Title

Semiparametric sieve maximum likelihood estimation under cure model with partly interval censored and left truncated data for application to spontaneous abortion

Journal

Lifetime Data Analysis, 25(3)

ISSN

1380-7870

Authors

Wu, Yuan
Chambers, Christina D
Xu, Ronghui

Publication Date

2019-07-01

DOI

10.1007/s10985-018-9445-4

Peer reviewed

HHS Public Access

Author manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

Lifetime Data Anal. 2019 July ; 25(3): 507–528. doi:10.1007/s10985-018-9445-4.

Semiparametric sieve maximum likelihood estimation under cure model with partly interval censored and left truncated data for application to spontaneous abortion

Yuan Wu[1], **Christina D. Chambers**[2,3], and **Ronghui Xu**[3,4]

[1]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

[2]Department of Pediatrics, University of California, San Diego, USA

[3]Department of Family Medicine and Public Health, University of California, San Diego, USA

[4]Department of Mathematics, University of California, San Diego, USA

Abstract

This work was motivated by observational studies in pregnancy with spontaneous abortion (SAB) as outcome. Clearly some women experience the SAB event but the rest do not. In addition, the data are left truncated due to the way pregnant women are recruited into these studies. For those women who do experience SAB, their exact event times are sometimes unknown. Finally, a small percentage of the women are lost to follow-up during their pregnancy. All these give rise to data that are left truncated, partly interval and right-censored, and with a clearly defined cured portion. We consider the non-mixture Cox regression cure rate model and adopt the semiparametric spline-based sieve maximum likelihood approach to analyze such data. Using modern empirical process theory we show that both the parametric and the nonparametric parts of the sieve estimator are consistent, and we establish the asymptotic normality for both parts. Simulation studies are conducted to establish the finite sample performance. Finally, we apply our method to a database of observational studies on spontaneous abortion.

Keywords

Empirical process; Generalized gradient projection algorithm; Spline function

1 Introduction

Our work was motivated by research work carried out at the Organization of Teratology Information Specialists (OTIS), which is a North American network of university or hospital based teratology services that counsel between 70,000 and 100,000 pregnant women every year. Research subjects are enrolled from the Teratology Information Services and through

Correspondence to: Yuan Wu.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10985-018–9445-4) contains supplementary material, which is available to authorized users.

Lemmas and Technical Proofs (.pdf file): This supplemental material contains the proofs of Theorems 1–4. This material also contains some technical lemmas including their proofs.

other methods of recruitment, where the mothers and their babies are followed over time. Recently it has been of interest to assess the effects of medication and vaccine exposures on spontaneous abortion (SAB). By definition SAB occurs before week 20 of gestation; any pregnancy loss after that is called still birth. Ultimately we would like to know if an exposure modifies the risk of SAB for a woman, which may be increased or decreased. It is known that in the population for clinically recognized pregnancies the rate of SAB is about 12%. On the other hand, in our database the empirical SAB rate is consistently lower than 10%. This is due to the fact that women may enter a study any time before week 20 in gestation. The fact that we don't observe the women from the start of their pregnancy is known as left truncation in survival analysis; it reflects the selection bias in that women who have early SAB events can be seen as less likely to be in our studies. In addition, a substantial portion of the SAB events do not have an exact known date, rather a window during which it occurred is typically available. This is known as interval censoring in survival analysis. Finally, the fact that the majority of the pregnant women are free of SAB is considered 'cured' in the time-to-event context.

Like in other clinical studies our data also have right-censoring due to loss to follow-up before 20 weeks of gestation. The typical survival analysis models assume that all subjects in the study population will eventually experience the event of interest, at least if they are not lost to follow-up. When this is not the case, in the literature researchers have proposed mixture and non-mixture cure models to deal with the situation. Mixture cure models have parts for the cure rate and the hazard function of the uncured subjects separately. The most popular semiparametric mixture cure rate model adopts logistic regression for the cure rate and Cox regression for the hazard rate. For example, Sy and Taylor (2000) proposed the estimation under this model for right-censored data. Ma (2010) proposed the estimation under the same model for interval censored data, and Lam and Xue (2005) and Hu and Xiang (2016) adopted the sieve approach to ease computation for interval censored data. While mixture cure models might have useful interpretation in practice, in our own experiences they also pose challenges because there are two parameters for each covariate from the two parts, including for the main exposure of interest. Such models are also computationally complex, and may have numerical identification issues with real data.

Non-mixture cure models have become popular for analyzing population with a well defined cured portion. An immediate advantage of such models is the simple formulation, and its similarity with commonly used survival methods such as the Cox regression model (Tsodikov 1998). Chen et al. (1999) showed that non-mixture cure models follow the biological process associated with cancer clinical trials, and proposed a semiparametric method for a non-mixture cure model based on the Cox regression for right-censored data. Zeng et al. (2006) further extended the Cox regression to general transformation models. For interval censored data, Liu and Shen (2009) proposed a semiparametric method under the non-mixture cure model structure and established consistency of their estimator, and Hu and Xiang (2013) adopted the sieve approach for the nonparametric part and besides consistency they also established the asymptotic normality for the parametric part of the model.

In practical data analysis using cure models, a predetermined follow-up time window is often used to identify the observed cured subjects, see for example, Sy and Taylor (2000)

and Zeng et al. (2006). The end point of the follow-up window is called cure threshold by Zeng et al. (2006), and it is assumed that most or all events will occur before the cure threshold. In some applications, the cure threshold may be naturally defined related to the events of interest. For example, spontaneous abortion (SAB) mentioned earlier is only defined as pregnancy loss before week 20 of gestation, and subjects without such events before week 20 are clearly "cured" for SAB. Therefore, the cure threshold is naturally defined as week 20 in this case. In this way, the cure model is also a natural candidate tool for analyzing this type of data.

The fact that the SAB data consist of both interval censored and exactly observed event times is referred as partly interval censored and actually occurs very often in practice. Another example of partly interval censored data is progression free survival (PFS) time in clinical trials, because PFS time is defined as the smaller of death and progression times which are usually right-censored and interval censored, respectively. Intuitively the asymptotic results for the maximum likelihood estimation (MLE) under the Cox model with partly interval censored data will be the same as those for the MLE with right-censored data in terms of convergence rate, since for both partly interval censored and right-censored data the likelihood function will be dominated by the term with observed events. However, Kim (2003b) pointed out that if the interval censored observations are naively ignored from the whole data set, both estimation bias and standard error will be enlarged. Hence, a method correctly addressing this type of complex data is needed. Unfortunately we have found no published work on cure rate model with partly interval censored data in the presence of left truncation, which is the case for the SAB data application that we will describe in more details in Sect. 7. We will consider the sieve approach which has shown efficiency in computation for both nonparametric and semiparametric survival analysis problems under smoothness assumptions, and has variance estimator readily available. Ramsay (1988) has observed that closely related to the well-known B-splines, there are so-called M-splines and I-splines, where the M-splines are the derivatives of the I-splines. In the following we will use the B-spline form for theoretical developments, and the M-spline and I-spline form for simplicity of computing.

To our best knowledge, this work is the first attempt to provide an approach for analyzing complex survival data that are partly interval censored, left truncated and with a cured portion. The paper is organized as follows. Section 2 proposes the semi-parametric sieve MLE for the non-mixture Cox model when data are left truncated, partly interval censored and with a cured portion. The reason for us to work with a non-mixture cure model is partially due to the relative simplicity of the model structure, makes computing and theoretical development more manageable for our complex data. Section 3 provides all the asymptotic results for both the parametric part and the nonparametric part including consistency and asymptotic normality. The asymptotic normality for the nonparametric part is established for a smooth functional of the sieve estimator. Section 4 describes the computational method for the proposed sieve MLE. Section 5 finds the estimator for the variance of both the parametric and the nonparametric part. Section 6 conducts simulation studies to verify the finite sample performance for the proposed method. Section 7 applies the proposed methodology to analyze an observational data set on spontaneous abortion. Section 8 summaries the theoretical and numerical results and discusses a few cases when

the data structure is simplified, and mentions some potential future work. In Supplementary Material we provide proofs for all theorems in this paper with necessary lemmas using modern empirical process theory.

2 Semiparametric sieve MLE.

Consider the non-mixture cure model proposed by Chen et al. (1999), in which the survival function of the event time T given covariates $\mathbf{Z} = \mathbf{z} \in \mathbb{R}^d$ is $S(t|z) = \exp\left\{-e^{\tilde{\boldsymbol{\beta}}'\tilde{z}}F(t)\right\}$, where $\tilde{\boldsymbol{\beta}} = \left(\varphi_0, \boldsymbol{\beta}'\right)'$ is a vector of regression parameter contains an intercept $\varphi 0$ and d-dimensional vector $\boldsymbol{\beta}, \tilde{z} = (1, z')'$, and $0 \le F(t) \le 1$ is a distribution function. Let $F(\boldsymbol{\tau}) = 1$ and in the following we focus on the case when $\boldsymbol{\tau} < \infty$; for example, $F(\cdot)$ could be a right truncated exponential distribution function or a uniform distribution. Since the survival function here does not decrease beyond $\boldsymbol{\tau}$, there are no subjects with $T > \boldsymbol{\tau}$ and the cure threshold is naturally equal to $\boldsymbol{\tau}$ (Zeng et al. 2006). Because $S(\boldsymbol{\tau}|z) < \infty$, there is a positive proportion of subjects who do not experience the event within the time window $(0, \boldsymbol{\tau})$, and these are considered cured subjects. That is, a subject is cured implies that $T = \boldsymbol{\tau}$ $(< \infty)$. For our SAB data described in the Introduction, $\boldsymbol{\tau}$ equals week 20 of gestation, i.e. if a woman's pregnancy survived to week 20 then she was 'cured' from SAB.

Write $\Lambda(t) = F(t)\exp(\varphi_0)$, which represents the baseline cumulative hazard for the non-mixture cure model. Note that Λ and (F, φ_0) have a one-to-one correspondence and $0 \le \Lambda(t) \le \exp(\varphi_0)$ since F is a distribution function and has maximum of one. The baseline cure rate is $S(\boldsymbol{\tau}|\mathbf{0}) = \exp\{-e^{\varphi 0}\}$.

In the following we rewrite the above non-mixture cure model as

$$S(t|z) = \exp\left\{-e^{\boldsymbol{\beta}'Z}\Lambda(t)\right\}, \quad (1)$$

where $0 \le \Lambda(t) \le A(\boldsymbol{\tau}) = \exp(\varphi 0)$. While (1) appears the same as the regular Cox regression model, it is different in that the baseline cumulative hazard in the regular Cox model approaches infinity as there are no cured subjects, but here $\Lambda(t)$ is bounded. It is easy to show that (1) is identifiable:Suppose $\exp\left\{-e^{\boldsymbol{\beta}'z}\Lambda(t)\right\} = \exp\left\{-e^{\tilde{\boldsymbol{\beta}}'z}\widetilde{\Lambda}(t)\right\}$ for all t. Let $z = \mathbf{0}$, we have $\Lambda(\cdot) = \widetilde{\Lambda}(\cdot)$. Then $\boldsymbol{\beta}'z = \tilde{\boldsymbol{\beta}}'z$. Since $\mathbf{z}$ is arbitrary, we also have $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$.

Let Q be the left truncation time on $[0, \boldsymbol{\tau}_1]$ with $0 < \boldsymbol{\tau}_1 \le t$. And let $[U, V]$ be the observation interval on $[Q, \boldsymbol{\tau}]$, where U and V may be both equal to $\boldsymbol{\tau}$. We assume that T and $[U, V]$ are independent given $\mathbf{Z}$ and Q, and T and Q are independent given $\mathbf{Z}$. With left truncated data the baseline cumulative hazard $\Lambda(\cdot)$ may not be reliably estimated due to the lack of observations near time zero. In this paper we will show that nonetheless the functional increase of the baseline cumulative hazard from a "nonzero" point can be still accurately estimated. We note that since $\exp(\varphi_0) = \Lambda(\boldsymbol{\tau})$, it will be also hard to estimate φ_0 and therefore the (baseline) cure rate with left truncated data.

Denote $\Delta_1 = I_{[U<T\leq V]}$ for interval censoring, $\Delta_2 = I_{[T>V]}$ for right-censoring and $\Delta_3 = I_{[T\leq U]}$ for observed events. Let $\tilde{X} = T\,\Delta_3 + (U,V)\,\Delta_1 + V\,\Delta_2$, then $\boldsymbol{X} = (\tilde{X}, Q, \boldsymbol{Z}, \Delta_1, \Delta_2, \Delta_3)$ is the observed vector of random variables. Let $\lambda(t)$ satisfy $\Lambda(t) = \int_0^t \lambda(u)du$. The log-likelihood of an i.i.d. sample $\boldsymbol{x}_i = (\tilde{x}_i, q_i, z_i, \delta_{1,i}, \delta_{3,i})$ with $\tilde{x}_i = t_i\delta_{i,3} + (u_i, v_i)\delta_{1,i} + v_i\delta_{2,i}$ for $i = 1, \ldots, n$, based on the cure model (1) is

$$l_n(\boldsymbol{\beta}, \lambda; \cdot) = \sum_{i=1}^{n} \delta_{1,i} \log\left(\exp\left[-e^{\boldsymbol{\beta}' Z_i}\left\{\Lambda(u_i) - \Lambda(q_i)\right\}\right] - \exp\left[-e^{\boldsymbol{\beta}' Z_i}\left\{\Lambda(v_i) - \Lambda(q_i)\right\}\right]\right) + \sum_{i=1}^{n} \delta_{2,i}\left[-e^{\boldsymbol{\beta}' Z_i}\left\{\Lambda(v_i) - \Lambda(q_i)\right\}\right] + \sum_{i=1}^{n} \delta_{3,i}\left[-e^{\boldsymbol{\beta}' Z_i}\left\{\Lambda(t_i) - \Lambda(q_i)\right\} + \boldsymbol{\beta}' z_i + \log\lambda(t_i)\right], \quad (2)$$

by omitting the additive terms that do not involve (β, λ).

The optimization of the above log-likelihood can be very challenging, as the semi-parametric MLE approach would discretize λ into point masses at each distinct observed event time, and under the continuous distribution assumption the number of distinct observations is comparable to the sample size. We will then have to maximize (2) with a very large number of parameters when the sample size is large. To ease the computational difficulties for these type of estimation problems, Geman and Hwang (1982) proposed a sieve maximum likelihood estimation procedure. The main idea of the sieve method is to maximize the likelihood with much fewer variables in a subclass that "approximates" to the original function space. In addition, Huang et al. (2008) established that the sieve method provides an easy way to compute the observed information matrix. In the following the sieve maximum likelihood estimation is proposed for the non-mixture cure model with partly interval censored and left truncated data.

Let the B-spline basis functions of order l be $\{B_j^l(t)\}_{j=1}^{p_n}$ with knot sequence $\{\xi_j\}_{j=1}^{p_n+l}$ satisfying

$$0 = \xi_1 = \cdots = \xi_l < \xi_{l+1} < \cdots < \xi_{p_n} < \xi_{p_n+1} = \cdots = \xi_{p_n+l} = \tau,$$

where $p_n = O(n^k)$ for $k < 1$. with $\{B_j^l(t)\}_{j=1}^{p_n}$, define

$$\Psi_n = \left\{ \lambda_n = \sum_{j=1}^{p_n} \alpha_j B_j^l : \alpha_j \geq 0 \text{ for } j = 1, \ldots, p_n \right\}.$$

The requirement for all coefficients being nonnegative will guarantee that Ψ_n only contains nonnegative function for approximating the space of smooth hazard functions on $[0, \tau]$.

If λ is replaced by λ_n in (2) we have the log-likelihood function as

$$l_n(\boldsymbol{\beta}, \lambda_n ; \cdot) = \sum_{i=1}^{n} \delta_{1,i} \log\left(\exp\left[-e^{\boldsymbol{\beta}' \mathcal{Z}_i} \left\{ \int_{qi}^{u_i} \sum_{j=1}^{p_n} \alpha_j B_j^l(t) dt \right\} \right] \right. \tag{3}$$
$$\left. - \exp\left[-e^{\boldsymbol{\beta}' \mathcal{Z}_i} \left\{ \int_{qi}^{v_i} \sum_{j=1}^{p_n} \alpha_j B_j^l(t) dt \right\} \right] \right) + \sum_{i=1}^{n} \delta_{2,i} \left[-e^{\boldsymbol{\beta}' \mathcal{Z}_i} \left\{ \int_{qi}^{v_i} \sum_{j=1}^{p_n} \alpha_j B_j^l(t) dt \right\} \right]$$
$$+ \sum_{i=1}^{n} \delta_{3,i} \left[-e^{\boldsymbol{\beta}' \mathcal{Z}_i} \left\{ \int_{qi}^{t_i} \sum_{j=1}^{p_n} \alpha_j B_j^l(t) dt \right\} + \boldsymbol{\beta}' z_i + \log\left\{ \sum_{j=1}^{p_n} \alpha_j B_j^l(t_i) \right\} \right].$$

The sieve maximum likelihood estimation is obtained through maximizing the log-likelihood function (3) in terms of $(\boldsymbol{\beta}, \lambda_n)$. Note that the sieve MLE could have good asymptotic properties if Ψ_n "approximates" the space of nonnegative functions.

3 Asymptotic properties

In this section, we describe the asymptotic properties of the proposed sieve semi-parametric MLE. Study of the asymptotic properties of the proposed sieve estimator needs empirical process theory and requires some regularity conditions, regarding the event time, the observation time, the truncation time and the covariates. Denote $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \lambda_0)$ as the true parameter and baseline hazard function. The following conditions sufficiently guarantee the results in the forthcoming theorems.

C1 Covariate variable $\mathbf{Z}$ is bounded, that is, there exists a scalar z_0 such that $|\mathbf{Z}| < z_0$. Here $|\cdot|$ denotes Euclidean norm. $\mathbb{B}$ is a compact set in $\mathbb{R}^d$ and includes $\boldsymbol{\beta}_0$ in its interior.

C2 For the true baseline cumulative hazard $\Lambda_0(\cdot)$ for T, let $\lambda_0(\cdot)$ satisfy $\Lambda_0(t) = \int_0^t \lambda_0 \, du$. Then $\lambda_0(\cdot)$ has a positive lower bound and bounded pth derivative $\lambda_0^{(p)}(\cdot)$ on $[0, \tau]$.

C3 For $0 < \tau_d < \tau$, (U, V) are continuously distributed for $U \in [Q, \tau - \tau_d]$ and $V \in [U + \tau_d, \tau]$. In addition, (U, V) has some probability mass at (τ, τ).

C4 The joint density function of $(U, V, Q, \mathbf{Z})$, $f_{U, V, Q, \mathbf{Z}}$ has a positive lower bound and an upper bound for $u \in [q, \tau - \tau_d]$ and $v \in [u + \tau_d, \tau]$, and for $0 \leq q \leq \tau$ and $q \leq \tilde{u} \leq \tau$

$$\lim_{h_q \to 0, \boldsymbol{h}_z \to \mathbf{0}} \frac{\Pr(q \le Q \le q + h_q, z \le \mathbf{Z} \le z + \boldsymbol{h}_z, U \ge \tilde{u})}{h_q \prod_{i=1}^{d} h_z^{(i)}} = \int_{\tilde{u}}^{\tau - \tau_d} \int_{u + \tau_d}^{\tau} f_{U,V,Q,\mathbf{Z}}(u, v, q, z) dv du + f^{(0)}_{Q,\mathbf{Z}}(q, z),$$

where $\boldsymbol{h}_z = (h_z^{(1)}, \ldots, h_z^{(d)})' f^{(0)}_{Q,z}$ is the joint probability density and mass function of $(Q, \mathbf{Z}, U)$ at $U = \boldsymbol{\tau}$ and has positive lower and upper bounds.

C5 For some $\eta \in (0, 1), \boldsymbol{a}' Var(\mathbf{Z} | T, Q) \boldsymbol{a} \le \eta \boldsymbol{a}' E(\mathbf{Z}\mathbf{Z}' | T, Q) \boldsymbol{a}$ for all $\boldsymbol{a} \in \mathbb{R}^d$.

Remark 1 Condition C2 implies that $\lambda_0(t)$ is bounded on $[0, \boldsymbol{\tau}]$ and hence the survival rate of T at $\boldsymbol{\tau}$ is not 0, which is assumed by the cure rate model. Condition C2 also implies that the first derivative of $\lambda_0(t)$ is bounded on $[0, \boldsymbol{\tau}]$, which is necessary to apply the result of Example 19.10 in Vaart (1998) in the proof of consistency. Condition C3 guarantees the interval censored term in the likelihood function to be bounded, and U and V both equal to $\boldsymbol{\tau}$ for a significant portion of observations. Condition C4 implies that the Q is continuous and the density function of $(Q, \mathbf{Z})$ is bounded on $[0, \boldsymbol{\tau}]$, though U has point mass at $\boldsymbol{\tau}$. Condition C1, C2 and C4 imply that the density functions of T, U and V all have positive lower bounds and hence the data structure is truly partly interval censored including significant portions of observed events, interval censored events and right-censored events. Condition C5 will be used similarly as C13 and C14 in Wellner and Zhang (2007).

Before stating our main theorems, we define some notations. For the knot sequence $\{\xi_j\}_{j=1}^{p_n + l}$ previously defined for Ψ_n with $p_n = O(n^k)$ for $\kappa < 1$, further let $\max_j \Delta_j = \max_{j=l,\ldots,p_n}(\xi_{j+1} - \xi_j)$ and $\min_j \Delta_j = \min_{j=l,\ldots,p_n}(\xi_{j+1} - \xi_j)$. Then, with $\{\xi_j\}_{j=1}^{p_n + l}$ we define

$$\mathfrak{F}_n = \left\{ \lambda_n \sum_{j=1}^{p_n} \alpha_j B_j^l : a_0 \le \alpha_j \le K \tau b_0 \text{ for } j = 1, \ldots, p_n, \frac{|\alpha_{j+1} - \alpha_j|}{\max_j \Delta_j} \le K^2 d_0 \text{ for } j = 1, \ldots, p_n - 1, \int_0^{\tau} \lambda_n(t) dt \le \tau b_0, \frac{\max_j \Delta_j}{\min_j \Delta_j} \text{ has a upper bound independent of } n \right\}, \quad (4)$$

where a_0, b_0 and d_0 satisfy $a_0 \le \lambda_0(t) \le b_0$ and $|\lambda_0'(t)| \le d_0$ on $[0, \boldsymbol{\tau}]$, K is a large positive number for relaxing the constraints on $\mathfrak{F}_n$ in finite sample computing as discussed in Sect. 4.

Note that a_0, b_0 and d_0 do exist given C2. Then $\mathfrak{F}_n \subset \Psi_n$. Note that Ψ_n is a general space of nonnegative spline functions, and for the theoretical developments some regularity

conditions are necessary to form $\mathfrak{F}_n$. We also let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)$ with $\boldsymbol{\beta} \in \mathbb{B}$ and $\lambda \in \mathfrak{F}_n$. Then $\boldsymbol{\theta} \in \Omega_n = (\mathbb{B}, \mathfrak{F}_n)$. We denote $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}, \hat{\lambda}_n)$ the maximize of $l_n(\boldsymbol{\theta},)$ over Ω_n.

Define $\| \cdot \|_{L_2(v)}$ the norm associated with the joint probability measure $v(t, q)$ for (T, Q) based on the fact that $T \geq Q$, as $\|f\|_{L_2(v)} = \left\{ \int_0^\tau \int_q^\tau f^2(t) dv(t, q) \right\}^{1/2}$. Then we could define the distance between $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \lambda_1)$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \lambda_2)$ as

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left(|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|^2 + \| \lambda_1 - \lambda_2 \|_{L_2(v)}^2 \right)^{1/2}.$$

For one single observation $\boldsymbol{x}$ from the random observation $\boldsymbol{X}$ and a general semi-parametric variable $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)$, the likelihood (after removing terms unrelated to $\boldsymbol{\theta}$) is given by

$$l(\boldsymbol{\theta}; \boldsymbol{x}) = \delta_1 \log\left(\exp\left[-e^{\boldsymbol{\beta}^T z} \{ \Lambda(u) - \Lambda(q) \} \right] - \exp\left[-e^{\boldsymbol{\beta}^T z} \{ \Lambda(v) - \Lambda(q) \} \right] \right) + \delta_2 \left[-e^{\boldsymbol{\beta}^T z} \{ \Lambda(v) - \Lambda(q) \} \right]$$
$$+ \delta_3 \left[-e^{\boldsymbol{\beta}^T z} \{ \Lambda(t) - \Lambda(q) \} + \boldsymbol{\beta}^T z + \log \lambda(t) \right].$$

We denote $\mathbb{M}(\theta) = Pl(\boldsymbol{\theta}; \boldsymbol{x})$ with P being the true joint probability measure of $\boldsymbol{X}$, and $\mathbb{M}_n(\boldsymbol{\theta}) = \mathbb{P}_n l(\boldsymbol{\theta}; \boldsymbol{x})$ with $\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(y_i)$ the empirical process indexed by f. In what follows, we first show the consistency of the proposed estimator and establish the rate of convergence.

Theorem 1 *Suppose that C1–C5 hold, then* $\hat{\boldsymbol{\theta}}_n$ *is a consistent estimator for* $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \lambda_0)$ *and*

$$d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\left(n^{-\min\{p\kappa, (1-\kappa)/2\}} \right).$$

Remark 2 This theorem implies that for $\kappa = 1/(1+2p)$, $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\left(n^{-p/(1+2p)} \right)$. For any fixed q and t with $0 < q \leq \tau_1$ and $q < t \leq \tau$, Lemma 4 in the supplemental material implies that $\left[\int_q^t \{ \hat{\lambda}_n(s) - \lambda_0(s) \}^2 ds \right]^{1/2} < c d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)$. Hence, the estimation for the baseline hazard at any "non-zero" point is fine and the functional increase $\Lambda_0(t) - \Lambda_0(q)$ can be consistently estimated by the proposed sieve MLE. This is similar to the theoretical result for the estimated baseline hazard based on left truncated interval censored data in Kim (2003a).

Now we present the asymptotic normality for the proposed estimator including the parametric part and the smooth functional of the nonparametric part. Consider a parametric smooth submodel with parameter $(\boldsymbol{\beta}, \lambda_{(s,h)})$, with $\lambda_{(s,h)} = \lambda + sh$, then $\lambda_{(0,h)} = \lambda$, $\left. \frac{\partial \lambda_{(s,h)}}{\partial s} \right|_{s=0} = h$ and $\left. \frac{\partial \phi(\lambda_{(s,h)})}{\partial s} \right|_{s=0} = \phi_\lambda(\lambda)[h]$ for a functional $\phi(\cdot)$. Let $\mathscr{H}$ be the

class of functions h defined by this equation. The score operator for λ with h is the directional derivative at λ along h:

$$l_{\lambda}(\boldsymbol{\theta};\boldsymbol{x})[h] = \frac{\partial}{\partial_s} l\left(\boldsymbol{\beta},\ \lambda_{(s,h)};\boldsymbol{x}\right)\Big|_{s=0} \equiv f_h(\boldsymbol{\beta},\ \lambda;\boldsymbol{x}).$$

And the two times directional derivative at λ along h_1 and h_2 is

$$l_{\lambda,\lambda}(\boldsymbol{\theta};\boldsymbol{x})[h_1][h_2] = \frac{\partial}{\partial_s} f h_1(\boldsymbol{\beta},\ \lambda_{(s,h_2)};\boldsymbol{x})\Big|_{s=0}.$$

In addition, for $\boldsymbol{h} = \left(h_{1,\ldots,} h_d\right)'$ with $h_s \in \mathscr{H}$ for $s = 1, \ldots, d$, let $l_{\lambda}(\boldsymbol{\theta}; \mathbf{x})[\boldsymbol{h}]$ be the d-dimensional vector with its sth element $l_{\lambda}(\boldsymbol{\theta}; \boldsymbol{x})[h_s]$. For $h_1 = (h_{1,1}, \ldots, h_{1,d})'$ and $h_2 = (h_{2,1}, \ldots, h_{2,d})'$, let $l_{\lambda,\lambda}(\boldsymbol{\theta}; \mathbf{x})[h_1][h_2]$ be the $d \times d$ matrix with its ith row j th column element $l_{\lambda,\lambda}(\boldsymbol{\theta}; \boldsymbol{x})[h_{1,i}][h_{2,j}]$.

For the d-dimensional $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_d\right)'$, let $l_{\boldsymbol{\beta}}(\boldsymbol{\theta};\boldsymbol{x}) = \left\{l_{\beta_1}(\boldsymbol{\theta};\boldsymbol{x}), \ldots, l_{\beta_d}(\boldsymbol{\theta};\boldsymbol{x})\right\}'$, where $l_{\beta s}(\theta; x)$ is the partial derivative of $l(\boldsymbol{\theta}; \boldsymbol{x})$ with respect to β_s, s = 1 , . . . , d. Denote $\phi_s(\boldsymbol{\theta}, h) = \left\{l_{\beta s}(\boldsymbol{\theta};\boldsymbol{x}) - l_{\lambda}(\boldsymbol{\theta};\boldsymbol{x})[h]\right\}^2$ for $s = 1$, . . . , d. If $h_s^* = \arg\min_{h \in \mathscr{H}} p\,\phi_s\left(\boldsymbol{\theta}_0,\ h\right)$, then by Theorem 1 on page 70 in Bickel et al. (1993) the efficient score for $\boldsymbol{\beta}_0$ is $l_{\boldsymbol{\beta}}(\boldsymbol{\theta}; \mathbf{x}) - l_{\lambda}(\boldsymbol{\theta}; \boldsymbol{x})[\boldsymbol{h}^*]$ with $\boldsymbol{h}^* = (h_1^*, \ldots, h_d^*)'$. Let $\boldsymbol{\phi}(\boldsymbol{\theta}, \boldsymbol{h}) = \{l_{\boldsymbol{\beta}}(\boldsymbol{\theta}; \boldsymbol{x}) - l_{\lambda}(\boldsymbol{\theta}; \mathbf{x})[\boldsymbol{h}]\}^{\otimes 2}$, where $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}'$ for a column vector $\boldsymbol{v}$. Then the information matrix for $\boldsymbol{\beta}_0$ is given by

$$I(\boldsymbol{\beta}_0) = P\phi(\boldsymbol{\theta}_0, \boldsymbol{h}^*). \quad (5)$$

Theorem 2 *Suppose that C1–C5 hold,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = n^{-1/2} I^{-1}(\boldsymbol{\beta}_0) \sum_{i=1}^{n} l^*(\boldsymbol{\theta}_0; \boldsymbol{x}_i) + o_P(1),$$

Where $l^*(\boldsymbol{\theta}; \boldsymbol{x}) = l_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0; \boldsymbol{x}) - l_{\lambda}(\boldsymbol{\theta}_0; \boldsymbol{x})[\boldsymbol{h}^*]$. *That is,* $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \to_d N\left(0,\ I^{-1}\left(\boldsymbol{\beta}_0\right)\right)$ *by the central limit theorem.*

Since the convergence rate we established is slower than $1/\sqrt{n}$, the asymptotic normality is not easy to obtain for $\hat{\lambda}_n(\cdot)$, the nonparametric part of the sieve MLE. However it can still be shown that the asymptotic normality is available for its smooth functional $\rho\left(\hat{\boldsymbol{\theta}}_n\right) = \int_q^t \hat{\lambda}_n$, which is the plug-in estimator of $\Lambda_0(t) - \Lambda_0(q)$ for any fixed q and t with $0 < q \leq \tau_1$ and $q < t \leq \tau$. Here $q > 0$ is chosen due to left truncation, when the parameter cannot be estimated reliably on the region close to zero. This corresponds to the consistency result for the nonparametric part we discussed in Remark 2.

The asymptotic normality of $\rho(\hat{\boldsymbol{\theta}}_n)$ is established using the idea in Shen (1997) and Chen et al. (2006).

Let $\mathfrak{F}_0 = \{\lambda : \lambda \text{ satisfies } C2\}$ and $\mathfrak{M}$ be the linear span of $(\mathbb{B}, \mathfrak{F}_0) - (\boldsymbol{\beta}_0, \lambda_0)$. Let $w = (w'_{\boldsymbol{\beta}}, w_\lambda)' \in \mathfrak{M}$, then the directional derivative along w of $l(\boldsymbol{\theta}, \boldsymbol{x})$ evaluated at $\boldsymbol{\theta}_0$ is given by

$$\left.\frac{dl(\boldsymbol{\theta}_0 + t\boldsymbol{w}; \boldsymbol{x})}{dt}\right|_{t=0} = \frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x})}{d\boldsymbol{\theta}}[\boldsymbol{w}] = l_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0; \boldsymbol{x})'\boldsymbol{w}_{\boldsymbol{\beta}} + l_\lambda(\boldsymbol{\theta}_0; \boldsymbol{x})[w_\lambda], \quad (6)$$

where $l_\lambda(\boldsymbol{\theta}_0; \boldsymbol{x})[w_\lambda]$ is as previously defined. Based on the directional derivative, the Fisher information inner product for $\boldsymbol{w}$ and $\tilde{\boldsymbol{w}}$ is defined as $\langle \boldsymbol{w}, \tilde{\boldsymbol{w}} \rangle = P\left\{\left(\frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x})}{d\boldsymbol{\theta}}[\boldsymbol{w}]\right)\left(\frac{dl(\boldsymbol{\theta}_0; \boldsymbol{x})}{d\boldsymbol{\theta}}[\tilde{\boldsymbol{w}}]\right)\right\}$ and the Fisher information norm for $\boldsymbol{w}$ is given by $\|\boldsymbol{w}\|^2 = \langle \boldsymbol{w}, \boldsymbol{w} \rangle$.

For any $\boldsymbol{w} \in \mathfrak{M}$, we write

$$\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}] = \lim_{t \to 0} \frac{\rho(\boldsymbol{\theta}_0 + t\boldsymbol{w}) - \rho(\boldsymbol{\theta}_0)}{t}.$$

Theorem 3 *Given that C1–C5 hold,*

$$\sqrt{n}\{\rho(\hat{\boldsymbol{\theta}}_n) - \rho(\boldsymbol{\theta}_0)\} \to_d N\left(0, \left\|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\right\|^2\right),$$

$$\text{where} \left\|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}\right\|^2 = \sup_{\boldsymbol{w} \in \mathfrak{M}, \|\boldsymbol{w}\| = 1} \left|\frac{d\rho(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}[\boldsymbol{w}]^2\right|.$$

4 Computing the sieve MLE

In the theoretical part we denoted the sieve MLE $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}, \hat{\lambda}_n)$ as the maximizer of $l_n(\boldsymbol{\beta}, \lambda_n; \cdot)$ defined by (3) over $\Omega_n = (\mathbb{B}, \mathfrak{F}_n)$. In finite sample computing, we consider to relax the conditions for the α_j 's in $\mathfrak{F}_n$ given in (4).

In what follows we first outline how to choose the spline knot sequence based on the observed data. Specifically, let

$$\mathcal{O} = \{q_i\}_{i=1}^n \cup \{u_i\}_{i:\delta_{1,i}=1} \cup \{v_i\}_{i:\delta 1, i+\delta_{2,i}=1, v_i \neq \tau} \cup \{t_i\}_{i:\delta_3, i=1},$$

that is, $\mathcal{O}$ contains observations of (Q, U, V, T) excluding $\{v_i\}_{i:\delta_{2,i}=1, v_i=\tau}$, which represent a significant portion of observations by C3. Then we let the number of the interior knots be $[n^{1/3}]$ (the closest integer to $n^{1/3}$) for sample size n and put interior knots at quantiles of $\mathcal{O}$.

In $\mathfrak{F}_n$ the condition $\frac{|\alpha_{j+1}-\alpha_j|}{\max_j \Delta_j} \le K^2 d_0$ implies that the difference between two adjacent B-spline coefficients is not large compared to $\max_j \Delta_j$, which will hold if $a_0 < \alpha_j < K\tau b_0$ for finite sample size and large K. In addition, by Lemma 5 in the supplemental material we have proved that $\frac{\max_j \Delta_j}{\max_j \Delta_j}$ based on $\mathcal{O}$ is asymptotically bounded.

Hence, we define

$$\mathfrak{F}_n' = \left\{ \lambda_n = \sum_{j=1}^{p_n} \alpha_j B_j^l : a_0 \le \alpha_j \le K\tau b_0, \quad \text{for } j = 1, \ldots, p_n, \int_0^\tau \lambda_n(t)dt \le \tau b_0 \right\}$$

with the knot sequence we just mentioned, as a simplified version of $\mathfrak{F}_n$ for computing and find the maximizer $\hat{\boldsymbol{\theta}}$ of (3) over $\Omega_n' = (\mathbb{B}, \mathfrak{F}_n')$. From the compactness of $\mathbb{B}$, we simply let $|\boldsymbol{\beta}| \le c_0$.

We observe that in (3) the integration of the B-spline basis functions are involved, which complicates the computing. As an alternative to the B-spline based sieve estimation, monotone I-spline technique for sieve estimation was first introduced by Ramsay (1988). In what follows we choose to adopt the monotone I-splines to approximate the baseline cumulative hazard $\Lambda_0(\cdot)$. Thus the integration of B-spline basis functions can be avoided. We note that Joly et al. (1998) also applied a similar computational approach for estimating the baseline hazard and the cumulative hazard functions in survival data with a penalty term in the likelihood, but with no theoretical results.

Let I_j^l and M_j^l be the I-spline and M-spline basis functions, respectively, as defined, by Ramsay (1988) and Schumaker (1981), with $M_j^l(t) = \frac{d I_j^l(t)}{dt}$. Wu and Zhang (2012) showed that $I_j^l(t) = \sum_{k=j+1}^{p_n+1} B_k^{l+1}(t)$ and $M_j^l(t) = \frac{l}{\xi_{j+l}-\xi_j} B_j^l(t)$, where ξ_{j+l}, ξ_j are two knots from the knot sequence $\{\xi_k\}_{k=1}^{p_n+l}$ associated with the according B-spline basis functions. Note that I_j^l has degree l, while both B_j^l and M_j^l have degree $l-1$.

Then we can show that $\Phi_{N,n} = \{\int_0 \lambda_n : \lambda_n \in \mathfrak{F}_n'\}$ is equivalent to $\mathfrak{F}_{I,n}$ with the I-spline function space $\mathfrak{F}_{I,n}$ defined as

$$\mathfrak{F}_{I,n} = \left\{ \Lambda_n = \sum_{j=1}^{p_n} \eta_j I_j^l : \sum_{j=1}^{p_n} \eta_j \le \tau b_0, a_0 \le \frac{l}{\xi_{j+1} - \xi_j} \eta_j \le K\tau b_0, \text{ for } j = 1, \ldots, pn \right\}.$$

Hence, the B-spline based estimation problem can be converted to an equivalent I-spline based estimation problem. As just discussed, for finite sample case with large K we could further simplify $\mathfrak{F}_{I,n}$ as

$$\Phi_{I,n} = \left\{ \Lambda_n = \sum_{j=1}^{p_n} \eta_j I_j^l : \sum_{j=1}^{p_n} \eta_j \le \tau b_0, \eta_j \ge m_j, \text{ for } j = 1, \ldots, p_n \right\}, \quad (7)$$

with each small positive number $m_j = \frac{\xi_{j+l} - \xi_j}{l} a_0$.

Now we write the likelihood with I-spline basis functions as

$$\begin{aligned} \tilde{l}_n(\boldsymbol{\beta}, \Lambda_n ; \cdot) = & \sum_{i=1}^{n} \delta_{1,i} \log \left(\exp\left[-e^{\boldsymbol{\beta}' z_i} \left\{ \sum_{j=1}^{p_n} \eta_j I_j^l(u_i) - \sum_{j=1}^{p_n} \eta_j I_j^l(qi) \right\} \right] \right. \quad (8) \\ & \left. - \exp\left[-e^{\boldsymbol{\beta}' z_i} \left\{ \sum_{j=1}^{p_n} \eta_j I_j^l(v_i) - \sum_{j=1}^{p_n} \eta_j I_j^l(q_i) \right\} \right] \right) \\ & + \sum_{i=1}^{n} \delta_{2,i} \left[-e^{\boldsymbol{\beta}' z_i} \left\{ \sum_{j=1}^{p_n} \eta_j I_j^l(v_i) - \sum_{j=1}^{p_n} \eta_j I_j^l(q_i) \right\} \right] \\ & + \sum_{i=1}^{n} \delta_{3,i} \left[-e^{\boldsymbol{\beta}' z_i} \left\{ \sum_{j=1}^{p_n} \eta_j I_j^l(t_i) - \sum_{j=1}^{p_n} \eta_j I_j^l(q_i) \right\} \right. \\ & \left. + \boldsymbol{\beta}' z_i + \log \left\{ \sum_{j=1}^{p_n} \eta_j M_j^l(t_i) \right\} \right]. \end{aligned}$$

In practice for the finite sample I-spline based computing,

We need to find the maximizer $\tilde{\boldsymbol{\xi}}_n = (\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_n) \in \widetilde{\Omega}_n = (\mathbb{B}, \Phi_{I,n})$ for $\tilde{l}_n(\boldsymbol{\beta}, \Lambda_n ; \cdot)$ as defined by (8) over $\widetilde{\Omega}_n$. Then by the aforementioned equivalency, we have $\tilde{l}_n(\tilde{\boldsymbol{\xi}}_n; \cdot) = l_n(\hat{\boldsymbol{\theta}}_n; \cdot)$. Since the constraints in $\Phi_{I,n}$ given by (7) is made by linear inequalities, the maximization of (8) over $\widetilde{\Omega}_n$ can be efficiently implemented by the generalized gradient projection algorithm (Jamshidian 2004), as done in Zhang et al. (2010) and Wu and Zhang (2012). More details about this algorithm can be found in these papers.

5 Variance estimation

In addition to the advantage in computing the MLE, it is also straightforward to obtain the consistent observed information matrix for $\boldsymbol{\beta}$ based on the proposed sieve MLE approach. Denote $\boldsymbol{B}^l = \left(B_1^l, \ldots, B_{p_n}^l\right)'$ as the vector of B-spline basis functions of order l, then

$$l_\lambda\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left[\boldsymbol{B}^l\right] = \left\{l_\lambda\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left[B_1^l\right], \ldots, l_\lambda\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left[B_{p_n}^l\right]\right\}'.$$

Let $\boldsymbol{A}_{11} = \mathbb{P}_n\left[\left\{l_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\right\}^{\otimes 2}\right], \boldsymbol{A}_{12} = \mathbb{P}_n\left[l_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left\{l_\lambda\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left[\boldsymbol{B}^l\right]\right\}'\right]$, $\boldsymbol{A}_{21} = \boldsymbol{A}_{12}'$ and $\boldsymbol{A}_{22}$ The $= \mathbb{P}_n\left[\left\{l_\lambda\left(\hat{\boldsymbol{\theta}}_n; \boldsymbol{x}\right)\left[\boldsymbol{B}^l\right]\right\}^{\otimes 2}\right]$.
observed information matrix is given by

$$\hat{\boldsymbol{O}} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}. \quad (9)$$

Theorem 4 *Given that C1–C5 hold,* $\hat{\boldsymbol{O}} \to_P I\left(\boldsymbol{\beta}_0\right)$, *where* $I\left(\boldsymbol{\beta}_0\right)$ *is given in* (5).

Next, we propose how to estimate the variance $\left\|\frac{d\rho\left(\boldsymbol{\theta}_0\right)}{d\boldsymbol{\theta}}\right\|^2$ for the plug-in estimator $\rho\left(\hat{\boldsymbol{\theta}}_n\right)$ of $\Lambda_0(t) - \Lambda_0(q)$. We consider a similar method as for the observed information matrix for $\boldsymbol{\beta}$. In what follows we adopt the idea described in Cheng et al. (2014). Let $\hat{\lambda}_n = \sum_{j=1}^{p_n} \hat{\alpha}_j \boldsymbol{B}_j^l$ with $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}, \hat{\lambda}_n\right)$. By the construction of $\boldsymbol{A}_{11}$, $\boldsymbol{A}_{12}$, $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22}$ above, we can treat $\tilde{\boldsymbol{O}} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ as the observed information matrix for the spline coefficient vector $\hat{\boldsymbol{\alpha}} = \left(\hat{\alpha}_1, \ldots, \hat{\alpha}_{p_n}\right)'$. Since $\rho\left(\hat{\boldsymbol{\theta}}_n\right) = \int_q^t \hat{\lambda}_n(s)ds = \int_q^t \sum_{j=1}^{p_n} \hat{\alpha}_j B_j^l(s)ds$, we have

$$\frac{\partial\rho\left(\hat{\boldsymbol{\theta}}_n\right)}{\partial\hat{\boldsymbol{\alpha}}} = \left\{\frac{\partial\rho\left(\hat{\boldsymbol{\theta}}_n\right)}{\partial\hat{\alpha}_1}, \ldots, \frac{\partial\rho\left(\hat{\boldsymbol{\theta}}_n\right)}{\partial\hat{\alpha}_{p_n}}\right\}' = \left\{\int_q^t B_1^l(s)ds, \ldots, \int_q^t B_{p_n}^l(s)ds\right\}' \equiv \tilde{\boldsymbol{\omega}}.$$

Hence, by delta method the variance for $\rho\left(\hat{\boldsymbol{\theta}}_n\right)$ can be estimated by $\tilde{\boldsymbol{\omega}}'\tilde{\boldsymbol{O}}^{-1}\tilde{\boldsymbol{\omega}}$.

6 Simulation studies

From our previous experiences as well as the literature cubic or quadratic splines are good enough to get satisfactory sieve estimations. In simulation studies we let all spline basis functions have order $l = 3$, that is, we will use quadratic B-spline and M-spline basis functions, cubic I-spline basis functions. We choose sample size as 200 and 500 with 1000 repetitions. The knot sequence for splines is chosen as described in Sect. 4.

Let $\boldsymbol{\beta}_0 = (\beta_{0,1}, \beta_{0,2})' = (0.7\ -0.5)'$ and let covariate $\boldsymbol{Z} = (Z_1, Z_2)'$, where Z_1 follows standard normal distribution and Z_2 follows Bernoulli distribution with probability 0.5 of $Z_2 = 1$.

We generate T with cumulative hazard function $\Lambda_{0,k}(\cdot|\boldsymbol{Z}) = \Lambda_{0,k}(\cdot)\exp\boldsymbol{\beta}_0'Z$ for $k = 1, 2, 3$, where $\Lambda_{0,1}(t) = e^{1.2}\frac{1-e^{-t}}{1-e^{-4}}$, $\Lambda_{0,2}(t) = \frac{1-e^{-t}}{1-e^{-4}}$ and $\Lambda_{0,3}(t) = e^{-1.2}\frac{1-e^{-t}}{1-e^{-4}}$ for $0 \le t \le 4$, and $\Lambda_{0,1}(t) = e^{1.2}$, $\Lambda_{0,2}(t) = 1$ and $\Lambda_{0,3}(t) = e^{-1.2}$ for $t > 4$. $\Lambda_{0,1}(\cdot)$ represents the situation with an average observed cure rate of 0.135 (small cure rate), $\Lambda_{0,2}(\cdot)$ with an average cure rate of 0.448 (medium cure rate) and $\Lambda_{0,3}(\cdot)$ with an average cure rate of 0.755 (large cure rate). How to generate T with a specific cumulative hazard function (or a survival function) is referred to Sect. 5.1 in Liu and Shen (2009), which is based on the connection between a non-mixture cure model and a mixture cure model as pointed out by Chen et al. (1999).

For all three baseline cumulative hazard functions we generate left truncation time Q and observation interval $[U, V]$ in two different ways: (1) Q is generated from Uniform [0, 1], and $[U, V]$ is generated from Uniform [1, 4.5];(2) Q is generated from Uniform [0, 4], and $[U, V]$ is generated from Uniform $[Q, 4.5]$. For both cases U and V are set to 4 if they are > 4, and $U = V - 0.005$ if $V - U < 0.005$. For (1) the resulting average truncation rates in the uncured subjects with $\Lambda_{0,1}(\cdot)$, $\Lambda_{0,2}(\cdot)$ and $\Lambda_{0,3}(\cdot)$ are 0.654, 0.501, and 0.421, respectively, and the resulting average overall censoring rates (including interval censoring and right censoring) in the remaining subjects after truncation are 0.108,0.163 and 0.195, respectively; for (2) the corresponding average truncation and average overall censoring rates are 0.894, 0.831, and 0.794, and 0.258, 0.323 and 0.359, respectively. We refer to the above two settings as relatively light truncation and censoring (TC) versus heavy TC in the following.

Since the final observed data sets for analysis are truncated, in Table 1 for each setting we have reported the average rates of cured, uncured and right-censored (URC), and uncured but not right-censored (UNRC) among the truncated data. In the observed data all cured subjects are right censored. Hence the observed right-censored subjects include all cured subjects and uncured but right-censored ones. Note that all aforementioned average rates are obtained from the simulation studies with sample size 500 (1000 repetitions).

Before discussing the simulation results, in the following we outline how to obtain a truncated data set with sample size n.

1. Generate a sample of $\boldsymbol{Z}$ with sample size $10 \times n$; then generate a sample of T given $\boldsymbol{Z}$;
2. Generate a sample of Q with sample size $10 \times n$;
3. Comparing each observation of Q with the corresponding observation of T, for observations with $Q \le T$, we generate corresponding observations of U and V;
4. Choose a random subset with sample size n from the generated sample of (T, Q, U, V, $\boldsymbol{Z}$) satisfying $Q \le T$.

In the simulation, we use the proposed sieve MLE method to estimate the parametric part $\boldsymbol{\beta}$ and the nonparametric part $\lambda_{0,k}(t)$ with its smooth functionals $\Lambda_{0,k}(t) - \Lambda_{0,k}(q)$ for $k = 1$, 2, and 3. Due to limitation of space we present here results for the small and large cure rates ($k = 1$ and 3), while the results for $k = 2$ are in-between of these two cases and are available from the authors. Tables 2 and 3 present the results for estimating $\boldsymbol{\beta}$ for $n = 200$ and 500, and Tables 4 and 5 present the results for the plug-in estimates of $\Lambda_{0,k}(t) - \Lambda_{0,k}(q)$ with $k = 1$ and 3, $q = 1$ and $t = 1.5$, 2.5, 3.5, respectively. The tables include the average point estimates, sample standard deviation (SD) and average estimated standard error based on the proposed estimated information matrix introduced in Sect. 5 (SE), and coverage probability of nominal 95% confidence intervals based on the estimated standard error (95% CP).

For the nonparametric part, we also show the estimation results of the baseline hazard function $\lambda_{0,k}(\cdot)$ with $k = 1$ and 3 on interval [0, 3.9] for sample size 200 and 500, which are averaged over 1000 curves. Figures 1 and 2 present the results for estimating $\lambda_{0,1}(\cdot)$ and $\lambda_{0,3}(\cdot)$, respectively.

From Tables 2 and 3 we can see that the simulation results for estimating both β and the increments of the baseline cumulative function from q to t in general become more accurate (i.e. smaller variances) when the sample size is increased from 200 to 500, or the truncation and censoring becomes less severe, with also more accurate standard errors compared to the sample standard deviations.

We also see that the coverage probabilities of the confidence intervals are generally acceptable. With increasing sample size the parametric estimate becomes less biased, and we discuss the nonparametric estimate separately below.

In Figs. 1 and 2, we see that the nonparametric estimate becomes more accurate for larger n in terms of "function distance" $\int_q^t \{\hat{\lambda}_n(s) - \lambda_0(s)\}^2 ds$, but not necessarily at each time point. These figures also show that the estimation close to the right end point $\tau = 4$ is not very accurate for light truncation and censoring and sample size 200, which is likely caused by the small number of the events around there. We note that the estimation near $\tau = 4$ seems to improve with heavier truncation and censoring, most likely because with heavier truncation more observations appear at later times (closer to $\tau = 4$). In addition, it is important to note that the baseline hazard becomes noticeably underestimated close to time zero when the cure rate is larger and truncation and censoring is more severe. This underestimation is likely caused by the reduced risk set sizes due to left truncation, and is consistent with our theoretical result that the estimation of the hazard function close to time zero is not reliable. However, we have noted earlier that the increments of the baseline cumulative hazard function can nonetheless be well estimated.

Finally for the plug-in estimation of $\Lambda_{0,k}$, Tables 4 and 5 show that the plugin estimation bias for the increments of the baseline cumulative hazard is not always decreased when the sample size is increased from 200 to 500. This is a not contradiction to the results in Figs. 2 and 3, where the accuracy for the nonparametric estimate $\hat{\lambda}_n(\cdot)$ is improved in terms of "function distance" when the sample size is increased or truncation and censoring is less

severe. The fact that $\int_q^t \left\{\hat{\lambda}_n(s) - \lambda_0(s)\right\}^2 ds$ is smaller (nonparametric estimate is better in terms of L_2 distance) cannot imply that $\int_q^t \left\{\hat{\lambda}_n(s) - \lambda_0(s)\right\} ds$ is smaller (plug-in point estimate is better). On the other hand, Tables 4 and 5 show that in all scenarios the nonparametric plug-in estimation bias is acceptably small, standard errors and coverage probabilities all appear acceptable.

Overall the simulations show that sample size 200 provides quite good estimation results which are comparable to those from sample size 500 in all settings.

7 Spontaneous abortion data analysis

We apply the proposed sieve MLE method to an observational data set on spontaneous abortion from the autoimmune disease in pregnancy database of the Organization of Teratology Information Specialists (OTIS) mentioned earlier. Our focus is to investigate the potential effect of autoimmune disease medication on (spontaneous abortion) SAB, which is defined as any spontaneous pregnancy loss occurring before week 20 of gestation.

Our study sample includes pregnant women who entered a research study between 2005 and 2012. It consists of 923 women who entered the study before week 20 of their gestation. Since some women in the population may experience the SAB event before having the chance to enter the study, we consider the study entry time as left truncation time. Among the 923 subjects 56 women experienced the SAB event and the exact SAB time is known, 10 women also experienced the SAB event but only a time window including the incidence is available, 2 women were lost to follow-up before week 20, the rest of the women did not experience the SAB event.

In our proposed method, the lost to follow-up subjects and the observed cured subjects (subjects did not experience the SAB events before the cure threshold of week 20) are both treated as right-censored in the likelihood function under the nonmixture cure model, the same as in Sy and Taylor (2000). This way in the study sample we have 56 subjects with exact observed event times, 10 interval censored event times, and the rest are treated as right-censored. So the data set is partly interval censored with left truncation, and also with a well defined cured portion. Since 10 interval censoring from all 66 women who experienced SAB is not an ignorable portion, the existing methods based on right-censoring alone is not applicable here. Therefore the proposed sieve MLE method can be a good choice for the analysis.

For the primary comparison groups, among the 923 women 481 were pregnant and with certain autoimmune diseases which were treated with medications under investigation, 262 were women with the same specific autoimmune diseases but who were not treated with the medications under investigation, and the rest were healthy pregnant women without autoimmune diseases who were not treated with the medications. We also include three important covariates: maternal age, prior therapeutic abortion (TAB; yes/no), and smoking (yes/no). The distributions of the covariates are given in Table 6. For the analysis, as in the simulation studies we use quadratic B-spline and M-spline basis functions, and cubic I-spline basis functions. The knot sequence for the splines is chosen as described in Sect. 4.

Table 7 presents the estimation results for our study sample based on the proposed sieve MLE approach. According to the results from Table 7 we do not have statistical evidence to show that the autoimmune disease drugs have any significant effects on the risk of SAB. We also see that older women have higher risk to experience the SAB events and smoking will increase the risk of the SAB. Table 7 also shows the proposed sieve MLE for $\Lambda_0(t)$ and $\Lambda_0(t) - \Lambda_0(q)$ with $t = 17, 18, 19$ and $q = 5$ (weeks). The standard errors of these estimates are consistent with our theoretical results and imply that while the direct estimate for the baseline cumulative hazard function for the SAB occurring time has too much variability due to left truncation, the functional increase from a point not close to zero can still be reliably estimated.

Figure 3 shows the estimated baseline hazard function based on the proposed sieve MLE, and implies that the highest risk period for women to experience the SAB events is between 5 and 10 weeks of gestation. This is consistent with existing scientific knowledge about spontaneous abortion. In addition, the baseline survival function conditional upon having survived 5 weeks of pregnancy is also plotted in Fig. 3.

Since the baseline hazard cannot be estimated accurately at early stage of pregnancy, the unconditional baseline survival is not shown here. Similarly, the baseline cure rate can be estimated by the baseline survival function at week 20 of pregnancy, which is 0.9825; but this estimate may be unreliable due to left truncation since the unconditional baseline survival function is used here.

8 Concluding remarks

In this paper we have proposed the semiparametric sieve MLE method to analyze complex survival data that are partly interval censored, left truncated and with a cured portion. The proposed approach is motivated by a spontaneous abortion data application with this type of complex structure, since no existing survival method is able to directly handle this type of survival data. Non-mixture cure model based on the Cox regression is used due to the relative simplicity of the likelihood computation. Using modern empirical process we have thoroughly studied the asymptotic properties for the proposed method: we have established that the proposed estimation is consistent; we have also established the asymptotic normality for both estimators of the parametric part and a functional of the nonparametric part. In addition, we have provided closed-form variance estimation for both the parametric and the nonparametric parts. In simulation studies we have showed that the finite sample performance of the proposed sieve MLE is satisfactory. Finally, the proposed model was successfully applied to analyze the SAB data set.

The proposed method is designed for relatively general survival data and usually applicable for simpler data structures. For different types of survival data, the proposed model may perform differently. For example, if partly interval censored data is replaced by right-censored only data, the proposed sieve MLE has the same asymptotic properties in terms of convergence rate and asymptotic normality as we mentioned in Sect. 1. However, if the data is purely interval censored, the estimation of hazard function will not be available based on the likelihood (since the third term in (2) disappears); separately by similar method as in

Zhang et al. (2010) it can be shown that the rate of estimation of the baseline cumulative hazard function will be slower than $\sqrt{n}$. In addition, if there is no left truncation, the baseline cumulative hazard function itself can be reliably estimated, as opposed to only its increments.

We have established that due to lack of data information around time zero for left truncated data, the nonparametric estimation around that region is not reliable. In the future we plan to tackle this issue and improve the estimation for the nonparametric part around time zero. Another potential work might be to replace the Cox model by the more general transformation model (Zeng et al. 2006) and develop the general semiparametric sieve MLE method. Related to this, checking whether our proposed model fits the data well could be a penitential research topic. Very recently Peng and Taylor (2017) proposed residual-based model diagnosis methods for the mixture cure model. But we found no work that has been published for model checking for the non-mixture cure models. For interval-censored data, Sun (1997) reviewed diagnostic methods for the Cox regression model, which were mostly residual and graphically based (Farrington 2000), and no theory existed to provide goodness-of-fit tests. Developing these techniques under the mixture cure model and with left truncated and mixed censoring types data will be non-trivial, but nonetheless very useful in practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements The research of Yuan Wu was supported in part by award number P01CA142538 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.

References

Bickel P, Klaassen C, Ritov Y, Wellner JA (1993) Efficient and adaptive estimation for semiparametric models. Johns Kopkins University Press, Baltimore

Chen MH, Ibrahim JG, Sinha D (1999) A new bayesian model for survival data with a survival fraction. J Am Stat Assoc 94:909-919-

Chen X, Fan Y, Tsyrennikov V (2006) Efficient estimation of semiparametric multivariate copula models.J Am Stat Assoc 475:1228-1240-

Cheng G, Zhou L, Chen X, Huang JZ (2014) Efficient estimation of semiparametric copula models for bivariate survival data. J Multivar Anal 123:330-344-

Farrington CP (2000) Residuals for proportional hazards models with interval-censored survival data. Biometrics 56:473-482-

Geman A, Hwang CR (1982) Nonparametric maximum likelihood estimation by the method of sieves. Ann Stat 10:401-414-

Hu T, Xiang L (2013) Efficient estimation for semiparametric cure models with interval-censored data. J Multivar Anal 121:139-151-

Hu T, Xiang L (2016) Partially linear transformation cure models for interval-censored data. Comput Stat Data Anal 93:257-269-

Huang J, Zhang Y, Hua L (2008) A least-squares approach to consistent information estimation in semi-parametric models. Technical report, Department of Biostatistics, University of Iowa

Jamshidian M (2004) On algorithms for restricted maximum likelihood estimation. Comput Stat Data Anal 45:137-157-

Joly P, Commenges D, Letenneur L (1998) A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. Biometrics 54:185-194 [PubMed: 9574965] -

Kim JS (2003a) Efficient estimation for the proportional hazards model with left-truncated and “case 1” interval-censored data. Stat Sin 13:519-537-

Kim JS (2003b) Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. J Roy Stat Soc B 65:489-502 -

Lam KF, Xue H (2005) A semiparametric regression cure model with current status data. Biometrika 92:573-586-

Liu H, Shen Y (2009) A semiparametric regression cure model for inverval-censored data. J Am Stat Assoc 104:1168-1178 [PubMed: 20354594] -

Ma S (2010) Mixed case interval censored data with a cured subgroup. Stat Sin 20:1165-1181 -

Peng Y, Taylor JMG (2017) Residual-based model diagnosis methods for mixture cure models. Biostatistics 73:495-505-

Ramsay JO (1988) Monotone regression splines in action. Stat Sci 3:425-441 -

Schumaker L (1981) Spline function: basic theory. John Wiley, New York

Shen X (1997) On methods of sieves and penalization. Ann Stat 25:2555-2591-

Sun J (1997) The statistical analysis of interval-censored failure time data, vol 2006. Springer, New York

Sy JP, Taylor JMG (2000) Estimation in a cox proportional hazards cure model. Biometrics 56:227-236 [PubMed: 10783800] -

Tsodikov A (1998) A proportional hazards model taking account of long-term survivors. Biometrics 54:1508-1516 [PubMed: 9883549] -

van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge

Wellner JA, Zhang Y (2007) Likelihood-based semiparametric estimation methods for panel count data with covariates. Ann Stat 35:2106-2142-

Wu Y, Zhang Y (2012) Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. Ann Stat 40:1609-1636-

Zeng D, Yin G, Ibrahim JG (2006) Semiparametric transformation models for survival data with a cure fraction. J Am Stat Assoc 101:670-684-

Zhang Y, Hua L, Huang J (2010) A spline-based semiparametric maximum likelihood estimation for the cox model with interval-censored data. Scand J Stat 37:338-354-

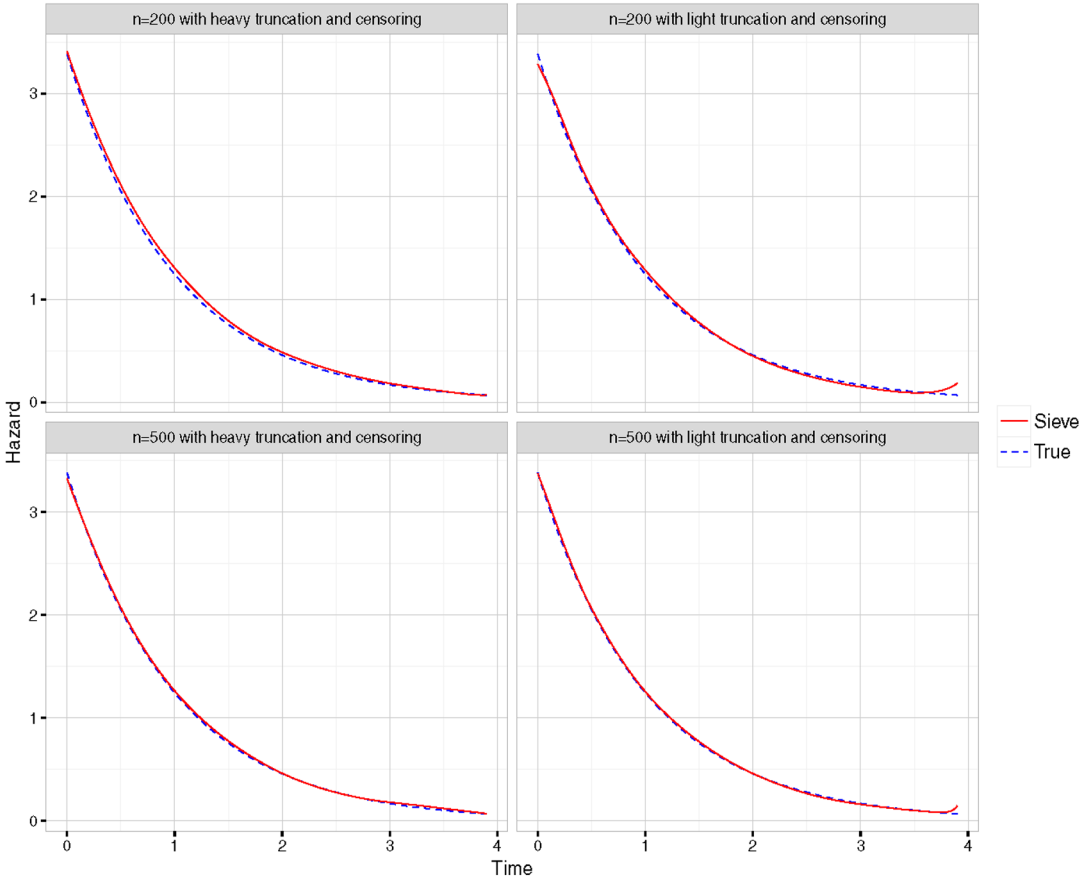

Fig.1.

True baseline hazard function (true) and its sieve MLE (sieve) with small cure rate

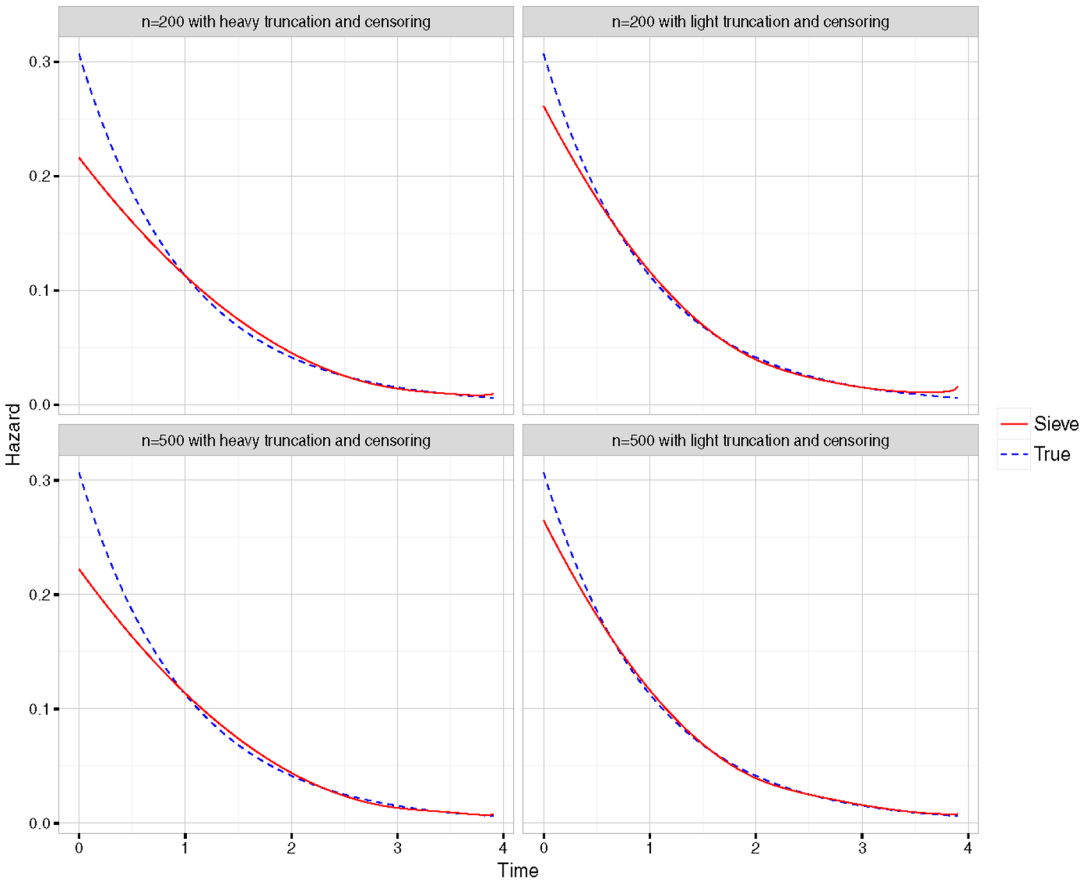

Fig.2.

True baseline hazard function (true) and its sieve MLE (sieve) with large cure rate

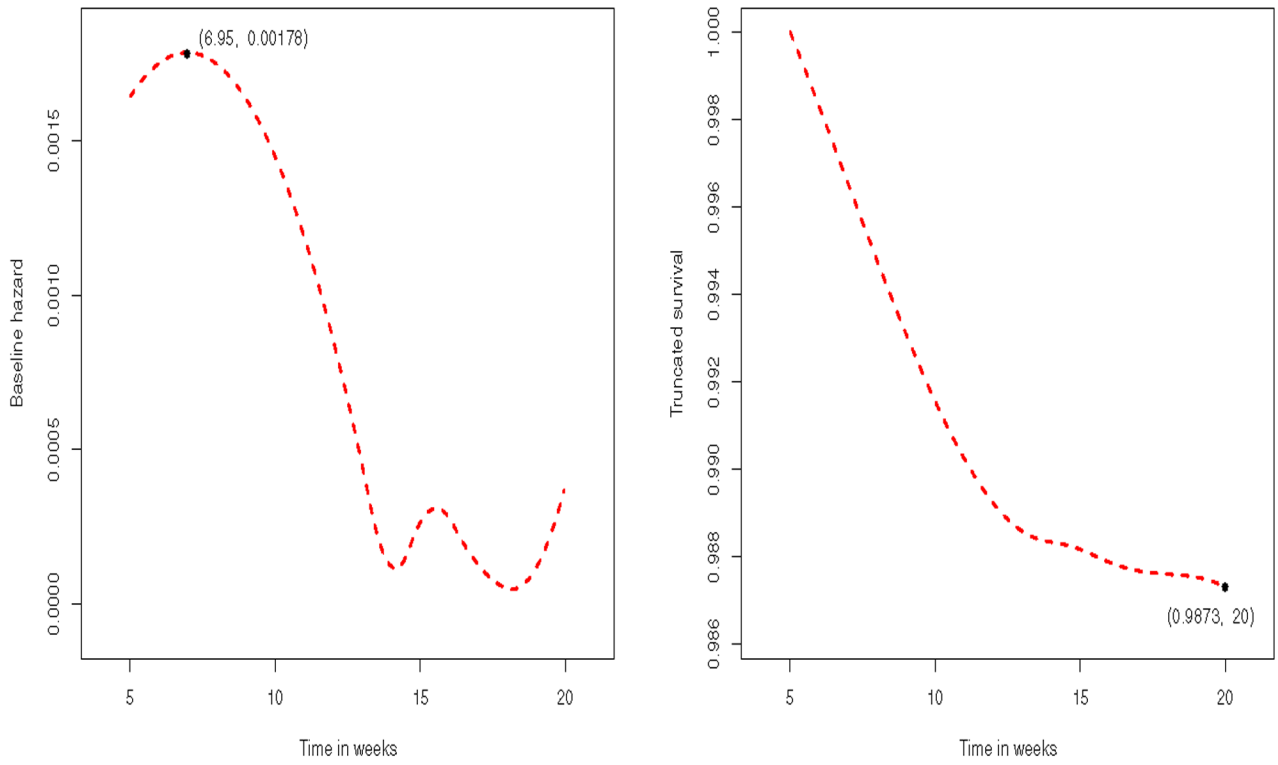

Fig.3.

Estimated baseline hazard (left) and baseline survival function conditional on having survived 5 weeks of pregnancy (right)

Table 1

Average rates of cured, URC and UNRC in truncated data sets for different simulation settings

	Cured	URC	UNRC	Cured	URC	UNRC
	Light TC			Heavy TC		
Small cure	0.313	0.018	0.669	0.597	0.021	0.382
Medium cure	0.621	0.015	0.364	0.827	0.013	0.160
Large cure	0.842	0.008	0.150	0.938	0.005	0.057

Table 2

Estimation of the parametric part with small cure rate

		True value	Estimate	SD	SE	95% CP (%)
Light truncation and censoring						
Size = 200	$\beta_{0,1}$	0.7	0.711	0.112	0.117	96.5
	$\beta_{0,2}$	−0.5	−0.503	0.177	0.186	96.0
Size = 500	$\beta_{0,1}$	0.7	0.709	0.069	0.071	95.7
	$\beta_{0,2}$	−0.5	−0.501	0.112	0.114	95.4
Heavy truncation and censoring						
Size = 200	$\beta_{0,1}$	0.7	0.736	0.153	0.163	96.4
	$\beta_{0,2}$	−0.5	− 0.525	0.244	0.256	94.7
Size = 500	$\beta_{0,1}$	0.7	0.716	0.098	0.097	95.7
	$\beta_{0,2}$	−0.5	−0.510	0.149	0.155	95.5

Table 3

Estimation of the parametric part with large cure rate

		True value	Estimate	SD	SE	95% CP (%)
Light truncation and censoring						
Size = 200	$\beta_{0,1}$	0.7	0.717	0.209	0.221	96.7
	$\beta_{0,2}$	−0.5	− 0.527	0.405	0.412	96.5
Size = 500	$\beta_{0,1}$	0.7	0.710	0.129	0.129	95.3
	$\beta_{0,2}$	−0.5	− 0.524	0.244	0.245	95.0
Heavy truncation and censoring						
Size = 200	$\beta_{0,1}$	0.7	0.736	0.381	0.468	96.8
	$\beta_{0,2}$	−0.5	−0.619	0.697	0.856	99.0
Size = 500	$\beta_{0,1}$	0.7	0.720	0.212	0.230	96.7
	$\beta_{0,2}$	−0.5	− 0.509	0.399	0.427	96.6

Table 4

Estimation of the baseline cumulative hazard difference with small cure rate

		True value	Estimate	SD	SE	95% CP (%)
Light truncation and censoring						
Size = 200	$\Lambda_{0,1}(1.5)-\Lambda_{0,1}(l)$	0.490	0.507	0.109	0.112	95.9
	$\Lambda_{0,1}(2.5)-\Lambda_{0,1}(l)$	0.967	0.980	0.207	0.224	96.8
	$\Lambda_{0,1}(3.5)-\Lambda_{0,1}(l)$	1.142	1.142	0.233	0.308	98.3
Size = 500	$\Lambda_{0,1}(1.5)-\Lambda_{0,1}(l)$	0.490	0.496	0.067	0.070	96.6
	$\Lambda_{0,1}(2.5)-\Lambda_{0,1}(l)$	0.967	0.974	0.126	0.133	95.5
	$\Lambda_{0,1}(3.5)-\Lambda_{0,1}(l)$	1.142	1.144	0.148	0.171	96.6
Heavy truncation and censoring						
Size = 200	$\Lambda_{0,1}(1.5)-\Lambda_{0,1}(l)$	0.490	0.515	0.133	0.143	96.2
	$\Lambda_{0,1}(2.5)-\Lambda_{0,1}(l)$	0.967	1.020	0.262	0.277	96.4
	$\Lambda_{0,1}(3.5)-\Lambda_{0,1}(l)$	1.142	1.210	0.295	0.328	97.7
Size = 500	$\Lambda_{0,1}(1.5)-\Lambda_{0,1}(l)$	0.490	0.500	0.087	0.087	94.4
	$\Lambda_{0,1}(2.5)-\Lambda_{0,1}(l)$	0.967	0.983	0.155	0.160	95.9
	$\Lambda_{0,1}(3.5)-\Lambda_{0,1}(l)$	1.142	1.170	0.178	0.185	95.5

Table 5

Estimation of the baseline cumulative hazard difference with large cure rate

		True value	Estimate	SD	SE	95% CP (%)
Light truncation and censoring						
Size = 200	$\Lambda_{0,3}(1.5)-\Lambda_{0,3}(l)$	0.044	0.046	0.013	0.015	96.9
	$\Lambda_{0,3}(2.5)-\Lambda_{0,3}(l)$	0.088	0.088	0.028	0.032	96.1
	$\Lambda_{0,3}(3.5)-\Lambda_{0,3}(l)$	0.104	0.104	0.032	0.074	99.1
Size = 500	$\Lambda_{0,3}(1.5)-\Lambda_{0,3}(l)$	0.044	0.046	0.008	0.009	96.1
	$\Lambda_{0,3}(2.5)-\Lambda_{0,3}(l)$	0.088	0.088	0.018	0.019	95.4
	$\Lambda_{0,3}(3.5)-\Lambda_{0,3}(l)$	0.104	0.104	0.020	0.030	98.6
Heavy truncation and censoring						
Size = 200	$\Lambda_{0,3}(1.5)-\Lambda_{0,3}(l)$	0.044	0.047	0.021	0.025	95.9
	$\Lambda_{0,3}(2.5)-\Lambda_{0,3}(l)$	0.088	0.094	0.039	0.056	97.6
	$\Lambda_{0,3}(3.5)-\Lambda_{0,3}(l)$	0.104	0.109	0.044	0.070	98.0
Size = 500	$\Lambda_{0,3}(1.5)-\Lambda_{0,3}(l)$	0.044	0.047	0.013	0.014	94.5
	$\Lambda_{0,3}(2.5)-\Lambda_{0,3}(l)$	0.088	0.092	0.025	0.029	96.1
	$\Lambda_{0,3}(3.5)-\Lambda_{0,3}(l)$	0.104	0.107	0.029	0.033	95.4

Table 6

Mean (SD) or *n* (%) of covariates by comparison groups

	Diseased treated (*N* = 481)	**Diseased control (*N* = 262)**	**Healthy control (*N* = 180)**
Maternal age	32.37 (4.85)	33.13 (4.75)	32.35 (5.04)
Prior Tab—yes	62 (12.89%)	28 (10.69%)	13 (7.22%)
Smoking—yes	61 (12.68%)	28 (10.69%)	5 (2.78%)

Table 7

Estimation of covariate effects and baseline cumulative hazard using the spontaneous abortion data

	Estimate	SE	*p* value
Maternal age	0.079	0.025	0.002
Prior tab	−0.358	0.436	0.411
Smoking	0.823	0.364	0.024
Healthy control	−0.303	0.479	0.527
Diseased control	0.236	0.279	0.398
$\Lambda_0(17)$	0.0173	0.020	–
$\Lambda_0(18)$	0.0174	0.020	–
$\Lambda_0(19)$	0.0174	0.020	–
$\Lambda_0(17)-\Lambda_0(5)$	0.0124	0.004	–
$\Lambda_0(18)-\Lambda_0(5)$	0.0125	0.004	–
$\Lambda_0(19)-\Lambda_0(5)$	0.0126	0.004	–