mathcal{C}_{\theta^2}$ under two different target populations. The coverage rate of $\mathcal{C}_{\theta^2}$ using $\hat{\sigma}_N^2$ surged marginally by increasing the censoring proportion. By comparison with $\hat{\sigma}_N^2$, $\mathcal{C}_{\theta^2}$ using σ_n^{2*} showed almost acceptable results under 15% and 30% censoring although the performance of the former was still markedly better, as the coverage probabilities in most of the scenarios were closer to 95%. In contrast, $\mathcal{C}_{\theta^2}$ using σ_n^{2*} performed poorly for 40% censoring as the coverage probabilities deviated significantly from the nominal level. Our numerical results exposed that the bootstrap method was inadequate for estimating the variance function close to the minimum of data.

All in all, although both performed very well, $\mathcal{C}_{\theta^1}$ exhibited slightly better results in terms of coverage probabilities and coverage of θ^1 close to the origin while $\mathcal{C}_{\theta^2}$ formed tighter estimation on the right tail. By increasing the sample size, the confidence bands were capable of covering θ^i close to the origin. Larger sample sizes also resulted in considerably narrower bands and thus more precise estimation. Estimating confidence bands using $\hat{\sigma}_N^i$ revealed more accurate results compared to σ_n^{i*}. Confidence bands using $\hat{\sigma}_N^i$ could preserve the confidence level well. In contrast, the procedures based on σ_n^{*i} could not maintain the nominal level over the same ranges in scenarios with higher levels of censoring in particular. Following a comment by the first reviewer, we scrutinized all the steps of our simulation process under different sampling settings and we realized

that the obstacle in the performance of the confidence bands using σ_n^* is that the bootstrap variance does not perform well for small values of time. The bootstrap method may still be used if we have large enough samples or we can disregard inefficiencies of the bootstrap variance for survival times close to the minimum of observations. It is of note that while we proved Theorem 5 under assumption (A2), which is not a restrictive condition in examples of practical importance, the target distributions that we considered did not satisfy this assumption. Accordingly, the proposed confidence bands were found to be robust against violation of assumption (A2) for distributions satisfying assumption (A1).

9. Real data application

The Canadian Study of Health and Aging is a nationwide multicentre epidemiological study in geriatrics, including dementia and other health problems in elderly population of Canada commenced in 1991. Over the first phase of the study, 10,263 subjects of age 65 years or older among a random sample selected throughout Canada accepted to participate and be screened for dementia. Those who were diagnosed with dementia were split into three separate categories, namely *probable Alzheimer's disease*, *possible Alzheimer's disease* and *vascular dementia*. The date of onset of dementia for the subjects under study were collected from responsible carers. The prevalent cases cross-sectionally recruited were then followed over time until 1966 when the primary phase ended. The date of death or censoring for the subjects within the cohort were recorded during the period of follow-up study. Individuals who were still alive at the end of study were considered to be censored. The collection of data comprised 823 subjects with dementia. As the diagnosis of dementia was made on prevalent cases, the challenge posed was that the data formed a non-representative sample of the target population of interest. It could be reasonably assumed that the incident rate of dementia remained constant, and thus the stationarity assumption held and the collected data were length-biased. The data was initially studied by [42]. Thereafter, several authors have analysed the CSHA data from different perspectives. The validity of the stationarity assumption was inspected and verified by [1, 6, 9]. Of 823 subjects followed in the study, there were 20 individuals who had survived for almost 20 years or more. We excluded these subjects from the set of data as it is very unlikely that these subjects had dementia. Over the course of the follow-up study, just above 22% of the cases recruited were censored. The share of censored subjects remained relatively constant among the subcategories of dementia. We used survival times (in month) along with the corresponding censoring indicators for our analysis. We applied the method proposed in Section 7 with $N = 500$ generated paths to approximate the variance functions $\hat{\sigma}_N^i$ $(i = 1, 2)$ for all the confidence bands presented in this section.

Figure 6 compares the NPMLEs of the cumulative hazard function (left) and the survival function (right) between different subgroups of dementia. Despite fluctuations, the hazard rate, also known as the mortality rate in medical sciences, of possible Alzheimer's was remained relatively unchanged until the end as the slope of the curve was constant. Vascular dementia accounted for the

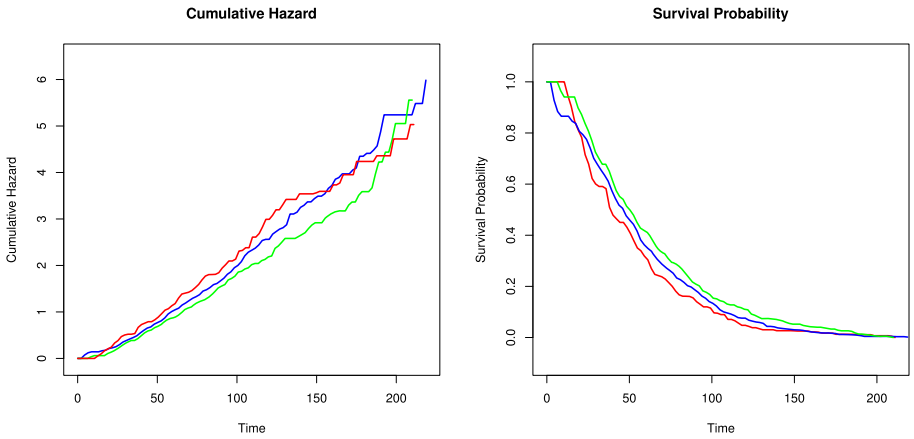

FIG 6. *NPMLEs of the cumulative hazard functions and the survival functions after adjustment for length-bis for cases with different types of dementia observed in CSHA: Possible Alzheimer's Disease* ——, *Probable Alzheimer's Disease* ——, *Vascular Dementia* ——.

highest rate of mortality among different types of dementia for 145 months after the onset of dementia, reaching the cumulative hazard of about 3.5. Thereafter, the subjects with vascular dementia experienced lower mortality rate as the curve was outstripped by possible Alzheimer's at around month 170 and later by probable Alzheimer's at the end. The NPMLE of probable Alzheimer's stood at last for almost first 200 months. The slope of the curve indicated cases in this group had the least mortality rate for 190 months when there was a substantial rise possibly because of natural causes of death in higher ages. Turning to the NPMLEs of the survival function, the overall trends of the estimated survival chance were aligned with those of the hazard rates. In spite of variations in first months, the probable and possible Alzheimer's had the respective first and second highest survival chance throughout the period of dementia, while vascular dementia had the least survival probability among these groups.

Figure 7 illustrates the NPMLEs along with the 95% confidence bands estimated for the cumulative hazard function and the survival function on CSHA on dementia including all subgroups. According to the hazard function diagram (left), the hazard rate were the lowest during the first 20 months after the onset of dementia. Afterwards, the hazard rate increased significantly and stood at the same level for the next 125 months as the slope of the curve remained almost constant in this period. After that, those subjects surviving roughly 145 months after the diagnosis of dementia, had experienced lower mortality rate until the month 180, approximately, when the hazard rate abruptly and dramatically increased. This was understandable to see the effect of the natural causes of demise due to aging in subject who survived beyond a certain point. Despite variations, this trend remained unchanged until the end. Considering the survival function diagram (right), the survival chance for subjects declined slowly for the first roughly 20 month before dropping sharply over the next

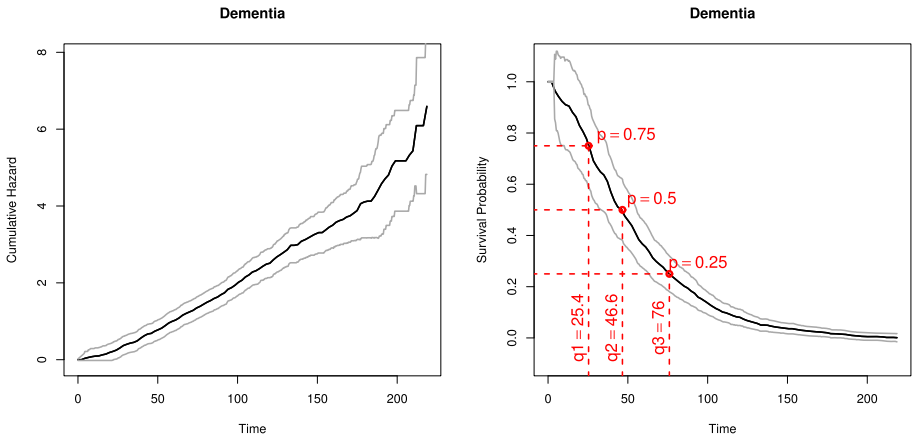

FIG 7. 95% *Confidence band and the NPMLE of the cumulative hazard function (left) and the survival function (right) for cases with dementia studied in CSHA after adjustment for length-bias: Confidence bands* ——, *NPMLES* ——.

more than 50 months. That was followed by a gradual fall until the end, hitting zero at about month 220. Furthermore, by locating the points on the NPMLE of survival curve intersecting lines survival probability (p) equal to 0.75, 0.5 and 0.25, we found the corresponding quartiles. According to our calculations, 25% of the cases at least died during the first 25.4 months after diagnosis with dementia ($q_1 = 25.4$), while only almost the same share of subject survived beyond 76 months ($q_3 = 76$). The second quartile, also known as the median of the sample, was equal to $q_2 = 46.6$ months. The confidence band for the hazard function widened steadily as the time increased, while that for the survival function dropped gradually. That was predictable because as the number of survivors decreases over time, the hazard rate vary more by each death observed, leading the variance of the NPMLE of the hazard function to grow.

The confidence bands presented in Figure 7 provide us with interval estimations for the lines θ^1 and θ^2. That in return may be applied to carry out the hypothesis tests $H_0 : \theta^1 = \theta_0^1$ and $H_0 : \theta^2 = \theta_0^2$ at the 5% significance level, for the Canadian elderly population with dementia. The confidence bands enable us to draw comparison between progression of dementia in different populations or to identify disease risk factors and assess the interventions in follow-up studies.

Figure 8 graphically presents 95% confidence bands for subjects in two separate groups of dementia, Alzheimer's disease and vascular dementia. The figure also illustrates confidence bands for Alzeimer's cases compartmentalised according to the clinical diagnoses of probable Alzheimer's and possible Alzheimer's. Comparing the NPMLE and the band for Alzheimer's with those for vascular dementia, the cases in the latter were revealed to experience a dramatically higher mortality rate. Accordingly, while the curve for cases with vascular dementia reached a peak of above 3.5 cumulative risk about 140 months after the diagnosis, that for the subjects with dementia gradually increased, hitting the

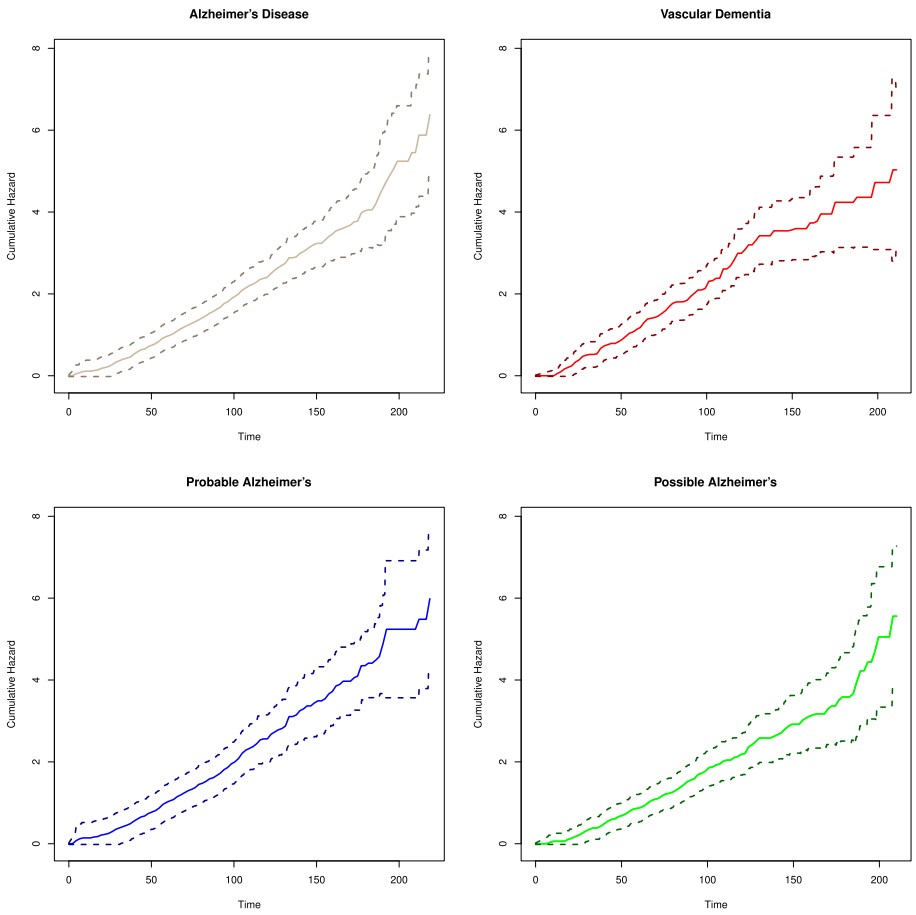

FIG 8. 95% *Confidence bands of the hazard functions for CSHA subjects from two main groups of dementia, Alzheimer's Disease* ____ *and Vascular Dementia* ____, *and two subgroups of Alzheimer's disease, Probable Alzheimer's* ____, *Possible Alzheimer's* ____.

same high after 175 months approximately. Then, although the hazard rate for vascular dementia decreased considerably, the curve for Alzheimer's continued the same trend for another 20 months before showing a remarkable increase in mortality rate for cases surviving beyond nearly 195 months due probably to the natural causes of death.

10. Proofs

To study the asymptotic behavior of the cumulative hazard function, the following representations are required. For any $x \in [0, t_k)$,

$$\hat{\Lambda}(x) - \Lambda(x) = \int_0^x \frac{(\hat{G} - G)(dt)}{t \int_t^\infty s^{-1} \hat{G}(ds)} - \int_0^x \frac{\int_t^\infty s^{-1}(\hat{G} - G)(ds)}{\int_t^\infty s^{-1} \hat{G}(ds)} \Lambda(dt). \qquad (14)$$

Also, for any $x \in [0, \tau)$,

$$\hat{\Lambda}(x) - \Lambda(x) = \int_0^x \frac{(\hat{G} - G)(dt)}{t \int_t^\infty s^{-1} G(ds)} - \int_0^x \frac{\int_t^\infty s^{-1}(\hat{G} - G)(ds)}{\int_t^\infty s^{-1} G(ds)} \hat{\Lambda}(dt). \tag{15}$$

Proof of Theorem 1. In view of (15), it is obtained by employing integration by parts for any arbitrary $x \in [0, \tau)$ that

$$\left|\hat{\Lambda}(x) - \Lambda(x)\right| \leq J_{1n}(x) + J_{2n}(x) + J_{3n}(x). \tag{16}$$

where

$$J_{1n}(x) = \left| \int_0^{x \wedge \gamma_{n_1,n_2}} \frac{\left(\hat{G} - G\right)(dt)}{t \int_t^\infty s^{-1} G(ds)} \right|,$$

$$J_{2n}(x) = \left| \int_{\gamma_{n_1,n_2}}^x \frac{\left(\hat{G} - G\right)(dt)}{t \int_t^\infty s^{-1} G(ds)} I_{[\gamma_{n_1,n_2},\tau)}(x) \right|$$

and

$$J_{3n}(x) = \left| \int_0^x \left(\frac{t^{-1}\left(\hat{G}(t) - G(t)\right) + \int_t^\infty \left(\hat{G}(s) - G(s)\right) d\left(\frac{1}{s}\right)}{\int_t^\infty s^{-1} G(ds)} \right) \hat{\Lambda}(dt) \right|.$$

To deal with $J_{1n}(x)$, we have

$$J_{1n}(x) \leq \left| \int_0^{\gamma_{n_1,n_2}} \frac{\mu_{X'} G(dt)}{t\left(1 - F_{X'}(t)\right)} \right| \leq \frac{F_{X'}(\gamma_{n_1,n_2})}{1 - F_{X'}(\gamma_{n_1,n_2})}.$$

Turning to $J_{2n}(x)$, it is derived that

$$\begin{aligned} J_{2n}(x) &\leq \left| \frac{\hat{G}(x) - G(x)}{x \int_x^\infty s^{-1} G(ds)} I_{[\gamma_{n_1,n_2},\tau)}(x) \right| + \left| \frac{\mu_{X'} G(\gamma_{n_1,n_2})}{\gamma_{n_1,n_2}\left(1 - F_{X'}(\gamma_{n_1,n_2})\right)} \right| \\ &\quad + \left| \int_{\gamma_{n_1,n_2}}^x \frac{\left|\hat{G}(t) - G(t)\right|}{\int_t^\infty s^{-1} G(ds)} d\left(\frac{1}{t}\right) I_{[\gamma_{n_1,n_2},\tau)}(x) \right| \\ &\quad + \left| \int_{\gamma_{n_1,n_2}}^x \frac{\left|\hat{G}(t) - G(t)\right|}{t^2 \left(\int_t^\infty s^{-1} G(ds)\right)^2} G(dt) I_{[\gamma_{n_1,n_2}.\tau)}(x) \right| \\ &\leq \left\|\hat{G} - G\right\|_{[0,\tau)} \left(\frac{I_{[0,\tau)(x)}}{1 - G(x)} \right) \left\{ \frac{2\tau + 1}{\gamma_{n_1,n_2}} + \|\Lambda\|_{[0,\tau)} \right\}. \end{aligned}$$

To deal with $J_{3n}(x)$, we need to prove the following relation in advance

$$\begin{aligned}
&\left|\frac{\int_x^\infty t^{-1}(G-\hat{G})(dt)}{\int_x^\infty t^{-1}\hat{G}(dt)}\right| \le 2\left(\frac{\tau-\gamma_{n_1,n_2}}{\gamma_{n_1,n_2}}\right)\left|\frac{I_{[\gamma_{n_1,n_2},t_k)}(x)}{1-\hat{G}(x)}\right| \|\hat{G}-G\|_{[0,\tau)} \\
+\quad &\tau\left\{\left|\int_x^{\gamma_{n_1,n_2}}\frac{1}{t}G(dt)\right|+\left|\frac{\int_{\gamma_{n_1,n_2}}^\infty t^{-1}\left(\hat{G}-G\right)(dt)}{1-\hat{G}(x)}\right|\right\} I_{[0,\gamma_{n_1,n_2})}(x) \\
\le\quad &\tau\left(\frac{F_{X'}(\gamma_{n_1,n_2})}{\mu_{X'}}\right)+2\left(\frac{\tau-\gamma_{n_1,n_2}}{\gamma_{n_1,n_2}}\right)\left|\frac{I_{[0,t_k)}(x)}{1-\hat{G}(x)}\right| \|\hat{G}-G\|_{[0,\tau)}. \qquad (17)
\end{aligned}$$

Given the definition of $J_{3n}(x)$, it can be shown that

$$\begin{aligned}
J_{3n}(x) \le\ &\tau I_{[0,t_k)}(x)\int_0^x\left(\frac{t^{-1}\left|\hat{G}(t)-G(t)\right|+\int_t^\infty\left|\hat{G}(s)-G(s)\right|s^{-2}ds}{t\int_t^\infty s^{-1}G(ds)\left(1-\hat{G}(t)\right)}\right)\hat{G}(dt) \\
&+\tau I_{[t_k,\tau)}(x) \\
&\times\int_0^x\left(\frac{t^{-1}\left|\hat{G}(t)-G(t)\right|+\int_t^\infty\left|\hat{G}(s)-G(s)\right|s^{-2}ds}{t\int_t^\infty s^{-1}\hat{G}(ds)\left(1-G(t)\right)}\right) I_{[0,t_k)}(t)\hat{G}(dt) \\
\le\quad &\tau\|\hat{G}-G\|_{[0,\tau)}\left(\frac{I_{[0,t_k)}(x)}{1-\hat{G}(x)}\right)\int_{\gamma_{n_1,n_2}}^x\frac{2}{t^2\int_t^\infty s^{-1}G(ds)}\hat{G}(dt) \\
&+\tau\|\hat{G}-G\|_{[0,\tau)}\left(\frac{I_{[t_k,\tau)}(x)}{1-G(x)}\right)\int_0^x\frac{2I_{[0,t_k)}(t)}{t^2\int_t^\infty s^{-1}\hat{G}(ds)}\hat{G}(dt) \\
\le\quad &2\tau\frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}}\left(\frac{I_{[0,t_k)}(x)}{1-\hat{G}(x)}\right)\left\{\left|\int_{\gamma_{n_1,n_2}}^x\frac{\left(\hat{G}-G\right)(dt)}{t\int_t^\infty s^{-1}G(ds)}\right|+\Lambda(x)\right\} \\
&+\tau\|\hat{G}-G\|_{[0,\tau)}\left(\frac{I_{[t_k,\tau)}(x)}{1-G(x)}\right) \\
&\times\int_0^x\frac{2I_{[0,t_k)}(t)}{t^2\int_t^\infty s^{-1}G(ds)}\left(1+\frac{\int_t^\infty s^{-1}(G-\hat{G})(ds)}{\int_t^\infty s^{-1}\hat{G}(ds)}\right)\hat{G}(dt) \\
\le\quad &2\tau\frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}}\left(\frac{I_{[0,t_k)}(x)}{1-\hat{G}(x)}\right)\left\{\|J_{2n}\|_{[0,\tau)}+\|\Lambda\|_{[0,\tau)}\right\} \\
&+2\tau\frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}}\left(\frac{I_{[t_k,\tau)}(x)}{1-G(x)}\right)\left\{\left|\int_{\gamma_{n_1,n_2}}^x\frac{\left(\hat{G}-G\right)(dt)}{t\int_t^\infty s^{-1}G(ds)}\right|+\Lambda(x)\right\} \\
&\times\left(1+\left\|\frac{\int_x^\infty t^{-1}(G-\hat{G})(dt)}{\int_x^\infty t^{-1}\hat{G}(dt)}\right\|_{[0,\tau)}\right)
\end{aligned}$$

$$\leq \quad 2\tau \frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}} \left(\frac{I_{[0,t_k)}(x)}{1-\hat{G}(x)} + \big(1+o(1)\big) \frac{I_{[t_k,\tau)}(x)}{1-G(x)} \right)$$
$$\times \Big\{ \|J_{2n}\|_{[0,\tau)} + \|\Lambda\|_{[0,\tau)} \Big\} \qquad a.s.,$$

where the last inequality is obtained by using (17). Considering (16), it is easy now to check that for any $b \in \mathcal{J}$

$$\|\hat{\Lambda}-\Lambda\|_{[0,b]} \quad \leq \quad \|J_{1n}\|_{[0,b]} + \|J_{2n}\|_{[0,b]} + \|J_{3n}\|_{[0,b]}. \tag{18}$$

Following that, owing to Theorem 1 of [8] alongside (16)–(18), it is obtained for any $b \in [0,\tau)$ and $b \in \mathcal{J}$ that

$$\|\hat{\Lambda}-\Lambda\|_{[0,b]} = \sup_{0\leq x\leq b} \left| \hat{\Lambda}(x) - \Lambda(x) \right| \xrightarrow{a.s.} 0.$$

Moreover, it is derived for any $b \in \mathcal{J}$ and $b \leq \tau - \epsilon$ that, if assumption (A1) holds,

$$\|\hat{\Lambda}-\Lambda\|_{[0,b]} = O\left(\frac{F_{X'}(\gamma_{n_1,n_2})}{1-F_{X'}(\gamma_{n_1,n_2})} + \gamma_{n_1,n_2}^{-1} \sqrt{\frac{\log\log n}{n}} \right) \qquad a.s.$$

Accordingly, it is seen that under assumption (A2)

$$\|\hat{\Lambda}-\Lambda\|_{[0,b]} \quad = \quad O\left(\sqrt{\frac{\log\log n}{n}} \right) \qquad a.s.,$$

where $b \in \mathcal{J}$ and $b \leq \tau - \epsilon$, and thus the desired results are concluded. □

Proof of Theorem 2. Considering (14), it is obtained by using Theorem 1 that

$$\hat{\Lambda}(x) - \Lambda(x) \quad = \quad \big(1+o_p(1)\big) \int_0^x \frac{\left(\hat{G}-G\right)(dt)}{t \int_t^\infty s^{-1} G(ds)}$$
$$- \int_0^x \frac{\int_t^\infty s^{-1} \left(\hat{G}-G\right)(ds)}{\int_t^\infty s^{-1} \hat{G}(ds)} \Lambda(dt). \tag{19}$$

uniformly over $[0,b]$, where $b \in \mathcal{J}$. To prove Theorem 2, we investigate the limiting behavior of the right side of (19) in two separate parts. As for the first part, let define the operator $\mathcal{P}_1 : \mathcal{D}_0[\gamma,\infty] \to \mathcal{D}_0[\gamma,\infty]$ with representation

$$\mathcal{P}_1(f)(x) := \frac{f(x)}{x} - \int_0^\infty f(t) I_{[\gamma,x)}(t) d\left(\frac{1}{t}\right) - \int_0^\infty \frac{f(t) I_{[\gamma,x)}(t)}{t^2 \int_{s\geq t} s^{-1} G(ds)} G(dt),$$

where $f \in \mathcal{D}_0[\gamma,\infty]$. $\mathcal{P}_1$ can be derived to be a bounded linear operator. To show that,

$$\|\mathcal{P}_1\| = \sup_{f:\|f\|\leq 1} \|\mathcal{P}_1(f)\|_{[\gamma,\infty]} \leq \left\{ \frac{2}{\gamma} + \frac{\Lambda(\tau)}{\gamma} \right\} < \infty,$$

and so $\mathcal{P}_1$ is a bounded linear operator. As a consequence, by substituting $\mu_{X'}\mathcal{B}_n/\left(1-F(\cdot)\right)$ for f in $\mathcal{P}_1$ and employing Theorem 2 of [8], it is reached through integration by parts that

$$\int_0^\infty \mathcal{I}_x^1(t,t)\mathcal{B}_n(dt) \quad \xrightarrow{\mathcal{W}} \quad \int_0^\infty \mathcal{I}_x^1(t,t)\mathcal{B}(dt), \tag{20}$$

where $\mathcal{B}_n$ and $\mathcal{B}$ where defined in Section 2.2.

Turning to the second part of the proof, let define the linear operator $\mathcal{P}_2$ such that $\mathcal{P}_2 : \mathcal{D}_0[\gamma,\infty] \to \mathcal{D}_0[\gamma,\infty]$ with representation

$$\mathcal{P}_2(f)(x) := \int_0^\infty \mathcal{H}_x(t)f(dt),$$

where $f \in \mathcal{D}_0[\gamma,\infty]$ and $\mathcal{H}_x$ is presented by

$$\mathcal{H}_x(t) := \frac{1}{t} I_{[x,\tau)}(t).$$

It then follows that

$$\begin{aligned}
\sup_{\gamma\le x\le a}\left|\int_0^\infty \mathcal{H}_x(t)f(dt)\right| &\le \sup_{\gamma\le x\le a}\left|\mathcal{H}_x(t)f(t)\right| + \sup_{\gamma\le x\le a}\int_0^\infty \left|\frac{\mathcal{H}_x(t)f(t)}{t}\right| dt \\
&\le \left\{\sup_{\gamma\le x\le a}\left|\mathcal{H}_x(t)\right| + \sup_{\gamma\le x\le a}\int_0^\infty \left|\frac{\mathcal{H}_x(t)}{t}\right| dt\right\} \|f\|_{[0,b]}.
\end{aligned}$$

Accordingly,

$$\begin{aligned}
\|\mathcal{P}_2\| &= \sup_{f:\|f\|\le 1} \|\mathcal{P}_2(f)\|_{[\gamma,\infty]} \\
&\le \left\{\sup_{\gamma\le x\le a}\left|\mathcal{H}_x(t)\right| + \sup_{\gamma\le x\le a}\int_0^\infty \left|\frac{\mathcal{H}_x(t)}{t}\right| dt\right\} < \infty,
\end{aligned} \tag{21}$$

that implies that $\mathcal{P}_2$ is a bounded linear operator. On the other hand, given Theorem 3, we have

$$\left|\hat{\mu}_{X'}^{-1}\hat{S}(\cdot)\right| \xrightarrow{\mathcal{P}} \left|\mu_{X'}^{-1}S(\cdot)\right| \tag{22}$$

uniformly over $[0,b]$, where $b \in \mathcal{J}$. Hence, (21) and (22) together imply that

$$g_n \xrightarrow{\mathcal{W}} g \qquad \text{in} \qquad \mathcal{D}_0[0,b], \tag{23}$$

where $b \in \mathcal{J}$,

$$g_n(x) := \frac{\hat{\mu}_{X'}}{\hat{S}(x)} \int_0^\infty \mathcal{H}_x(t)\mathcal{B}_n(dt)$$

and

$$g(x) = \frac{\mu_{X'}}{S(x)} \int_0^\infty \mathcal{H}_x(t)\mathcal{B}(dt).$$

Additionally, consider $\mathcal{P}_3 : \mathcal{D}_0[\gamma,\tau] \to \mathcal{D}_0[\gamma,\tau]$ such that

$$\mathcal{P}_3(f)(x) := \int_0^\infty f(t) I_{[0,x)}(t) \Lambda(dt),$$

and $f \in \mathcal{D}_0[\gamma,\tau]$. It is apparent that $\|\mathcal{P}_3\| \leq \|\Lambda\|_{[0,\tau]}$, and so $\mathcal{P}_3$ is also a bounded linear operator. Following that, let identify f with g_n in $\mathcal{P}_3$. Then, given (23), it is observed that

$$\mathcal{P}_3(g_n) \xrightarrow{\mathcal{W}} \mathcal{P}_3(g) \qquad \text{in} \quad \mathcal{D}_0[\gamma,b], \tag{24}$$

where $b \in \mathcal{J}$. Given (19), (20) and (24), it is revealed by means of integration by parts that

$$U_n^1 \xrightarrow{\mathcal{W}} U^1 \qquad \text{in} \qquad \mathcal{D}_0[\gamma,b],$$

where $b \in \mathcal{J}$ and U^1 is defined in Theorem 2, and so the desired result is achieved. □

Remark 1. *It was possible to use the Hadamard differentiability we prove latter in this section or apply G-Dosker classes given in [33] to prove the weak convergence presented in Theorem 2. However, such methods could not prove the weak convergence over the whole interval that the distribution function defined on. Also, the P-Glivenko-Cantelli classes could be applied to prove uniform strong consistency presented in Theorem 1. However, it would not prove result over the whole interval* $[0,\tau)$.

Proof of Theorem 3. To prove the strong uniform consistency of $\hat{S}$, it is derived under assumption (A1) for any $x \in [0,\tau)$ that

$$\begin{aligned}
\left|\hat{S}(x) - S(x)\right| &\leq \left|\int_x^\infty \frac{\mu_{X'}}{t}\left(\hat{G}-G\right)(dt)\right| + \left|\hat{S}(x)\int_0^\infty \frac{\mu_{X'}}{t}\left(\hat{G}-G\right)(dt)\right| \\
&\leq \left(1 + I_{[0,\gamma_{n_1,n_2}]}(x)\right)\left\{F_{X'}(\gamma_{n_1,n_2}) + 2\mu_{X'}\frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}}\right\} \\
&\quad + I_{[\gamma_{n_1,n_2},\tau)}(x) 2\mu_{X'}\frac{\|\hat{G}-G\|_{[0,\tau)}}{\gamma_{n_1,n_2}}.
\end{aligned} \tag{25}$$

Accordingly, the desired result is concluded from (25) along with Theorem 1 of [8]. □

Proof of Theorem 4. Let $\mathbb{D}_{\phi^1}$ denote the set of all nondecreasing cadlag functions $\bar{f} : [0,\eta] \to \mathbb{R}$ with $\bar{f}(0) = 0$ and $1 - \bar{f}(\eta) \geq \epsilon > 0$ for some $\epsilon > 0$. Under assumption (A2), we initially show that the map $\phi^1 : \mathbb{D}_{\phi^1} \subset \mathcal{D}_0[0,\eta] \to \mathcal{D}_0[0,\eta]$ that is defined by $\phi^1(G) = \theta^1$ in Section 4 is Hadamard differentiable. For this purpose, it is sufficient to prove the Hadamard differentiablity of the following maps:

$$G \to F_{X'} \to (F_{X'}, 1 - F_{X'}) \to \left(F_{X'}, \frac{1}{1-F_{X'}}\right) \to \int_0^\eta \frac{dF_{X'}}{1-F_{X'}}$$

Given the inverse of the representation (4), the Hadamard differentiablity of the first map can be shown under the assumption (A1). The differentiablity of the second and third maps is easy to prove, while that for the last map follows from Lemma 20.10 of [34]. By definition of θ^1, the Hadamard differentiablity of ϕ^1 is inferred by using the chain rule along with differentiablity of the each of the for maps given above.

Turning to θ^2, let $\mathbb{D}_{\phi^2}$ denote the set of all nondecreasing cadlag functions $\bar{f} : [0,\tau] \to [0,1]$ with $\bar{f}(0) = 0$ and $\bar{f}(\tau) = 1$. Recall the map ϕ^2 which was defined by $\phi^2(G) = \theta^2$ in Section 4. The Hadamard differentiablity of the map $\phi^2 : \mathbb{D}_{\phi^2} \subset \mathcal{D}_0[0,\tau] \to \mathcal{D}_0[0,\tau]$ under assumption (A2) is derived using analogous arguments to those used in the above paragraph:

$$G \to 1 - F_{X'}.$$

Hence, it is proved that $\hat{\theta}^i = \phi^i(\hat{G})$ for $i = 1$ and 2, are two Hadamard differentiable maps of the Vardi estimators $\hat{G}$ on the respective domains $[0,\eta]$ and $[0,\tau]$. Following that, by employing Theorem 25.47 of [34], the proof is complete. □

Proof of Theorem 5. Owing to the fact that σ_n^i is uniformly a strong consistent estimator of σ^i, the desired result for $i = 1$ is achieved by Theorem 2 in this paper along with Theorem 18.10 of [34]. Given (10), an analogous argument can be used to prove the theorem for $i = 2$. □

Proof of Corollary 1. Given Theorem 5 along with the definition of $\mathcal{S}^i(x_0, \theta^i(x_0))$ and $\sigma^i(x_0)$, it is apparent $\mathcal{S}^i(x_0, \theta^i(x_0))$ follows the standard normal distribution. □

Proof of Theorem 6. To prove this theorem, we need the following representation.

$$\begin{aligned}
\left|\hat{V}(x) - V(x)\right| :=& \left|\left(\hat{p}^{1/2} - p^{1/2}\right) B_1\left(\hat{G}^{\mathcal{U}}(x)\right)\right| \\
+ & \left|p^{1/2}\left[B_1\left(\hat{G}^{\mathcal{U}}(x)\right) - B_1\left(G^{\mathcal{U}}(x)\right)\right]\right| \\
+ & (1-p)^{1/2}\left[\left(\frac{1-\hat{p}}{1-p}\right)^{1/2}\hat{f}_r(x) - f_r(x)\right]\int_{0<t\le x} B_2\left(\hat{G}^{\mathcal{C}}(t)\right) d\frac{1}{\hat{f}_r(t)} \\
+ & (1-p)^{1/2} f_r(x)\left\{\int_{0<t\le x} B_2\left(\hat{G}^{\mathcal{C}}(t)\right) d\left[\frac{1}{\hat{f}_r(t)} - \frac{1}{f_r(t)}\right]\right. \\
+ & \left.\int_{0<t\le x}\left[B_2\left(\hat{G}^{\mathcal{C}}(t)\right) - B_2\left(G^{\mathcal{C}}(t)\right)\right] d\frac{1}{f_r(t)}\right\} \\
+ & \left[\left(\frac{\hat{p}}{1-\hat{p}}\right)^{1/2} - \left(\frac{p}{1-p}\right)^{1/2}\right]\left(\hat{G}^{\mathcal{U}}(x) - \hat{G}(x)\right) Z \\
+ & \left(\frac{p}{1-p}\right)^{1/2}\left[\left(\hat{G}^{\mathcal{U}}(x) - G^{\mathcal{U}}(x)\right) - \left(\hat{G}(x) - G(x)\right)\right] Z. \qquad (26)
\end{aligned}$$

Using Glivenko–Cantelli lemma, it is easy to check that $\hat{G}^{\mathcal{U}}$ and $\hat{G}^{\mathcal{C}}$ are uniformly consistent estimators of the respective distribution functions $G^{\mathcal{U}}$ and $G^{\mathcal{C}}$. It is then proved through Theorem 1.4.1 of [12] that, as $n \to \infty$,

$$\left\| B_1(\hat{G}^{\mathcal{U}}) - B_1(G^{\mathcal{U}}) \right\|_{[0,b]} \xrightarrow{a.s.} 0. \tag{27}$$

and

$$\left\| B_2(\hat{G}^{\mathcal{C}}) - B_2(G^{\mathcal{C}}) \right\|_{[0,b]} \xrightarrow{a.s.} 0. \tag{28}$$

Given (26), the proof is completed by employing (27) and (28) along with the uniformly strong consistency of the estimators $\hat{G}$, $\hat{f}_r$, $\hat{G}^{\mathcal{U}}$, $\hat{G}^{\mathcal{C}}$, and the strong consistency of $\hat{p}$ for G, f_r, $G^{\mathcal{U}}$, $G^{\mathcal{C}}$, and p, respectively. □

Proof of Theorem 7. To show that $\hat{g}$ is a uniformly consistent estimator of g, we need to define the following density function first,

$$g_K(x) := \frac{1}{h} \int_0^\infty K\left(\frac{x-t}{h}\right) G(dt).$$

The kernel density estimator $\hat{g}$ is proved to be a uniformly consistent estimator of g_K. For this aim, we have

$$\begin{aligned} \left|\hat{g}(x) - g_K(x)\right| &= \left| \frac{1}{h} \int_0^\infty \left(\hat{G}(t) - G(t)\right) dK\left(\frac{x-t}{h}\right) \right| \\ &\leq \|\hat{G} - G\|_{[0,\tau]} V_K, \end{aligned}$$

and hence, as $n \longrightarrow \infty$,

$$\|\hat{g} - g_K\|_{[0,\tau]} \xrightarrow{a.s.} 0. \tag{29}$$

Besides, considering the continuity, and thus the uniform continuity, of the density function g on $[0, \tau]$, the dominated convergence theorem can be applied to derive that, as $n \longrightarrow \infty$,

$$\|g_K - g\|_{[0,\tau]} \xrightarrow{a.s.} 0. \tag{30}$$

The proof is therefore completed by having (29) and (30). □

In order to prove Theorem 8, we need to present the following lemmas first. The first lemma shows that $\hat{\mathcal{F}}$ is a bounded and invertible linear operator with a bounded inverse.

Lemma 1. *If $b \in \mathcal{J}$, then for all sufficiently large enough n,*

i) $\hat{\mathcal{F}}$ is a bounded linear operator on $\mathcal{D}_0[0,b]$ such that

$$\|\hat{\mathcal{F}}\| \leq 1 \qquad a.s.$$

ii) $\hat{\mathcal{F}}$, under the conditions of Theorem 7, is an invertible linear operator on $\mathcal{D}_0[0,b]$ with

$$\|\hat{\mathcal{F}}^{-1}\| \leq \frac{(2/\alpha(b) - 1/(1-\beta))}{1 - 2(2/\alpha(b) - 1/(1-\beta))\beta} \qquad a.s.,$$

where β and the function α are defined in Section 2.3.

Proof. i) First, let define $\hat{\alpha}(x) := (1-\hat{\beta})I_{[0,t_1]}(x)+I_{[t_1,t_k]}(x)\hat{p}\hat{g}^{\mathcal{U}}(x)/\hat{g}(x)$. Then,

$$\begin{aligned}
\left|\hat{\alpha}(x)-\alpha(x)\right| \leq & \left|(\hat{p}-p)\frac{g^{\mathcal{U}}(x)}{g(x)}\left(\frac{g(x)-\hat{g}(x)}{\hat{g}(x)}+1\right)\right. \\
\times & \left(\frac{\hat{g}^{\mathcal{U}}(x)-g^{\mathcal{U}}(x)}{g^{\mathcal{U}}(x)}+1\right)I_{[t_1,t_k]}(x) \\
+ & \left. p\frac{g^{\mathcal{U}}(x)}{g(x)}\left[\left(\frac{g(x)-\hat{g}(x)}{\hat{g}(x)}+1\right)\left(\frac{\hat{g}^{\mathcal{U}}(x)-g^{\mathcal{U}}(x)}{g^{\mathcal{U}}(x)}+1\right)I_{[t_1,t_k]}(x)-1\right]\right| \\
+ & \left|(1-\hat{\beta})-\alpha(x)\right|I_{[0,t_1]}(x).
\end{aligned}$$

Given the uniform consistency of $\hat{g}^{\mathcal{U}}$ and $\hat{g}$ in Theorem 7, it is revealed that

$$\|\hat{\alpha}-\alpha\|_{[0,b]} \xrightarrow{a.s.} 0,$$

as $n \to \infty$, where $b \in \mathcal{J}$. Besides,

$$\hat{\mathcal{G}}_1(u)(x) = \int_{0<y\leq x} \hat{\alpha}(y)u(dy).$$

Since $\hat{g}$ and $\hat{g}^{\mathcal{U}}$ are functions of bounded variations and always positive on the domin of $\hat{\alpha}$, it is then resulted that $\hat{\alpha}$ is also a function of bounded variations. Given that $\hat{\alpha}$ is a uniform consistent estimator of α, using integration by parts, we have for sufficiently large n that

$$\|\hat{\mathcal{G}}_1\| \leq (1-\beta) \qquad a.s.$$

On the other hand, it can be shown by employing the integration by parts that

$$\begin{aligned}
\left|\hat{\mathcal{G}}_2(u)(x)\right| =& (1-\hat{p})\left|\int_{0<z\leq x}\left(1-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\left(\int_{z\leq s}\frac{u(s)}{s^2}ds\right)\frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)}dz\right. \\
& \left.+\int_{0<z\leq x}\left(1-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)}\left(\frac{u(z)}{z}\right)dz\right| \\
\leq & \left(\frac{1-\hat{p}}{1-p}\right)\left\|1+\left(\frac{f_r-\hat{f}_r}{\hat{f}_r}\right)\right\|_{[0,\infty]}\left\|1+\left(\frac{\hat{g}^{\mathcal{C}}-g^{\mathcal{C}}}{g^{\mathcal{C}}}\right)\right\|_{[0,\infty]}\|u\|_{[0,\infty]} \\
& \times\left|\beta+(1-p)\int_{0<z\leq x} z\left(\int_{z\leq s}\frac{1}{s^2}ds\right)d\left[\left(1-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\frac{g^{\mathcal{C}}(z)}{f_r(z)}\right]\right|.
\end{aligned}$$

Hence, considering the uniformly strong consistency of $\hat{f}_r$ and $\hat{g}^{\mathcal{C}}$, it is derived for sufficiently large n that

$$\|\hat{\mathcal{G}}_2\| \leq 2\beta \qquad a.s.$$

which completes the proof of part i).

ii) Turning to the second part of the lemma, we have

$$\hat{\mathcal{G}}_1^{-1}(u)(x) = \int_{0<y\leq x} \frac{I_{[0,t_k]}(y)}{\hat{\alpha}(y)} u(dy).$$

In consequence, it is derived under the conditions of Theorem 7 via Theorem C of [30] and an argument similar to that used in i) that, for sufficiently large n,

$$\|\hat{\mathcal{G}}_1^{-1}\| \leq \left(\frac{2}{\alpha(b)} - \frac{1}{1-\beta}\right) \qquad a.s.$$

Note that $\mathcal{L}(\mathcal{D}_0[0,b], \mathcal{D}_0[0,b])$ is a Banach algebra, and thus $\hat{\mathcal{F}}$ is invertible almost surely for large n. Actually,

$$\hat{\mathcal{F}} = \hat{\mathcal{G}}_1(I + \hat{\mathcal{G}}_1^{-1}\hat{\mathcal{G}}_2)$$

and hence

$$\hat{\mathcal{F}}^{-1} = (I + \hat{\mathcal{G}}_1^{-1}\hat{\mathcal{G}}_2)^{-1}\hat{\mathcal{G}}_1^{-1}.$$

It is then follows for sufficiently large n that

$$\begin{aligned} \|\hat{\mathcal{F}}^{-1}\| &\leq \|(I + \hat{\mathcal{G}}_1^{-1}\hat{\mathcal{G}}_2)^{-1}\|\|\hat{\mathcal{G}}_1^{-1}\| \\ &\leq \frac{(2/\alpha(b) - 1/(1-\beta))}{1 - 2\beta(2/\alpha(b) - 1/(1-\beta))}, \qquad a.s. \quad \square \end{aligned}$$

Lemma 2. *Suppose that $b \in \mathcal{J}$, it is then derived for sufficiently large n that*

$$\|\hat{\mathcal{F}}^{-1} - \mathcal{F}^{-1}\| \xrightarrow{a.s.} 0.$$

Proof. On the one hand, we have

$$\begin{aligned} \left|\hat{\mathcal{G}}_1(u)(x) - \mathcal{G}_1(u)(x)\right| &= \left|\int_{0<y\leq x} \hat{\alpha}(y)u(dy) - \int_{0<y\leq x} \alpha(y)u(dy)\right| \\ &\leq \|u\|_{[0,\infty]} \left|\hat{\alpha}(x) - \alpha(x)\right| + \left|\int_{0<y\leq x} u(y)(\alpha - \hat{\alpha})(dy)\right|. \end{aligned}$$

Given the characteristics of both α and $\hat{\alpha}$, it is obtained that

$$\|\hat{\mathcal{G}}_1 - \mathcal{G}_1\| \xrightarrow{a.s.} 0. \tag{31}$$

Besides, it is derived that

$$\begin{aligned} \left|\hat{\mathcal{G}}_2(u)(x) - \mathcal{G}_2(u)(x)\right| \leq |p - \hat{p}| \Bigg| &\int_{0<z\leq x} \left(1 - \frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\left(\int_{z\leq s} \frac{u(s)}{s^2} ds\right) \frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)} dz \\ + &\int_{0<z\leq x} \left(1 - \frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right) \frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)} \left(\frac{u(z)}{z}\right) dz \Bigg| \end{aligned}$$

$$
\begin{aligned}
&+(1-p)\left|\int_{0<z\leq x}\left(\int_{z\leq s}\frac{u(s)}{s^2}ds\right)\left(\frac{f_r(x)}{f_r(z)}-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\frac{g^{\mathcal{C}}(z)}{f_r(z)}dz\right.\\
&\left.+\int_{0<z\leq x}\left(\frac{f_r(x)}{f_r(z)}-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\frac{g^{\mathcal{C}}(z)}{f_r(z)}\left(\frac{u(z)}{z}\right)dz\right|\\
&+(1-p)\left|\int_{0<z\leq x}\left(\int_{z\leq s}\frac{u(s)}{s^2}ds\right)\left(1-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\left[\frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)}-\frac{g^{\mathcal{C}}(z)}{f_r(z)}\right]dz\right.\\
&\left.+\int_{0<z\leq x}\left(1-\frac{\hat{f}_r(x)}{\hat{f}_r(z)}\right)\left[\frac{\hat{g}^{\mathcal{C}}(z)}{\hat{f}_r(z)}-\frac{g^{\mathcal{C}}(z)}{f_r(z)}\right]\left(\frac{u(z)}{z}\right)dz\right|\\
\leq\ &\left[\left|\frac{p-\hat{p}}{1-\hat{p}}\right|\|\hat{\mathcal{G}}_2\|+(1-p)\left(\beta\|S-\hat{S}\|_{[0,\infty]}+\left\|\frac{\hat{g}^{\mathcal{C}}}{\hat{f}_r}-\frac{g^{\mathcal{C}}}{f_r}\right\|_{[0,\infty]}\right)\right]\|u\|_{[0,\infty]}.
\end{aligned}
$$

Following that, we have

$$\|\hat{\mathcal{G}}_2-\mathcal{G}_2\|\longrightarrow 0. \tag{32}$$

Given (31) and (32), it is derived that

$$\|\hat{\mathcal{F}}-\mathcal{F}\|\longrightarrow 0. \tag{33}$$

Furthermore, it is concluded using Proposition 2.3 on Page 11 of [46] that

$$\|\hat{\mathcal{F}}^{-1}-\mathcal{F}^{-1}\|\leq 2\|\mathcal{F}\|^2\|\hat{\mathcal{F}}-\mathcal{F}\|.$$

The above representation along with (33) and Lemma 1 complete the proof. □

Given Lemma 1 and 2, we can now present the proof of Theorem 8.

Proof of Theorem 8. To prove this theorem, it can be shown that

$$\|\hat{\mathcal{F}}^{-1}(\hat{V})-\mathcal{F}^{-1}(V)\|_{[0,b]}\ \leq\ \left\|\hat{\mathcal{F}}^{-1}\right\|\|\hat{V}-V\|_{[0,b]}+\left\|\hat{\mathcal{F}}^{-1}-\mathcal{F}^{-1}\right\|\|V\|_{[0,b]}.$$

Given the above inequality, the proof is therefore completed by employing Lemma 1 and 2, along with Theorem 6. □

Acknowledgments

The authors thank the Co-Editor-in-Chief, Professor Grace Y. Yi, and the anonymous referees for their insightful and constructive comments.

References

[1] Addona, V., Asgharian, M. and Wolfson, D. B. (2009). On the incidence-prevalence relation and length-biased sampling. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **37** 206–218. MR2531827

[2] Amiri, N., Fakoor, V., Sarmad, M. and Shariati, A. (2022). Empirical likelihood analysis for accelerated failure time model using length-biased data. *Statistics* **56** 578–597. MR4446622

[3] Andersen, P. E. R. K., Borgan, O., Gill, R. D. and Keiding, N. (1996). *Statistical Models Based on Counting Processes. Springer Series in Statistics*. Springer New York. MR1198884

[4] Asgharian, M. (2003). Biased sampling with right censoring: A note on sun, Cui & Tiwari (2002). *Canadian Journal of Statistics* **31** 349–350. https://doi.org/10.2307/3316091. MR2030129

[5] Asgharian, M., Carone, M. and Fakoor, V. (2012). Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.* **40** 159–187. MR3013183

[6] Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-Biased Sampling with Right Censoring: An Unconditional Approach. *Journal of the American Statistical Association* **97** 201–209. MR1947280

[7] Asgharian, M., Wolfson, C. and Wolfson, D. B. (2014). Analysis of biased survival data: the canadian study of health and aging and beyond. *Stat Action Can Outlook* **193**. MR3241976

[8] Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics* **33** 2109–2131. MR2211081

[9] Asgharian, M., Wolfson, D. B. and Zhang, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine* **25** 1751–1767. MR2227351

[10] Bergeron, P.-J., Asgharian, M. and Wolfson, D. B. (2008). Covariate Bias Induced by Length-Biased Sampling of Failure Times. *Journal of the American Statistical Association* **103** 737–742. MR2524006

[11] Cox, D. R. (1962). *Renewal Theory. Methuen's monographs on applied probability and statistics*. Methuen. MR0153061

[12] Csörgo, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics. Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. Academic Press. https://doi.org/10.1016/B978-0-12-198540-0.50003-2. MR0666546

[13] de Uña-Álvarez, J. (2002). Product-limit Estimation for Length-biased Censored Data. *Test* **11** 109–125. MR1915780

[14] Efromovich, S. and Chu, J. (2018). Hazard rate estimation for left truncated and right censored data. *Annals of the Institute of Statistical Mathematics* **70** 889–917. MR3830291

[15] Gilbert, P. B., Lele, S. R. and Vardi, Y. (1999). Maximum Likelihood Estimation in Semiparametric Selection Bias Models with Application to AIDS Vaccine Trials. *Biometrika* **86** 27–43. MR1688069

[16] Gill, R. D., Vardi, Y. and Wellner, J. A. (1988). Large Sample Theory of Empirical Distributions in Biased Sampling Models. *Ann. Statist.* **16** 1069–1112. MR0959189

[17] Huang, Y. and Wang, M.-C. (1995). Estimating the Occurrence Rate

for Prevalent Survival Data in Competing Risks Models. *Journal of the American Statistical Association* **90** 1406–1415. MR1379484

[18] Hyde, J. (1980). Survival Analysis with Incomplete Observations. *Biostatistics Casebook, R. G. Miller, B. Efron, B. W. Brown, and L. E. Moses, eds.* **30** 31–46.

[19] Lagakos, S. W., Barraj, L. M. and Gruttola, V. D. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75** 515–523. MR0967591

[20] Lynden-Bell, D. (1971). A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars. *Monthly Notices of the Royal Astronomical Society* **155** 95–118.

[21] McVittie, J. H. and Asgharian, M. (2022). A Nonparametric Test for Equality of Survival Medians Using Right-Censored Prevalent Cohort Survival Data. *to be appeared in: Statistical Methods in Medical Research.* MR4513309

[22] Nadaraya, E. A. (1965). On Non-Parametric Estimates of Density Functions and Regression Curves. *Theory of Probability & Its Applications* **10** 186–190. MR0172400

[23] Ning, J., Qin, J. and Shen, Y. (2010). Non-parametric tests for right-censored data with biased sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 609–630. MR2758238

[24] Parthasarathy, K. R. (1967). *Probability measures on metric spaces.* Academic Press New York. MR0226684

[25] Parzen, E. (1980). Quantile Functions, Convergence in Quantile, and Extreme Value Distribution Theory. Technical Report, Texas A and M University College Station Institute of Statistics.

[26] Rabhi, Y. and Asgharian, M. (2017). Inference under biased sampling and right censoring for a change point in the hazard function. *Bernoulli* **23** 2720–2745. MR3648043

[27] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* **27** 832–837. MR0079873

[28] Schuster, E. F. (1969). Estimation of a Probability Density Function and Its Derivatives. *The Annals of Mathematical Statistics* **40** 1187–1195. MR0247723

[29] Shen, J. and He, S. (2006). Empirical likelihood for the difference of two survival functions under right censorship. *Statistics & Probability Letters* **76** 169–181. https://doi.org/10.1016/j.spl.2005.07.018. MR2233389

[30] Silverman, B. W. (1978). Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. *The Annals of Statistics* **6** 177–184. MR0471166

[31] Tsai, W.-Y., Jewell, N. P. and Wang, M.-C. (1987). A Note on the Product-Limit Estimator Under Right Censoring and Left Truncation. *Biometrika* **74** 883–886.

[32] Turnbull, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal*

Statistical Society. Series B (Methodological) **38** 290–295. MR0652727

[33] VAN DER VAART, A. W. (1996). New Donsker Classes. *The Annals of Probability* **24** 2128–2140. MR1415244

[34] VAN DER VAART, A. W. (2000). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press. MR1652247

[35] VARDI, Y. (1982). Nonparametric Estimation in Renewal Processes. *The Annals of Statistics* **10** 772–785. MR0663431

[36] VARDI, Y. (1985). Empirical Distributions in Selection Bias Models. *The Annals of Statistics* **13** 178–203. MR0773161

[37] VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

[38] VARDI, Y. and ZHANG, C.-H. (1992). Large Sample Study of Empirical Distributions in a Random-Multiplicative Censoring Model. *The Annals of Statistics* **20** 1022–1039. MR1165604

[39] WANG, M. C. (1991). Nonparametric Estimation from Cross-Sectional Survival Data. *Journal of the American Statistical Association* **86** 130–143. MR1137104

[40] WANG, M. C., BROOKMEYER, R. and JEWELL, N. P. (1993). Statistical Models for Prevalent Cohort Data. *Biometrics* **49** 1–11. MR1221402

[41] WANG, M.-C., JEWELL, N. P. and TSAI, W.-Y. (1986). Asymptotic Properties of the Product Limit Estimate Under Random Truncation. *Ann. Statist.* **14** 1597–1605. MR0868322

[42] WOLFSON, C., WOLFSON, D. B., ASGHARIAN, M., M'LAN, C. E., ØSTBYE, T., ROCKWOOD, K. and HOGAN, D. B. (2001). A Reevaluation of the Duration of Survival after the Onset of Dementia. *New England Journal of Medicine* **344** 1111–1116. PMID: 11297701.

[43] ZELEN, M. (2004). Forward and Backward Recurrence Times and Length Biased Sampling: Age Specific Models. *Lifetime Data Analysis* **10** 325–334. MR2125419

[44] ZELEN, M. and FEINLEIB, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56** 601–614. MR0258224

[45] ZHOU, M. (2016). *Empirical likelihood method in survival analysis. Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, Florida. MR3616660

[46] ZHU, K. (1993). *An introduction to operator algebras* **9**. CRC press. MR1257406

[47] ZHU, S., YANG, Y. and ZHOU, M. (2015). A Note on the Empirical Likelihood Confidence Band for Hazards Ratio with Covariate Adjustment. *Biometrics* **71** 859–863. MR3402622