 DOI: 10.1002/bimj.202000173

A product-limit estimator of the conditional survival function when cure status is partially known

Wende Clarence Safari[1] | **Ignacio López-de-Ullibarri**[2] | **María Amalia Jácome**[3]

[1] Faculty of Computer Science, Department of Mathematics, University of A Coruña, CITIC, A Coruña, Spain
[2] Department of Mathematics, Escuela Universitaria Politécnica, University of A Coruña, Ferrol, Spain
[3] Faculty of Science, Department of Mathematics, University of A Coruña, CITIC, A Coruña, Spain

Correspondence
Wende Clarence Safari, Faculty of Computer Science, University of A Coruña, Elviña, 15071 A Coruña, Spain.
Email: wende.safari@udc.es

Abstract
We introduce a nonparametric estimator of the conditional survival function in the mixture cure model for right-censored data when cure status is partially known. The estimator is developed for the setting of a single continuous covariate but it can be extended to multiple covariates. It extends the estimator of Beran, which ignores cure status information. We obtain an almost sure representation, from which the strong consistency and asymptotic normality of the estimator are derived. Asymptotic expressions of the bias and variance demonstrate a reduction in the variance with respect to Beran's estimator. A simulation study shows that, if the bandwidth parameter is suitably chosen, our estimator performs bet-ter than others for an ample range of covariate values. A bootstrap bandwidth selector is proposed. Finally, the proposed estimator is applied to a real dataset studying survival of sarcoma patients.

Keywords
bootstrap bandwidth, censoring, cure model, kernel estimator, Nadaraya-Watson weights

1 | INTRODUCTION

The standard survival model assumes that, if there is no censoring, at some point all individuals will experience the event of interest. However, cure models have been developed because there are many situations in which this assumption is not appropriate. In clinical settings, for example, it is very unlikely to have any recurrence of some tumors later than a certain period after radiation treatment. Such examples can be found in many other disciplines: some people will never get married, one-child mothers will never have a second child, some workers will never get a career shift, etc. In most literature, subjects in which an event will never take place are referred to as cured subjects.

The mixture cure model, originally proposed by Boag (1949), has received much attention in recent years. It assumes that the population is a mixture of cured and susceptible individuals. Note that here a "cured" individual is defined as being free of experiencing the event of interest, not necessarily cured in medical terms. The goal is to model the probability of cure and the survival function of the uncured subjects, also called latency. There has been substantial work on the mixture cure model, mostly with a (semi)parametric approach (see Amico & Van Keilegom, 2018; Maller & Zhou, 1996; Patilea & Van Keilegom, 2020, and references therein). These models are constructed under different (semi)parametric frameworks for the proportion of long-term survivors and/or the latency. However, when the underlying functions cannot be well

approximated by the assumed (semi)parametric structures, applying those models will lead to biased estimates. Therefore, it is important to have completely nonparametric methods to model survival data with a cure fraction. Maller and Zhou (1992) proposed a consistent nonparametric estimator of the cure rate but their method cannot handle covariates. Based on the estimator of the conditional survival function in Beran (1981), Xu and Peng (2014), López-Cheda, Cao, et al. (2017), López-Cheda, Jácome, et al. (2017), and López-Cheda et al. (2020) developed nonparametric methods for the mixture cure model in the presence of covariates.

Absence of an individual's cure status (i.e., cured, uncured) is an important challenge for cure models. A subject who experiences the event is known to be uncured. However, censoring prevents from observing whether a censored subject would experience the event eventually. This hinders the classification of the censored observations as cured or uncured. In this situation, it is customary to assume no additional information on the cure status of the censored individuals, thus, to model the cure status as a latent variable. Nonetheless, there are situations in which some of the censored individuals can be identified to be immune to the event of interest, that is, to be cured. For example, based on the result of a diagnosis procedure, some patients could be assumed to be cured from a given disease. Also, for some types of cancer it is extremely unlikely to have any recurrence later than a given fixed time after treatment, known as a cure threshold. Another example of a situation with individuals known to be cured is the analysis of hospital bed and intensive care unit (ICU) occupancy. In this, it is important to estimate the distribution of time a patient will be in the hospital ward or ICU, specifically, modeling the time a patient stays in the hospital ward until admitted to the ICU. In the language of cure models, all patients who have died or have been discharged from the hospital bed without entering the ICU are censored and are known to be cured from the ICU admission. This is of great interest to hospital management, particularly in outbreaks of epidemic diseases such as the novel coronavirus disease.

Few authors have explored cure models when the cure status is known for some censored observations. Laska and Meisner (1992) and Betensky and Schoenfeld (2001) discussed nonparametric cure rate estimation with cure status available, but neither of them considered the presence of covariates. Nieto-Baraja and Yin (2008) proposed a Bayesian semiparametric approach for estimating a survival function with a cure fraction in the presence of covariates. A semiparametric approach based on a Cox proportional hazards cure model when cure information is partially known was studied by Wu et al. (2014). Bernhardt (2016) proposed a flexible cure rate model with potentially known cure threshold and showed that ignoring a known cure threshold may lead to biased estimates. Recently, Chen and Du (2018) developed a nonparametric approach to modeling the covariate effects under the framework of promotion time. They considered a fixed cure threshold, so that observations censored at times larger than that are assumed to correspond to cured subjects. Contrary to the methods mentioned, in this paper we develop a completely nonparametric mixture cure model with covariates that can be applied in general situations, in which the identification of the cured individuals does not depend on a fixed cured threshold. Examples of situations in which a fixed cure threshold cannot be assumed were mentioned above: a study in which a diagnostic procedure is used to discriminate between cured and uncured subjects, or a study of time to ICU admission of hospital inpatients, in which discharge or death can occur before ICU admission. Therefore, we propose a generalized product-limit estimator of the survival function that extends Beran's estimator when cure status information is available. From the proposed survival function estimator, further methods for the estimation of the cure rate and latency functions can be derived, in the spirit of Xu and Peng (2014), López-Cheda, Cao, et al. (2017), López-Cheda, Jácome, et al. (2017), and López-Cheda et al. (2020).

This paper is organized as follows. In Section 2, after specifying the model notations, new estimators of the conditional cumulative hazard and survival functions are proposed, and some asymptotic results for them are given. For the choice of the bandwidth we propose a bootstrap procedure in Section 3. In Section 4, we study the efficiency of the estimator of the survival function with a simulation study in which our estimator is compared to Beran's estimator, which ignores the available cure status information, as well as to the semiparametric estimator proposed by Bernhardt (2016). In Section 5, the estimator is applied to estimate the distribution of the time to death from sarcoma cancer of 233 patients from the University Hospital of Santiago de Compostela, Spain. Section 6 contains a discussion and thoughts for future work.

2 | MIXTURE CURE MODEL WHEN CURE STATUS IS PARTIALLY KNOWN

2.1 | Model notation

Let Y be the survival time, C the random censoring time, and $\mathbf{X}$ a vector of covariates. Assume that the survival time Y is subject to random right censoring, so that instead of observing Y, only $T = \min(Y, C)$ and $\delta = \mathbf{1}(Y \leq C)$ can be observed.

The random variables Y and C are assumed to be conditionally independent given $\mathbf{X} = \mathbf{x}$. Let $F(t|\mathbf{x}) = P(Y \leq t|\mathbf{X} = \mathbf{x})$ denote the conditional distribution function of Y and $G(t|\mathbf{x}) = P(C \leq t|\mathbf{X} = \mathbf{x})$ denote the conditional distribution function of C. It is assumed that X, Y, and C are absolutely continuous. We set $Y = \infty$ if the subject is cured. Let $\nu = \mathbf{1}(Y = \infty)$ be an indicator of being cured. Note that ν is partially observed because $\delta = 1$ implies $\nu = 0$. In addition, when the cure status is partially known, $\nu = 1$ is also observed for some censored individuals. Suppose that ξ indicates whether the cure status is known ($\xi = 1$) or not ($\xi = 0$). Hence, the observations $\{(\mathbf{X}_i, T_i, \delta_i, \xi_i, \xi_i\nu_i) : i = 1, \ldots, n\}$ can be classified into three groups: (a) the individual is observed to have experienced the event and therefore known to be uncured ($\mathbf{X}_i, T_i = Y_i, \delta_i = 1, \xi_i = 1, \xi_i\nu_i = 0$); (b) the lifetime is censored and the cure status is unknown ($\mathbf{X}_i, T_i = C_i, \delta_i = 0, \xi_i = 0, \xi_i\nu_i = 0$); and (c) the lifetime is censored and the individual is known to be cured ($\mathbf{X}_i, T_i = C_i, \delta_i = 0, \xi_i = 1, \xi_i\nu_i = 1$). The probability of cure is $1 - p(\mathbf{x}) = P(Y = \infty|\mathbf{X} = \mathbf{x})$, and the conditional survival function of the uncured individuals, also known as latency, is $S_0(t|\mathbf{x}) = P(Y > t \mid Y < \infty, \mathbf{X} = \mathbf{x})$. The mixture cure model writes the survival function $S(t|\mathbf{x}) = 1 - F(t|\mathbf{x}) = P(Y > t|\mathbf{X} = \mathbf{x})$ as

$$S(t \mid \mathbf{x}) = 1 - p(\mathbf{x}) + p(\mathbf{x})S_0(t \mid \mathbf{x}). \tag{1}$$

Assuming model (1), the cure rate and the latency can be written in terms of the survival function $S(t|\mathbf{x})$ as follows:

$$1 - p(\mathbf{x}) = \lim_{t\to\infty} S(t \mid \mathbf{x}) > 0, \quad S_0(t \mid \mathbf{x}) = \frac{S(t \mid \mathbf{x}) - \{1 - p(\mathbf{x})\}}{p(\mathbf{x})}.$$

Therefore, the availability of a suitable estimator of $S(t|\mathbf{x})$ would yield appropriate estimators of the cure probability and the latency directly.

One key issue in cure models is identifiability. This arises because of the lack of cure status information at the end of the follow-up period, hence resulting in difficulties in distinguishing models with high incidence of susceptibles and long tails of the latency distribution from low incidence of susceptibles and short tails of the latency distribution (Li et al., 2001). Following the argumentation of Hanin and Huang (2014), who discussed in detail the identifiability of the mixture cure model, model (1) is identifiable if the latency function is proper. Thus, we assume that $\lim_{t\to\infty} S_0(t|\mathbf{x}) = 0$ for all $\mathbf{x}$. This condition is similar to the zero-tail constraint in Taylor (1995), López-Cheda, Cao, et al. (2017), and other papers.

2.2 | Proposed estimators

Without loss of generality, for simplicity we only consider a single continuous covariate X with density function $m(x)$. As shown in the Appendix, an estimator of the conditional cumulative hazard function of Y, $\Lambda(t|x)$, when the cure status is partially known is

$$\widehat{\Lambda}_h^c(t \mid x) = \sum_{i=1}^{n} \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}, \tag{2}$$

where $X_{[i]}$, $\delta_{[i]}$, $\xi_{[i]}$, and $\nu_{[i]}$ are the concomitants of the ordered observed times $T_{(1)} \leq \cdots \leq T_{(n)}$; $B_{h[i]}(x)$ are the Nadaraya–Watson weights,

$$B_{h[i]}(x) = \frac{K_h\left(x - X_{[i]}\right)}{\sum_{j=1}^{n} K_h\left(x - X_j\right)};$$

and $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function $K(\cdot)$ rescaled with bandwidth h. We work with Nadaraya–Watson kernel estimates because it is the natural choice for random design regression.

The corresponding product-limit estimator of the conditional survival function $S(t|x)$ when the cure status is partially known, is

$$\widehat{S}_h^c(t \mid x) = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right\}. \tag{3}$$

An important feature of these estimators is that subjects who are known to be cured below time $T_{(i)}$ remain in the risk set, that is, they are counted in the denominator. In the following, we also refer to this estimator as $1-\widehat{F}_h^c(t|x)$. A motivation for estimators (2) and (3) is given in the Appendix.

Proposition 1. *The proposed estimator $\widehat{S}_h^c(t|x)$ has the following general properties.*

1. *When there are no censored observations known to be cured, that is, $\xi_i\nu_i = 0$ for $i = 1, \dots, n$, $\widehat{S}_h^c(t|x)$ reduces to Beran's estimator:*

$$\widehat{S}_h\left(t \mid x\right) = \prod_{i=1}^{n}\left\{1 - \frac{\delta_{[i]}B_{h[i]}\left(x\right)\mathbf{1}\left(T_{(i)} \le t\right)}{\sum_{j=i}^{n}B_{h[j]}\left(x\right)}\right\}. \tag{4}$$

2. *In the specific case when some individuals are observed as cured when their survival time exceeds a known fixed cure threshold, $\widehat{S}_h^c(t|x)$ also reduces to Beran's estimator in (4).*
3. *When there is no censoring, $\widehat{S}_h^c(t|x)$ reduces to the kernel-type estimator of the conditional survival function (Nadaraya, 1964):*

$$\tilde{S}_h\left(t \mid x\right) = \sum_{i=1}^{n} B_{h[i]}\left(x\right)\mathbf{1}\left(T_{(i)} > t\right).$$

4. *In an unconditional setting, the proposed estimator is*

$$\widehat{S}_n^c\left(t\right) = \prod_{i=1}^{n}\left\{1 - \frac{\delta_{[i]}\mathbf{1}\left(T_{(i)} \le t\right)}{n - i + 1 + \sum_{j=1}^{i-1}\mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}\right\}.$$

In the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, $\widehat{S}_n^c(t)$ reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992).

The proof of these properties is outlined in the Appendix.

Proposition 2. *The $1-\widehat{F}_h^c(t|x)$ estimator in (3) is the nonparametric local maximum likelihood estimator of $1-F(t|x)$.*

The proof of Proposition 2 is given in the Appendix.

2.3 | Asymptotic results

In this section, we investigate the asymptotic properties of $\widehat{\Lambda}_h^c(t|x)$ and $\widehat{S}_h^c(t|x)$. In order to prove our asymptotic results, we consider the following (sub)distribution functions:

$$\begin{aligned}
H\left(t \mid x\right) &= P\left(T \le t \mid X = x\right),\\
H^1(t \mid x) &= P\left(T \le t, \delta = 1 \mid X = x\right),\\
H^{11}(t \mid x) &= P\left(T \le t, \xi = 1, \nu = 1 \mid X = x\right),\\
J(t \mid x) &= 1 - H\left(t \mid x\right) + H^{11}\left(t \mid x\right),
\end{aligned}$$

and Assumptions 1–8. Assumptions such as these have been commonly used in literature; see, for example, Iglesias-Pérez and González-Manteiga (1999).

Theorems 1 and 2 present the asymptotic representations of $\widehat{\Lambda}_h^c(t|x)$ and $1-\widehat{F}_h^c(t|x)$, respectively. Based on these results, in Corollary 1 we show that $\widehat{\Lambda}_h^c(t|x)$ and $1-\widehat{F}_h^c(t|x)$ are strongly consistent estimators of $\Lambda(t|x)$ and $1-F(t|x)$, respectively. The asymptotic normality of $1-\widehat{F}_h^c(t|x)$ is proved in Theorem 3.

Theorem 1. *Suppose that Assumptions 1–8 hold, and the bandwidth $h=(h_n)$ satisfies $h\to 0, \log n/(nh)\to 0$ and $nh^5/\log n = O(1)$ as $n\to\infty$. Then, for $x\in I, t\in[a,b]$ we have*

$$\widehat{\Lambda}_h^c(t\mid x)-\Lambda(t\mid x)=\sum_{i=1}^{n}\tilde{B}_{hi}(x)\,\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)+R_{n1}(t,x),$$

with

$$\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)=\frac{\mathbf{1}(T_i\le t,\delta_i=1)}{J(T_i^-\mid x)}-\int_0^t\{\mathbf{1}(T_i\ge v)+\mathbf{1}(T_i<v,\xi_i\nu_i=1)\}\frac{dH^1(v\mid x)}{J^2(v^-\mid x)}, \tag{5}$$

$$\tilde{B}_{hi}(x)=\frac{1}{m(x)}\frac{1}{nh}K\left(\frac{x-X_i}{h}\right), \tag{6}$$

where $R_{n1}(t,x)$ satisfies

$$\sup_{a\le t\le b,x\in I}\mid R_{n1}(t,x)\mid = O\left\{(nh)^{-3/4}(\log n)^{3/4}\right\}\ \text{almost surely.}$$

Theorem 2. *Suppose that Assumptions 1–8 hold, and the bandwidth $h=(h_n)$ satisfies $h\to 0, \log n/(nh)\to 0$ and $nh^5/\log n = O(1)$ as $n\to\infty$. Then, for $x\in I, t\in[a,b]$ we have*

$$\widehat{F}_h^c(t\mid x)-F(t\mid x)=\{1-F(t\mid x)\}\sum_{i=1}^{n}\tilde{B}_{hi}(x)\,\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)+R_{n2}(t,x),$$

where $\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)$ is defined in (5), $\tilde{B}_{hi}(x)$ in (6) and $R_{n2}(t,x)$ satisfies

$$\sup_{a\le t\le b,x\in I}\mid R_{n2}(t,x)\mid = O\left\{(nh)^{-3/4}(\log n)^{3/4}\right\}\ \text{almost surely.} \tag{7}$$

The sketch of the proofs of Theorems 1 and 2 is outlined in the Supporting Information. The detailed proofs follow that of Theorem 2 of Iglesias-Pérez and González-Manteiga (1999) for Beran's estimator. As an immediate consequence of these theorems, the following corollary on the strong consistency of the estimators $\widehat{\Lambda}_h^c(t|x)$ and $1-\widehat{F}_h^c(t|x)$ is obtained.

Corollary 1. *Suppose that Assumptions 1–8 hold, and the bandwidth $h=(h_n)$ satisfies $h\to 0, \log n/(nh)\to 0$ and $nh^5/\log n = O(1)$ as $n\to\infty$. Then, for $x\in I, t\in[a,b]$, we have*

$$\sup_{a\le t\le b,x\in I}\mid \widehat{\Lambda}_h^c(t\mid x)-\Lambda(t\mid x)\mid = O\left\{(nh)^{-1/2}(\log n)^{1/2}\right\}\ \text{almost surely,}$$

and

$$\sup_{a\le t\le b,x\in I}\mid \widehat{F}_h^c(t\mid x)-F(t\mid x)\mid = O\left\{(nh)^{-1/2}(\log n)^{1/2}\right\}\ \text{almost surely.}$$

The proof of Corollary 1 is outlined in the Appendix.

Proposition 3. *Suppose that Assumptions 1–8 hold, and the bandwidth $h = (h_n)$ satisfies $h \to 0, \log n/(nh) \to 0$ and $nh^5/\log n = O(1)$ as $n \to \infty$. Then, the bias and variance of $1 - \widehat{F}_h^c(t|x)$ are, respectively,*

$$\mu_{h,c}(t,x) = h^2 B_c(t,x) + O\left(h^4\right), \quad \sigma^2_{h,c}(t,x) = (nh)^{-1} s_c^2(t,x) + O(n^{-1}h), \tag{8}$$

with

$$B_c(t,x) = \frac{\{1 - F(t \mid x)\}\{2\Phi_c'(x,t,x)\, m'(x) + \Phi_c''(x,t,x)\, m(x)\} d_K}{2m(x)}, \tag{9}$$

$$s_c^2(t,x) = \frac{\{1 - F(t \mid x)\}^2 \Phi_1^c(x,t,x)\, c_K}{m(x)}, \tag{10}$$

where $d_K = \int v^2 K(v)dv, c_K = \int K^2(v)dv$,

$$\Phi_c(y,t,x) = E\{\zeta(T,\delta,\xi,\nu,t,x) \mid X = y\}, \quad \Phi_1^c(y,t,x) = E\left\{\zeta^2(T,\delta,\xi,\nu,t,x) \mid X = y\right\},$$

with $\zeta(T,\delta,\xi,\nu,t,x)$ given in (5). Besides, $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ are the first and second derivatives of $\Phi_c(y,t,x)$ with respect to y.

The proof of Proposition 3 is outlined in the Appendix. The following theorem, whose proof is in the Appendix, establishes the asymptotic normality of $1 - \widehat{F}_h^c(t \mid x)$.

Theorem 3. *Suppose that Assumptions 1–8 hold, then, for $x \in I$ and $t \in [a,b]$ it follows that*

(i) if $nh^5 \to 0$ and $(\log n)^3/(nh) \to 0$, then

$$(nh)^{1/2}\left\{\widehat{F}_h^c(t \mid x) - F(t \mid x)\right\} \to N(0, s_c^2(t,x)) \text{ in distribution;}$$

(ii) if $nh^5 \to C^5 > 0$, then

$$(nh)^{1/2}\left\{\widehat{F}_h^c(t \mid x) - F(t \mid x)\right\} \to N(C^{5/2}B_c(t,x), s_c^2(t,x)) \text{ in distribution,}$$

with $B_c(t,x)$ given in (9), $s_c^2(t,x)$ in (10) and C is constant.

2.4 | Effect of ignoring the cure status

In this section, we make a theoretical comparison between the proposed estimator $1 - \widehat{F}_h^c(t|x)$ and Beran's estimator. The asymptotic properties of Beran's estimator were obtained by Iglesias-Pérez and González-Manteiga (1999) and Van Keilegom and Veraverbeke (1997), among others. More precisely, in order to understand the effect of ignoring the cure status, the dominant terms of the bias and variance of Beran's estimator are compared with those of the proposed estimator. The asymptotic bias and variance of Beran's estimator are, respectively,

$$\mu_h(t,x) = h^2 B(t,x) + O\left(h^4\right) \text{ and } \sigma_h^2(t,x) = (nh)^{-1} s^2(t,x) + O(n^{-1}h), \tag{11}$$

with

$$B(t,x) = \frac{\{1 - F(t \mid x)\}\{2\Phi'(x,t,x)\, m'(x) + \Phi''(x,t,x)\, m(x)\} d_K}{2m(x)} \tag{12}$$

and

$$s^2(t,x) = \frac{\{1 - F(t \mid x)\}^2 \Phi_1(x,t,x)\, c_K}{m(x)}, \tag{13}$$

where, see Lemmas 4 and 5 in López-Cheda, Jácome, et al. (2017),

$$\Phi(y,t,x) = \int_0^t \frac{dH^1(v \mid y)}{1 - H(v^- \mid x)} - \int_0^t \frac{1 - H(v^- \mid y)}{\{1 - H(v^- \mid x)\}^2} dH^1(v \mid x),$$
$$\Phi_1(x,t,x) = \int_0^t \frac{dH^1(v \mid x)}{\{1 - H(v^- \mid x)\}^2}, \tag{14}$$

and $\Phi'(y,t,x)$ and $\Phi''(y,t,x)$ are the first and the second derivatives of $\Phi(y,t,x)$ with respect to y. Expressions (11)–(13) for Beran's estimator are equivalent to the bias and variance terms (8)–(10) for $\widehat{S}_h^c(t \mid x)$, replacing $\Phi_c(x,t,x)$ and $\Phi_1^c(x,t,x)$ with $\Phi(x,t,x)$ and $\Phi_1(x,t,x)$, respectively. From Lemmas 2 and 5 in the Supporting Information, we have

$$\Phi_c(y,t,x) = \int_0^t \frac{dH^1(v \mid y)}{1 - H(v^- \mid x) + H^{11}(v^- \mid x)} - \int_0^t \frac{1 - H(v^- \mid y) + H^{11}(v^- \mid y)}{\{1 - H(v^- \mid x) + H^{11}(v^- \mid x)\}^2} dH^1(v \mid x),$$
$$\Phi_1^c(x,t,x) = \int_0^t \frac{dH^1(v \mid x)}{\{1 - H(v^- \mid x) + H^{11}(v^- \mid x)\}^2}.$$

As for the variance, when the cure status information is ignored then $H^{11}(t|x) = 0$ for all t and x. Therefore, $\Phi_1^c(x,t,x) \leq \Phi_1(x,t,x)$. Note that when the same bandwidth is used for both estimators, ignoring the cure status increases asymptotically the variance of the estimator.

Returning to the bias, by applying Lemma 3 in the Supporting Information, we have

$$\Phi_c'(x,t,x) = \Phi'(x,t,x) = -\frac{S'(t^- \mid x)}{S(t^- \mid x)},$$

where $S'(t|x)$ is the derivative of $S(t|x)$ with respect to x, meaning that the effect of knowing the cure status on the bias is given by $\Phi_c''(x,t,x)$. From Lemma 4 in the Supporting Information,

$$\Phi_c''(x,t,x) = 2\int_0^t \frac{G_c'(v^- \mid x)}{1 - G_c(v^- \mid x)} \frac{d}{ds}\left\{\frac{S'(s \mid x)}{S(s \mid x)}\right\}\Big|_{s=v^-} dv - \frac{S''(t^- \mid x)}{S(t^- \mid x)}, \tag{15}$$

with

$$1 - G_c(t \mid x) = 1 - G(t \mid x) + \pi_1(t,x)\{1 - p(x)\}G_1(t \mid x),$$

where

$$\pi_1(t,x) = P(\xi = 1 \mid \nu = 1, C \leq t, X = x), \quad G_1(t \mid x) = P(C \leq t \mid \nu = 1, X = x), \tag{16}$$

and $S'(t|x)$, $S''(t|x)$ and $G'(t \mid x)$ refer to the derivatives with respect to x. If the cure status is ignored, that is, $\pi_1(x,t) = 0$ for all t and x, then (15) reduces to

$$\Phi''(x,t,x) = 2\int_0^t \frac{G'(v^- \mid x)}{1 - G(v^- \mid x)} \frac{d}{ds}\left\{\frac{S'(s \mid x)}{S(s \mid x)}\right\}\Big|_{s=v^-} dv - \frac{S''(t^- \mid x)}{S(t^- \mid x)}.$$

In terms of bias, the advantage of knowing the cure status is not straightforward as it depends on the derivative with respect to x of the cure probability $1 - p(x)$ and the functions $\pi_1(t, x)$ and $G_1(t, x)$ in (16). This implies that there is no guarantee that there will be a gain in terms of bias for the proposed estimator with respect to Beran's estimator.

3 | BANDWIDTH SELECTION

Bootstrap procedures have been successfully used to address the issue of bandwidth selection in the context of the mixture cure model (López-Cheda, Cao, et al., 2017; López-Cheda, Jácome, et al., 2017). Next, we propose a bootstrap bandwidth selector to choose the smoothing parameter h of the proposed estimator $\widehat{S}^c_h(t|x)$. The bootstrap bandwidth, h^*_x, is the bandwidth minimizing the bootstrap version of the mean integrated squared error (MISE). This bootstrap MISE can be approximated using Monte Carlo by

$$\text{MISE}^*_x(h) \simeq \frac{1}{B}\sum_{b=1}^{B}\int \left\{\widehat{S}^{c,*b}_h(v \mid x) - \widehat{S}^c_{g_x}(v \mid x)\right\}^2 \omega(v, x)dv, \tag{17}$$

where $\widehat{S}^{c,*b}_h(t|x)$ is the proposed estimator computed with the bth bootstrap resample and a bandwidth h, and $\widehat{S}^c_{g_x}(t|x)$ is the same estimator computed with the original sample and with a pilot bandwidth g_x. Note that $\omega(v, x)$ is a nonnegative weight function, intended to give lower weight in the right tail of the distribution. The algorithm to compute the bootstrap bandwidth for a fixed covariate value x, is as follows:

Step 1. *With the original sample and the pilot bandwidth g_x, compute $\widehat{S}^c_{g_x}(t|x)$.*

Step 2. *Choose a dense enough grid of L bandwidths $\{h_1, \dots, h_L\}$.*

Step 3. *Generate B bootstrap resamples $\{(X^{(b)}_i, T^{*(b)}_i, \delta^{*(b)}_i, \xi^{*(b)}_i, \xi^{*(b)}_i \nu^{*(b)}_i) : i = 1, \dots, n\}$, for $b = 1, \dots, B$.*

Step 4. *For the bth bootstrap resample and the bandwidths h_l, for $l = 1, \dots, L$, compute $\widehat{S}^{c,*b}_{h_l}(t|x)$.*

Step 5. *For $h_l, l = 1, \dots, L$, compute the Monte Carlo approximation of $MISE^*_x(h_l)$ given by (17).*

Step 6. *The bootstrap bandwidth, h^*_x, is the bandwidth of the grid $\{h_1, \dots, h_L\}$ that minimizes the approximation of $MISE^*_x(h)$ in (17).*

The bootstrap resamples in Step 3 are generated as follows: fix x, for $i = 1, \dots, n$, set $X^*_i = X_i$ and generate a 4-tuple $(T^*_i, \delta^*_i, \xi^*_i, \xi^*_i \nu^*_i)$ from the weighted empirical conditional distribution of $\{(T_1, \delta_1, \xi_1, \xi_1\nu_1), \dots, (T_n, \delta_n, \xi_n, \xi_n\nu_n)\}$:

$$\widehat{F}_{g_x}(t, d, w, z \mid x) = \sum_{i=1}^{n} B_{g_x i}(x)\mathbf{1}\left(T_i \le t, \delta_i \le d, \xi_i \le w, \xi_i\nu_i \le z\right),$$

where $B_{g_x i}(x)$ are the Nadaraya–Watson weights with bandwidth g_x.

The pilot bandwidth g_x should tend to 0 at a slower rate than h^*_x. This oversmoothing pilot bandwidth is required for the bootstrap integrated squared bias and variance to be asymptotically efficient estimators of the integrated squared bias and variance terms. For practical applications we recommend to use $g_x = c_x n^{-1/9}$, as suggested by Li and Datta (2001), which coincides with the optimal order obtained by Cao and González-Manteiga (1993) for the uncensored case. Simulation results in the Supporting Information (see also López-Cheda, Cao, et al., 2017; López-Cheda, Jácome, et al., 2017) show that the choice of the pilot bandwidth has a small effect on the selected bootstrap bandwidth. We propose to use the same local pilot bandwidth as in López-Cheda, Cao, et al. (2017) and López-Cheda, Jácome, et al. (2017):

$$g_x = \frac{d^+_k(x) + d^-_k(x)}{2} 100^{1/9} n^{-1/9},$$

where $d_k^+(x)$ and $d_k^-(x)$ are the distances from x to the kth nearest neighbor on the right and left, and k is a suitably chosen integer depending on the sample size. If there are not at least k neighbors on the right (or left), we use $d_k^+(x) = d_k^-(x)$ (or $d_k^-(x) = d_k^+(x)$). Following López-Cheda, Cao, et al. (2017) and López-Cheda, Jácome, et al. (2017), we suggest setting $k = [n/4]$.

4 | SIMULATION STUDY

We studied the practical performance of $\widehat{S}_h^c(t|x)$ through a simulation study. We considered the conditional survival function $S(t|x) = 1 - p(x) + p(x)S_0(t|x)$, where

$$S_0(t \mid x) = \begin{cases} \dfrac{\exp(-\alpha(x)\,t) - \exp(-\alpha(x)\,4.605)}{1 - \exp(-\alpha(x)\,4.605)} & 0 \le t \le 4.605 \\ & t > 4.605 \end{cases}, \qquad \alpha(x) = \exp\left(\frac{x+20}{40}\right).$$

We simulated two scenarios given by the cure rates:

$$1 - p_1(x) = 1 - \frac{\exp(0.476 + 0.358x)}{1 + \exp(0.476 + 0.358x)}, \quad 1 - p_2(x) = 0.5 - \frac{1}{16000}x^3.$$

The censoring variable C was generated from an exponential distribution with mean $10/3$. The covariate X was uniformly distributed on the interval $[-20, 20]$. The percentage of censoring was 54% and the average cure probability 0.467 in Scenario 1, whereas in Scenario 2, 61% of the observations were censored and the average cure probability was 0.5. In both scenarios, the proportion of the identified cured individuals was $\pi = 0.2, 0.8$ and 1. Data were generated so that the censoring times C and the lifetimes Y were independent conditionally on X. We generated 1000 datasets of sample sizes $n = 50, 100$, and 200. This section contains the results for $\pi = 0.8$ and $n = 100$; the rest of the results can be found in the Supporting Information.

Our first goal was to evaluate the performance of $\widehat{S}_h^c(t|x)$ in terms of the MISE. It was approximated over a grid of bandwidths equispaced in a logarithmic scale, from $h_1 = 3$ to $h_{100} = 20$ in Scenario 1, and from $h_1 = 4$ to $h_{101} = 100$ in Scenario 2. For the weight function we chose $\omega(t, x) = \mathbf{1}(a_x \le t \le b_x)$ where $a_x = 0$ and $b_x = \tau_x$, the 90th percentile of $S_0(t|x)$. We compared $\widehat{S}_h^c(t|x)$ computed in a grid of bandwidths with Beran's estimator, $\widehat{S}_h(t|x)$, computed with the optimal bandwidth. The semiparametric estimator by Bernhardt (2016), which fits a logistic regression for the cure probability and seminonparametric accelerated failure time model for the latency function, was also considered for comparison. The semiparametric estimator is expected to perform well in Scenario 1. We chose the Epanechnikov kernel to compute $\widehat{S}_h^c(t|x)$ and $\widehat{S}_h(t|x)$.

Figure 1 shows the MISE curves of the three estimators. In Scenario 1, as expected, the semiparametric estimator behaves well. Nevertheless, both $\widehat{S}_h^c(t|x)$ and $\widehat{S}_h(t|x)$ are quite competitive for suitable values of the bandwidth, even beating the semiparametric estimator for some values of X close to 0 and 20. In Scenario 2, both nonparametric estimators outperform the semiparametric estimator. Taking into account the known cure status gives either similar or better results than ignoring it for most values of X, especially in Scenario 2 (see Figure 1). In Table 1, the performance of the estimators is compared in terms of the integrated squared bias, integrated variance and MISE for the covariate values $x = -10, 0$, and 10. In both scenarios, at $x = -10$, the proposed estimator has smaller integrated squared bias and variance than Beran's estimator. On the contrary, for $x = 10$, the integrated squared bias and variance of Beran's estimator is smaller compared to $\widehat{S}_h^c(t|x)$ estimator. As expected, the integrated squared bias and variance estimates for the semiparametric estimator are larger in Scenario 2.

The performance of the bootstrap bandwidth selector was assessed using $B = 1000$ resamples and an increased grid of bandwidths from 1.5 to 100 for both scenarios. Figure 2 displays the quartiles of the selected bootstrap bandwidths together with the optimal bandwidth. Corresponding contour plots in Figure 3 show the density of the bootstrap bandwidths and the MISE of $\widehat{S}_h^c(t|x)$ as a function of the bandwidth h and the covariate value x. Figure 4 shows the MISE of $\widehat{S}_h^c(t|x)$ as a function of the bandwidth h, for four values of the covariate. Figures 2 and 3 illustrate that the bootstrap bandwidth approximates quite well the optimal bandwidth. Note that in Figure 3 vertical contour lines indicate that, given x, the MISE of $\widehat{S}_h^c(t|x)$ tends to be constant as a function of h. Therefore, different bandwidths would yield approximately the same

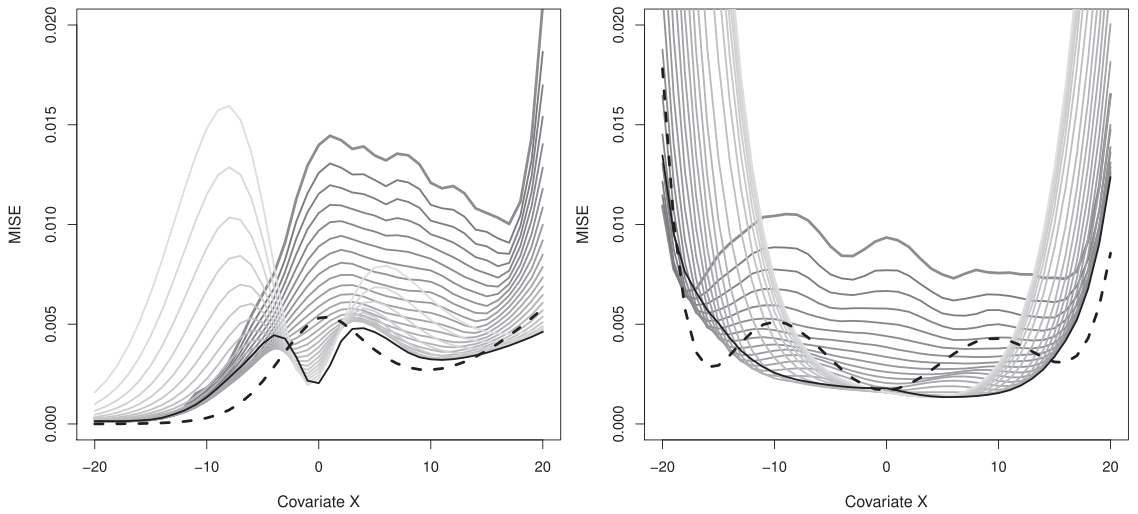

FIGURE 1 MISE of the proposed estimator $\widehat{S}_h^c(t|x)$ for a selection of 25 bandwidths from the lowest (darkest gray line) to the highest (lightest gray line) in Scenario 1 (left) and Scenario 2 (right). Also $\widehat{S}_h(t|x)$ computed with the optimal bandwidth (solid black line), and of the estimator by Bernhardt (2016) (dashed black line)

TABLE 1 Integrated squared bias (Ibias^2), integrated variance (Ivar), and MISE of the proposed estimator, $\widehat{S}_h^c(t|x)$, Beran's estimator, $\widehat{S}_h(t|x)$ (both computed with the optimal bandwidth), and the semiparametric estimator by Bernhardt (2016)

		Proposed				Beran				Semiparametric		
			Ibias^2	Ivar	MISE		Ibias^2	Ivar	MISE	Ibias^2	Ivar	MISE
Scenario	x	h	$\times 10^3$	$\times 10^3$	$\times 10^3$	h	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$
1	−10	6.582	0.119	1.022	1.141	6.334	0.163	1.177	1.340	0.002	0.299	0.301
	0	20.000	0.371	1.834	2.205	20.000	0.119	1.927	2.046	0.035	5.274	5.310
	10	12.152	0.375	2.902	3.277	12.387	0.355	2.885	3.240	0.372	2.337	2.709
2	−10	25.874	0.076	2.206	2.282	23.492	0.065	2.501	2.566	2.795	2.330	5.125
	0	36.867	0.058	1.517	1.575	30.392	0.151	1.632	1.783	0.037	1.652	1.689
	10	26.721	0.103	1.474	1.577	28.497	0.058	1.492	1.550	2.171	2.109	4.280

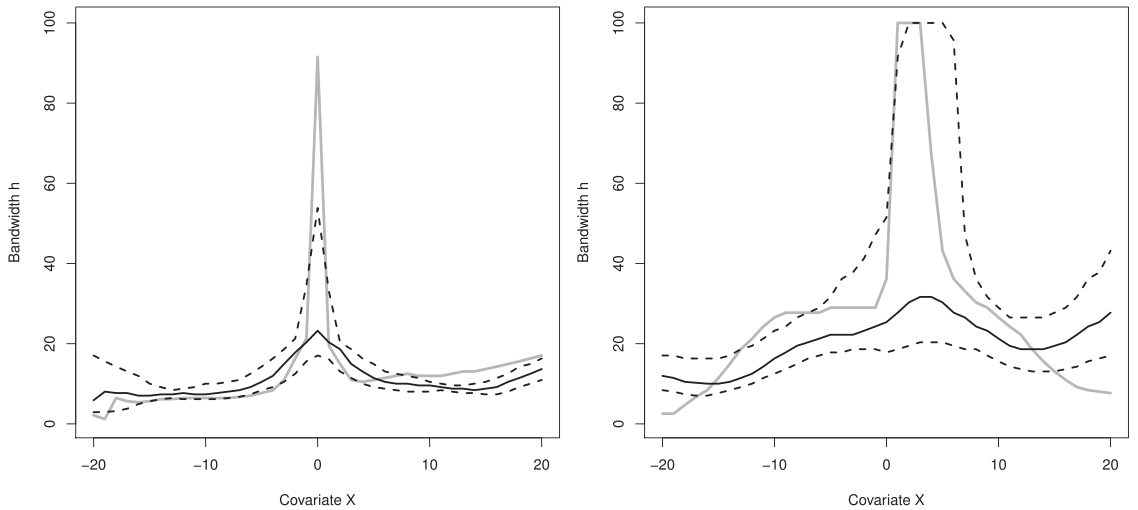

FIGURE 2 Median (solid black line) and first and third quartiles (dashed lines) of the bootstrap bandwidths for $\widehat{S}_h^c(t|x)$ in Scenario 1 (left) and Scenario 2 (right). The optimal bandwidth (solid gray line) is displayed as reference

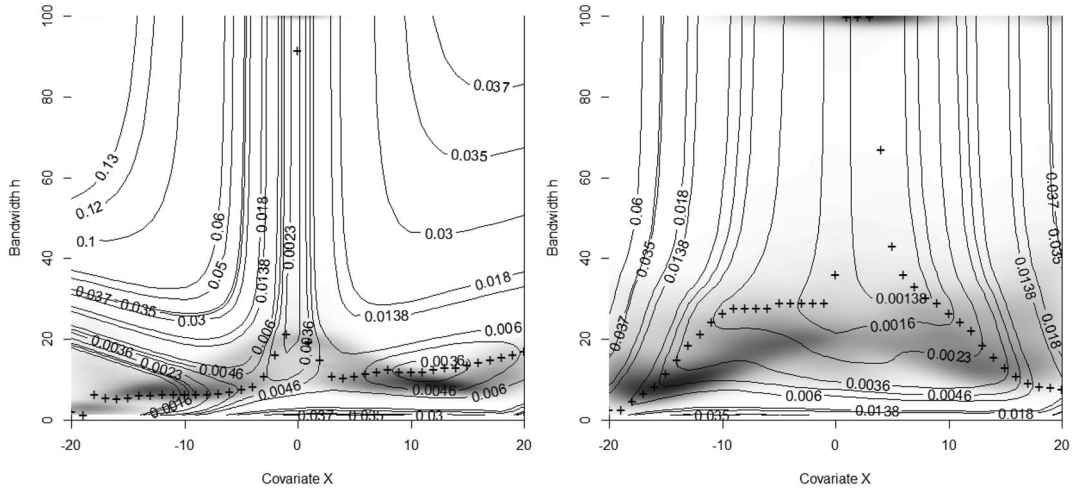

FIGURE 3 Contour plots of the MISE of $\widehat{S}_h^c(t|x)$ as a function of the bandwidth h and the covariate value x in Scenario 1 (left) and Scenario 2 (right). For each value of covariate, the optimal bandwidth is marked with a cross. The density of the bootstrap bandwidths is shown in gray scale

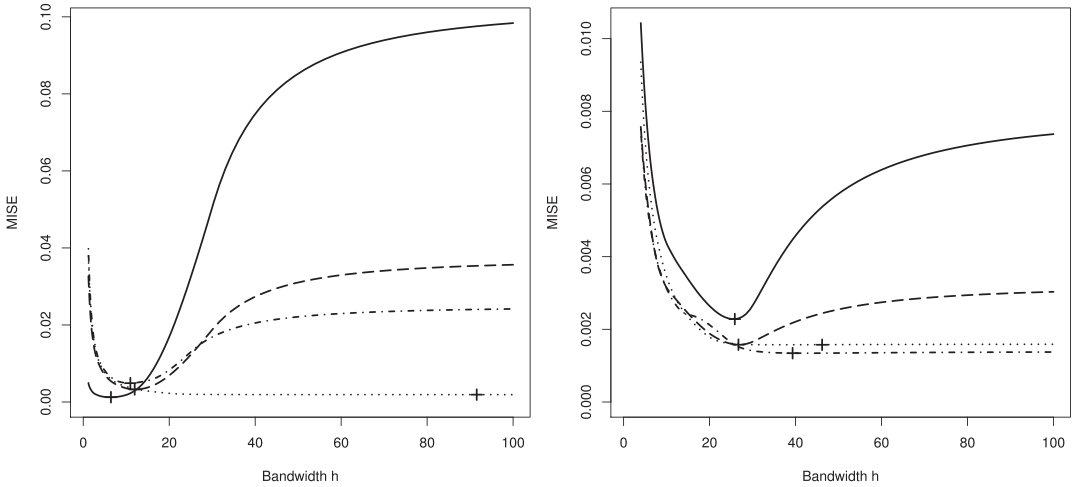

FIGURE 4 MISE of $\widehat{S}_h^c(t|x)$ as a function of the bandwidth h for four different values of the covariate $x = -10$ (solid line), $x = 0$ (dotted), $x = 5$ (dot-dashed) and $x = 10$ (long dash) in Scenario 1 (left) and Scenario 2 (right). For each value of the covariate, the optimal bandwidth where the minimum MISE is reached is marked with a cross

MISE. In those cases, the bootstrap bandwidth being far from the optimal bandwidth does not imply a loss of efficiency. Similar results are observed in Figure 4. For example, let us consider $x = 0$ in Scenario 2, we see that the MISE initially decreases as the bandwidth increases, although afterward it becomes constant.

5 | APPLICATION TO REAL DATA

To illustrate the practical performance of $\widehat{S}_h^c(t|x)$, we considered a dataset of 233 patients of sarcoma cancer aged 20–90 from the University Hospital of Santiago de Compostela, Spain (CHUS). Sarcoma is a rare type of cancer that represents 1% of all adult solid malignancies. If a tumor can be surgically removed to render the patient with sarcoma free of detectable disease, 5 years is the survival time at which sarcoma oncologists assume long-term remissions (Choy, 2014). Overall, 59 patients died from sarcoma, and the remaining 174 patients were censored. Among censored patients, 18 patients were tumor free for more than 5 years. Hence, they were assumed to be long-term survivors. The aim was to estimate the

TABLE 2 Descriptive demographic and clinical characteristics of sarcoma patients stratified by age, sex, location of the sarcoma, metastatic, and the margin status

Characteristics	*n* (%)	Death	Censored	
			Cured	Unknown
Age[a]				
<60	105 (45.3)	25	9	71
≥60	127 (54.7)	33	9	85
Sex				
Male	100 (42.9)	25	7	68
Female	133 (57.1)	34	11	88
Tumor site[a]				
Retroperitoneal	86 (37.2)	28	4	54
Extremities	70 (30.3)	14	5	51
Other sites	75 (32.5)	16	9	50
Metastatic[a]				
No	112 (67.1)	11	9	92
Yes	55 (32.9)	32	3	20
Margin status[a]				
Negative	133 (65.8)	26	12	95
Positive	69 (34.2)	17	3	49

Note: In addition, the total number of patients for each subgroup (n), the number of patients died of sarcoma (death), those who were known to be cured (cured), and those with unknown cure status (unknown) are given.

[a]A few missing data.

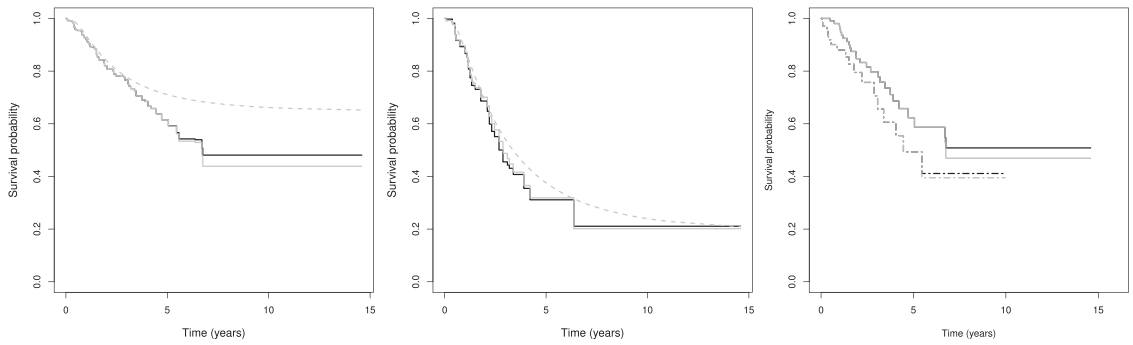

FIGURE 5 Survival estimates for sarcoma patients aged 40 (left) and 90 (center) years are obtained with the proposed estimator $\widehat{S}_h^c(t|x)$ (solid black line) and Beran's estimator $\widehat{S}_h(t|x)$ (solid gray line), both computed with the corresponding bootstrap bandwidth, and the semi-parametric estimator of Bernhardt (2016) (dashed gray line). The right figure shows survival estimates stratified by the margin status, negative margin (solid lines) versus positive margin (dashed lines). These estimates are computed using the proposed estimator $\widehat{S}_n^c(t)$ (black lines) and the Kaplan–Meier estimator (gray lines)

survival time of the patients until death from sarcoma as a function of covariates such as the age at diagnosis, sex, tumor site, cancer spread (metastasis), and the margin status. The variables selected for estimating the survival probabilities were previously reported to be related to long-term sarcoma survival (Carbonnaux et al., 2019; Daigeler et al., 2014, among others).

Table 2 shows the descriptive demographic and clinical characteristics of sarcoma patients by age, sex, and relevant clinical factors. Of the 233 patients, 100 (42.9%) were males. Tumor site was categorized as retroperitoneal, extremities, or other. Most tumors were found in the retroperitoneum (37%) and in the extremities (30%), with other areas of the body accounting for about 33%. Fifty-five (32.9%) patients were diagnosed of metastatic sarcoma.

Figure 5 compares the results obtained using the proposed estimator $\widehat{S}_h^c(t|x)$, which takes into account the 18 long-term survivors, with Beran's estimator $\widehat{S}_h(t|x)$, which ignores individuals known to be cured and treats them as simply

censored observations. Both estimators were computed with the corresponding bootstrap bandwidth. The semiparametric estimator of Bernhardt (2016) was also considered as reference. All estimators show that the survival curve decreases when age increases from 40 to 90 years. We find the largest differences between the proposed estimator and Beran's estimator at the right tail of the distribution, where the survival curve for $\widehat{S}_h^c(t|x)$ is slightly higher. As the cure probability can be obtained as the limit of $S(t|x)$ when $t \to \infty$ using the proposed estimator of the survival curve will yield in higher estimates of the probability of cure.

On the other hand, the survival curve estimated by the semiparametric estimator of Bernhardt (2016) tends to decrease much slower than those obtained with the nonparametric estimators $\widehat{S}_h^c(t|x)$ and $\widehat{S}_h(t|x)$, suggesting that further testing is required to provide evidence that assumptions in the semiparametric model are fulfilled.

Figure 5 on the right shows the survival curves of sarcoma patients stratified by the margin status. In this case, the proposed estimator in an unconditional setting $\widehat{S}_n^c(t)$ is applied and the Kaplan and Meier (1958) estimator is considered as reference. The survival curves tend to decrease with time in both subgroups. The positive margin survival curve decreases slightly faster than the negative survival curve. In addition, the distinction between $\widehat{S}_n^c(t)$ and the Kaplan–Meier estimator is found at the right tail of the distribution with the survival curves estimated by $\widehat{S}_n^c(t)$ being slightly higher than the Kaplan–Meir curves. For example, the survival probability, at the tail of the distribution, for patients with negative margins is around 0.51 when estimated by $\widehat{S}_n^c(t)$, while it is around 0.47 when estimated by the Kaplan–Meier estimator. Again, the estimated probability of cure is slightly higher when the survival curve is fitted taking into account the known cured subjects.

6 | DISCUSSION

The proposed nonparametric estimator of the survival function takes advantage of the additional cure status information that Beran's estimator ignores. As a further step, it could be used to derive nonparametric estimators for the cure probability and the latency function.

Thus far, the estimation procedure was discussed involving a single continuous covariate. It would be of interest to extend our estimator to the case of multiple covariates, with $\mathbf{X}$ a vector of mixed discrete, categorical, and/or continuous variables. One possibility is to consider product kernels (Li & Racine, 2008). Another possibility is to use dimension reduction techniques like a single-index model. Specifically, the idea is to apply the proposed estimator of the survival function with a new covariate given by an estimator of the index $\tilde{\mathbf{X}} = \beta^T \mathbf{X}$, with β a parameter vector of the same dimension of $\mathbf{X}$. Semiparametric index estimation of the conditional distribution in the presence of right censoring was considered recently by Li and Patilea (2018).

Although the proposed estimator utilizes the cure status information and shows good results both theoretically and practically, it is not without limitations. It is competitive over Beran's estimator in terms of the MISE, showing a general better behavior. But for some values of the covariate, it does not result in an improvement but a slightly worse MISE performance. The clear gain in terms of the integrated variance could be cancelled out by the integrated squared bias, which depends on the cure probability, the conditional censoring distribution, and the conditional probability of observed cured individuals. For the semiparametric estimator by Bernhardt (2016), our numerical experience indicates that if the sample size is small (<100), it is challenging to obtain stable values for the model parameters.

The `R` package `npcure` by López-de-Ullibarri et al. (2020) provides the nonparametric estimation and testing procedures in mixture cure models proposed by López-Cheda, Cao, et al. (2017), López-Cheda, Jácome, et al. (2017), and López-Cheda et al. (2020), including Beran's estimator. The situation when cure status is partially known is not currently supported by the package but will be considered in future versions. Further, the estimator of the conditional survival function introduced in this paper and subsequent estimators of the cure rate and latency functions will be incorporated in the upgraded package.

ACKNOWLEDGMENTS

We are grateful to Dr. Ángel Díaz-Lagares, Head of Cancer Epigenomics Lab, Translational Medical Oncology Group (IDIS, CHUS), for providing with the sarcoma dataset obtained from the public The Cancer Genome Atlas (TCGA) program. This work has been supported by MINECO grant MTM2017-82724-R, the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015), and the Centro de Investigación de Galicia "CITIC," funded by Xunta de Galicia and the European Union (European Regional Development Fund-Galicia 2014-2020 Program), by grant ED431G 2019/01, all of them through the ERDF.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The sarcoma dataset supporting the reproducibility of the findings is provided as Supplementary Information.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to their computational complexity.

ORCID

Wende Clarence Safari https://orcid.org/0000-0003-4639-7552
Ignacio López-de-Ullibarri https://orcid.org/0000-0002-3438-6621
María Amalia Jácome https://orcid.org/0000-0001-7000-9623

REFERENCES

Amico, M., & Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, *5*, 311–342.

Beran, R. (1981). *Nonparametric regression with randomly censored survival data* (Technical Report). Berkeley, CA: University of California.

Bernhardt, P. W. (2016). A flexible cure rate model with dependent censoring and a known cure threshold. *Statistics in Medicine*, *35*(25), 4607–4623.

Betensky, R. A., & Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics*, *57*(1), 282–286.

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B*, *11*(1), 15–53.

Cao, R., & González-Manteiga, W. (1993). Bootstrap methods in regression smoothing. *Journal of Nonparametric Statistics*, *2*, 379–388.

Carbonnaux, M., Brahmi, M., Schiffler, C., Meeus, P., Sunyach, M.-P., Bouhamama, A., Karanian, M., Tirode, F., Pissaloux, D., Vaz, G., Ray-Coquard, I., Blay, J.-Y., & Dufresne, A. (2019). Very long-term survivors among patients with metastatic soft tissue sarcoma. *Cancer Medicine*, *8*(4), 1368–1378.

Chen, T., & Du, P. (2018). Promotion time cure rate model with nonparametric form of covariate effects. *Statistics in Medicine*, *37*(10), 1625–1635.

Choy, E. (2014). Sarcoma after 5 years of progression-free survival: Lessons from the French sarcoma group. *Cancer*, *120*(19), 2942–2943.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Annals of Statistics*, *17*, 1157–1167.

Daigeler, A., Zmarsly, I., Hirsch, T., Goertz, O., Steinau, H., Lehnhardt, M., & Harati, K. (2014). Long-term outcome after local recurrence of soft tissue sarcoma: a retrospective analysis of factors predictive of survival in 135 patients with locally recurrent soft tissue sarcoma. *British Journal of Cancer*, *110*(6), 1456–1464.

Hanin, L., & Huang, L.-S. (2014). Identifiability of cure models revisited. *Journal of Multivariate Analysis*, *130*, 261–274.

Iglesias-Pérez, M. C., & González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *Journal of Nonparametric Statistics*, *10*(3), 213–244.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Laska, E. M., & Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics*, *48*(4), 1223–1234.

Li, C.-S., Taylor, J. M., & Sy, J. P. (2001). Identifiability of cure models. *Statistics & Probability Letters*, *54*(4), 389–395.

Li, G., & Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Annals of the Institute of Statistical Mathematics*, *53*, 708–729.

Li, Q., & Racine, J. S. (2008). Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business Economics & Statistics*, *26*(4), 423–434.

Li, W., & Patilea, V. (2018). A dimension reduction approach for conditional Kaplan–Meier estimators. *TEST*, *27*(2), 295–315.

López-Cheda, A., Cao, R., Jácome, M.A., & Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis*, *105*, 144–165.

López-Cheda, A., Jácome, M.A., & Cao, R. (2017). Nonparametric latency estimation for mixture cure models. *TEST*, *26*(2), 353–376.

López-Cheda, A., Jácome, M. A., Van Keilegom, I., & Cao, R. (2020). Nonparametric covariate hypothesis tests for the cure rate in mixture cure models. *Statistics in Medicine*, *39*(17), 2291–2307.

López-de-Ullibarri, I., López-Cheda, A., & Jácome, M. A. (2020). *npcure: Nonparametric estimation in mixture cure models* (R Package Version 0.1-5). https://cran.r-project.org/web/packages/npcure/

Maller, R. A., & Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, *79*(4), 731–739.
Maller, R. A., & Zhou, S. (1996). *Survival analysis with long-term survivors*. Chichester: Wiley.
Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory of Probability & Its Applications*, *9*(3), 497–500.
Nieto-Baraja, L. E., & Yin, G. (2008). Bayesian semiparametric cure rate model with an unknown threshold. *Scandinavian Journal of Statistics*, *35*(3), 540–556.
Patilea, V., & Van Keilegom, I. (2020). A general approach for cure models in survival analysis. *Annals of Statistics*, *48*(4), 2323–2346.
Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, *51*(3), 899–907.
Van Keilegom, I., & Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *49*(3), 467–491.
Wu, Y., Lin, Y., Lu, S.-E., Li, C.-S., & Shih, W. J. (2014). Extension of a Cox proportional hazards cure model when cure information is partially known. *Biostatistics*, *15*(3), 540–554.
Xu, J., & Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canadian Journal of Statistics*, *42*(1), 1–17.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Safari WC, López-de-Ullibarri I, Jácome MA. A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*. 2021;*63:*984–1005. https://doi.org/10.1002/bimj.202000173

APPENDIX A

The following assumptions are made.

Assumption 1.

(i) Let $I = [x_1, x_2]$ be an interval contained in the support of the density function of X, $m(x)$, such that

$$0 < \gamma = \inf_{x \in I_\varepsilon} m(x) < \sup_{x \in I_\varepsilon} m(x) = \Gamma < \infty$$

for some $I_\varepsilon = [x_1 - \varepsilon, x_2 + \varepsilon]$ with $\varepsilon > 0$ and $0 < \varepsilon\Gamma < 1$. And for all $x \in I$, Y, C are conditionally independent at $X = x$.

(ii) There exist $a, b \in \mathbb{R}$, with $a < b$ satisfying $J(t|x) \geq \theta > 0$ for $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 2. The first derivative with respect to x of function $m(x)$ exists and is continuous in $x \in I_\varepsilon$, and the first derivatives with respect to x of functions $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ exist and are continuous and bounded in $(t, x) \in [0, \infty) \times I_\varepsilon$.

Assumption 3. The second derivative with respect to x of function $m(x)$ exists and is continuous in $x \in I_\varepsilon$, and the second derivatives with respect to x of functions $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ exist and are continuous and bounded in $(t, x) \in [0, \infty) \times I_\varepsilon$.

Assumption 4. The first derivatives with respect to t of $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 5. The second derivatives with respect to t of $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 6. The first derivative with respect to x and the second derivative with respect to t of $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 7. The (sub)densities corresponding to the (sub)distribution functions $H(t|x)$, $H^1(t|x)$, and $H^{11}(t|x)$ are bounded away from 0 in $[a,b]\times I_\varepsilon$.

Assumption 8. The kernel function $K(v)$ is a symmetrical density with zero mean, vanishing outside $(-1,1)$, and the total variation is less than $\lambda<\infty$.

Motivation of the proposed estimators. The cumulative hazard function $\Lambda(t|x)$ can be written as follows:

$$\Lambda(t\mid x)=\int_0^t\frac{dF(v\mid x)}{1-F(v^-\mid x)}=\int_0^t\frac{\left\{1-G(v^-\mid x)+G^{11}(v^-\mid x)\right\}dF(v\mid x)}{\{1-G(v^-\mid x)+G^{11}(v^-\mid x)\}\{1-F(v^-\mid x)\}},\tag{A1}$$

where $G^{11}(t|x)=P(C\le t,\xi=1,\nu=1|X=x)$ is the conditional censoring subdistribution of the individuals observed to be cured. The numerator in (A1) is $dH^1(t\mid x)$:

$$\begin{aligned}&\int_0^t\left\{1-G(v^-\mid x)+G^{11}(v^-\mid x)\right\}dF(v\mid x)\\&=\int_0^t P(C\ge v\mid X=x)dF(v\mid x)+\int_0^t P(C<v,\xi=1,\nu=1\mid X=x)dF(v\mid x)\\&=P(Y\le t,C\ge Y\mid X=x)+P(Y\le t,C<Y,\xi=1,Y=\infty\mid X=x)\\&=P(T\le t,\delta=1\mid X=x)=H^1(t\mid x).\end{aligned}\tag{A2}$$

Similarly, the denominator in (A1) is $J(t^-|x)$:

$$\begin{aligned}&\left\{1-G(t^-\mid x)+G^{11}(t^-\mid x)\right\}\{1-F(t^-\mid x)\}\\&=\{P(C\ge t\mid X=x)+P(C<t,\xi=1,\nu=1\mid X=x)\}P(Y\ge t\mid X=x)\\&=P(Y\ge t,C\ge t\mid X=x)+P(Y\ge t,C<t,\xi=1,\nu=1\mid X=x)\\&=P(T\ge t\mid X=x)+P(T<t,\xi=1,\nu=1\mid X=x)\\&=1-H(t^-\mid x)+H^{11}(t^-\mid x)=J(t^-\mid x).\end{aligned}\tag{A3}$$

Taking (A2) and (A3) into account, (A1) can be written as

$$\Lambda(t\mid x)=\int_0^t\frac{dH^1(v\mid x)}{J(v^-\mid x)}.\tag{A4}$$

Consider the Nadaraya–Watson kernel estimates of $H^1(t|x)$ and $J(t^-|x)$:

$$\widehat{H}_h^1(t\mid x)=\sum_{i=1}^n B_{hi}(x)\mathbf{1}\left(T_i\le t,\delta_i=1\right),\tag{A5}$$

$$\widehat{J}_h(t^-\mid x)=\sum_{i=1}^n B_{hi}(x)\mathbf{1}\left(T_i\ge t\right)+\sum_{i=1}^n B_{hi}(x)\mathbf{1}\left(T_i<t,\xi_i\nu_i=1\right).\tag{A6}$$

The estimator of $\Lambda(t|x)$ when the cure status is partially known, $\widehat{\Lambda}_h^c(t|x)$, is obtained by plugging in (A4) the estimates (A5) and (A6). As for the estimator of the survival function, it can be shown that $S(t|x)=\exp\{-\Lambda(t|x)\}$. By considering a Taylor's expansion of the exponential function around 0 and evaluating it at each increment of $\widehat{\Lambda}_h^c(t|x)$, the estimator $\widehat{S}_h^c(t|x)$ in (3) is obtained.

Proof of Proposition 1. The estimator $\widehat{S}_h^c(t|x)$ has the following properties:

1. If there is no known cure status, $\widehat{S}_h^c(t|x)$ reduces to $\widehat{S}_h(t|x)$.

   Proof. It is straightforward because $\xi_i \nu_i = 0, i = 1, \dots, n$. □

2. In the specific case when some individuals are observed as cured when their survival time exceeds a known fixed cure threshold, $\widehat{S}_h^c(t|x)$ reduces to $\widehat{S}_h(t|x)$.

   Proof. Assume there exists a common specific known cure threshold $d_i = d$ for $i = 1, \dots, n$. This implies that in the ordered sample, $\{(X_{[i]}, T_{(i)}, \delta_{[i]}, \xi_{[i]}, \xi_{[i]}\nu_{[i]}) : i = 1, \dots, n\}$, the n_1 first observations correspond to individuals with $T_{(i)} < d$ either not cured or with unknown cure status ($\xi_{[i]}\nu_{[i]} = 0$), and the remaining m observations are cured individuals with $T_{(i)} \geq d$ and $\xi_{[i]}\nu_{[i]} = 1$. Therefore,

$$\widehat{S}_h^c(t \mid x) = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n_1} B_{h[j]}(x) + \sum_{j=n_1+1}^{n} B_{h[j]}(x)} \right\} = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right\} = \widehat{S}_h(t \mid x).$$

   This completes the proof. □

3. When there is no censoring, the estimator $\widehat{S}_h^c(t|x)$ reduces to the kernel type estimator of the conditional survival function.

   Proof. Without censoring, $T_i = Y_i, \delta_i = 1$ and the cure status is always observed $\xi_i = 1$. In this situation, the $n = n_1 + m$ observations can be ordered and split into the n_1 uncured individuals with finite lifetimes Y_i, and the m cured individuals with lifetime $Y_i = \infty$. Thus,

$$\begin{aligned}
\widehat{S}_h^c(t \mid x) &= \prod_{i=1}^{n} \left\{ 1 - \frac{B_{h[i]}(x) \mathbf{1}\left(Y_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)} \right\} \\
&= \prod_{i=1}^{n} \left\{ 1 - \frac{B_{h[i]}(x) \mathbf{1}\left(Y_{(i)} \leq t\right)}{\sum_{j=i}^{n_1} B_{h[j]}(x) + \sum_{j=n_1+1}^{n} B_{h[j]}(x)} \right\} = \prod_{i:Y_{(i)} \leq t} \left\{ \frac{\sum_{j=i+1}^{n} B_{h[j]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right\}.
\end{aligned}$$

   Note that the kernel estimator of the survival function $\tilde{S}_h(t|x) = \sum_{i=1}^{n} B_{h[i]}(x) \mathbf{1}(Y_{(i)} > t)$ is a step function with jumps $B_{hi}(x)$ at the observations, Y_i. By defining $k = \max\{i : Y_{(i)} \leq t\}$ i.e., $Y_{(k)} \leq t$ and $Y_{(k+1)} > t$, one can write

$$\prod_{i:Y_{(i)} \leq t} \left\{ \frac{\sum_{j=i+1}^{n} B_{h[j]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right\} = \prod_{i:Y_{(i)} \leq t} \left\{ \frac{\tilde{S}_h(Y_{(i)} \mid x)}{\tilde{S}_h(Y_{(i-1)} \mid x)} \right\} = \frac{\tilde{S}_h(Y_{(1)} \mid x)}{1} \frac{\tilde{S}_h(Y_{(2)}|x)}{\tilde{S}_h(Y_{(1)} \mid x)} \cdots \frac{\tilde{S}_h(Y_{(k)} \mid x)}{\tilde{S}_h(Y_{(k-1)} \mid x)}$$

$$= \tilde{S}_h(Y_{(k)} \mid x) = \sum_{i=1}^{n} B_{h[i]}(x) \mathbf{1}(Y_{(i)} > t).$$

   This completes the proof. □

4. In an unconditional setting the proposed estimator is

$$\widehat{S}_n^c(t) = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} \mathbf{1}\left(T_{(i)} \leq t\right)}{n - i + 1 + \sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right\}.$$

Proof. In unconditional setting the weights are $1/n$ for $i = 1, \ldots, n$. Thus, the proposed estimator becomes

$$\widehat{S}_n^c(t) = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} \frac{1}{n} \mathbf{1}\left(T_{(i)} \leq t\right)}{\frac{1}{n}(n-i+1) + \frac{1}{n}\sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right\}.$$

In the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, with $n = n_1 + m$ observations, when m are identified as cured, the ordered observed lifetimes are $T_{(1)} \leq \cdots \leq T_{(n_1)}$ strictly lower than d, and the m cured individuals with $T_{(i)} \geq d$. Thus, $\widehat{S}_h^c(t|x)$ reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992):

$$\widehat{S}_n^c(t) = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} \frac{1}{n} \mathbf{1}\left(T_{(i)} \leq t\right)}{\frac{1}{n}(n_1-i+1) + \frac{1}{n}m} \right\} = \prod_{i=1}^{n} \left\{ 1 - \frac{\delta_{[i]} \mathbf{1}(T_{(i)} \leq t)}{n-i+1} \right\}.$$

This completes the proof. □

Proof of Proposition 2. The proof follows the argument in Theorem 2 in López-Cheda, Cao, et al. (2017) and Theorem 1 in Laska and Meisner (1992). To derive the expression of the local likelihood of the mixture cure model, we consider the three potential cases for the ith observation:

1. *Case 1*: $\delta_i = 1$. The event is observed and the individual is not cured. We observe $Y_i = t_i, \nu_i = 0$, with probability:

$$\begin{aligned} P\left(Y_i = t_i, C_i > t_i, \nu_i = 0 \mid X = x\right) &= P\left(C_i > t_i \mid Y_i = t_i, \nu_i = 0, X = x\right) \\ &\quad \times P\left(Y_i = t_i \mid \nu_i = 0, X = x\right) P\left(\nu_i = 0 \mid X = x\right) \\ &= S_{C|Y,X,\nu=0}\left(t_i \mid x\right) \left\{ S_0(t_i^- \mid x) - S_0(t_i \mid x) \right\} p\left(x\right), \end{aligned}$$

   where $S_{C|Y,X,\nu=0}(t \mid x)$ is the conditional survival function for the censoring variable C for uncured individuals.
2. *Case 2*: $(\delta_i = 0, \xi_i\nu_i = 0)$. The individual is censored and the cure status is unknown. We observe $C_i = t_i$, and ν_i is unknown, with probability:

$$\begin{aligned} P\left(Y_i > t_i, C_i = t_i \mid X = x\right) &= P\left(Y_i > t_i, C_i = t_i \mid \nu_i = 1, X = x\right) P(\nu_i = 1 \mid X = x) \\ &\quad + P\left(Y_i > t_i, C_i = t_i \mid \nu_i = 0, X = x\right) P(\nu_i = 0 \mid X = x) \\ &= f_{C|X,\nu=1}\left(t_i \mid x\right)\{1 - p\left(x\right)\} + f_{C|X,\nu=0}\left(t_i \mid x\right) S_0\left(t_i \mid x\right) p\left(x\right), \end{aligned}$$

   where $f_{C|X,\nu=1}(t \mid x)$ and $f_{C|X,\nu=0}(t \mid x)$ are the conditional density functions for the censoring variable C of the cured and uncured individuals, respectively.
3. *Case 3*: $(\delta_i = 0, \xi_i\nu_i = 1)$. The individual is censored and known to be cured. We observe $C_i = t_i, \nu_i = 1$, with probability

$$\begin{aligned} P\left(Y_i > t_i, C_i = t_i, \nu_i = 1 \mid X = x\right) &= P\left(C_i = t_i \mid Y_i > t_i, \nu_i = 1, X = x\right) \\ &\quad \times P\left(Y_i > t_i \mid \nu_i = 1, X = x\right) P\left(\nu_i = 1 \mid X = x\right) \\ &= f_{C|X,\nu=1}\left(t_i \mid x\right)\{1 - p\left(x\right)\}. \end{aligned}$$

In the absence of specification of the distribution of X, the terms in the log-likelihood are weighted with the kernel weights $B_{h[i]}(x)$. Then, the local likelihood of the data is

$$L\left(X, T, \delta, \xi, \nu\right) = \prod_{i=1}^{n} \left[S_{C|Y,X,\nu=0}\left(T_{(i)} \mid x\right) \left\{ S_0(T_{(i)}^- \mid x) - S_0(T_{(i)} \mid x) \right\} p\left(x\right) \right]^{B_{h[i]}(x)\mathbf{1}(\delta_{[i]}=1)}$$

$$\times \left[f_{C|X,\nu=1}\left(T_{(i)} \mid x\right)\{1-p(x)\} + f_{C|X,\nu=0}\left(T_{(i)} \mid x\right) S_0\left(T_{(i)} \mid x\right) p(x) \right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}$$

$$\times \left[f_{C|X,\nu=1}\left(T_{(i)} \mid x\right)\{1-p(x)\} \right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}.$$

If the distribution of the censoring variable C is conditionally independent of Y and the cure status ν given the covariate X, then

$$L(X,T,\delta,\xi,\nu) = \prod_{i=1}^{n} [q_i(x)p(x)]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} \left\{1-p(x)+S_0\left(T_{(i)} \mid x\right)p(x)\right\}^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}$$

$$\times \{1-p(x)\}^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)} \left(1-\sum_{j=1}^{i-1} g_j(x)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} g_i(x)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0\right)} \tag{A7}$$

where, for $i = 1, \ldots, n$, $q_i(x) = S_0(T_{(i)}^{-}|x) - S_0(T_{(i)}|x)$ are the increments of $S_0(t|x)$, and $g_i(x) = G(T_{(i)}|x) - G(T_{(i)}^{-}|x)$ the increments of $G(t|x)$. Let $P_i(x) = p(x)q_i(x)$ be the increments of $S(t|x)$, then $\sum_{i=1}^{n} P_i(x) = p(x)$. Maximizing (A7) is equivalent to maximizing the likelihood

$$L(X,T,\delta,\xi,\nu) = \prod_{i=1}^{n} P_i(x)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} \left(1-\sum_{j=1}^{i-1} P_j(x)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)} \left(1-\sum_{j=1}^{n} P_j(x)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}. \tag{A8}$$

Further, consider the functions $\lambda_i(x) = P_i(x)/\{1-\sum_{j=1}^{i-1} P_j(x)\}$ satisfying

$$1-\sum_{j=1}^{k} P_j(x) = \prod_{j=1}^{k}\{1-\lambda_j(x)\}. \tag{A9}$$

Then, the increments $P_i(x)$ can be written in terms of $\lambda_i(x)$:

$$P_i(x) = \lambda_i(x)\left\{1-\sum_{j=1}^{i-1} P_j(x)\right\} = \lambda_i(x)\prod_{j=1}^{i-1}\left\{1-\lambda_j(x)\right\}. \tag{A10}$$

By substituting (A9) and (A10) in (A8), the likelihood (A8) is

$$L(X,T,\delta,\xi,\nu;p,S_0) = \prod_{i=1}^{n} \lambda_i(x)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} \prod_{i=1}^{n}\left[\prod_{j=1}^{i-1}\left\{1-\lambda_j(x)\right\}\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}$$

$$\times \prod_{i=1}^{n}\left[\prod_{j=1}^{i-1}\left\{1-\lambda_j(x)\right\}\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)} \prod_{i=1}^{n}\left[\prod_{j=1}^{n}\left\{1-\lambda_j(x)\right\}\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}.$$

Taking into account that $\prod_{i=1}^{n}[\prod_{j=1}^{i-1} a_j]^{b_i} = \prod_{i=1}^{n} a_i^{\sum_{j=i+1}^{n} b_j}$, where a_i and b_i, $i = 1, \ldots, n$, are arbitrary sequences of non-negative numbers, the likelihood becomes

$$L(X,T,\delta,\xi,\nu;p,S_0) = \prod_{i=1}^{n} \lambda_i(x)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} \prod_{i=1}^{n}\{1-\lambda_i(x)\}^{\sum_{j=i+1}^{n} B_{h[j]}(x)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=0\right)+\sum_{j=1}^{n} B_{h[j]}(x)\mathbf{1}\left(\delta_{[j]}=0,\xi_{[j]}\nu_{[j]}=1\right)}.$$

Maximizing the likelihood $L(X,T,\delta,\xi,\nu;p,S_0)$ is equivalent to maximizing the local log-likelihood:

$$\Psi\{\lambda_1(x),\ldots,\lambda_n(x)\}=\sum_{i=1}^{n}\left[B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right),\log\lambda_i(x)\right.$$
$$\left.+\left\{\sum_{j=i+1}^{n}B_{h[j]}(x)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=0\right)+\sum_{j=1}^{n}B_{h[j]}(x)\mathbf{1}\left(\delta_{[j]}=0,\xi_{[j]}\nu_{[j]}=1\right)\right\},\log(1-\lambda_i)\right]$$

subject to $\prod_{i=1}^{n}\{1-\lambda_i(x)\}=1-p(x)$. The maximizer $\lambda_i(x)$ of the log-likelihood is

$$\widehat{\lambda}_i(x)=\frac{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}{\sum_{j=i}^{n}B_{h[j]}(x)+\sum_{j=1}^{i-1}B_{h[j]}(x)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=1\right)}.$$

In virtue of (A10), the estimator $\widehat{S}_h^c(t|x)$ computed by forming the product of $\widehat{\lambda}_i$'s such that $T_{(i)}\leq t$ is the nonparametric maximum likelihood estimator of $S(t|x)$. This completes the proof of Proposition 2. □

Proof of Corollary 1. The dominant part of $\widehat{\Lambda}_h^c(t|x)-\Lambda(t|x)$ in Theorem 1 verifies

$$\sum_{i=1}^{n}B_{hi}(x)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)=\int_0^t\frac{d\widehat{H}_h^1(v\mid x)}{J(v^-\mid x)}-\int_0^t\frac{\widehat{J}_h(v^-\mid x)}{J^2(v^-\mid x)}dH^1(v\mid x)$$
$$=\int_0^t\frac{d\widehat{H}_h^1(v\mid x)-dH^1(v\mid x)}{J(v^-\mid x)}-\int_0^t\frac{\widehat{J}_h(v^-\mid x)-J(v^-\mid x)}{J^2(v^-\mid x)}dH^1(v\mid x)$$
$$=\left[\frac{\widehat{H}_h^1(v\mid x)-H^1(v\mid x)}{J(v^-\mid x)}\right]_0^t+\int_0^t\frac{\widehat{H}_h^1(v\mid x)-H^1(v\mid x)}{J^2(v^-\mid x)}dJ(v\mid x)$$
$$-\int_0^t\frac{\widehat{J}_h(v^-\mid x)-J(v^-\mid x)}{J^2(v^-\mid x)}dH^1(v\mid x)$$
$$\leq\frac{1}{\theta}\sup_{a\leq t\leq b,x\in I}|\widehat{H}_h^1(t\mid x)-H^1(t\mid x)|+\frac{1}{\theta}\sup_{a\leq t\leq b,x\in I}|\widehat{H}_h^1(t\mid x)-H^1(t\mid x)|$$
$$-\frac{1}{\theta^2}\sup_{a\leq t\leq b,x\in I}|\widehat{J}_h(t\mid x)-J(t\mid x)|.$$

The last three terms in the inequality are bounded by applying Lemma 5 in Iglesias-Pérez and González-Manteiga (1999), which holds not only for conditional survival functions like $1-H(t|x)$, but also for conditional subdistribution functions as $H^1(t|x)$ and $H^{11}(t|x)$ (see Remark 2 in Iglesias-Pérez and González-Manteiga, 1999, and the proof of Theorem 2.1 in Dabrowska, 1989). As a consequence, the dominant term of $\widehat{\Lambda}_h^c(t|x)-\Lambda(t|x)$ is bounded by

$$\sup_{a\leq t\leq b,x\in I}\left|\sum_{i=1}^{n}B_{hi}(x)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right|=O\left\{(nh)^{-1/2}(\log n)^{1/2}\right\}.$$

Using the results of Theorem 2 it is straightforward to prove the second part of this corollary. □

Proof of Proposition 3. From Theorem 2, the bias of the nonparametric estimator $1-\widehat{F}_h^c(t|x)$ is asymptotically equal to the expected value

$$\frac{(nh)^{-1}\{1-F(t\mid x)\}}{m(x)}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)=I+II,\tag{A11}$$

where

$$I = \frac{(nh)^{-1}\{1-F(t\mid x)\}}{m(x)}\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x) - \mathrm{E}\left\{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right\}\right], \quad \text{(A12)}$$

$$II = \frac{(nh)^{-1}\{1-F(t\mid x)\}}{m(x)}\mathrm{E}\left\{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right\}. \quad \text{(A13)}$$

As $E(I)=0$, the asymptotic bias of the estimator $1-\widehat{F}_h^c(t\mid x)$ is II. Using Lemmas 1 and 2 in the Supporting Information,

$$II = \frac{h^2\{1-F(t\mid x)\}(\Phi_c''(x,t,x)\,m(x)+2\Phi_c'(x,t,x)\,m'(x))d_K}{2m(x)} + O\left(h^4\right),$$

with $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ the first and second derivatives of $\Phi_c(y,t,x)$ with respect to y. Recalling (A11), the asymptotic variance of $1-\widehat{F}_h^c(t|x)$ is

$$\mathrm{Var}(I) = \frac{\{1-F(t\mid x)\}^2}{m^2(x)}(V_1-V_2), \quad \text{(A14)}$$

where

$$V_1 = \frac{1}{nh^2}\mathrm{E}\left\{K^2\left(\frac{x-X}{h}\right)\zeta^2(T,\delta,\xi,\nu,t,x)\right\},\quad V_2 = \frac{1}{nh^2}\left[\mathrm{E}\left\{K\left(\frac{x-X}{h}\right)\zeta(T,\delta,\xi,\nu,t,x)\right\}\right]^2.$$

From Lemmas 1 and 2 in the Supporting Information, V_2 reduces to

$$V_2 = \frac{1}{4}\frac{h^2}{n}d_K^2\left\{\frac{\Phi_c''(x,t,x)\,m(x)+2\Phi_c'(x,t,x)\,m'(x)}{m(x)}\right\}^2 + O\left(\frac{h^4}{n}\right). \quad \text{(A15)}$$

As for V_1, let us define $\Phi_1^c(y,t,x)=E(\zeta^2(T,\delta,\xi,\nu,t,x)|X=y)$. Then, after a change of variable and a Taylor's expansion (as in the proof of Lemma 1 in the Supporting Information) we obtain

$$V_1 = \frac{1}{nh}\Phi_1^c(x,t,x)\,m(x)\,c_K + \frac{1}{2}\frac{h}{n}e_K\frac{d^2}{dy^2}\left\{\Phi_1^c(y,t,x)\,m(y)\right\}|_{y=x} + O\left(n^{-1}h^3\right), \quad \text{(A16)}$$

where $e_K=\int v^2K^2(v)dv$. The proof concludes by substituting (A15) and (A16) into (A14). □

Proof of Theorem 3. From Theorem 2, we consider

$$(nh)^{1/2}\left\{\widehat{F}_h^c(t\mid x)-F(t\mid x)\right\} = (nh)^{1/2}\{1-F(t\mid x)\}\sum_{i=1}^{n}\tilde{B}_{hi}(x)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x) + (nh)^{1/2}R_{n2}(t,x)$$

with $\zeta(T,\delta,\xi,\nu,t,x)$ and $R_{n2}(t,x)$ given in (5) and (7), respectively. The condition $(\log n)^3/(nh)\to 0$ implies that $(nh)^{1/2}(\log n/(nh))^{3/4}\to 0$, so the remainder term $(nh)^{1/2}R_{n2}(t,x)$ is negligible. Consequently, the asymptotic distribution of $(nh)^{1/2}\{\widehat{F}_h^c(t\mid x)-F(t\mid x)\}$ is that of

$$(nh)^{1/2}\frac{1-F(t\mid x)}{m(x)}\sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{x-X_i}{h}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x) = (nh)^{1/2}(I+II), \quad \text{(A17)}$$

where I and II are given in (A12) and (A13). Under the assumption $nh^5 \to 0$, we have $(nh)^{1/2} II = o(1)$. Therefore, the asymptotic distribution of (A17) is that of $(nh)^{1/2} I$. Let us define $(nh)^{1/2} I = \sum_{i=1}^{n} \eta_{i,h}(t,x)$, where

$$\eta_{i,h}(t,x) = \frac{(nh)^{-1/2}\{1 - F(t \mid x)\}}{m(x)} \left[K\left(\frac{x - X_i}{h}\right) \zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) - \mathrm{E}\left\{ K\left(\frac{x - X_i}{h}\right) \zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) \right\} \right]$$

is a sequence of n independent random variables with mean 0. Note that

$$\mathrm{Var}(\eta_{i,h}(t,x)) = h\mathrm{Var}(I) = \frac{1}{n} \frac{\{1 - F(t \mid x)\}^2}{m(x)} \Phi_1^c(x,t,x) c_K + O\left(\frac{h^2}{n}\right) = \frac{1}{n} s_c^2(t,x) + O\left(\frac{h^2}{n}\right)$$

with $\mathrm{Var}(I)$ in (A14) and $s_c^2(t,x)$ in (13). As $\mathrm{Var}(\eta_{i,h}(t,x)) < \infty$ for $i = 1, \ldots, n$ and $\mathrm{Var}(\eta_h(t,x)) = \sum_{i=1}^{n} \mathrm{Var}(\eta_{i,h}(t,x))$ is positive, then we can apply Lindeberg's theorem (Billingsley, 1968) to obtain

$$\frac{\sum_{i=i}^{n} \eta_{i,h}(t,x)}{s_c^2(t,x)} \to N(0,1) \quad \text{in distribution.}$$

Therefore, $(nh)^{1/2}\{\widehat{F}_h^c(t|x) - F(t|x)\} \to N(0, s_c^2(t,x))$ in distribution. This proves (i). In parallel to the proof (i) we can prove (ii) as follows, note that if $nh^5 = C^5$ then the bias term is $(nh)^{1/2} II = (nh)^{1/2}\{h^2 B_c(t,x) + O(h^4)\} = (nh^5)^{1/2} B_c(t,x) + O\{(nh^9)^{1/2}\}$ with $B_c(t,x)$ in (9). Thus, $(nh)^{1/2}\{\widehat{F}_h^c(t|x) - F(t|x)\} \to N(C^{5/2} B_c(t,x), s_c^2(t,x))$ in distribution. This completes the proof. □