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

Wende C. Safari[1] | **Ignacio López-de-Ullibarri**[2] | **M. Amalia Jácome**[3]

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

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

Abstract
An error was detected in the derivation of the expression of the cumulative hazard function in a recently published paper by Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2021), A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*, *63*(5), 984–1005, https://doi.org/10.1002/bimj.202000173. This short article aims to correct this error. There are some changes in the model notation in Section 2, the derivation of the expression of the cumulative hazard function in the Appendix, and the proofs of Lemmas 3 and 4 in the Supporting Information. Moreover, there is a small change in the generation of the values of the censoring variable C^* in the simulation study. As a consequence, the simulation results in Section 4 are affected. A corrected version of these sections is given in the Supporting Information.

Keywords
correction, censoring, cure models, kernel estimator, Nadaraya–Watson weights

1 | CORRECTIONS

The corrections to be made in Safari et al. (2021, pp. 985–986, 990–994, 999, 1001–1002; https://doi.org/10.1002/bimj.202000173) are outlined as follows: In Section 2 (pp. 985–986), we introduce the notation in the mixture cure model when cure status is partially known. Let Y be the survival time with distribution function $F(t|\mathbf{x}) = P(Y \leq t|\mathbf{X} = \mathbf{x})$ and $\mathbf{X}$ a vector of covariates, and C^* is the censoring time with distribution function $G(t|\mathbf{x}) = P(C^* \leq t|\mathbf{X} = \mathbf{x})$. The random variables Y and C^* are assumed to be conditionally independent given $\mathbf{X} = \mathbf{x}$. Assume that the survival time Y is subject to random right censoring, so that instead of observing Y, only $T^* = \min(Y, C^*)$ and $\delta = \mathbf{1}(Y < C^*)$ would be observed. Let $\nu = \mathbf{1}(Y = \infty)$ be the indicator of being cured from the event. Note that ν is partially observed because $\delta = 1$ implies $\nu = 0$, but ν is usually unknown for the censored observations. In the mixture cure model when the cure status is partially known, $\nu = 1$ is also observed randomly for some censored individuals.

Suppose that ξ indicates whether the cure status is known ($\xi = 1$) or not ($\xi = 0$). To accommodate for this situation, let the censoring distribution be an improper distribution function $G(t|\mathbf{x}) = \{1 - \pi(\mathbf{x})\}G_0(t|\mathbf{x})$, so with probability $\pi(\mathbf{x})$ the censoring variable is $C^* = \infty$, and with probability $\{1 - \pi(\mathbf{x})\}$ the value of the censoring variable C^* corresponds to the value of a random variable C with proper continuous distribution function $G_0(t|\mathbf{x})$. A cured individual is identified with probability $P(\xi = 1|\nu = 1, \mathbf{X} = \mathbf{x}) = P(C^* = \infty|\mathbf{X} = \mathbf{x}) = \pi(\mathbf{x})$. In this context, we observe the quintuple $\{(\mathbf{X}_i, T_i, \delta_i, \xi_i, \xi_i\nu_i) : i = 1, \ldots, n\}$, where, for individuals not identified as cured the observed time is $T_i = T_i^* = \min(Y_i, C_i^*)$, while the corresponding observed time for those known to be cured ($T_i^* = \infty$) is recorded as $T_i = C_i$.

The following corrects the derivation of the motivation of the estimator of the cumulative hazard function in Safari et al. (2021, p. 999). Without loss of generality, for simplicity we only consider a single continuous covariate X. The cumulative hazard function $\Lambda(t|x)$ can be written as follows:

$$\Lambda(t \mid x) = \int_0^t \frac{dF(v \mid x)}{1 - F(v^- \mid x)} = \int_0^t \frac{\{1 - G(v^- \mid x)\}dF(v \mid x)}{\{1 - G(v^- \mid x)\}\{1 - F(v^- \mid x)\}}. \tag{A1}$$

Observe that

$$\begin{aligned}\{1 - F(v^- \mid x)\}\{1 - G(v^- \mid x)\} &= P(Y \geq v, C^* \geq v \mid X = x)\\ &= P(Y \geq v, C^* \geq v, \xi\nu = 0 \mid X = x) + P(Y \geq v, C^* \geq v, \xi\nu = 1 \mid X = x)\\ &= P(T^* \geq v, \xi\nu = 0 \mid X = x) + P(\xi\nu = 1 \mid X = x).\end{aligned}$$

Note that, by definition,

$$\begin{aligned}&P(T^* \geq v, \xi\nu = 0 \mid X = x) + P(\xi\nu = 1 \mid X = x)\\ &= P(T \geq v \mid X = x) + P(T < v, \xi\nu = 1 \mid X = x)\\ &= 1 - H(v^- \mid x) + H^{11}(v^- \mid x). \end{aligned} \tag{A2}$$

On the other hand,

$$\int_0^t \{1 - G(v^- \mid x)\}dF(v \mid x) = P(C^* \geq Y, Y \leq t \mid X = x) = P(T^* \leq t, \delta = 1 \mid X = x),$$

where

$$P(T^* \leq t, \delta = 1 \mid X = x) = P(T \leq t, \delta = 1 \mid X = x) = H^1(t \mid x). \tag{A3}$$

Although the correction affects some lines in the proof of Lemmas 3 and 4 in the Supporting Information of Safari et al. (2021), these modifications do not affect the theoretical results.

The correction does, however, affect the simulation results in Section 4 of Safari et al. (2021, pp. 992–994) and in the Supporting Information. In the simulation study, there is a small change in the generation of the values of the censoring variable C^*. In consequence, the simulations results have been greatly improved and the gains are reflected in our proposed estimator.

For convenience reference, readers of Safari et al. (2021) are referred to a modified version of the paper which is provided as Supporting Information.

ACKNOWLEDGMENTS

We are grateful to Ingrid Van Keilegom and Jad Beyhum, who pointed out the incorrectness of the previous proof.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

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 are partially reproducible due to their computational complexity.

ORCID

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

REFERENCE

Safari, W. C., López-de-Ullibarri, I., & Jácome, M. A. (2021). A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*, *63*(5), 984–1005. https://doi.org/10.1002/bimj.202000173.

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, W. C., López-de-Ullibarri, I., & Jácome, M. A. (2021). Correction: A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*, *63,* 1544–1546. https://doi.org/10.1002/bimj.202100156