Nonparametric kernel estimation of the cure probability in a mixture cure model when cure status is partially known

Wende Clarence Safari* | Ignacio López-de-Ullibarri | M. Amalia Jácome

Department of Mathematics, Faculty of Computer Science, University of A Coruña, CITIC, A Coruña, Spain
Department of Mathematics, Escuela Universitaria Politécnica, University of A Coruña, Ferrol, Spain
Department of Mathematics, Faculty of Science, University of A Coruña, CITIC, A Coruña, Spain

Summary

We propose a nonparametric estimator of the cure probability in the mixture cure model under right random censoring when the cure status is available for some censored individuals. The estimator extends an estimator previously proposed in the literature for the general case when the cure status is unavailable due to censoring. The new estimator is shown to be strongly consistent and asymptotically normally distributed. Moreover, we demonstrate that the asymptotic variance of the proposed estimator is smaller or equal than the estimator for the general case when the cure status is unavailable. A bootstrap procedure for bandwidth selection is proposed. Moreover, as an alternative, we study a multiply imputed Nadaraya-Watson estimator of the cure probability when the cure status is partially observed. We perform simulations to evaluate the finite sample performance of both estimators, and apply them to the analysis of two datasets related to survival of breast cancer patients and length of hospital stay of COVID-19 patients requiring intensive care.

KEYWORDS: bootstrap bandwidth, censoring, cure models, kernel estimators, local weights.

1 | INTRODUCTION AND MOTIVATING EXAMPLES

A common assumption in standard survival modeling of time-to-event data is that all individuals will experience the event of interest if observed for enough time. However, such assumption may be not true in practice. For example, due to advances in oncological therapy, cancer may never recur in some patients. These patients are called long-term survivors, or simply cured from the cancer recurrence event. Using standard survival analysis for such data is not suitable because it does not allow to estimate the probability of cure. Cure models are appropriate in this setting as they explicitly allow for the cure fraction to be estimated. The mixture cure model assumes that the population is divided into two groups: cured and uncured. Substantial work has been done on the mixture cure model, mostly with a (semi)parametric approach.

In standard cure models, a subject that experiences the event is known to be uncured. However, the challenge of identifying the cure status of a censored individual is inescapable because it is unknown if the event will happen eventually. As a consequence, the cure status is only observed for the uncensored individuals (who are uncured), and it is usually considered a latent variable. Nonetheless, there are situations where some of the censored individuals can be identified to be unsusceptible to the event of interest, that is, to be cured. For example, in some types of cancer patients can be assumed to be cured if they are free of disease and their survival time is larger than a threshold (e.g., 5 years). Similarly, diagnostic tests can provide evidence on whether a cancer patient is cured or not. Another example can be found in the analysis of length of stay until certain hospital outcomes,
such as the time in a hospital ward until requiring admission to an intensive care unit (ICU). In this case, all patients who die or are discharged from the hospital ward without entering ICU are censored and are known to be cured from ICU admission.

Few authors have studied cure models when the cure status is known for some censored observations. Laska and Meisner\(^5\) and Betensky and Schoenfeld\(^6\) discussed nonparametric cure rate estimation with cure status available, but neither of them considered the presence of covariates. Nieto-Baraja and Yin\(^7\) 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.\(^3\). Bernhardt\(^9\) proposed a flexible semiparametric cure rate model with potentially known cure threshold and showed that ignoring a known cure threshold may lead to biased estimates. Recently, we proposed a generalized product limit estimator of the conditional survival function in the mixture cure model when the cure status is partially known.\(^10\)

In the absence of cure status information for the censored observations, a nonparametric estimator of the cure probability was proposed by Xu and Peng\(^11\), and studied also by López-Cheda et al.\(^12\). In this paper, we extend that estimator to the case when some censored observations are known to be cured, so the cure status is partially known for the censored individuals. Before introducing the proposed estimator, we describe two studies from the medical domain to motivate our estimator.

The first example is a breast cancer study that was conducted using clinicopathological data from The Cancer Genome Atlas (TCGA) program. The data include \(n = 898\) women diagnosed with breast cancer between 1988 and 2013. The event of interest was death from breast cancer. The observed time was considered as censored if the patient was alive at the end of the study, she was cancer-free (either alive or dead), or she was lost to follow-up. Information on demographic and clinical characteristics was collected at baseline. In this study, patients who have been cancer-free for at least 10 years were defined to be long-term survivors\(^13,14\) (in our context, also referred to as cured).

The second example uses data collected by the Galician Healthcare Service\(^15\) on patients hospitalized with COVID-19 in Galicia (North-West of Spain) during the first weeks of the outbreak (between March 6 and May 7, 2020). The event of interest was admission to ICU from hospital ward. This database includes demographic characteristics such as sex and age, and the dates of different medical outcomes like admission to ICU, discharge or death. A total of \(n = 2453\) patients were admitted to the hospital bed and 197 of them needed ICU care. Event times of patients who died or were discharged from the hospital before entering ICU were considered as censored and cured from the admission to the ICU. We provide more details on these examples in Section 6.

An informal way to identify the possible presence of cured individuals is to look at a plateau on the right tail of the survival curve.\(^16\) In Figure 1, Kaplan-Meier estimates of the survival curves of the breast cancer and COVID-19 examples are presented. We observe that the survival curves level off to a non-zero probability at the end of the follow-up time, indicating the existence of individuals cured from the event of interest. For large times, the curve for older patients (≥ 55 years) lies above that for younger ones (< 55 years), suggesting that the probability of cure may increase with age. Also the cross marks on the plot represent the presence of patients known to be cured in the data. We can see that the times of known cured patients in the breast cancer data are distributed at larger times while in COVID-19 data they are distributed randomly.

The rest of the paper is organized as follows. In Section 2, the new estimator of the cure probability is defined and its asymptotic properties are studied. For bandwidth selection we propose a bootstrap procedure in Section 3. In Section 4 we study the multiply imputed Nadaraya-Watson estimator\(^17\) as an alternative regression-based method to estimate the cure probability. In Section 5, a simulation study is performed to compare our estimator to (a) the corresponding nonparametric cure probability estimator ignoring the available information about the cure status,\(^11,12\) (b) the semiparametric estimator proposed by Bernhardt\(^9\), and (c) the alternative estimator of Section 4. In Section 6, the different cure probability estimators are applied to the breast cancer and COVID-19 datasets. We conclude in Section 7 with 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, \(X\) a vector of covariates and \(F(t \mid x)\) the distribution function of \(Y\) conditional on \(X = x\). Let \(C\) be a random censoring time with conditional distribution function \(G(t \mid x)\). So, instead of observing \(Y\), only \(T = \min(Y, C)\) and \(\delta = 1(Y \leq C)\) can be observed. The random variables \(Y\) and \(C\) are assumed to be conditionally independent given \(X = x\). We set \(Y = \infty\) if the subject is cured. Let \(\nu = 1(Y = \infty)\) be the indicator of being cured. Note that \(\nu\) is partially observed because \(\delta = 1\) implies \(\nu = 0\), but in the general situation \(\nu\) is unknown when \(\delta = 0\). When the cure status is partially known, \(\nu = 1\) is
FIGURE 1 Kaplan-Meier survival curves of breast cancer patients (left) and COVID-19 patients (right), both stratified by age groups: patients aged younger than 55 years (black line) and 55 years or older (gray line). Crosses correspond to patients known to be cured from the event of interest.

also observed for some censored individuals. Suppose that \( \xi \) indicates whether the cure status is known (\( \xi = 1 \)) or not (\( \xi = 0 \)). Hence, the observations \( \{(X_i, T_i, \delta_i, \xi_i, v_i) : i = 1, \ldots, n\} \) can be classified into three groups: (a) the individual is observed to have experienced the event and therefore known to be uncured \( (X_i, T_i = Y_i, \delta_i = 1, \xi_i = 1, v_i = 0) \); (b) the lifetime is censored and the cure status is unknown \( (X_i, T_i = C_i, \delta_i = 0, \xi_i = 0, v_i = 0) \); and (c) the lifetime is censored and the individual is known to be cured \( (X_i, T_i = C_i, \delta_i = 0, \xi_i = 1, v_i = 1) \). In standard cure models when the cure status is unknown for the censored observations, only groups (a) and (b) are considered.

The probability of cure is \( 1 - p(x) = P(Y = \infty | X = x) \), and the conditional survival function of the uncured individuals, also known as latency, is \( S_0(t | x) = P(Y > t | Y < \infty, X = x) \). The mixture cure model specifies the survival function \( S(t | x) = P(Y > t | X = x) \) as

\[
S(t | x) = 1 - p(x) + p(x) S_0(t | x). \tag{1}
\]

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. Following the argumentation of Hanin and Huang, 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 \( \lim_{t \to \infty} S_0(t | x) = 0 \) for all \( x \). This condition is similar to the zero-tail constraint in López-Cheda et al., Taylor and other papers.

2.2 Proposed estimator

Without loss of generality, let us consider a continuous covariate \( X \) with density function \( m(x) \). Assuming model (1), the cure probability can be written in terms of the survival function as follows:

\[
1 - p(x) = \lim_{t \to \infty} S(t | x). \tag{2}
\]
To introduce a nonparametric estimator of the cure probability when the cure status is partially known we consider the relationship (2) and the generalized product limit estimator\(^{10}\) of the conditional survival function \(S(t \mid x)\),

\[
\hat{S}_h^c(t \mid x) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_{[i]} B_{h_{[i]}}(x) 1(T_{(i)} \leq t)}{\sum_{j=1}^{n} B_{h_{[j]}}(x) + \sum_{j=1}^{i-1} B_{h_{[j]}}(x) 1(\xi_{[j]} v_{[j]} = 1)} \right),
\]

where \(X_{[i]}, \delta_{[i]}, \xi_{[i]}\) and \(v_{[i]}\) are the concomitants of the ordered observed times \(T_{(1)} \leq \ldots \leq T_{(n)}\), \(B_{h_{[i]}}(x)\) are the Nadaraya-Watson weights,

\[
B_{h_{[i]}}(x) = \frac{K_h(x - X_{[i]})}{\sum_{j=1}^{n} K_h(x - X_{[j]})},
\]

and \(K_h(\cdot) = K(\cdot/h)/h\) is a kernel function \(K(\cdot)\) rescaled with bandwidth \(h \to 0\) as \(n \to \infty\). The proposed estimator of the cure probability \(1 - p(x)\) is

\[
1 - \hat{p}_h^c(x) = \hat{S}_h^c(T_{(n)}^1 \mid x) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_{[i]} B_{h_{[i]}}(x)}{\sum_{j=1}^{n} B_{h_{[j]}}(x) + \sum_{j=1}^{i-1} B_{h_{[j]}}(x) 1(\xi_{[j]} v_{[j]} = 1)} \right),
\]

(3)

where \(T_{(n)}^1 = \max T_{(i)}\) is the largest uncensored observed lifetime. An important feature of this estimator is that subjects who are known to be cured before time \(T_{(i)}\) remain in the risk set, i.e., their weights are included in the denominator.

**Proposition 1.** The proposed estimator \(1 - \hat{p}_h^c(x)\) has the following general properties.

1. When there are no censored observations known to be cured, i.e., \(\xi_{[i]} v_{[i]} = 0\) for \(i = 1, \ldots, n\), \(1 - \hat{p}_h^c(x)\) reduces to the nonparametric cure probability estimator in Xu and Peng\(^{11}\):

\[
1 - \hat{p}_h(x) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_{[i]} B_{h_{[i]}}(x)}{b_{[i]} B_{h_{[i]}}(x)} \right).
\]

2. In the specific case when some individuals are observed as cured when their survival time exceeds a known fixed cure threshold, \(1 - \hat{p}_h^c(x)\) also reduces to the nonparametric cure probability estimator of Xu and Peng\(^{11}\).

3. When there is no censoring, all the cure status \(v_i\) are observed \((\xi_{[i]} = 1, i = 1, \ldots, n)\). In that case, \(1 - \hat{p}_h^c(x)\) reduces to the Nadaraya-Watson (NW)\(^{21,22}\) estimator of the cure probability:

\[
1 - \hat{p}_h^{NW}(x) = \sum_{i=1}^{n} B_{h_{[i]}}(x) 1(v_{[i]} = 1) = \frac{\sum_{i=1}^{n} K_h(x - X_{[i]}) v_{[i]}}{\sum_{j=1}^{n} K_h(x - X_{[j]})}.
\]

(4)

It must be kept in mind that when there is no censoring the nonparametric cure probability estimator by Xu and Peng\(^{11}\) will be zero.

4. In an unconditional setting, the proposed estimator is

\[
1 - \hat{p}_n^c = \prod_{i=1}^{n} \left( 1 - \frac{\delta_{[i]}}{n - i + 1 + 1(\sum_{j=1}^{i-1} \xi_{[j]} v_{[j]} = 1)} \right).
\]

(5)

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 \(d\), \(1 - \hat{p}_n^c\) reduces to the generalized maximum likelihood estimator in Laska and Meisner\(^{5}\).

The proof of these properties is outlined in the Supplementary Material.

**Proposition 2.** The estimator \(1 - \hat{p}_h^c(x)\) in (3) is the nonparametric local maximum likelihood estimator of \(1 - p(x)\).

The proof of Proposition 2 follows the proof of Proposition 2 in Safari et al.\(^{10}\) and it is thus omitted.
2.3 Asymptotic results

In this section, we study the asymptotic properties of the proposed estimator, \( 1 - \hat{P}_h(x) \). We begin by defining the following (sub)distribution functions:

\[
H(t \mid x) = P(T \leq t \mid X = x),
\]
\[
H^1(t \mid x) = P(T \leq t, \delta = 1 \mid X = x),
\]
\[
H^{11}(t \mid x) = P(T \leq t, \xi = 1, \nu = 1 \mid X = x),
\]
\[
J(t \mid x) = 1 - H(t \mid x) + H^{11}(t \mid x).
\]

The following assumptions will be made.

**Assumption 1.** \( X, Y \) and \( C \) are absolutely continuous random variables.

**Assumption 2.**

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

\[
0 < \gamma = \inf_{x \in I_x} m(x) < \sup_{x \in I_x} m(x) = \Gamma < \infty
\]

for some \( I_x = [x_1 - \epsilon, x_2 + \epsilon] \) with \( \epsilon > 0 \) and \( 0 < \epsilon \Gamma < 1 \). 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 \( J(t \mid x) \geq \theta > 0 \) for \( (t, x) \in [a, b] \times I_x \).

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

**Assumption 4.** The second derivative with respect to \( x \) of function \( m(x) \) exists and is continuous in \( x \in I_x \), and the second derivatives with respect to \( x \) of functions \( H(t \mid x) \), \( H^1(t \mid x) \) and \( H^{11}(t \mid x) \) exist and are continuous and bounded in \( (t, x) \in [0, \infty) \times I_x \).

**Assumption 5.** The first derivatives with respect to \( t \) of \( H(t \mid x) \), \( H^1(t \mid x) \) and \( H^{11}(t \mid x) \) exist and are continuous in \( (t, x) \in [a, b] \times I_x \).

**Assumption 6.** The second derivatives with respect to \( t \) of \( H(t \mid x) \), \( H^1(t \mid x) \) and \( H^{11}(t \mid x) \) exist and are continuous in \( (t, x) \in [a, b] \times I_x \).

**Assumption 7.** The first derivative with respect to \( x \) and the second derivative with respect to \( t \) of \( H(t \mid x) \), \( H^1(t \mid x) \) and \( H^{11}(t \mid x) \) exist and are continuous in \( (t, x) \in [a, b] \times I_x \).

**Assumption 8.** The (sub)densities corresponding to the (sub)distribution functions \( H(t \mid x) \), \( H^1(t \mid x) \) and \( H^{11}(t \mid x) \) are bounded away from 0 in \( [a, b] \times I_x \).

**Assumption 9.** The kernel function \( K(v) \) is a symmetrical density with zero mean, vanishing outside \((-1, 1)\), and the total variation is less than \( \lambda < \infty \).

We also assume that the censoring distribution function \( G(t \mid x) \) is a proper survival function. Further we define \( \tau_H(x) = \inf \{ t : H(t \mid x) = 1 \} \), \( \tau_0(x) = \inf \{ t : S_0(t \mid x) = 0 \} \) and \( \tau_G(x) = \inf \{ t : G(t \mid x) = 1 \} \). Note that \( \tau_H(x) = \tau_G(x) \) since \( T = \min(Y, C) \) and \( S(t \mid x) > 0 \) for any \( t \). Let \( \tau_0 = \sup_{x \in I_x} \tau_0(x) \), then we require:

\[
\tau_0 < \tau_G(x) \quad \text{for any } x \text{ with probability 1.} \tag{6}
\]

The condition (6) relies on the assumption that all censored observations after the largest uncensored observed lifetime correspond to cured subjects, such that the susceptible subjects will experience the event within the follow-up period. Besides, this condition guarantees that the proposed estimator does not overestimate the true cure probability. To verify this condition in practice, one can test a null hypothesis \( H_0 \) by looking at the difference between the largest observed time \( T_n \) and \( T_1 \). If this interval is large, then, there is sufficient followed-up time in which the risk of developing the event of interest is negligible. A test statistic for this hypothesis without covariates was developed by Maller and Zhou. In case of covariates, we may divide a given dataset into several subgroups according to the values of \( x \) and apply this test in each subgroup. A similar condition was used by Laska and Meisner, Xu and Peng, López-Cheda et al., among others. Xu and Peng pointed out that, if \( G(t \mid x) \)
has a heavier tail than $S_0(t \mid x)$, the assumption of finiteness of $r_0(x)$ can be relaxed because the proposed cure probability estimator will tend to have smaller bias regardless of the value of $r_0(x)$. At this point it is important to note that, if the maximum follow-up time is uncensored observed lifetime, there is no need of applying condition (6).

The next theorem establishes an asymptotic representation for $1 - \hat{P}_h(x)$. Based on that result, we prove the strong consistency and asymptotic normality of $1 - \hat{P}_h(x)$. Furthermore, in Section 2.4 we provide evidence that $1 - \hat{P}_h(x)$ has smaller asymptotic variance than the nonparametric cure probability estimator $1 - \hat{P}_h(x)$ of Xu and Peng\textsuperscript{11}.

**Theorem 1.** Suppose that assumptions 1–9 and condition (6) hold, and the bandwidth $h = (h_n)$ satisfies $h \to 0$, $\log n/nh \to 0$ and $nh^2 / \log n = O(1)$ as $n \to \infty$. Then, for $x \in I$ we have

$$(1 - \hat{P}_h(x)) - (1 - p(x)) = (1 - p(x)) \sum_{i=1}^{n} \tilde{B}_{hi}(x) \zeta \left( T_i, \delta_i, \xi_i, \nu_i, r_0, x \right) + R_n(x)$$

where

$$\zeta \left( T_i, \delta_i, \xi_i, \nu_i, t, x \right) = \frac{1}{J(T_i - x)} - \int_0^{1} \left( 1 \left( T_i \geq v \right) + 1 \left( T_i < v, \xi_i = 1 \right) \right) \frac{dH_1(v \mid x)}{J^2(v - x)},$$

and $R_n(x)$ satisfies

$$\sup_{x \in I} | R_n(x) | = O \left( (\log n)^{3/4} (nh)^{-3/4} \right) \text{ a.s.}$$

The proof of Theorem 1 is outlined in the Supplementary Materials.

The following corollary on the strong consistency of the estimator $1 - \hat{P}_h(x)$ is deduced from Theorem 1.

**Corollary 1.** Suppose that assumptions 1–9 and condition (6) hold, and the bandwidth $h = (h_n)$ satisfies $h \to 0$, $\log n/nh \to 0$ and $nh^2 / \log n = O(1)$ as $n \to \infty$. Then, for $x \in I$, we have

$$\sup_{x \in I} | \hat{P}_n(x) - p(x) | = O \left( (nh)^{-1/2} (\log n)^{1/2} \right) \text{ a.s.}$$

The corollary is proved by considering the asymptotic representation in Theorem 1 and following similar arguments as those used in the proof of Corollary 1 in Safari et al.\textsuperscript{10} for $t = T_{1n}^\top$.

The next proposition establishes the asymptotic bias and variance expression of the estimator $1 - \hat{P}_h(x)$.

**Proposition 3.** Suppose that assumptions 1–9 and condition (6) hold, and the bandwidth $h = (h_n)$ satisfies $h \to 0$, $\log n/nh \to 0$ and $nh^2 / \log n = O(1)$ as $n \to \infty$. Then, the asymptotic bias and variance of $1 - \hat{P}_h(x)$ are, respectively,

$$\mu_{h,c}(x) = h^2 B_c(x) + O \left( h^4 \right) \quad \text{and} \quad \sigma_{h,c}^2(x) = \frac{1}{nh} s^2_c(x) + O \left( \frac{h}{n} \right),$$

where the dominant term of the bias is $B_c(x) = (c_{1,c}(x) + c_{2,c}(x)) \, d_K$,

$$c_{1,c}(x) = \frac{2(1 - p(x))' m'(x) + (1 - p(x))'' m(x)}{2m(x)}$$

and

$$c_{2,c}(x) = (1 - p(x)) \int_0^{r_0} \left[ \frac{G(v \mid x) - \pi_1(v, x)(1 - p(x))G(v \mid x)}{1 - G(v \mid x) + \pi_1(v, x)(1 - p(x))G(v \mid x)} ds \left( S'(s \mid x) \right) \right]_{s=x} dv,$$

where $\pi_1(t, x) = P(\xi = 1 \mid v = 1, C \leq t, X = x), \quad G_1(t \mid x) = P(C \leq t \mid v = 1, X = x).$

Here $p'(x), p''(x), S'(t \mid x)$ and $G'(t \mid x)$ refer to the derivatives with respect to $x$. The dominant term of the variance is

$$s^2_c(x) = \frac{(1 - p(x))^2}{m(x)} \int_0^{r_0} \frac{dH_1(v \mid x)}{(1 - H(v \mid x) + H_1^1(v \mid x))^2} c_K,$$

with

$$d_K = \int v^2 K(v) dv \quad \text{and} \quad c_K = \int K^2(v) dv.$$

The proof of Proposition 3 is outlined in the Supplementary Material. The following theorem, whose proof is in the Supplementary Material, establishes the asymptotic normality of $1 - \hat{P}_h(x)$. 


**Theorem 2.** Suppose that assumptions 1–9 and condition (6) are satisfied, then for \( x \in I \) it follows that:

- (i) If \( nh^3 \to 0 \) and \( (log n)^3/nh \to 0 \), then
  \[
  (nh)^{1/2} \left( \frac{\hat{p}_h'(x) - p(x)}{\hat{p}_h''(x)} \right) \to N(0, s^2_c(x))
  \]
  in distribution.

- (ii) If \( C \) is a constant, then
  \[
  (nh)^{1/2} \left( \frac{\hat{p}_h(x) - p(x)}{\hat{p}_h''(x)} \right) \to N(C^{5/2} B_c(x), s^2_c(x))
  \]
  in distribution, where \( B_c(x) \) is defined in (9) and (10), \( s^2_c(x) \) in (12).

### 2.4 Effect of ignoring the cure status

The use of the information given by the cure status has an impact on both the bias and variance of the proposed cure probability estimator \( 1 - \hat{p}_h(x) \). When the cure status is ignored in the estimation procedure, the asymptotic expressions of the bias and variance of the nonparametric cure probability estimator \( 1 - \hat{p}_h(x) \) are\(^{10}\):

\[
\mu_h(x) = h^2 B(x) + O(h^4) \quad \text{and} \quad \sigma^2_h(x) = \frac{1}{nh} s^2(x) + O \left( \frac{h}{n} \right),
\]

where \( B(x) = (c_1(x) + c_2(x)) d_K \), with \( c_1(x) \) in (9), and

\[
c_2(x) = (1 - p(x)) \int_0^{\tau_0} \frac{G'(v^- | x)}{1 - G(v^- | x)} \frac{d}{ds} \left( \frac{S'(s | x)}{S(s | x)} \right) \bigg|_{s=v^-} dv.
\]

The dominant term of the variance of \( 1 - \hat{p}_h(x) \) is

\[
s^2(x) = \frac{(1 - p(x))^2}{m(x)} \int_0^{\tau_0} \frac{dH^1(v^- | x)}{(1 - H(v^- | x))^2} c_K.
\]

When the cure status information is ignored, then \( H^{11}(t | x) = 0 \), and therefore \( s^2(x) \leq s^2(x) \) for all \( x \). As a consequence, when the known cure status is used for estimating the cure probability, the variance of the proposed estimator decreases asymptotically with respect to that of the estimator by Xu and Peng\(^ {11}\).

The cure status information affects the bias of the proposed estimator only in the second term \( c_{2x}(x) \) in (10) and, specifically, through the derivative with respect to \( x \) of \( \pi_1(t, x)(1 - p(x))G_1(t | x) \). Therefore, in terms of bias, the gain of knowing the cure status is not straightforward, since it depends on the conditional probability of observing cures \( \pi_1(t, x) \), the conditional censoring distribution of the cures \( G_1(t | x) \) in (11), and the probability of cure \( 1 - p(x) \) itself.

### 3 Bandwidth Selection

Different approaches for bandwidth selection of kernel estimators are available in the literature, cross-validation (CV), plug-in and bootstrap methods being the most common. In the context of mixture cure models, the finite-sample behavior of the CV bandwidth selector is unsatisfactory, as it is highly variable and tends to undersmooth.

The plug-in bandwidth selector is the estimator of the asymptotically optimal local bandwidth, in the sense of minimizing the asymptotic mean squared error (MSE) of \( 1 - \hat{p}_h(x) \), that is

\[
h_\alpha = \left( \frac{s^2_c(x)}{4B_c(x)} \right)^{1/5} n^{-1/5}, \tag{13}
\]

where \( B_c(x) \) and \( s^2_c(x) \) are given in Proposition 3. The asymptotically optimal bandwidth \( h_\alpha \) cannot be easily estimated in practice, due to the implicit dependency on unknown hard-to-estimate quantities, tending to an endless process which seems to be more complicated than the problem of the cure probability estimation.

We follow López-Cheda at al.\(^ {10,12}\) ideas and propose a bandwidth selector based on the bootstrap. The principle is to select the bandwidth, \( h_\alpha^* \), which minimizes the bootstrap version of the MSE, \( MSE_\alpha^*(h) \), approximated by Monte Carlo as:

\[
MSE_\alpha^*(h) \approx \frac{1}{B} \sum_{b=1}^{B} (\hat{p}_h(x) - \hat{p}_h^*(x))^2, \tag{14}
\]
where \(1 - \hat{P}_h^{c,b}(x)\) is the proposed estimator computed with the \(b\)th bootstrap resample and a bandwidth \(h\). In addition, \(1 - \hat{P}_{g_x}(x)\) is the cure probability estimate computed with the original sample and a given pilot bandwidth \(g_x\). 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 \(1 - \hat{P}_{g_x}(x)\).

**Step 2:** Choose a dense enough grid of \(L\) bandwidths \(\{h_1, \ldots, h_L\}\).

**Step 3:** Generate \(B\) bootstrap resamples \(\{(X_i^{(b)}, T_i^{(b)}, \delta_i^{(b)}, \xi_i^{(b)}, \gamma_i^{(b)}, v_i^{(b)}): i = 1, \ldots, n\}\), for \(b = 1, \ldots, B\).

**Step 4:** With the \(b\)th bootstrap resample and the bandwidth \(h_l\), compute \(1 - \hat{P}_{h_l}^{c,b}(x)\), for \(l = 1, \ldots, L\).

**Step 5:** For \(h_l, l = 1, \ldots, L\), compute the Monte Carlo approximation of \(\text{MSE}_x(h_l)\) given by (14).

**Step 6:** The bootstrap bandwidth, \(h_x^\ast\), is the bandwidth of the grid \(\{h_1, \ldots, h_L\}\) that minimizes the approximation of \(\text{MSE}_x(h)\) in (14).

The bootstrap resamples in Step 3 are generated as follows. Fix \(x\), for \(i = 1, \ldots, n\), set \(X_i^{(b)} = X_i\) and generate a 4-tuple \((T_i^{(b)}, \delta_i^{(b)}, \xi_i^{(b)}, \gamma_i^{(b)}, v_i^{(b)})\) from the weighted empirical distribution function

\[
\hat{F}_{g_x}(t, d, u, v \mid x) = \sum_{i=1}^n B_{g_x,i}(x) \mathbb{1}(T_i \leq t, \delta_i \leq d, \xi_i \leq u, \gamma_i \leq v),
\]

where \(B_{g_x,i}(x)\) are the NW weights with pilot bandwidth \(g_x\). López-Cheda et al.\(^{12,25}\) and Safari et al.\(^{10}\), following suggestions by Li and Datta\(^{26}\), proposed using a local pilot bandwidth \(g_x\), of the covariate \(X\):

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

where \(d^+_k(x)\) and \(d^-_k(x)\) are the distances from \(x\) to the \(k\)th 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 et al.\(^{12,25}\), we suggest setting \(k\) to be integer part of \(n/4\). Moreover, the simulations in López-Cheda et al.\(^{12}\) demonstrated that the choice of pilot bandwidth has small effect on the final bootstrap bandwidth.

### 4 ALTERNATIVE ESTIMATOR

The proposed estimator of the cure probability is based on the relation in (2) between the cure probability \(1 - p(x)\) and the survival function \(S(t \mid x)\). Nonetheless, the cure probability can also be written as \(1 - p(x) = E(\psi \mid X = x)\), i.e., as the conditional expectation of the cure status \(\psi\). Equivalently, in an unconditional setting the cure probability is \(1 - p = E(\psi)\).

The unconditional cure probability \(1 - p\) may easily be estimated using an empirical estimator \(1 - \hat{p} = \sum_{i=1}^n \psi_i / n\). For the estimation of the cure probability \(1 - p(x)\), the NW estimator\(^{21,22}\) in (4) is one of the most frequently used estimators in nonparametric regression. However, these methods estimate the cure probability assuming that the cure status \(\psi\) is completely observed. There has been extensive work dealing with estimating the unconditional and conditional mean when the response variable is only partially observed\(^{27-29}\). Aerts et al.\(^{17}\) developed a fully nonparametric local multiple imputation procedure to estimate the unconditional mean of a variable in the presence of missing response data. We apply their methodology for the estimation of \(1 - p\) and extend their idea to propose an alternative estimator for the cure probability \(1 - p(x)\) when the cure status is partially observed.

### 4.1 Multiply imputed Nadaraya-Watson estimator

Consider the \(\xi\) indicator, with \(\xi = 1\) if the cure status is observed. We assume that the cure status is missing at random\(^{30}\), which implies that, given \(\psi_i\) and \(X_i\), the probability that the cure status is missing depends only on the observed information \(X_i\), not on the unobserved \(\psi_i\). That is,

\[
\pi(X_i) = E(\xi_i \mid X_i) = E(\xi_i \mid X_i, \psi_i).
\]
The main idea in the approach of Aerts et al. is to use the assumed regression relationship between $X$ and $v$ imputing locally the missing observations of $v$. Here, the term local makes reference to the region of the covariate $X$ that is close to the observations with missing $v$. An outline of the algorithm is:

Step 1: (Resampling step) Fix $m$ between 1 and $M$, we perform a nonparametric resampling of the observed data. That is, for each observation $i = 1, \ldots, n$, if the cure status is observed ($\xi_i = 1$) generate $v_i^{*(m)}$ from the distribution $\mathcal{L}(X_i)$ with cumulative distribution function

$$
\sum_{j=1}^{n} B_{g_1,j}(X_i) \mathbb{1}(v_j \leq u)
$$

where $B_{g_1,j}(x)$ are the NW weights with bandwidth $g_1$.

Step 2: (Imputation step) Given the resampled data from Step 1, the missing values of $v$ are imputed using local resampling. More specifically, conditionally on the resampled data $((X_i, v_i^{*(m)}, \xi_i) : i = 1, \ldots, n)$, a second distribution $\mathcal{L}^*(X_i)$ is constructed, with cumulative distribution function

$$
\sum_{j=1}^{n} B_{g_2,j}(X_i) \mathbb{1}(v_j^{*(m)} \leq u)
$$

where the NW weights $B_{g_2,j}(x)$ are computed with a second bandwidth $g_2$. Then, if $v_i$ is missing, that is $\xi_i = 0$, generate $v_i^{+m}$ from $\mathcal{L}^*(X_i)$.

Step 3: (Computation of the final estimator) For $\tilde{v}^m_i = \xi_i v_i + (1-\xi_i) v_i^{+m}$, let $1 - \hat{p}_n^m = (1/n) \sum_{i=1}^{n} \tilde{v}^m_i$ be the empirical estimator of the cure probability with the $m$th augmented dataset. The final multiply imputation (MI) estimator for the cure probability $1 - p$ is

$$
1 - \hat{p}_n^{MI} = \frac{1}{M} \sum_{m=1}^{M} (1 - \hat{p}_n^m).
$$

Analogously, let us define $1 - \hat{p}_h^m(x) = \sum_{i=1}^{n} B_{g_1}(x) \tilde{v}_i^m$ as the NW estimator in (4) computed with bandwidth $h$ and the $m$th augmented dataset. The final multiply imputed Nadaraya-Watson (MI-NW) estimator for the cure probability $1 - p(x)$ is

$$
1 - \hat{p}_h^{MI-NW}(x) = \frac{1}{M} \sum_{m=1}^{M} (1 - \hat{p}_h^m(x)).
$$

In our context, it is natural to use the NW weights in the empirical distributions in Steps 1 and 2. Note that Step 1 is needed to fully account for all uncertainty in predicting the missing values by adding extra variability into the multiply imputed values; as we can never know the true values of the missing data. Under conditions similar to those in Cheng, Aerts et al showed that the proposed estimator of $1 - p$ is consistent. Next we derive the asymptotic expressions of the bias and variance for the MI-NW estimator, directly as in Aerts et al. The following regularity assumption will be needed.

**Assumption 10.** The function $\pi(x)$ has at least two bounded derivatives.

**Proposition 4.** Suppose that assumptions 2a, 3–4, 8 and 10 hold. Also, the bandwidths $h$, $g_1$, $g_2$ satisfy $h \to 0$, $g_1 \to 0$, $g_2 \to 0$, $nh \to \infty$, $ng_1 \to \infty$ and $ng_2 \to \infty$ as $n \to \infty$. The asymptotic bias of $1 - \hat{p}_n^{MI-NW}(x)$ is

$$
\mu_{h,g_1,g_2}^{MI-NW}(x) = h^2 c_{1,c}(x) + (g_1^2 + g_2^2) c_{2,MI-NW}(x) + O \left( (h^2 + g_1^2 + g_2^2)^2 \right),
$$

where $c_{1,c}(x)$ is defined in (9), and

$$
c_{2,MI-NW}(x) = \frac{(1 - \pi(x)) [\pi(x) (1 - p(x)) m(x)]''}{2m(x) \pi(x)} d_{\alpha}.
$$
If the bandwidths are \( g_1/h \rightarrow C_1 \) and \( g_2/h \rightarrow C_2 \), then the asymptotic variance is

\[
\sigma^2_{\text{MI-NW}}(x) = \frac{c_k}{Mn_h} \cdot \frac{1 - \pi(x) \cdot p(x) \cdot (1 - p(x))}{\pi(x) \cdot m(x)} + \frac{1}{n_1} \cdot \frac{1 - p(x)}{m(x)} \left[ \pi(x) \cdot c_k + (1 - \pi(x)) \left( c_{K,C_1} \cdot c_2 + \frac{1 - \pi(x)}{\pi(x)} \cdot d_{K,C_1} \cdot c_2 \right) \right] + \frac{1}{n_2} \cdot \frac{1 - p(x)}{m(x)} \left[ \pi(x) \cdot c_k + 2 \cdot c_{K,C_1} \cdot c_2 + \frac{1 - \pi(x)}{\pi(x)} \left( c_{K,C_1} \cdot c_2 + 2 \cdot d_{K,C_1} \cdot c_2 \right) \right] + \frac{1}{n_3} \cdot \frac{1}{m(x)} \left( \frac{1}{Mn_h} \right) + \left( \frac{1}{n_1} \right) + \left( \frac{1}{n_2} \right) + \left( \frac{1}{n_3} \right).
\]

where \( c_{K,C} = \int \int K(u)K(v)K(u + C_1)du \cdot dv \cdot c_{K,C_1,C_2} = \int \int \int K(u)K(v)K(u + C_1)K(u + C_2)du \cdot dv \cdot dw \) and \( d_{K,C_1,C_2} = \int \int \int K(u)K(v)K(u + C_1)K(u + C_2)du \cdot dv \cdot dw \).

The proof of Proposition 4 is deferred to the Supplementary Material. As for the bias, the term \( c_{1,c}(x) \) is the dominant term of the bias of the NW estimator\(^2\)\(^2\)\(^2\) of \( 1 - \hat{p}_x(x) = E(v|X = x) \), while \( c_{2,MN}(x) \) stems from the multiple imputation procedure in Steps 1 and 2. Back to the variance part, note that if \( C_1 = C_2 = 0 \) then \( c_{K,C_1} = c_{K,C_1,C_2} = d_{K,C_1} = c_k \), whereas if \( C_1 = \infty \) or \( C_2 = \infty \), then \( c_{K,C_1} = c_{K,C_1,C_2} = d_{K,C_1} \cdot c_2 = 0 \). It is easy to prove that in the case of no missingness, the dominant term of the bias reduces to that of the NW estimator \( c_{1,c}(x) \), whereas the leading term of the variance becomes \( (1/nh)(\sigma^2(x) + \mu^2(x))/m(x) \), where \( \sigma^2(x) = \text{var}(v|X = x) = p(x)(1 - p(x)) \) and \( \mu(x) = E(v|X = x) = 1 - p(x) \). The comparison in terms of bias of the proposed estimator of the cure probability and the MI-NW estimator is a trade-off between the terms \( c_{2,c}(x) \) in (10) and \( c_{2,MN}(x) \) in (17), while the comparison in terms of variance is not straightforward.

5 | SIMULATION STUDY

A simulation study was conducted to assess the finite sample performance of the proposed cure probability estimator, \( 1 - \hat{p}_x(x) \). Data were generated from the mixture cure model in (1). The latency part is modeled using a truncated exponential distribution and specified as

\[
S_0(t|x) = \begin{cases} 
    \exp(-a(x)t) - \exp(-a(x)4.605) & \text{if } 0 \leq t \leq 4.605 \\
    1 - \exp(-a(x)4.605) & \text{if } t > 4.605 
\end{cases}, \quad \text{where } a(x) = \exp\left(\frac{x + 20}{40}\right).
\]

Six different scenarios characterized by the cure probability function, \( 1 - p(x) \), were considered. As can be seen in Table 1, the cure probability displays a wide range of different functions.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1 - ( p(x) )</th>
<th>Censoring percentage</th>
<th>Cure percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Logistic function ( (1 + \exp(0.476 + 0.358x))^{-1} )</td>
<td>57.2</td>
<td>46.7</td>
</tr>
<tr>
<td>2</td>
<td>Cubic function ( 0.5 - \frac{x}{16000} \times x^3 )</td>
<td>60.9</td>
<td>50.0</td>
</tr>
<tr>
<td>3</td>
<td>Linear function ( 0.5 - 0.025x )</td>
<td>60.3</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>Low constant ( 0.2 )</td>
<td>38.8</td>
<td>20.0</td>
</tr>
<tr>
<td>5</td>
<td>High constant ( 0.8 )</td>
<td>84.7</td>
<td>80.0</td>
</tr>
<tr>
<td>6</td>
<td>Convex function ( 0.0025x^2 )</td>
<td>49.0</td>
<td>33.3</td>
</tr>
</tbody>
</table>

The censoring time \( C \) was generated, independently of \( X \) and \( Y \), from a Weibull distribution with shape parameter \( \alpha = 2 \) and scale parameter \( \beta = 4 \). The covariate \( X \) was uniformly distributed on the interval \([-20, 20]\). Notice that \( S_0(t | x) \) is truncated at \( r_0 = 4.605 \), so that the support for \( C \) is larger than the support of \( Y \) in order to fulfill condition (6). Depending on the scenario, the percentage of censored observations ranged from 38.8% (where 20% of observations were cured) to 84.7% (where 80% of observations were cured). For all scenarios the proportion \( \pi \) of cured individuals that are observed was set to 0.2, 0.8 and 1.
Two different designs were considered. They differ with respect to the distribution of the observed times of the individuals known to be cured. In the first design (Design 1), the observed lifetimes of the patients known to be cured were simulated to be falling within the largest censoring times. This design intended to reflect the pattern of the observed lifetimes of the patients known to be cured in the breast cancer data. In the second design (Design 2) the distribution of the observed times of the known cured patients in COVID-19 data is mimicked. In this case the known cured observations were simply chosen at random among the censored observations. For each scenario we simulated 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 Supplementary Material.

Our first goal was to evaluate the small sample size performance of \( 1 - \hat{p}_h^*(x) \) in terms of the squared bias, variance and MSE. For this purpose, we compared the estimator \( 1 - \hat{p}_h^*(x) \) with (a) the estimator \( 1 - \hat{p}_h(x) \) by Xu and Peng \(^{11} \); (b) the estimator \( 1 - \hat{p}_h^{\text{MI-NW}}(x) \) in (16) with \( M = 5 \) multiple imputations, and (c) the estimator \( 1 - p(x; \hat{\lambda}) \) by Bernhardt \(^{9} \), which considers a logistic regression model to fit the probability of cure with an expectation–maximization algorithm for estimating the model parameter \( \gamma \).

For all the nonparametric estimators, the search for the optimal bandwidth \( h \) was performed in a grid of 21 values ranging from 1.5 to 100 equispaced on a logarithmic scale. Besides, as for the MI-NW estimator, the pilot bandwidths for the local resampling step \((g_1)\) and the imputation step \((g_2)\) were searched in a grid of 11 bandwidths equispaced from 1.5 to 100 on a logarithmic scale. It should be noted that the Epanechnikov kernel function was used in the nonparametric estimations.

The MSE of \( 1 - \hat{p}_h^*(x) \) as well as \( 1 - \hat{p}_h(x) \) and \( 1 - \hat{p}_h^{\text{MI-NW}}(x) \), all of them computed with the corresponding optimal bandwidths, and the MSE of \( 1 - p(x; \hat{\lambda}) \) for Design 1 are illustrated in Figure 2. As expected, in Scenario 1, the semiparametric estimator behaves well since it fits a logistic regression for the cure probability. However, it is not surprising that in Scenarios 2 – 6 the semiparametric estimator performs worse. This is due to the fact that the logistic distribution assumption for the cure probability is violated. The proposed estimator, Xu and Peng’s estimator and the MI-NW estimator are quite competitive when the optimal bandwidth is used, even beating the semiparametric estimator in Scenario 1 for values of \( X \) close to 0. In all the simulated scenarios, the proposed estimator behaves similarly or outperforms Xu and Peng’s estimator for all values of the covariate. Further, the proposed estimator is competitive over the MI-NW estimator, showing in general a better behavior.

We obtained similar results for Design 2, see Figure 1 in the Supplementary Materials. In this design the differences in the squared bias between \( 1 - \hat{p}_h^*(x) \) and the other estimators across all scenarios are quite apparent. As it can be seen, in almost all the scenarios the performance of \( 1 - \hat{p}_h^*(x) \) and \( 1 - \hat{p}_h(x) \) is a compromise; \( 1 - \hat{p}_h^*(x) \) outperforms when the cure probability is high (Scenario 5) and underperforms when the cure probability is small (Scenario 4). Table 2 provides information on the MSE, the squared bias and variance of \( 1 - \hat{p}_h^*(x), 1 - \hat{p}_h(x), 1 - \hat{p}_h^{\text{MI-NW}}(x) \) and \( 1 - p(x; \hat{\lambda}) \), for both Design 1 and 2. We see that the proposed estimator has smaller squared bias when the cure probability is high \( (x = 5) \) but it seems to be slightly biased when the cure probability is small \( (x = 5) \), although this bias disappears rapidly for increasing \( n \). For \( n = 200 \) (see Supplementary Material Figure 2 and Table 1) we see that the differences in bias are much smaller. Regarding the variance, the proposed estimator performs better in all scenarios and in both designs. Similar results are observed when the proposed estimator is compared with the MI-NW estimator. The MI-NW estimator performs quite well at high proportions of observed cured observations. However, its performance worsens as the proportion \( \pi \) of cured individuals that are observed decreases (see the results for \( \pi = 0.2 \) in Supplementary Material Figure 3 and Table 2).

An additional simulation study was conducted to evaluate the practical performance of the proposed bandwidth selector. For each scenario 1000 sets of simulated datasets of sample size \( n = 100 \) were generated. The number of bootstrap samples was \( B = 1000 \) and \( \kappa = 0.8 \). In Figure 3 we plot the quartiles of the selected bootstrap bandwidth \( h^{*}_c \) for the six scenarios with Design 2. We also computed the optimal bandwidth \( h^{*}_c \) as defined in (13) and compare with \( h^{*}_c \). Figure 4 shows the corresponding contour plots, illustrating the density of the bootstrap bandwidths and the MSE of \( 1 - \hat{p}_h(x) \) as a function of the bandwidth \( h \) and the covariate value \( x \). The performance of \( h^{*}_c \) varies depending on the scenario, but in general seems to perform well in all scenarios. The choice of bandwidth seems to be important in Scenarios 1 – 3 and 6, as different bandwidths would result to slightly different MSE. In Scenarios 4 and 5, different bandwidths yield approximately the same MSE. In this case the bootstrap bandwidth being far from the optimal bandwidth does not entail a significant loss of efficiency.
FIGURE 2 The MSE of the proposed estimator 1 – \( \hat{p}_h(x) \) (solid black line), Xu and Peng’s estimator 1 – \( \hat{p}_h(x) \) (dashed black line) and the MI-NW estimator 1 – \( \hat{p}_{MI-NW}(x) \) (dashed grey line), all of them computed with the optimal bandwidth, and the semiparametric estimator 1 – \( p(x; \hat{\gamma}) \) (grey line) in the simulated scenarios and when the observed times for known cures are falling within the largest censoring times (Design 1).

6 | REAL DATA ANALYSIS

6.1 | Breast cancer data

We illustrate the practical performance of 1 – \( \hat{p}_h(x) \) with an analysis of the breast cancer data presented in Section 1. A total of 42 (4.7%) patients died from cancer within the follow-up period. The observed times for the remaining patients were right-censored. Among the censored patients, 32 (3.8%) were alive with cancer and 20 (2.4%) were cancer free patients with observed time larger than 10 years, suggesting the presence of a cured group in the data. Often, when analysing the survival of breast cancer patients, it is of great interest to study the clinical effect of well-established clinicopathologic prognostic factors on the probability of cure and the survival of those who are not cured. The aim of this study was to estimate the probability of cure from breast cancer depending on the cancer stage, number of positive lymph nodes, menopausal status, margins and age at diagnosis.

The probability of cure from breast cancer 1 – \( p \) for different groups of patients according to the aforementioned categorical factors was estimated using (a) the empirical estimator 1 – \( \hat{p} \) that discards the patients with unknown cure status, (b) the MI estimator in (15) with \( M = 20 \), (c) the unconditional nonparametric estimator 1 – \( \hat{p}_n \) by Xu and Peng, and (d) the proposed estimator in (5). The results are given in Table 3.

The empirical estimator 1 – \( \hat{p} \) seems to underestimates the true probability of cure as patients with unknown cure status are excluded. In addition, when the known cures are distributed among the largest observed times (Design 1), as in this dataset, its biased behavior is stronger towards too small estimated probabilities. The MI estimator takes into consideration the unknown cure status, but it still appears to be performing poorly, because, 93.1% of patients have missing cure status (the proportion of
TABLE 2 Squared bias (bias²), variance (var) and MSE of \(1 - \hat{p}_h(x)\), \(1 - \hat{p}(x)\), \(1 - \hat{p}_{\text{MI-NW}}(x)\), all computed with the optimal MSE bandwidth \((h)\), and \(1 - p(x; \hat{\gamma})\) in the simulated scenarios, when the observed times for known cures are falling within the largest censoring times (Design 1) or are randomly among the censored times (Design 2).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>x</th>
<th>Design 1</th>
<th>1 - (\hat{p}_h(x))</th>
<th>(1 - \hat{p}(x))</th>
<th>1 - (\hat{p}_{\text{MI-NW}}(x))</th>
<th>1 - (p(x; \hat{\gamma}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-5</td>
<td>6.523</td>
<td>6.892</td>
<td>5.149</td>
<td>9.044</td>
<td>8.200</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>12.247</td>
<td>4.820</td>
<td>7.275</td>
<td>3.930</td>
<td>3.553</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>5.288</td>
<td>5.519</td>
<td>5.430</td>
<td>4.390</td>
<td>3.914</td>
</tr>
<tr>
<td>2</td>
<td>-5</td>
<td>100</td>
<td>0.001</td>
<td>2.674</td>
<td>2.672</td>
<td>2.756</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
<td>0.025</td>
<td>2.674</td>
<td>2.572</td>
<td>2.656</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>43.174</td>
<td>0.001</td>
<td>2.674</td>
<td>2.732</td>
<td>2.756</td>
</tr>
<tr>
<td>3</td>
<td>-5</td>
<td>18.64</td>
<td>0.052</td>
<td>3.225</td>
<td>3.272</td>
<td>4.572</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>100</td>
<td>0.014</td>
<td>2.674</td>
<td>2.732</td>
<td>3.030</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>18.64</td>
<td>0.201</td>
<td>3.535</td>
<td>3.736</td>
<td>2.858</td>
</tr>
<tr>
<td>4(a)</td>
<td>-5</td>
<td>0.10</td>
<td>0.781</td>
<td>1.845</td>
<td>1.923</td>
<td>2.461</td>
</tr>
<tr>
<td>4(a)</td>
<td>0</td>
<td>43.174</td>
<td>0.107</td>
<td>1.844</td>
<td>1.916</td>
<td>2.228</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>53.262</td>
<td>0.109</td>
<td>1.843</td>
<td>1.913</td>
<td>2.457</td>
</tr>
<tr>
<td>4(b)</td>
<td>-5</td>
<td>81.06</td>
<td>0.021</td>
<td>1.572</td>
<td>1.593</td>
<td>2.009</td>
</tr>
<tr>
<td>4(b)</td>
<td>0</td>
<td>81.06</td>
<td>0.020</td>
<td>1.573</td>
<td>1.596</td>
<td>2.009</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>53.262</td>
<td>0.020</td>
<td>1.574</td>
<td>1.594</td>
<td>2.009</td>
</tr>
<tr>
<td>5</td>
<td>-5</td>
<td>6.523</td>
<td>0.711</td>
<td>3.369</td>
<td>4.080</td>
<td>7.590</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4.286</td>
<td>0.302</td>
<td>1.888</td>
<td>1.489</td>
<td>112.083</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>6.523</td>
<td>0.656</td>
<td>3.604</td>
<td>4.260</td>
<td>76.138</td>
</tr>
</tbody>
</table>

Design 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>x</th>
<th>Design 2</th>
<th>1 - (\hat{p}_h(x))</th>
<th>(1 - \hat{p}(x))</th>
<th>1 - (\hat{p}_{\text{MI-NW}}(x))</th>
<th>1 - (p(x; \hat{\gamma}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-5</td>
<td>6.523</td>
<td>6.485</td>
<td>7.672</td>
<td>7.149</td>
<td>9.044</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>12.247</td>
<td>3.527</td>
<td>8.228</td>
<td>4.963</td>
<td>7.119</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>5.288</td>
<td>5.185</td>
<td>6.148</td>
<td>5.430</td>
<td>4.390</td>
</tr>
<tr>
<td>2</td>
<td>-5</td>
<td>100</td>
<td>0.074</td>
<td>2.585</td>
<td>2.659</td>
<td>2.739</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
<td>0.207</td>
<td>2.585</td>
<td>2.732</td>
<td>3.228</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>34.996</td>
<td>0.010</td>
<td>2.639</td>
<td>2.648</td>
<td>3.553</td>
</tr>
<tr>
<td>3</td>
<td>-5</td>
<td>18.64</td>
<td>0.002</td>
<td>3.095</td>
<td>3.097</td>
<td>3.553</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>100</td>
<td>0.151</td>
<td>2.593</td>
<td>2.744</td>
<td>3.553</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>18.64</td>
<td>0.539</td>
<td>3.518</td>
<td>4.057</td>
<td>3.553</td>
</tr>
<tr>
<td>4(a)</td>
<td>-5</td>
<td>100</td>
<td>0.364</td>
<td>1.933</td>
<td>2.297</td>
<td>2.628</td>
</tr>
<tr>
<td>4(a)</td>
<td>0</td>
<td>43.174</td>
<td>0.350</td>
<td>1.932</td>
<td>2.282</td>
<td>2.628</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>34.996</td>
<td>0.304</td>
<td>1.956</td>
<td>2.260</td>
<td>2.628</td>
</tr>
<tr>
<td>4(b)</td>
<td>-5</td>
<td>81.06</td>
<td>0.092</td>
<td>1.492</td>
<td>1.584</td>
<td>2.009</td>
</tr>
<tr>
<td>4(b)</td>
<td>0</td>
<td>81.06</td>
<td>0.090</td>
<td>1.494</td>
<td>1.584</td>
<td>2.009</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.089</td>
<td>1.496</td>
<td>1.585</td>
<td>2.009</td>
</tr>
<tr>
<td>5</td>
<td>-5</td>
<td>6.523</td>
<td>1.177</td>
<td>3.464</td>
<td>4.644</td>
<td>82.248</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4.286</td>
<td>0.328</td>
<td>1.219</td>
<td>1.547</td>
<td>82.248</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>6.523</td>
<td>1.022</td>
<td>3.694</td>
<td>4.716</td>
<td>82.248</td>
</tr>
</tbody>
</table>

missingness is higher). The estimator by Xu and Peng \(1 - \hat{p}_n\) does consider the censored observations, however, it dismisses the cure status information so it still underestimates the true cure probabilities. The proposed estimator \(1 - \hat{p}_h\) makes use of the available information of the cure status giving more accurate estimates. Nonetheless, in a Design 1 scenario, the results given by \(1 - \hat{p}_h\) and \(1 - \hat{p}_n\) are very similar, with the estimates from the proposed estimator being, as stated earlier, slightly larger.

The estimated probability of cure as a function of a continuous covariate, as the age, is given in Figure 5 (left). The proposed estimator \(1 - \hat{p}_h(x)\) was obtained using the bootstrap bandwidth selector in Section 3. It is compared with Xu and Peng’s estimator \(1 - \hat{p}(x)\), computed using a bootstrap bandwidth as well, the semiparametric estimator \(1 - p(x; \hat{\gamma})\) and the MI-NW estimator \(1 - \hat{p}_{\text{MI-NW}}(x)\) computed with \(M = 20\). To the best of our knowledge, there is no any specifically tailored bandwidth selector
FIGURE 3 Median (solid black line) and first and third quartiles (dashed lines) of the $B = 1000$ bootstrap bandwidths for the proposed estimator $1 - \hat{p}_n(x)$ in the simulated scenarios with $n = 100$, and when the observed times for known cures are randomly among the censored times (Design 2). The optimal MSE bandwidth (solid grey line) is displayed as reference.

for $1 - \hat{p}_n^{\text{MI-NW}}(x)$. Thus, in this data analysis, the pilot bandwidths $g_1$ and $g_2$ for Steps 1 and 2 were selected using the cross-validation selector of Bowman et al.\textsuperscript{35} available in the R package \texttt{kerdiest}\textsuperscript{36}, and the bandwidth $h$ in Step 3 was chosen via an improved cross-validation\textsuperscript{37} bandwidth selector for the Nadaraya-Watson estimator using the R package \texttt{np}\textsuperscript{38}.

Although the semiparametric estimator $1 - p(x; \hat{\gamma})$ indicates that the cure probability relatively increases with age, the curves from the proposed estimator and Xu and Peng's estimator indicate an increment in the cure probability for younger to middle age patients, and indicate no effect of age on the cure probability for elderly patients. This suggests that the logistic model assumed by the semiparametric estimator might not be appropriate. Observe that the cure probability given by $1 - \hat{p}_n(x)$, estimator that disregards the information of the known cure status, is equal or lower than the probability estimated with the proposed method. This implies that not taking into account that some individuals are cured from the event could underestimate the probability of cure. On the other hand, the MI-NW estimator shows a similar trend as $1 - \hat{p}_n^{\text{MI-NW}}(x)$ and $1 - \hat{p}_n(x)$, although the estimated probabilities are substantially smaller. As pointed out before, the performance of $1 - \hat{p}_n^{\text{MI-NW}}(x)$ worsens significantly as the proportion of missing data increases.

### 6.2 COVID-19 data

During the COVID-19 outbreak in 2020, countries around the world experienced a large number of incident cases, with many patients requiring hospitalization wards. Although most infected people presented with mild disease, there were many severe cases that required long stays in ICU, overwhelming the healthcare systems with critical consequences on the disease mortality. An accurate knowledge of the duration of hospitalization, and the prediction of the probability that a hospitalized patient would
require a bed in ICU, were key for understanding the hospital demand for beds and crucial for decision-making and suitable planning.

As mentioned in Section 1, the database contains the 10454 confirmed COVID-19 cases reported in Galicia (Spain) between March 6 and May 7, 2020. The time of interest is the length of stay in hospital ward until admission to ICU, and the goal is to estimate the probability of requiring admission to ICU from hospital ward. Of the 2484 hospitalized cases, 104 (4.2%) patients were excluded from analysis because they were admitted and discharged on the same date resulting to a follow-up time of 0 days or were admitted directly to the ICU. For the remaining 2380 hospitalized patients for at least one day, 1210 (50.7%) were aged 73 years of age or above and 1262 (53%) were males. A total of 1638 (68.8%) patients were discharged alive before entering ICU, and 328 (13.7%) had died before entering ICU; all of them can be considered as cured from admission to ICU. Finally, 197 patients required admission to ICU, which gives an empirical estimated probability of admission to ICU of 0.0828. In addition, 217 (9.1%) continued to receive care in the hospital bed. Some of these censored observations might eventually need to stay in the ICU. This shows that 0.0828 might be an underestimation of the probability of admission to ICU, motivating the use of the alternative estimators of that probability than can handle censoring such as the proposed estimator. The aim of this analysis was to estimate the probability of admission to ICU from hospital ward given age and sex as covariates of interest.

Table 4 shows the estimated probabilities of requiring ICU, as $1 - p$ is the probability of being cured from admission to ICU, given by the empirical estimator $\hat{p}$, the MI estimator $\hat{p}_{n}^{MI}$ with $M = 20$, Xu and Peng’s estimator $\hat{p}_{n}$ and the proposed estimator $\hat{p}_{n}^{c}$. It should be noted that in this data only 9.1% patients have missing cure status. Therefore, when the proportion of individuals with observed cure status is higher, the MI estimator is expected to perform nicely, and the biased performance of the empirical estimator towards low cure probability fades away. As it can be seen in Table 4, the estimated probabilities of admission to ICU given by the empirical estimator, the MI estimator and the proposed estimator are very similar. On the other hand, the estimated

**FIGURE 4** Contour plots of the MSE of the proposed estimator $1 - \hat{p}_{n}^{c}(x)$ as a function of the bandwidth $h$ and the value $x$ of the covariate in the simulated scenarios, when the known cures are randomly among the censored observations (Design 2). The optimal bandwidth $h_{x}$ is marked with a cross. The density of the bootstrap bandwidths is shown in grey scale.
TABLE 3 Characteristics of the patients of the breast cancer dataset, and the estimated cure probability $1 - p$, using the empirical estimator $1 - \hat{p}$, the MI estimator $1 - \hat{p}^\text{MI}_n$, Xu and Peng’s estimator $1 - \hat{p}_n$ and the proposed estimator $1 - \hat{p}^c_n$.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Count (%)</th>
<th>Uncured</th>
<th>Censored</th>
<th>Estimated probability of cure $(1 - p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dead</td>
<td>Cured</td>
<td>Unknown</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 55 years</td>
<td>371 (41.3)</td>
<td>21</td>
<td>11</td>
<td>339</td>
</tr>
<tr>
<td>≥ 55 years</td>
<td>527 (58.7)</td>
<td>21</td>
<td>9</td>
<td>497</td>
</tr>
<tr>
<td>Pathologic stages†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>164 (19.8)</td>
<td>5</td>
<td>4</td>
<td>155</td>
</tr>
<tr>
<td>II</td>
<td>514 (62.0)</td>
<td>20</td>
<td>10</td>
<td>484</td>
</tr>
<tr>
<td>III</td>
<td>151 (18.2)</td>
<td>10</td>
<td>4</td>
<td>137</td>
</tr>
<tr>
<td>Menopausal status†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>185 (22.3)</td>
<td>10</td>
<td>0</td>
<td>175</td>
</tr>
<tr>
<td>Peri</td>
<td>63 (7.6)</td>
<td>7</td>
<td>11</td>
<td>45</td>
</tr>
<tr>
<td>Post</td>
<td>581 (70.1)</td>
<td>17</td>
<td>9</td>
<td>555</td>
</tr>
<tr>
<td>Number of positive lymph nodes †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>392 (50.5)</td>
<td>12</td>
<td>11</td>
<td>369</td>
</tr>
<tr>
<td>1–3</td>
<td>299 (38.5)</td>
<td>16</td>
<td>6</td>
<td>277</td>
</tr>
<tr>
<td>&gt; 3</td>
<td>86 (11.1)</td>
<td>11</td>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>Margin status †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>761 (92.5)</td>
<td>21</td>
<td>14</td>
<td>726</td>
</tr>
<tr>
<td>Positive</td>
<td>62 (7.5)</td>
<td>4</td>
<td>3</td>
<td>55</td>
</tr>
</tbody>
</table>

† contains missing observations

Event probabilities given by Xu and Peng’s estimator $\hat{p}_n$ seem to be higher. The reason is that, as derived from the simulation results, in Design 2 when the known cures are distributed uniformly among the observed lifetimes, as in this data set, the results from Xu and Peng’s estimator underperforms when the cure probability $1 - p$ is high, or equivalently, when the probability of the event $p$ is low.

TABLE 4 Characteristics of COVID-19 patients in Galicia (Spain), and estimated probability of requiring ICU ($p$), when the probability of cure from admission to ICU $(1 - p)$ is estimated using the empirical estimator $1 - \hat{p}$, the MI estimator $1 - \hat{p}^\text{MI}_n$, Xu and Peng’s estimator $1 - \hat{p}_n$ and the proposed estimator $1 - \hat{p}^c_n$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Count (%)</th>
<th>Uncured ICU admission</th>
<th>Censored</th>
<th>Estimated probability of requiring ICU ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dead</td>
<td>Discharged</td>
<td>Unknown</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–24 years</td>
<td>22 (0.9)</td>
<td>0</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>25–54 years</td>
<td>359 (15.1)</td>
<td>25</td>
<td>316</td>
<td>13</td>
</tr>
<tr>
<td>55–64 years</td>
<td>354 (14.9)</td>
<td>41</td>
<td>271</td>
<td>18</td>
</tr>
<tr>
<td>65–74 years</td>
<td>582 (24.5)</td>
<td>96</td>
<td>411</td>
<td>40</td>
</tr>
<tr>
<td>75–84 years</td>
<td>571 (24.0)</td>
<td>34</td>
<td>377</td>
<td>59</td>
</tr>
<tr>
<td>85 years and over</td>
<td>492 (20.7)</td>
<td>126</td>
<td>246</td>
<td>82</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1118 (47)</td>
<td>55</td>
<td>822</td>
<td>105</td>
</tr>
<tr>
<td>Male</td>
<td>1262 (53)</td>
<td>142</td>
<td>816</td>
<td>112</td>
</tr>
</tbody>
</table>

Figure 5 (right) shows the estimated probability of being admitted to the ICU depending on the age, obtained using the proposed estimator $1 - \hat{p}^c_n(x)$ and Xu and Peng’s estimator $1 - \hat{p}_n(x)$, both computed using the bootstrap bandwidth, the semiparametric estimator $1 - p(x; \hat{p})$, and the MI-NW estimator $1 - \hat{p}^\text{MI-NW}_n(x)$ computed using the same bandwidths selectors.
FIGURE 5 Estimation of the probability of cure from breast cancer (left) and probability of admission to ICU for COVID-19 patients (right) estimated using the proposed estimator \(1 - \hat{p}_h(x)\) (solid black line), Xu and Peng’s estimator \(1 - \hat{p}_h(x)\) (dashed black line), both computed with the bootstrap bandwidth, the MI-NW estimator \(1 - \hat{p}_{\text{MI-NW}}(x)\) (dashed grey line) computed using the cross-validation bandwidths, and the semiparametric estimator \(1 - p(x; \hat{\tau})\) (solid grey line).

as in the breast cancer example. Although the semiparametric estimator suggests a uniformly decreasing effect of the age on the probability of admission to the ICU, the other three estimators indicate that the logistic assumption for the cure probability might not be acceptable, as the curve patterns are characterized by a constant to a slightly increasing probability of admission to the ICU for younger patients (below 55 years), a sharp increase of the probability for middle age patients (from 55 to 69 years) and a decrease for elderly patients (70 years and older).

For the same reason as in the estimation of the unconditional probability \(p\) of requiring a bed in ICU, Xu and Peng’s estimator seems to overestimate the probability of ICU admission as a function of the age. On the other hand, the pattern of the MI-NW estimator is consistent with the proposed estimator. However, it seems to underestimate the probability of admission to ICU for young-to-middle age patients. This is due to the low percentage of observed admissions to ICU in patients of those ages. These results are consistent with the results in Table 4.

7 CONCLUSIONS

In this paper, we have highlighted the role that additional information given by the cure status of some censored observations plays in the mixture cure model under the right random censoring. It definitely makes a difference in the estimation of the probability of cure when the cure status information is considered, not only theoretically but also in the finite sample estimation.

We have demonstrated that the proposed estimator has always smaller or equal asymptotic variance than the nonparametric cure probability estimator that ignores the cure status. The beneficial effect of knowing the cure status on the bias is not straightforward, as it depends on the censoring distribution, the conditional probability of observing cured individuals and the probability of cure. However, based on the numerical study there seems to be an overall improvement in terms of the bias, especially when the cure probability is high.

When some individuals are observed to be cured from the event of interest, the empirical estimator of the unconditional cure probability, which disregards the censored observations, clearly underestimates the true probability and it cannot handle continuous covariates. While the multiply imputed Nadaraya-Watson estimator improves the estimation via imputing the partially
observed response, it is computationally quite expensive, particularly when the sample size is large. Moreover, when there is heavy missingness it performs poorly and cannot be used if there is no information on cured observations. It is, indeed, well known that the semiparametric estimator by Bernhardt\textsuperscript{9} heavily suffer when the logistic distribution assumption is violated. The nonparametric estimator by Xu and Peng\textsuperscript{11}, which ignores the cure status, and the proposed estimator, which makes use on the available information of the cure status, solve these inconveniences.

The probability of cure can be written not only in terms of the survival function but also as a regression function with the cure indicator as the response. Based on this approach, we are aware of the existence of other estimators that can be used to estimate the cure probability when the cure status is partially missing. To name one of them, the problem can also be addressed using inverse probability weighting method\textsuperscript{39}. However, the effectiveness of this method depends largely on the level of missingness in the data.

References


10. Safari W, López-de-Ullibarri I, Jácome A. A product limit estimator of the conditional survival function when cure status is partially known. Submitted 2020.


**How to cite this article:** W. Safari, I. López-de-Ullibarri, M.A. Jácome (2020), Nonparametric kernel estimation of the cure probability in a mixture cure model when cure status is partially known, *Statistics in Medicine.*
Supplementary Material for Nonparametric kernel estimation of
the cure probability in a mixture cure model when cure status is
partially known

Wende Clarence Safari* | Ignacio López-de-Ullibarri2 | M. Amalia Jácome3

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

1 | PROOFS

Proof of Proposition 1.

1. It is straightforward since \( \xi_i, v_i = 0, i = 1, \ldots, n \).

2. Assume there exists a common specific known cure threshold \( d_i = d \) for \( i = 1, \ldots, n \). This implies that in the ordered sample, \( \{ (X_i, T_i, \delta_i, \xi_i, v_i) : i = 1, \ldots, n \} \), the \( n_i \) first observations correspond to individuals with \( T_i < d \) either not cured or with unknown cure status \( (\xi_i, v_i) = 0 \), and the remaining \( m \) observations are cured individuals with \( T_i \geq d \) and \( \xi_i, v_i = 1 \). Therefore,

\[
1 - \hat{\rho}_n(x) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_i B_{h|i}(x)}{\sum_{j=i}^{n_i} B_{h|j}(x) + \sum_{j=n_i+1}^{n} B_{h|j}(x)} \right) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_i B_{h|i}(x)}{\sum_{j=i}^{n_i} B_{h|j}(x)} \right) = 1 - \hat{\rho}_h(x).
\]

3. 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,

\[
1 - \hat{\rho}_n(x) = \prod_{i=1}^{n} \left( 1 - \frac{B_{h|i}(x)}{\sum_{j=i}^{n_i} B_{h|j}(x) + \sum_{j=1}^{n} B_{h|j}(x) \mathbb{I}(v_j = 1)} \right) \times \prod_{j=1}^{n} \left( 1 - \frac{B_{h|j}(x)}{\sum_{j=1}^{n} B_{h|j}(x) \mathbb{I}(v_j = 1)} \right) \times \ldots \times \prod_{j=1}^{n} \left( 1 - \frac{B_{h|j}(x) \mathbb{I}(v_j = 1)}{\sum_{j=1}^{n} B_{h|j}(x) \mathbb{I}(v_j = 1)} \right) = \prod_{j=1}^{n} \frac{B_{h|j}(x) \mathbb{I}(v_j = 1)}{\sum_{j=1}^{n} B_{h|j}(x) \mathbb{I}(v_j = 1)}.
\]

4. In an unconditional setting the weights are \( 1/n \) for \( i = 1, \ldots, n \). Thus, the proposed estimator becomes

\[
1 - \hat{\rho}_n(x) = \prod_{i=1}^{n} \left( 1 - \frac{\delta_i x_{i+1}^{(n)}}{n^{-i} \sum_{j=1}^{n-i} (\xi_j v_j = 1)} \right).
\]
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 \( T_{(i)} \geq d \). Besides, the weights are \( 1/n \) for \( i = 1, \ldots, n \). Then the proposed estimator reduces to the one in Laska and Meisner:

\[
1 - \hat{\rho}_n = \prod_{i=1}^{n} \left( 1 - \frac{\delta_{i}}{n_1 - i + 1 + m} \right).
\]

**Proof of Theorem 1.** Consider the following decomposition

\[
(1 - \hat{\rho}_n) = (1 - \rho(x)) = \left( 1 - F(T_{(1)} | x) \right) - (1 - F(T_{(1)} | x)) = \exp \left( -\hat{\Lambda}_n(T_{(1)} | x) \right) - \exp \left( -\hat{\Lambda}_n(T_{(1)} | x) \right) + R_1(x)
\]

where \( R_1(x) = \left( 1 - \hat{F}_n(T_{(1)} | x) \right) - \exp \left( -\hat{\Lambda}_n(T_{(1)} | x) \right) \), and

\[
\hat{\Lambda}_n(t | x) = \sum_{i=1}^{n} \frac{\delta_{i} B_{h_{ij}}(x)}{\sum_{j=1}^{n} B_{h_{ij}}(x) + \sum_{j=1}^{n} B_{h_{ij}}(x) \mathbf{1} (\xi_{ij} v_{ij} = 1)}
\]

is the estimator of the cumulative hazard function corresponding to the estimator of the survival function \( \hat{S}_n(t | x) \). First notice that, by a Taylor's expansion of the exponential function around \( -\Lambda(t_0 | x) \), (S1) becomes

\[
(1 - \hat{\rho}_n) = (1 - \rho(x)) = \left( 1 - \rho(x) \right) \left( \hat{\Lambda}_n(T_{(1)} | x) - \Lambda(t_0 | x) \right) + R_1(x) + R_2(x),
\]

where \( R_2(x) = -\frac{1}{2} \exp( -\Lambda(t^* | x) \right) \left( \hat{\Lambda}_n(T_{(1)} | x) - \Lambda(t_0 | x) \right)^2 \) and \( \Lambda(t^* | x) \) is a value between \( \hat{\Lambda}_n(T_{(1)} | x) \) and \( \Lambda(t_0 | x) \). Further, we can decompose (S2) as

\[
(1 - \hat{\rho}_n) = (1 - \rho(x)) = \left( \hat{\Lambda}_n(T_{(1)} | x) - \Lambda(t_0 | x) + \hat{\Lambda}_n(t_0 | x) \right) \left( 1 - \rho(x) \right) + R_1(x) + R_2(x) + R_3(x)
\]

where

\[
R_3(x) = -\frac{1}{2} \exp( -\Lambda(t^* | x) \right) \left( \hat{\Lambda}_n(T_{(1)} | x) - \Lambda(t_0 | x) \right)^2 \).
\]

Arguing similarly as in the proof of Theorem 2 in Iglesias-Pérez and González-Manteiga, given \( t = T_{(1)} \), then

\[
\sup_{x \in I} | R_1(x) | = O \left( n^{-1} h^{-1} \right) \text{ a.s.}
\]

For the second term \( R_2(x) \) we make use of Lemma 5 in López-Cheda et al,

\[
n^a \left( t_0 - T_{(1)} \right) \to 0 \text{ a.s. for any } a \in (0, 1),
\]

and the strong consistency results for the estimator \( \hat{\Lambda}_n(t | x) \) in Corollary 1 of Safari et al.\(^4\) for \( t = t_0 \), then

\[
\sup_{x \in I} | R_2(x) | = O \left( n^{-1} h^{-1} \log n \right) \text{ a.s.}
\]

The third term is bounded as follows,

\[
\sup_{x \in I} | R_3(x) | \leq | T_{(1)} - t_0 \hat{\Lambda}_n(T^* | x) (1 - \rho(x)),
\]

where \( \hat{\Lambda}_n(t^* | x) \) is a value between \( \hat{\Lambda}_n(T_{(1)} | x) \) and \( \hat{\Lambda}_n(t_0 | x) \). From (S3) for a sequence of bandwidths satisfying \( h \to 0 \), we have

\[
t_0 - T_{(1)} = O \left( (\log n)^{3/4} (nh)^{-3/4} \right) \text{ a.s.}
\]

and as a consequence

\[
\sup_{x \in I} | R_3(x) | = O \left( (\log n)^{3/4} (nh)^{-3/4} \right) \text{ a.s.}
\]

The proof concludes by applying Theorem 1 of Safari et al.\(^4\) for \( t = t_0 \in [a, b] \) under condition (6).
Proof of Proposition 3. From Theorem 1, the bias of the nonparametric estimator $1 - \widehat{\rho}_h(x)$ is asymptotically equal to the expected value of
\[
\frac{(nh)^{-1}(1 - p(x))}{m(x)} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \xi \left( T_i, \delta_i, \xi_i, v_i, r_0, x \right) = I + II,
\]
where
\[
I = \frac{(nh)^{-1}(1 - p(x))}{m(x)} \left( \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \xi \left( T_i, \delta_i, \xi_i, v_i, r_0, x \right) - E \left( \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \xi \left( T_i, \delta_i, \xi_i, v_i, r_0, x \right) \right) \right),
\]
\[
II = \frac{(nh)^{-1}(1 - F(t \mid x))}{m(x)} E \left( \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \xi \left( T_i, \delta_i, \xi_i, v_i, r_0, x \right) \right).
\]
Since $E(I) = 0$, the asymptotic bias of the estimator $1 - \widehat{\rho}_h(x)$ is $II$. Using Lemma 1 and Lemma 2 of Safari et al.\(^4\) when $t = r_0$, we get
\[
II = \frac{\hat{h}^2(1 - p(x))(\Phi_c'(x, r_0, x) m(x) + 2\Phi_c'(x, r_0, x) m'(x))dK}{2m(x)} + O(h^4),
\]
with $\Phi_c(y, t, x)$ and $\Phi_c''(y, t, x)$ the first and the second derivatives of $\Phi_c(y, t, x)$ with respect to $y$. By applying Lemma 3 in Safari et al.\(^4\) for $t = r_0$, we have
\[
\Phi_c'(x, r_0, x) = -\frac{p'(x)}{1 - p(x)},
\]
Besides, from Lemma 4 in Safari et al.\(^4\).
\[
\Phi_c''(x, r_0, x) = 2 \int_{0}^{x} \frac{G_c'(v \mid x)}{1 - G_c(v \mid x)} d \left( \frac{S'(s \mid x)}{S(s \mid x)} \right) \bigg|_{s=v} \frac{d - p''(x)}{1 - p(x)},
\]
where
\[
1 - G_c(t \mid x) = 1 - G(t \mid x) + \pi_1(t, x) (1 - p(x)) G_1(t \mid x),
\]
with
\[
\pi_1(t, x) = P(\xi = 1 \mid v = 1, C \leq t, X = x), \quad G_1(t \mid x) = P(C \leq t \mid v = 1, X = x).
\]
The expression of the asymptotic bias of $1 - \widehat{\rho}_h(x)$ derives from plugging (S8) and (S9) in (S7). Recalling (S4), the asymptotic variance of $1 - \widehat{\rho}_h(x)$ is
\[
\text{var}(I) = \frac{(1 - p(x))^2}{m^2(x)} (V_1 - V_2),
\]
where
\[
V_1 = \frac{1}{nh^2} E \left( K^2 \left( \frac{x - X}{h} \right) \xi^2 \left( T, \delta, \xi, v, r_0, x \right) \right), \quad V_2 = \frac{1}{nh^2} \left( E \left( K \left( \frac{x - X}{h} \right) \xi \left( T, \delta, \xi, v, r_0, x \right) \right) \right)^2.
\]
From Lemmas 1 and 2 of Safari et al.\(^4\) for $t = r_0$, $V_2$ reduces to
\[
V_2 = \frac{1}{4} \frac{h^2}{n} d_K \left( \frac{\Phi_c'(x, r_0, x) m(x) + 2\Phi_c'(x, r_0, x) m'(x)}{m(x)} \right)^2 + O \left( \frac{h^4}{n} \right) = O \left( \frac{h^2}{n} \right).
\]
As for $V_1$, let us define $\Phi_1(y, t, x) = E \left( \xi^2 \left( T, \delta, \xi, v, r_0, x \right) \mid X = y \right)$. Then, after a change of variable and a Taylor's expansion we obtain
\[
V_1 = \frac{1}{nh} \Phi_1(x, r_0, x) m(x) c_K + O \left( \frac{h}{n} \right).
\]
From Lemma 5 in Safari et al.\(^4\), the function $\Phi_1(y, t, x)$ can be written as:
\[
\Phi_1(y, t, x) = \int_{0}^{1} \frac{d H^1(v \mid x)}{(1 - H(v \mid x) + H^{11}(v \mid x))^2}.
\]
The proof concludes by substituting (S11) and (S12) into (S10) and using (S13).
Proof of Theorem 2. From Theorem 1, we consider

\[ (nh)^{1/2}[(1 - \hat{\theta}_h(x)) - (1 - p(x))] = (nh)^{1/2} \left( (1 - p(x)) \sum_{i=1}^{n} \tilde{B}_{hi}(x) \zeta \left( T_i, \delta_i, \xi_i, \nu_i, \tau_0, x \right) + R_h(x) \right), \]

with \( \zeta(T, \delta, \xi, v, t, x) \) and \( R_h(x) \) given in (7) and (8) of the main paper, 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 \((nh)^{1/2} R_h(x)\) is negligible. Consequently, the asymptotic distribution of \((nh)^{1/2}[(1 - \hat{\theta}_h(x)) - (1 - p(x))]\) is that of \((I + II)\), where \(I\) and \(II\) are given in (S5) and (S6). Under the assumption \(nh^4 \to 0\), we have \((nh)^{1/2} II = o(1)\). Therefore, the asymptotic distribution of (S14) is that of \(nh^{1/2} I\). Let \(nh^{1/2} I = \sum_{i=1}^{n} \eta_{i,h}(x)\), where

\[ \eta_{i,h}(x) = \frac{(nh)^{-1/2}(1 - p(x))}{m(x)} \left( K \left( \frac{x - X_i}{h} \right) \zeta \left( T_i, \delta_i, \xi_i, \nu_i, \tau_0, x \right) - E \left( K \left( \frac{x - X_i}{h} \right) \zeta \left( T_i, \delta_i, \xi_i, \nu_i, \tau_0, x \right) \right) \right). \]

Lindeberg’s theorem for triangular arrays\(^5\) can be applied to obtain

\[ \sum_{i=1}^{n} \eta_{i,h}(x) \rightarrow N(0,1) \text{ in distribution.} \]

Therefore, \((nh)^{1/2}[(1 - \hat{\theta}_h(x)) - (1 - p(x))] \rightarrow N(0, s^2(x))\) in distribution. This proves (i). The proof of (ii) is similar, noting that if \(nh^3 = C^3\), then the bias term is non-negligible.

Proof of Proposition 4. Observe that the expectation of \(1 - \hat{\theta}_h^{ML-NW} (x)\) is

\[ E \left( 1 - \hat{\theta}_h^{ML-NW} (x) \right) = \frac{1}{m} \sum_{m=1}^{M} \sum_{i=1}^{n} E \left( B_{hi} (x) \hat{\nu}_i^m \right) \]

\[ = \frac{1}{m} \sum_{m=1}^{M} \sum_{i=1}^{n} E \left( B_{hi} (x) \xi_i \nu_i \right) + \frac{1}{m} \sum_{m=1}^{M} \sum_{i=1}^{n} E \left( B_{hi} (x) (1 - \xi_i) \nu_i^m \right) \]

\[ = E_1 (x) + E_2 (x). \] \hspace{1cm} (S15)

The NW weights verify

\[ B_{hi} (x) = \frac{2}{n m(x)} K_h \left( \frac{x - X_i}{h} \right) - \frac{1}{n^2} \frac{m^2(x)}{m(x)} \sum_{j=1}^{n} K_h \left( \frac{x - X_j}{h} \right) + \frac{n^{-1} K_h \left( \frac{x - X_i}{h} \right) \left( m(x) - \hat{m}_h(x) \right)^2}{m(x)}, \]

where \(\hat{m}_h(x)\) is the kernel estimator\(^6,7\) of \(m(x)\). Then, for the summands in \(E_1(x)\) we have

\[ E \left( B_{hi} (x) \xi_i \nu_i \right) = \frac{1}{n m(x)} E \left( K_h \left( \frac{x - X_i}{h} \right) \xi_i \nu_i \right) - \frac{1}{n^2} \frac{m^2(x)}{m(x)} E \left( K_h \left( \frac{x - X_i}{h} \right) \sum_{j=1}^{n} K_h \left( \frac{x - X_j}{h} \right) \xi_i \nu_i \right) + O \left( \frac{1}{n} \left( h^4 + \frac{1}{nh} \right) \right). \]

Under MAR assumption,

\[ E \left( \xi_i \nu_i | X_i \right) = \pi \left( X_i \right) (1 - p \left( X_i \right)) . \] \hspace{1cm} (S16)

Thus, using (S16), Taylor expansions and change of variable, the first expectation in (S15) is

\[ E_1 (x) = \frac{n+1}{n} \pi (x) (1 - p(x)) \quad \quad + \frac{1}{2} \frac{h^2}{m(x)} \left( \frac{n+1}{n} \pi (x) (1 - p(x)) m(x) \right) d_k + O \left( \frac{1}{n} \left( h^4 + \frac{1}{nh} \right) \right). \]

The terms of \(E_2(x)\) are

\[ E \left( B_{hi} (x) (1 - \xi_i) \nu_i^m \right) = E \left( B_{hi} (x) (1 - \xi_i) E \left( \nu_i^m \right) \left| \text{observed data} \right. \right) \]

\[ = E \left( B_{hi} (x) (1 - \xi_i) \sum_{k=1}^{n} B_{g,k} (X_i) \sum_{j=1}^{n} B_{g,j} (X_k) \nu_j \right). \]
where
\[
B_{ij}(x) = \frac{\xi_j K_\delta(x - X_j)}{\sum_{k=1}^n \xi_k K_\delta(x - X_k)} \approx \frac{1}{ng} \frac{\xi_j K\left(\frac{x - X_j}{\delta}\right)}{\pi(x) m(x)} (1 + o_\delta(1))
\]
are the NW weights used in Step 1 and 2. Then,
\[
E\left(B_{hi}(x) (1 - \xi_i) v_{i^+}^m\right) \approx \frac{1}{n^2 g_1 g_2} \sum_{j=1}^n E\left(\frac{\pi(X_j) \pi(X_k)}{m(X_j) m(X_k)} K\left(\frac{X_i - X_k}{g_1}\right) K\left(\frac{X_j - X_k}{g_2}\right) B_{hi}(x) (1 - \xi_i) \xi_j x_k v_j\right).
\]
Note that \(1 - \xi_i) x_k x_j = 0\) if \(i = j\) or \(i = k\). So, there are two cases to be considered: (a) \(i \neq j, i \neq k, k = j\), and (b) \(i \neq j, i \neq k, k \neq j\). Using again Taylor expansions and change of variable, for case (a) we have
\[
E\left(B_{hi}(x) (1 - \xi_i) v_{i^+}^m\right) = \frac{1}{n^2 g_1 g_2} \sum_{j=1}^n E\left(\frac{h^2 d_K(1 - \pi(x))}{m(x)} \left((1 - \pi(x))(1 - p(x)) m(x)\right)^m - m''(x)(1 - \pi(x))(1 - p(x))\right)
\]
Using similar arguments in case (b)
\[
E\left(B_{hi}(x) (1 - \xi_i) v_{i^+}^m\right) = \frac{(n-1)(n-2)}{n^3} \left((1 - \pi(x))(1 - p(x))\right)
\]
Taking into account (S18) and (S19), then the second expectation in (S15) is
\[
E_2(x) = \frac{(n-1)(n-2)}{n^3} \left((1 - \pi(x))(1 - p(x))\right)
\]
Conditioning on the observed data (O) and the resampling data in Step 1 (R), the variance of \(1 - \tilde{p}_{h,\text{MI-NW}}(x)\) is
\[
\sigma^2_{h,\text{MI-NW}}(x) = E\left(\var{1 - \tilde{p}_{h,\text{MI-NW}}(x) | O, R}\right) + \var{E\left(1 - \tilde{p}_{h,\text{MI-NW}}(x) | O, R\right)} = V_1 + V_2.
\]
By the definition of \(1 - \tilde{p}_{h,\text{MI-NW}}(x)\), the first term in (S21) is
\[
V_1 = \frac{1}{M} E\left[\var\left(\sum_{i=1}^n \frac{1}{nh} \frac{1}{m(x)} K\left(\frac{x - X_i}{h}\right) (\xi_i v_i + (1 - \xi_i) v_{i^+}^1) | O, R\right)\right]
\]
By straightforward calculations we get that
\[
V_{11} = \frac{c_K}{MNh} \frac{(1 - \pi(x))(1 - p(x))}{m(x)} \left(1 + O\left(g_1^2 + g_2^2 + h^2\right)\right)
\]
and
\[
V_{12} = \frac{c_K}{MNh} \frac{(1 - \pi(x))(1 - p(x))^2}{m(x)} \left(1 + O\left(g_1^2 + g_2^2 + h^2\right)\right).
\]
Thus,

\[ V_1 = \frac{c_K}{Mn} \frac{(1 - \pi(x))p(x)(1 - p(x))}{m(x)} \left(1 + O \left(n h^2 + s_1^2 + s_2^2 + h^2 \right) \right). \]  

(S22)

Next, we turn to the second term of (S21):

\[ V_2 = \text{var} \left(E \left( E \left( 1 - \hat{h}_{\text{ML-NW}}^2(x) | O, R | O \right) \right) \right) + E \left( \text{var} \left(E \left( 1 - \hat{h}_{\text{ML-NW}}^2(x) | O, R | O \right) \right) \right) = V_{21} + V_{22}. \]  

(S23)

Then the first term in (S23) is

\[
V_{21} = \frac{1}{n^2 h^2 m^2(x)} \frac{1}{m^2(x)} \sum_{i=1}^{n} \left[ K \left( \frac{x - X_i}{h} \right) \xi_i v_i + (1 - \xi_i) E \left( v_{i,1}^+ | O, R \right) \right] \\
= \frac{1}{n h^2 m^2(x)} \sum_{i=1}^{n} \left[ K \left( \frac{x - X_i}{h} \right) \xi_i v_i \right] + \frac{1}{n h^2 m^2(x)} \sum_{i=1}^{n} \left[ (1 - \xi_i) E \left( v_{i,1}^+ | O, R \right) \right] \\
+ 2 \frac{1}{n h^2 m^2(x)} \text{cov} \left[ K \left( \frac{x - X_i}{h} \right) \xi_i v_i, K \left( \frac{x - X_i}{h} \right) (1 - \xi_i) E \left( v_{i,1}^+ | O, R \right) \right] \\
+ \frac{1}{n h^2 m^2(x)} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov} \left[ K \left( \frac{x - X_i}{h} \right) \xi_i v_i + (1 - \xi_i) E \left( v_{i,1}^+ | O, R \right) \right] \\
= I_1 + I_2 + I_3 + I_4.
\]

After considerable computations it can be shown that

\[ I_1 = \frac{c_K \pi(x)(1 - p(x))}{n h m(x)} + O \left( \frac{1}{n} \right), \]  

(S24)

\[ I_2 = \frac{c_K (1 - \pi(x))(1 - p(x))^2}{n h m(x)} + O \left( \frac{1}{n} \right), \]  

(S25)

\[ I_3 = -2 \frac{\pi(x)(1 - \pi(x))(1 - p(x))^2}{m(x)} + O \left( \frac{h^2}{n} \right) = O \left( \frac{1}{n} \right). \]  

(S26)

The term \( I_4 \) is simplified to

\[ I_4 = (1 - p(x))^2 \pi^2(x - 1) + 2 \frac{n - 1}{n h^2 m^2(x)} I_{41} + \frac{n - 1}{n h^2 m^2(x)} I_{42} + O(h^2), \]  

(S27)

where

\[ I_{41} = E \left[ K \left( \frac{x - X_1}{h} \right) K \left( \frac{x - X_2}{h} \right) \xi_1 v_i \left( 1 - \xi_2 \right) E \left( v_{i,1}^+ | O, R \right) \right], \]

\[ I_{42} = E \left[ K \left( \frac{x - X_1}{h} \right) K \left( \frac{x - X_2}{h} \right) \left( 1 - \xi_1 \right) \left( 1 - \xi_2 \right) E \left( v_{i,1}^+ | O, R \right) E \left( v_{i,1}^+ | O, R \right) \right]. \]

After applying changes of variables and Taylor expansions, we obtain that \( I_{41} \) yields the following result depending on the bandwidths. If \( g_1/h \to C_1 \geq 0 \) and \( g_2/h \to C_2 \geq 0 \), then

\[
2 \frac{n - 1}{n h^2 m^2(x)} I_{41} = 2 \pi(x)(1 - \pi(x))(1 - p(x))^2 \\
+ 2 \frac{1}{n h} \left[ \frac{(1 - \pi(x))(1 - p(x))^2}{m(x)} \right] c_{K,C_2} + \left[ \frac{(1 - \pi(x))(1 - p(x))}{m(x)} \right] c_{K,C_2} \\
+ 2 \frac{1}{n g_1} \left[ \frac{(1 - \pi(x))(1 - p(x))^2}{m(x)} \right] K(0) + O \left( \frac{1}{n h^2} \right) + O \left( \frac{1}{n g_1} \right), \]  

(S28)

where \( c_{K,C_2} = \iint K(u)K(v)K(w)K(u + C_1 v + C_2 w) du dv dw \) and \( c_{K,C_2} = \iint K(u)K(v)K(u + C_2 v) du dv \). Note that if \( C_1 = C_2 = 0 \) then \( c_{K,C_2} = c_{K,C_2} = c_K = \iint K^2(v) dv \). On the other hand, if \( C_1 = \infty \) then \( c_{K,C_2} = 0 \), and \( C_1 = \infty \) implies \( c_{K,C_2} = 0 \).
The result for $I_{42}$ can be proved in the same spirit as above

$$
\frac{n - 1}{nh^2} \frac{1}{m(x)} I_{42} = (1 - \pi(x))^2(1 - p(x))^2
$$

$$
+ \frac{1}{nh} \left[ \frac{(1 - \pi(x))^2(1 - p(x))^2}{\pi(x)m(x)} (c_{K,C_1,C_2} + 2d_{K,C_1,C_2}) + \frac{(1 - \pi(x))^2(1 - p(x))}{\pi(x)m(x)} d_{K,C_1,C_2} \right]
$$

$$
+ \frac{2}{ng_1} \frac{(1 - \pi(x))^2(1 - p(x))^2}{\pi(x)m(x)} K(0) + O \left( \frac{1}{nh} \right) + O \left( \frac{1}{ng_1} \right).
$$

(S29)

where $d_{K,C_1,C_2} = \int u \int v K(u)K(v)K(u + C_1v + C_2(u + v))dudvdw$. Again, if $C_1 = C_2 = 0$ then $d_{K,C_1,C_2} = c_K$, whereas if $C_1 = \infty$ or $C_2 = \infty$ then $d_{K,C_1,C_2} = 0$.

Finally, rejoining (S24) – (S29), we arrive to the following expression for $V_{21}$:

$$
V_{21} = \frac{1}{nh} \frac{1 - p(x)}{m(x)} \left[ \pi(x)c_K + (1 - \pi(x)) \left( c_{K,C_1,C_2} + \frac{1 - \pi(x)}{\pi(x)} d_{K,C_1,C_2} \right) \right]
$$

$$
+ (1 - p(x)) \left( c_K + 2c_{K,C_1,C_2} + \frac{1 - \pi(x)}{\pi(x)} (c_{K,C_1,C_2} + 2d_{K,C_1,C_2}) \right) \right]
$$

$$
+ \frac{2}{ng_1} \frac{(1 - \pi(x))(1 - p(x))^2}{\pi(x)m(x)} K(0) + O \left( \frac{1}{nh} \right) + O \left( \frac{1}{ng_1} \right).
$$

(S30)

The term $V_{22}$ may also be checked in a relatively similar manner:

$$
V_{22} = \frac{1}{Mnh} \frac{1 - \pi(x)}{m(x)} \left( \frac{1 - \pi(x)}{\pi(x)} p(x)(1 - p(x)) + O \left( \frac{1}{Mnh} \right) \right).
$$

(S31)

Note that adding $V_1$ in (S22), $V_{21}$ in (S30) and $V_{22}$ in (S31) gives (18) of the main paper. This concludes the proof.

## 2 FURTHER SIMULATION RESULTS

This section presents the complete results of the simulation study described in Section 5 in the main paper. The simulation results when $n = 100$, $\pi = 0.8$ and the known cures are randomly distributed among the censored observations (Design 2) are shown in Figure 1. Summarizing all the scenarios, the proposed estimator $1 - \hat{p}_h(x)$ has smaller MSE than Xu and Peng’s estimator $1 - \hat{p}_h(x)$ when the cure probability is high. When the cure probability is small, the differences between the MSE for these estimators is minimal. For the covariate value $x = -5$, the bias of the proposed estimator $1 - \hat{p}_h(x)$ is smaller than $1 - \hat{p}_h(x)$ for $n = 50$ and 100. For covariate values $x = 0$ and $x = 10$, the squared bias of $1 - \hat{p}_h(x)$ is slightly larger than $1 - \hat{p}_h(x)$, see Table 1. Their differences decreases when the sample size $n$ increases.

Figure 2 shows the effect of the sample size on the behavior of the estimators, when $n = 50$ and 200, for $\pi = 0.8$ in Scenario 1 and Design 2. As the sample size increases, the differences in the MSE of all estimators decreases. Table 1 shows detailed of the sample size on the MSE, the squared bias and variance of $1 - \hat{p}_h(x)$, $1 - \hat{p}_h(x)$, $1 - \hat{p}_h^{MI-NW}(x)$ and $1 - p(x; \tilde{\gamma})$.

Figure 3 and Table 2 provide some insight about the effect of $\pi$ on the estimators. While the performance of $1 - \hat{p}_h(x)$ is not affected by $\pi$, since the known cured observations are treated as censored, the behavior of the proposed estimator, the MI-NW estimator $1 - \hat{p}_h^{MI-NW}(x)$ and the semiparametric estimator $1 - p(x; \tilde{\gamma})$ improves in general as $\pi$ increases. Here the conclusion we draw from Table 2 is that the variance reduction in the proposed estimator from $\pi = 0.2$ and $\pi = 1$ is clear, but the effect on bias depends on the cure probability. As expected, the MI-NW estimator performs poorly when $\pi = 0.2$. 


FIGURE 1 The MSE of the proposed estimator $\hat{\beta}(x)$ (solid black line), Xu and Peng’s estimator $\hat{\beta}(x)$ (dashed black line) and the MI-NW estimator $\hat{\beta}_{\text{MI-NW}}(x)$ (dashed grey line), all of them computed with the optimal bandwidth, and the semiparametric estimator $\hat{\beta}(x; \hat{\gamma})$ (grey line) in the simulated scenarios with $n = 100$, $\pi = 0.8$, and when the observed known cures are randomly distributed among the censored times (Design 2).

TABLE 1 Squared bias ($\text{bias}^2$), variance (var) and MSE of the proposed estimator, Xu and Peng’s estimator, the MI-NW estimator, all computed with the optimal MSE bandwidth ($h_x$), and the semiparametric estimator by Bernhardt, in Scenario 1 for sample sizes $n = 50$, $100$, $200$, $\pi = 0.8$ and when the known cures are randomly distributed among the censored observations (Design 2).
FIGURE 2 The minimum MSE of the proposed estimator $1 - \hat{p}_h(x)$ (solid black line), Xu and Peng’s estimator $1 - \hat{p}_h(x)$ (dashed black line), and the MI-NW estimator $1 - \hat{p}_{\text{MI-NW}}(x)$ (dashed grey line), all of them computed with the optimal bandwidth for sample sizes $n = 50$ (left), $n = 200$ (right) and $\pi = 0.8$ in Scenario 1 and Design 2. Also shown are the MSE of the semiparametric estimator $1 - p(x; \hat{y})$ (grey line).

FIGURE 3 The minimum MSE of the proposed estimator $1 - \hat{p}_h(x)$ (solid black line), Xu and Peng’s estimator $1 - \hat{p}_h(x)$ (dashed black line) and the MI-NW estimator $1 - \hat{p}_{\text{MI-NW}}(x)$ (dashed grey line), all of them computed with the optimal bandwidth for sample size $n = 100$ and $\pi = 0.2$ (left) and $\pi = 1$ (right). Also shown are the MSE of the semiparametric estimator $1 - p(x; \hat{y})$ (grey line) in Scenario 1 and Design 2.
TABLE 2 Squared bias ($\text{bias}^2$), variance (var) and MSE of the proposed estimator, Xu and Peng’s estimator, the MI-NW estimator, all computed with the optimal bandwidth ($h_x$), and the semiparametric estimator, in Scenario 1 for sample size $n = 100$, and $\pi = 0.2, 100, 1$, and when the known cures are randomly distributed among the censored observations (Design 2).

<table>
<thead>
<tr>
<th>$\pi$</th>
<th>$\chi$</th>
<th>Proposed</th>
<th>NP</th>
<th>MI-NW</th>
<th>Bernhardt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$h_x$</td>
<td>bias$^2$</td>
<td>var</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.288</td>
<td>0.562</td>
<td>4.951</td>
<td>5.513</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>12.247</td>
<td>2.505</td>
<td>4.887</td>
<td>7.392</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.288</td>
<td>0.562</td>
<td>4.951</td>
<td>5.513</td>
</tr>
<tr>
<td>0.8</td>
<td>0</td>
<td>12.247</td>
<td>3.527</td>
<td>4.701</td>
<td>8.228</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.288</td>
<td>0.99</td>
<td>5.158</td>
<td>6.148</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.288</td>
<td>0.778</td>
<td>4.835</td>
<td>5.613</td>
</tr>
</tbody>
</table>
References


**How to cite this article:** W. Safari, I. López-de-Ullibarri, M.A. Jácome (2020). Nonparametric kernel estimation of the cure probability in a mixture cure model when cure status is partially known, *Statistics in Medicine*. 