AIMS Mathematics, 10(9): 22249–22264.
DOI: 10.3934/math.2025991
Received: 26 May 2025
Revised: 22 July 2025
Accepted: 31 July 2025
Published: 26 September 2025

https://www.aimspress.com/journal/Math

Research article

Reparameterized flexible cure models with direct interpretation via the lambert W function

Diego I. Gallardo[1], **Yolanda M. Gómez**[1], **Marcelo Bourguignon**[2] **and Héctor J. Gómez**[3,*]

[1] Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción, Chile
[2] Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[3] Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería , Universidad Católica de Temuco, Temuco, Chile

* **Correspondence:** Email: hgomez@uct.cl.

Abstract: This article discusses the practical limitations of the flexible cure rate model proposed by Milienos (2022) in [1] [On a reparameterization of a flexible family of cure models, Statistics in Medicine 41, 4091–4111]. Although the model is structurally flexible, it has problems with interpretability and identifiability when covariates are included. To address these problems, we propose a novel reparameterization that directly links the cure fraction to meaningful parameters based on cancer cell counting mechanisms. This approach improves interpretability and resolves the identification problems present in the original model. Our new models facilitate straightforward inference within covariate structures while maintaining desirable flexibility. Parameter estimation is conducted using maximum likelihood methods. Through extensive simulation studies, we demonstrate that reparameterized models exhibit superior empirical performance. Their practical applicability is further illustrated through two melanoma datasets, where our models significantly outperform classical approaches in terms of fit and interpretability.

Keywords: cure rate models; reparameterization; melanoma data set; Lambert function
Mathematics Subject Classification: 62E10, 62F10

1. Introduction

The cure model, sometimes called “cure rate models”, refers to a class of models for censored survival data from subjects when some of them will not develop the event of interest, regardless of how long they are followed. Those who are not going to establish an event of interest are often referred to as “cured subjects” or “long-term survivors”. Clinical studies in cancer are a situation where there is a strong rationale for the existence of cured subjects because if the treatment is successful, the original

cancer is removed. The subject will not experience recurrence of the disease. This is particularly true for patients in the early stages of cancer. Cured subjects can also be found in other disciplines, such as economics, social studies, and education. Cure rate models are appropriate tools for problems in which the event of interest may not be observed in all individuals in the population. In the seminal work of [2], it is assumed that the population consists of two groups:, susceptible and cured, each with probabilities p and $1-p$, respectively. If $S(t;\boldsymbol{\xi})$ denotes the survival function (SF) of susceptible individuals (which is a proper function in the sense that $S(\infty;\boldsymbol{\xi}) = 0$), then the population SF is given by

$$S_{pop}(t) = 1 - p + pS(t;\boldsymbol{\xi}),$$

where $\boldsymbol{\xi}$ is a vector of parameters. In this case, and by construction, $p = S_{pop}(\infty)$ directly represents the cure rate of the model. This is an important property of the model because it allows the introduction of covariates directly through p, facilitating the interpretation of the regression coefficients. Later, [3] proposed an alternative cure rate model considering that each individual has a latent quantity of concurrent causes (say, M), each of which can produce the event of interest. For example, in a cancer context, M represents carcinogenic cells. In that proposal, the authors consider $M \sim \text{Po}(\theta)$, i.e., the Poisson distribution with mean θ. Denoting $W_1, W_2, \ldots$, a sequence of independent and identically distributed random variables (also independent of M) with common survival function $S(\cdot;\boldsymbol{\xi})$, it is possible to show that the population SF is given by

$$S_{pop}(t;\theta,\boldsymbol{\xi}) = G_M\left(S(t;\boldsymbol{\xi});\theta\right), \tag{1.1}$$

where $G_M(s;\theta) = \mathbb{E}_\theta\left(s^M\right)$ denotes the probability generating function of M. From that work, numerous alternative proposals have been presented in the literature, considering various distributions for M. To name a few distributions, we mention [4], where the authors proposed a unification of long-term survival models, proposing the negative binomial for the concurrent causes, the COM-Poisson [5], the power series [6, 7], and the polylogarithm [8], to name a few models. In this paper, we focus on the recently alternative proposal of [1]. In his proposal, he named flexible cure rate (Fcr) model, and the population SF is given by

$$S_{pop}(t) = \left(1 + \gamma\theta c^{\gamma\theta} F^{\lambda}(t;\boldsymbol{\xi})\right)^{-1/\gamma}, \tag{1.2}$$

where $\gamma \in \mathbb{R}$, $\theta, \lambda > 0$, $c = \mathrm{e}^{\mathrm{e}^{-1}}$, and F is a cumulative distribution function (CDF), which depends on the parameter vector $\boldsymbol{\xi} \in \Xi$, i.e., $\lim_{t\to\infty} F(t;\boldsymbol{\xi}) = 1$. In this case, the cured fraction is given by $p = \lim_{t\to\infty} S_{pop}(t) = (1 + \gamma\theta c^{\gamma\theta})^{-1/\gamma} \in [0, 1)$. The model has great appeal because it is very flexible. However, a first disadvantage of the model is that it is not associated with an underlying distribution for M, and therefore the population SF cannot be viewed as Eq (1.1), except for some combinations of parameters that correspond to particular cases as distribution for M, such as the Poisson or negative binomial model, among others. Unfortunately, nothing can be done to correct this defect in the model. A second disadvantage of the model is its limited interpretability of the regression coefficients. In the initial proposal by [1], it is not possible to obtain a meaningful interpretation of the model's regression parameters because θ, the parameter into which these regression coefficients are incorporated, does not represent any useful quantity, such as the cure fraction of the model, the mean of the competing risks, or any other measure of interest. However, this can be corrected by taking a convenient reparameterization of the model. Specifically, we propose a new reparameterization of the model that links the cure fraction directly to interpretable quantities derived from the cancer cell destruction mechanism. This

new formulation not only enhances the interpretability of the model's parameters, but also resolves identifiability issues that arise when covariates are included.

The paper is organized as follows. Section 2 outlines the model related to the population survival function described in Eq (1.2), including a reparameterized version of the model. In Section 3, we discuss the heterogeneous case concerning the proposed parametric model and examine its identifiability. Section 4 presents two simulation studies. In Section 5, we provide two real-data applications. Finally, the discussion and conclusions are presented in Section 6.

2. The model

In this section, we present the original and reparameterized versions of the model discussed in [1].

2.1. The original model

The cure fraction related to the population SF in Eq (1.2) is given by $p = \lim_{t\to\infty} S_{pop}(t) = (1 + \gamma\theta c^{\gamma\theta})^{-1/\gamma} \in [0, 1)$. In the following sections, we discuss in more detail some specific cases of the model.

i) For $\gamma > 0$, the negative binomial cure rate (NBcr) model [9] with population SF $S_{pop}(t) = \left(\frac{1-q}{1-q(1-F(t;\boldsymbol{\xi}))}\right)^r$, where $r = 1/\gamma$ and probability $q = \gamma\theta c^{\gamma\theta}/(1 + \gamma\theta c^{\gamma\theta})$. In addition, the geometric cure rate (Gcr) model is recovered for $\gamma = 1$, where r and q are the parameters of the NBcr.
ii) For $-1/\gamma \in \mathbb{N}$, the binomial cure rate model (Bcr) with population SF $S_{pop}(t) = (1 - qF(t;\boldsymbol{\xi}))^r$, where $r = -1/\gamma \in \mathbb{N}$ and success probability $q = -\gamma\theta c^{\gamma\theta}$. Furthermore, the mixture cure rate (Mcr) model of [2] is recovered for $\gamma = -1$.
iii) The promotion time cure rate (PTcr) model [3] with population SF $S_{pop}(t) = \exp(-\theta F(t;\boldsymbol{\xi}))$, is recovered for $\gamma \to 0$.

It should be noted that for $\gamma < 0$ with $-1/\gamma \notin \mathbb{N}$, no underlying distribution exists. It can be used with traditional tests, such as the LR and Wald (SW) tests to decide between the hypotheses $H_0 : \gamma = 1/r$, $H_0 : \gamma = -1/r$, for $r \in \mathbb{N}$ and $H_0 : \gamma = 0$ versus the corresponding alternative hypotheses. In addition, for $\zeta = -\mathrm{e}$, the population SF of the model is reduced to

$$S_{pop}(t) = \left(1 - F^{\lambda}(t;\boldsymbol{\xi})\right)^{\theta^*},$$

where $\theta^* = \theta/\mathrm{e}$. [1] highlights this case, because in this case S_{pop} corresponds to a proper SF, i.e., for the ordinary case (without cured individuals in the population).

The author also discusses an alternative parameterization for the model considering $\zeta = \gamma/\theta$, i.e.,

$$S_{pop}(t) = \left(1 + \zeta c^{\zeta} F^{\lambda}(t;\boldsymbol{\xi})\right)^{-\theta/\zeta}. \tag{2.1}$$

For this parameterization, the three mentioned cases are recovered as

i) The NBcr model, for $\zeta > 0$, $r = \theta/\zeta$ and $q = \zeta c^{\zeta}/(1 + \zeta c^{\zeta})$;
ii) The Bcr model, for $\zeta < 0$, $r = -\theta/\zeta \in \mathbb{N}$ and $q = -\zeta c^{\zeta}$;
iii) The PTcr model, for $\zeta \to 0$.

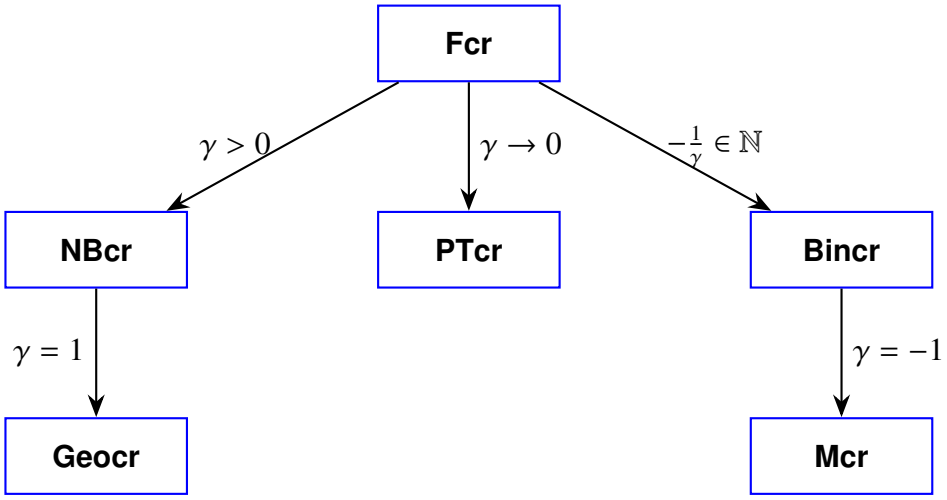

Figure 1. Diagram of relationships between the Fcr model and its particular cases.

Figure 1 illustrates the hierarchical relationships between the Fcr model and its particular cases. The diagram illustrates how specific values of the parameter γ result in distinct submodels, including the NBcr, PTcr, and Bincr models. Further special cases, including the Geocr and Mcr models, emerge under particular parameter constraints. This structure helps to understand the flexibility and generality of the Fcr modeling framework.

To avoid identifiability problems, [1] proposed including covariates taking $\theta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$, where $\mathbf{x}_i^\top$ is the vector of covariates for the ith individual, $\mathbf{x}_i^\top = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, and $\boldsymbol{\beta}^\top = (\beta_0, \ldots, \beta_p)$ is the vector of coefficients. In addition, under this parameterization, the hypothesis to decide between the general model and the NBcr or Bcr models in i) and ii) cannot be raised because the hypothesis r depends on θ_i in both cases. The exception is the PTcr model, where the hypothesis does not depend on covariates. Although the model appears to be quite flexible due to its numerous specific cases, it lacks a clear interpretation of the associated regression parameters when covariates are present. This makes it difficult to compare these coefficients with those from other cure models discussed in the literature.

Furthermore, the parameter θ does not have any useful interpretation except for the PTcr model ($\gamma \to 0$). Consequently, it is not possible to obtain an interpretation for the regression coefficients, whereas a comparison with other cure models in the literature for such coefficients is not feasible. [1] also proposed an alternative parametrization, such as

$$S_{pop}(t) = \left(1 - (1 - p^\gamma) F^\lambda(t; \boldsymbol{\xi})\right)^{1/\gamma}, \tag{2.2}$$

in which $\lim_{t\to\infty} S_{pop}(t) = p$ is directly the cure rate of the model. However, in this reparametrization, the author mentions that if $p = 0$, the parameter γ cannot be negative (that is, restricting the parametric space), introducing additional computational problems and potentially affecting the asymptotic distribution of the likelihood ratio (LR) test. In the following subsection, we introduce two alternative reparameterizations for the model in Eq (1.2) directly in terms of the cure rate, which addresses part of the aforementioned problems.

2.2. The proposed reparametrization

Consider p as the cure rate of the model. Therefore,

$$p = (1 + \gamma\theta c^{\gamma\theta})^{-1/\gamma} \Leftrightarrow \mathrm{e}^{-1}\left(p^{-\gamma} - 1\right) = \gamma\theta \mathrm{e}^{-1} \exp(\gamma\theta \mathrm{e}^{-1}). \tag{2.3}$$

Note that

$$\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right) \in \begin{cases} (0, \infty) & , \text{if } \gamma > 0; \\ (-\mathrm{e}^{-1}, 0) & , \text{if } \gamma < 0. \end{cases}$$

To solve the above equation, the Lambert function [10] (which we will denote as W) will be very useful. This function is the solution for b in equation $b\,\mathrm{e}^{b} = a$. For $a > 0$, a unique solution is obtained and determined by $b = W_0(a)$, where W_0 is known in the literature as the principal branch of the Lambert function. For $a \in (-\mathrm{e}^{-1}, 0)$, two solutions are obtained: $b = W_0(a)$ and $b = W_{-1}(a)$. Finally, for $a < -\mathrm{e}^{-1}$, such equation does not have a solution.

Therefore, using those results, we obtain that the solution for Eq (2.3) is given by $\theta = (\mathrm{e}/\gamma)W_0\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right)$, for $\gamma > 0$, and for $\gamma < 0$, there are two solutions for θ given by $(\mathrm{e}/\gamma)W_0\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right)$ and $(\mathrm{e}/\gamma)W_{-1}\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right)$, Finally, $\gamma \to 0$ is recovered the PTcr model with cure rate $p = \mathrm{e}^{-\theta}$, and it is straightforward that $\theta = -\log(p)$ in this case. In short, we obtain the following result:

$$\theta = \begin{cases} (\mathrm{e}/\gamma)W_0\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) & , \text{if } \gamma > 0; \\ -\log(p) & , \text{if } \gamma = 0; \\ (\mathrm{e}/\gamma)W_0\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) \quad \text{or} \quad (\mathrm{e}/\gamma)W_{-1}\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) & , \text{if } \gamma < 0. \end{cases}$$

Also, it is possible to show that $\lim_{\gamma\to 0}(\mathrm{e}/\gamma)W_k\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) = -\log(p)$, for $k = 0, -1$. Therefore, the two solutions for θ for the case $\gamma < 0$ converges to $-\log(p)$, as $\gamma \to 0$. Based on these results, we can reparameterize the model presented in Eq (1.2) in two ways.

First, considering

$$\theta = \begin{cases} (\mathrm{e}/\gamma)W_0\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) & , \text{if } \gamma \in \mathbb{R} - \{0\}; \\ -\log(p) & , \text{if } \gamma = 0. \end{cases}$$

and secondly, considering

$$\theta = \begin{cases} (\mathrm{e}/\gamma)W_{-1}\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right) & , \text{if } \gamma < 0; \\ -\log(p) & , \text{if } \gamma = 0. \end{cases}$$

We refer to these cases as flexible cure rate models of first (Fcr_0) and second (Fcr_{-1}) types, respectively. Note that the Bcr and PTcr models are particular cases of these models. Therefore, under these reparameterizations, the population SF for the Fcr_k model, $k = 0$ or $k = -1$, is given by

$$S_{pop}(t) = \begin{cases} \left(1 + W_k\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right)\exp\left\{1 + W_k\left(\mathrm{e}^{-1}\left(p^{-\gamma} - 1\right)\right)\right\}F^{\lambda}(t;\boldsymbol{\xi})\right)^{-1/\gamma} & , \text{if } \gamma \in \Gamma_k; \\ p^{F^{\lambda}(t;\boldsymbol{\xi})} & , \text{if } \gamma = 0, \end{cases}$$

where $\Gamma_0 = \mathbb{R} - \{0\}$ and $\Gamma_{-1} = (-\infty, 0)$. The hazard function for the model is

$$h_{pop}(t) = \begin{cases} \dfrac{\lambda W_k\left(e^{-1}\left(p^{-\gamma}-1\right)\right)\exp\left\{1+W_k\left(e^{-1}\left(p^{-\gamma}-1\right)\right)\right\} f(t;\boldsymbol{\xi})F^{\lambda-1}(t;\boldsymbol{\xi})}{\left(1+W_k\left(e^{-1}\left(p^{-\gamma}-1\right)\right)\exp\left\{1+W_k\left(e^{-1}\left(p^{-\gamma}-1\right)\right)\right\}F^{\lambda}(t;\boldsymbol{\xi})} & \text{, if } \gamma \in \Gamma_k; \\ -\log(p)\lambda F^{\lambda-1}(t;\boldsymbol{\xi})f(t;\boldsymbol{\xi}) & \text{, if } \gamma = 0. \end{cases}$$

We highlight that the Lambert function is implemented in many statistical/mathematical software. For example, packages `lamW` [11] and `pracma` [12] from [13] contain this function.

3. The Fcr$_k$ regression model

In this section, we discuss the heterogeneous case for the proposed reparametrized model. Consider a set of covariates $\mathbf{x}_i^\top = (x_{1i}, \ldots, x_{pi})$, $i = 1, \ldots, n$, which can be incorporated into parameter p using the classical logit link, i.e., $p_i = (1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}))^{-1}$, where $\boldsymbol{\beta}^\top = (\beta_1, \ldots, \beta_p)$ is the vector of unknown regression coefficients.. In order to compare the regression coefficients, for the Gcr, Bcr and PTcr (which also can be reparameterized in terms of the cure rate terms) the same link should be used. See [14] for details such reparametrized models. Let $\mathbf{x}_i^\top$ and $\mathbf{x}_i^{(j)\top} = (x_{1i}, \ldots, x_{ji} + 1, \ldots, x_{pi})$ be two vectors of covariates such as $\mathbf{x}_i^{(j)\top}$ is the same $\mathbf{x}_i^\top$, but with the j-th term increase in 1 unit. Note that in this case, the odds ratio for patients with covariates $\mathbf{x}_i^{(j)\top}$ and $\mathbf{x}_i^\top$ is given by

$$\frac{\frac{p_i^*}{1-p_i^*}}{\frac{p_i}{1-p_i}} = \frac{\exp(\mathbf{x}_i^{(j)\top}\boldsymbol{\beta})}{\exp(\mathbf{x}_i^\top\boldsymbol{\beta})} = \exp((\mathbf{x}_i^\top\boldsymbol{\beta} + \beta_j) - \mathbf{x}_i^\top\boldsymbol{\beta}) = \exp(\beta_j).$$

Therefore, $\exp(\beta_j)$ represents the increment in the odds ratio for patients where only the j-th covariate is incremented in 1 unit. This also applies to dichotomous covariates.

3.1. Estimation

Assume that the data are obtained under a right-censoring scheme. Thus, the observed data for the i-th individual can be represented by $T_i = \min(T_i^*, C_i)$ and $\delta_i = I(T_i^* \leq C_i)$, $1, \ldots, n$, where T_i^* and C_i denote failure and censoring time for each individual, respectively. Let $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\xi}, \gamma, \lambda)$ the vector of parameters for the Fcr$_k$ ($k = 0, -1$) model. For $\gamma \neq 0$, the log-likelihood function for the Fcr$_k$ model is given by

$$\begin{aligned}\ell(\boldsymbol{\psi}) = \sum_{i=1}^{n}\Bigg[&\left(-\frac{1}{\gamma}-\delta_i\right)\log\left(1+W_k\left(e^{-1}\left(p_i^{-\gamma}-1\right)\right)\exp\left\{1+W_k\left(e^{-1}\left(p_i^{-\gamma}-1\right)\right)\right\}F^{\lambda}(t_i;\boldsymbol{\xi})\right)\\ &+\log\left(W_k\left(e^{-1}\left(p_i^{-\gamma}-1\right)\right)\right)+1+W_k\left(e^{-1}\left(p_i^{-\gamma}-1\right)\right)\\ &+\log(\lambda)+(\lambda-1)\log(F(t_i;\boldsymbol{\xi}))+\log(f(t_i;\boldsymbol{\xi}))\Bigg],\end{aligned}$$

whereas for $\gamma = 0$, the log-likelihood function is given by

$$\ell(\boldsymbol{\psi}) = \sum_{i=1}^{n}[F(t_i;\boldsymbol{\xi})^{\lambda}\log(p_i) + \delta_i\{\log(-\log(p_i)) + \log(\lambda) + (\lambda-1)\log(F(t_i;\boldsymbol{\xi})) + \log(f(t_i;\boldsymbol{\xi}))\}].$$

Note that the parameter space for $\boldsymbol{\psi}$ is $\Theta_0 = (\mathbb{R}^p, \Xi, \mathbb{R}, \mathbb{R}^+)$ for the Fcr_0 model and $\Theta_{-1} = (\mathbb{R}^p, \Xi, \mathbb{R}_0^-, \mathbb{R}^+)$ for the Fcr_{-1} model, with $\mathbb{R}_0^- = (-\infty, 0]$. In [1], the estimators were obtained using a classical approach based on the maximum likelihood (ML) estimation procedure, specifically the expectation-maximization (ME) algorithm. However, for this particular problem, and denoting $\boldsymbol{D}_c$ as the complete data, the EM algorithm produces a complete log-likelihood, say $\ell_c(\widehat{\boldsymbol{\psi}}; \boldsymbol{D}_c)$, which cannot be written as $\ell_c(\boldsymbol{\psi}; \boldsymbol{D}_c) = \ell_c(\boldsymbol{\psi}_1; \boldsymbol{D}_c) + \ell_c(\boldsymbol{\psi}_2; \boldsymbol{D}_c)$, where $\boldsymbol{\psi}_1 \cup \boldsymbol{\psi}_2 = \boldsymbol{\psi}$, $\boldsymbol{\psi}_1 \cap \boldsymbol{\psi}_2 = \emptyset$, $\boldsymbol{\psi}_1 \neq \emptyset$, and $\boldsymbol{\psi}_2 \neq \emptyset$, except for the case corresponding to the Mcr model. Usually the application of the EM algorithm allows this scheme because it allows the maximizations of step M to be carried out independently for $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, allowing for a more robust estimation process. Therefore, the maximization problem of the observed log-likelihood and the complete log-likelihood has the same difficulty. We opt for maximizing the observed log-likelihood, i.e., the ML estimator of $\boldsymbol{\psi}$ is the value $\widehat{\boldsymbol{\psi}} = \left(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\xi}}^\top, \widehat{\gamma}, \widehat{\lambda}\right)^\top$ that maximizes $\ell(\boldsymbol{\psi})$. To obtain ML estimates of the model parameters, we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton method (See Mittelhammer et al. [15], p. 199) implemented in the subroutine `optim(·)` of the [13] software. The advantage of a reparametrized version of the model is that the values obtained for another model (for instance, the Mcr model) can be used as initial points for the maximization procedure. Under some regularity conditions [16] that are fulfilled for the proposed model whenever the parameters are in the interior of the parameter space, the asymptotic distribution of the ML estimator $\widehat{\boldsymbol{\psi}}$ follows asymptotically a multivariate normal distribution $\widehat{\boldsymbol{\psi}} \dot{\sim} \text{N}_{p+r+2}(\boldsymbol{\psi}, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi}))$, where r is the length of $\boldsymbol{\xi}$, $\dot{\sim}$ means 'approximately distributed', and $\boldsymbol{\Sigma}(\boldsymbol{\psi})$ is the Fisher information matrix, which is given by

$$\boldsymbol{\Sigma}(\boldsymbol{\psi}) = \mathbb{E}[-\partial \ell(\boldsymbol{\psi}) / \partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top],$$

which can be consistently estimated by the observed Fisher information matrix, given by

$$\boldsymbol{K}(\widehat{\boldsymbol{\psi}}) = -\partial \ell^2(\boldsymbol{\psi}) / \partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}.$$

3.2. On the particular cases of the Fcr_k model

For the Fcr_0 and Fcr_{-1} models, the Mcr, PTcr, and Gcr models appear as particular cases for $\gamma = -1$, $\gamma = 0$, and $\gamma = 1$, and LR and SW tests can be applied in a simple way, fitting the Fcr_k model ($k = 0, -1$) and the corresponding model (Mcr, PTcr, and Gcr model). Let $\widehat{\psi}^{(k)}$ and $\widehat{\psi}_{\gamma_0}^{(k)}$ be the maximum likelihood estimators for $\boldsymbol{\psi}$ under $\boldsymbol{\Theta}_k$ (the unrestricted parameter space) and under $\Theta_k^{(\gamma_0)} = (\mathbb{R}^p, \Xi, \mathbb{R}^+)$ (the restricted parameter space with $\gamma = \gamma_0$, $\gamma_0 = -1, 0, 1$ for Mcr, PTcr, and Gcr models, respectively). Therefore, statistics related to the LR and SW tests to decide among the hypotheses $H_0 : \gamma = \gamma_0$ and $H_1 : \gamma \neq \gamma_0$ are given by

$$\text{S}_{LR} = -2\left(\ell(\widehat{\boldsymbol{\psi}}^{(k)}) - \ell(\widehat{\boldsymbol{\psi}}_{\gamma_0}^{(k)})\right) \quad \text{and} \quad \text{S}_{SW} = (\widehat{\gamma}^{(k)} - \gamma_0) / \sqrt{\text{Var}(\widehat{\gamma}^{(k)})},$$

respectively. Under H_0, S_{LR} and S_{SW} follow asymptotically a $\chi^2_{(1)}$ distribution.
On the other hand, in the original parameterization discussed in Eq (2.2), for $\zeta = -e$, the population SF of the model is reduced to

$$S_{pop}(t) = \left(1 - F^\lambda(t; \boldsymbol{\xi})\right)^{\theta^*}, \tag{3.1}$$

where $\theta^* = \theta/\text{e}$. The author highlights this case because S_{pop} corresponds to a proper SF, i.e., for the ordinary case (without cured individuals in the population). However, we are very critical of this particular case mainly for two reasons:

- Usually, cure rate models are ad hoc tools for the case where the context of the problem suggests their use and not based on a statistical test. For example, in the cancer context, there is medical evidence that in some cancers, there is some treatment that allows patients to lead a life comparable to that of cancer-free patients, at least in longevity.
- The SF in Eq (3.1) corresponds to apply the Lehmann-type of distributions twice [17]. The first is applied with the baseline CDF $F(t;\boldsymbol{\xi})$ to obtain $F^{\lambda}(t;\boldsymbol{\xi})$. Then, it is applied for the survival function related to the last CDF, resulting in the form in Eq (3.1). This scheme should produce the identifiabilty problem. Without going further, for the Weibull distribution with SF $S(t;\boldsymbol{\xi}) = \exp(-\mathrm{e}^{\alpha_1}t^{\alpha_2})$, $\alpha_1 \in \mathbb{R}$ and $\alpha_2 > 0$, and $\lambda = 1$, the SF is reduced to

$$S_{pop}(t;\boldsymbol{\xi}) = S^{\theta^*}(t;\boldsymbol{\xi}) = \exp(-\mathrm{e}^{\alpha_1}\theta^* t^{\alpha_2}), \quad \alpha_1 \in \mathbb{R}, \alpha_2, \theta^* > 0,$$

  where it is clear that α_1 and θ^* are non-identifiable. We identify similar problems for the Fréchet and Gompertz distributions, among others.

Note that in the Fcr_k model ($k = 0$ or $k = -1$), the non-cure rate model case corresponds to $W_k\left(\mathrm{e}^{-1}\left(p^{-\gamma}-1\right)\right) = \mathrm{e}/\gamma^2$. Of course, in the presence of covariates, this cannot be tested. Therefore, Fcr_k does not inherit this property of the original model, but, as we argue in the last section, this is not a problem.

3.3. Identifiability

For the original model, [1] briefly discussed the identifiability issue of the model (which is present for all the cure rate models). The author discusses that the model is identifiable if at least one covariate with three or more levels is included as $\theta_i = \exp(\mathbf{x}_i^{\top}\boldsymbol{\beta})$. For the Fcr_k model, $k = 0$ or $k = -1$ is parameterized directly in terms of the cure rate, and the identifiability of the model follows in a simple way. In this case, in order to take into account covariates, we consider $p_i = p(\mathbf{x}_i,\boldsymbol{\beta}) = g(\mathbf{x}_i^{\top}\boldsymbol{\beta})$, where $g : \mathbb{R} \to (0,1)$ is a monotone function. Here, we consider $g(u) = \log(u/(1-u))$, i.e., the logit function; however, the result is valid for any monotone function g. For instance, for the usual probit, log-log and complementary log-log links. We also consider that a set of r covariates $\boldsymbol{x} \in \mathcal{X}$ (with at least one term different from the intercept) is available and introduced in p as mentioned previously. The population SF for the Fcr_k model can be written as

$$S_{pop}(t;\boldsymbol{\psi}) = p(\mathbf{x},\boldsymbol{\beta}) + (1 - p(\mathbf{x},\boldsymbol{\beta})) \times S(t;\xi^*),$$

where

$$S(t;\xi^*) = \frac{\left(1 + W_k\left(\mathrm{e}^{-1}\left([p(\mathbf{x},\boldsymbol{\beta})]^{-\gamma}-1\right)\right)\exp\left\{1 + W_k\left(\mathrm{e}^{-1}\left([p(\mathbf{x},\boldsymbol{\beta})]^{-\gamma}-1\right)\right)\right\}F(t;\boldsymbol{\xi})^{\lambda}\right)^{-1/\gamma}}{(1 - p(\mathbf{x},\boldsymbol{\beta}))},$$

i.e., the SF for the Fcr_k can be written as a mixture model with cure rate $p(x,\boldsymbol{\beta})$ and proper SF $S(t;\boldsymbol{\xi}^*)$. Therefore, as $\sup\{p(\mathbf{x},\boldsymbol{\beta}) : \mathbf{x} \in \mathcal{X}, \boldsymbol{\beta} \in \mathbb{R}^r\} = 1$, by Theorem 1, point 1 in [18], the model is identifiable.

4. Simulation studies

In this section, we present two simulation studies. The first is devoted to assessing the performance of the ML estimators in finite samples for the Fcr_0 and Fcr_{-1} regression models. The second assesses

the performance of the LR and SW tests for different values of γ to be tested in the Fcr_0 regression model. In all the cases, for the basal SF, we consider the Weibull with SF $S(t;\boldsymbol{\xi}) = \exp(-e^{\alpha_1}t^{\alpha_2})$, $\alpha_1 \in \mathbb{R}$ and $\alpha_2 > 0$.

4.1. Performance of the ML estimators

We consider a scheme similar to the second real-data application. We draw two covariates: x_1 (from the Bernoulli distribution with success probability 0.44) and x_2 (from the exponential distribution with rate 0.34). Such covariates were included through the cure rate as $p_i = \exp(\mathbf{x}_i^\top\boldsymbol{\beta})/[1+\exp(\mathbf{x}_i^\top\boldsymbol{\beta})]$, where $\mathbf{x}_i^\top = (1, x_{1i}, x_{2i})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. To draw failure times (say T_i^*) from the regression models Fcr_0 and Fcr_{-1}, we draw U_i from the standard uniform distribution and set $T_i^* = S_{pop}^{-1}(U_i;\boldsymbol{\psi})$, if $U_i \leq p_i$ (a time for susceptible individuals) and $T_i^* = \infty$ (related to cured individuals), while the censoring times were fixed at $C_i = 15$ (i.e., a type I censoring scheme). Therefore, the observed times and failure indicators are defined by $T_i = \min(T_i^*, C_i)$ and $\delta_i = I(T_i^*, C_i)$, respectively. The parameters were considered as $\boldsymbol{\beta}^\top = (1.4978, -1.1714, -0.1090)$, $(\alpha_1, \alpha_2) = (-13.7623, 6.4145)$, $\lambda = 0.4561$ and $\gamma = 6.5619$ for the Fcr_0 model, and $\boldsymbol{\beta}^\top = (1.8345, -1.5018, -0.2001)$, $(\alpha_1, \alpha_2) = (-1.4558, 1.0740)$, $\lambda = 1.9691$, and $\gamma = -0.0006$ for the Fcr_{-1} model. We also considered three sample sizes: 200, 400 and 1000. For each case, we draw 1000 replicates from the corresponding model, and compute the estimators based on the ML estimation procedure, and their corresponding estimated standard errors. Table 1 summarizes the results considering bias, the mean of the estimated standard errors (SE), and the root of the estimated mean squared error (RMSE). Note that the bias and RMSE terms are reduced for each parameter as the sample size increases. In addition, the SE and RMSE terms are closer, especially when the sample size increases, suggesting that the estimation of standard errors is acceptable.

Table 1. Summary of the recovery parameters of the ML estimation in the Fcr_0 and Fcr_{-1} models based on 1000 Monte Carlo replications.

		$n = 200$			$n = 400$			$n = 1000$		
model	parameter	bias	SE	RMSE	bias	SE	RMSE	bias	SE	RMSE
Fcr_0	β_0	0.058	0.247	0.174	0.044	0.179	0.132	0.019	0.109	0.076
	β_1	−0.064	0.277	0.186	−0.042	0.197	0.142	−0.024	0.122	0.090
	β_2	−0.005	0.034	0.025	−0.002	0.025	0.017	−0.002	0.015	0.010
	α_1	0.962	2.585	2.149	0.745	1.015	0.976	0.233	0.778	0.657
	α_2	−0.589	1.131	0.911	−0.420	0.881	0.775	−0.132	0.433	0.377
	λ	0.239	0.253	0.167	0.123	0.180	0.119	0.056	0.082	0.068
	γ	0.114	1.246	1.160	0.087	0.972	0.881	0.046	0.039	0.021
Fcr_{-1}	β_0	0.101	0.482	0.218	0.068	0.272	0.151	0.044	0.188	0.105
	β_1	−0.064	0.422	0.230	−0.043	0.300	0.153	−0.032	0.247	0.106
	β_2	−0.017	0.135	0.046	−0.012	0.061	0.027	−0.008	0.051	0.020
	α_1	−0.437	0.872	0.711	−0.241	0.598	0.412	−0.146	0.416	0.331
	α_2	−0.233	0.369	0.196	−0.061	0.244	0.152	−0.014	0.178	0.115
	λ	0.318	0.693	0.592	0.167	0.644	0.469	0.103	0.533	0.380
	γ	0.092	1.123	1.045	0.039	0.539	0.333	0.027	0.292	0.205

Table 2. Summary of LR and SW tests in the Fcr_0 model based on 1000 Monte Carlo replications.

			LR test							SW test						
significance	n	$\gamma_0 \downarrow$ / true $\gamma \rightarrow$	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5
5%	200	−1.5	0.068	0.093	0.125	0.231	0.329	0.424	0.551	0.066	0.041	0.045	0.141	0.296	0.441	0.589
		−1.0	0.083	0.071	0.072	0.142	0.234	0.315	0.407	0.105	0.043	0.040	0.084	0.208	0.331	0.489
		−0.5	0.137	0.098	0.076	0.100	0.153	0.233	0.320	0.171	0.088	0.036	0.068	0.133	0.254	0.388
		0.0	0.231	0.141	0.099	0.067	0.089	0.125	0.214	0.275	0.128	0.056	0.035	0.086	0.156	0.303
		0.5	0.322	0.247	0.150	0.084	0.072	0.094	0.133	0.367	0.211	0.093	0.045	0.043	0.117	0.196
		1.0	0.493	0.363	0.251	0.152	0.084	0.071	0.092	0.533	0.318	0.164	0.065	0.037	0.069	0.130
		1.5	0.615	0.500	0.403	0.225	0.143	0.084	0.077	0.636	0.455	0.261	0.105	0.037	0.053	0.081
	400	−1.5	0.069	0.079	0.166	0.284	0.469	0.624	0.738	0.086	0.049	0.115	0.268	0.462	0.669	0.781
		−1.0	0.091	0.064	0.103	0.170	0.285	0.433	0.615	0.115	0.063	0.064	0.137	0.287	0.490	0.681
		−0.5	0.202	0.092	0.056	0.100	0.163	0.326	0.456	0.254	0.104	0.055	0.083	0.171	0.380	0.531
		0.0	0.342	0.187	0.091	0.057	0.078	0.179	0.294	0.404	0.222	0.092	0.046	0.078	0.222	0.365
		0.5	0.498	0.320	0.193	0.098	0.068	0.077	0.160	0.546	0.345	0.187	0.077	0.060	0.104	0.238
		1.0	0.696	0.508	0.357	0.190	0.112	0.063	0.077	0.735	0.530	0.326	0.139	0.063	0.065	0.120
		1.5	0.809	0.719	0.516	0.372	0.172	0.119	0.067	0.839	0.704	0.490	0.288	0.118	0.068	0.074
	1000	−1.5	0.052	0.098	0.324	0.582	0.821	0.933	0.981	0.061	0.084	0.296	0.584	0.835	0.952	0.988
		−1.0	0.125	0.052	0.108	0.320	0.587	0.817	0.935	0.175	0.058	0.097	0.331	0.628	0.856	0.954
		−0.5	0.300	0.124	0.055	0.120	0.349	0.584	0.804	0.365	0.157	0.058	0.119	0.379	0.637	0.856
		0.0	0.588	0.322	0.130	0.059	0.122	0.313	0.568	0.638	0.369	0.147	0.067	0.147	0.376	0.637
		0.5	0.844	0.645	0.348	0.143	0.055	0.119	0.306	0.874	0.662	0.363	0.143	0.056	0.155	0.378
		1.0	0.954	0.870	0.663	0.355	0.140	0.055	0.099	0.961	0.884	0.664	0.328	0.120	0.065	0.139
		1.5	0.992	0.970	0.884	0.670	0.380	0.123	0.039	0.991	0.965	0.883	0.659	0.328	0.097	0.047
10%	200	−1.5	0.132	0.157	0.189	0.343	0.450	0.545	0.669	0.124	0.096	0.130	0.267	0.443	0.570	0.715
		−1.0	0.143	0.133	0.134	0.233	0.340	0.434	0.527	0.179	0.104	0.097	0.191	0.325	0.492	0.611
		−0.5	0.208	0.167	0.156	0.170	0.232	0.344	0.441	0.260	0.162	0.099	0.137	0.222	0.366	0.521
		0.0	0.343	0.231	0.173	0.133	0.162	0.219	0.303	0.394	0.238	0.138	0.097	0.155	0.240	0.381
		0.5	0.450	0.346	0.233	0.153	0.140	0.160	0.227	0.527	0.338	0.186	0.109	0.106	0.193	0.289
		1.0	0.622	0.503	0.367	0.258	0.139	0.129	0.162	0.663	0.491	0.297	0.170	0.098	0.133	0.203
		1.5	0.718	0.624	0.514	0.332	0.224	0.155	0.138	0.752	0.599	0.446	0.229	0.116	0.105	0.132
	400	−1.5	0.130	0.137	0.254	0.411	0.590	0.743	0.823	0.133	0.107	0.227	0.402	0.604	0.775	0.854
		−1.0	0.162	0.119	0.160	0.257	0.393	0.572	0.721	0.189	0.122	0.145	0.232	0.412	0.615	0.766
		−0.5	0.298	0.163	0.107	0.166	0.250	0.445	0.563	0.342	0.188	0.108	0.166	0.262	0.491	0.620
		0.0	0.457	0.297	0.152	0.118	0.137	0.267	0.393	0.495	0.333	0.150	0.106	0.152	0.318	0.469
		0.5	0.611	0.455	0.286	0.165	0.126	0.136	0.265	0.653	0.490	0.283	0.136	0.121	0.166	0.324
		1.0	0.795	0.643	0.473	0.288	0.166	0.128	0.137	0.814	0.659	0.476	0.255	0.130	0.117	0.187
		1.5	0.880	0.800	0.644	0.491	0.271	0.188	0.122	0.896	0.806	0.626	0.442	0.218	0.150	0.131
	1000	−1.5	0.100	0.184	0.432	0.692	0.902	0.963	0.994	0.117	0.175	0.427	0.704	0.918	0.966	0.992
		−1.0	0.208	0.107	0.180	0.434	0.711	0.885	0.965	0.250	0.119	0.184	0.457	0.749	0.908	0.976
		−0.5	0.406	0.201	0.119	0.194	0.459	0.684	0.873	0.458	0.234	0.130	0.199	0.492	0.732	0.906
		0.0	0.701	0.459	0.201	0.115	0.203	0.423	0.686	0.726	0.489	0.218	0.123	0.228	0.464	0.744
		0.5	0.905	0.738	0.480	0.228	0.109	0.205	0.419	0.920	0.759	0.484	0.237	0.115	0.234	0.486
		1.0	0.976	0.920	0.761	0.479	0.208	0.107	0.174	0.978	0.919	0.767	0.466	0.196	0.121	0.222
		1.5	0.996	0.987	0.939	0.780	0.498	0.198	0.087	0.995	0.984	0.931	0.768	0.465	0.181	0.094

4.2. Assessing the LR and SW tests

We also performed a second simulation study to assess the performance of the LR and SW tests to decide between the hypotheses $H_0 : \gamma = \gamma_0$ and $H_1 : \gamma \neq \gamma_0$. In this case, we focus on the Fcr_0 regression model. We considered the same set of parameters as in the previous study. However, we consider the true γ and the tested γ_0 in the set $\{-1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5\}$. For each combination of sample size, true γ, and γ_0, we draw 1000 replicates and compute the LR and SW statistics, along with their p-values, based on the asymptotic $\chi^2_{(1)}$ distribution. Table 2 presents the percentage of times in which the corresponding null hypothesis is rejected with two significance levels: 5% and 10% of significance. As expected, when the true γ is equal to γ_0, these percentages are closer to 0.05 and

0.10 for both cases, whereas these percentages should increase when the difference between the true γ and γ_0 increases. The results suggest that both tests are suitable for deciding among the referred hypotheses.

5. Applications to melanoma datasets

In this section, we present two real-data applications for the Fcr_0 and Fcr_{-1} models. For comparative purposes, we also present the results for the Mcr, Gcr, and PTcr models, which were also parameterized directly in the cure rate term. These models were compared using the AIC [19] and BIC [20] criteria.

5.1. Melanoma data set 1

This dataset was analyzed in [1], but using the FCR without our proposed reparameterization. The information pertains to 416 patients with cutaneous melanoma, among whom 54.6% of patients were censored. The unique covariate included in the analysis was the nodal category, indicating the lymph nodes involved in the disease, with values 0, 1, 2, or 3 (3 indicates three or more positive nodes). For this reason, we consider $p_i = (1 + \exp(\beta_0 + \beta_1 \times \text{nodal category}_i))^{-1}$, $i = 1, \ldots, 416$. Table 3 shows the results for the five cure rate models. As expected, the regression coefficients β_0 and β_1 are much closer for all models. For this particular data set, the log-likelihood values for Fcr_0 and Fcr_{-1} are very close to the ones obtained by the Gcr and PTcr models, respectively. This suggests that $\gamma = 1$ and $\gamma = 0$ in Fcr_0 and Fcr_{-1}, respectively, which is confirmed by the corresponding hypothesis tests in Table 4. Note that, based on the results for the Fcr_0 model, the estimated cure rates are 0.63, 0.53, 0.41, and 0.31 for patients with nodules 0, 1, 2, and 3, respectively. In other words, as $\exp(\beta_1) \approx 0.64$ (according to the Fcr_0 model), the odds ratio is decreased in 36% for each stage of the nodal category that the disease becomes worse.

Table 3. Estimates and standard error (s.e.) for different cure rate models parameterized in the cure rate term in the melanoma data set 1.

	Fcr_0		Fcr_{-1}		Mcr		Gcr		PTcr	
parameters	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.
β_0	0.5349	0.1819	0.6346	0.1831	0.7942	0.1842	0.6350	0.1573	0.6350	0.1828
β_1	−0.4396	0.0990	−0.5278	0.1026	−0.4253	0.1027	−0.5279	0.0865	−0.5279	0.1021
α_1	−0.8316	0.8553	0.2375	0.5192	0.3451	0.3721	0.2324	0.0905	0.2324	0.3905
α_2	1.2462	0.5649	0.7339	0.3011	0.6893	0.2502	0.9175	0.0689	0.7374	0.2697
λ	1.1847	0.6182	2.0353	0.3011	2.1986	1.3572	1.5616	0.1537	2.0195	1.2026
γ	2.4249	1.2310	−0.0010	0.8977	-		-		-	
log-likelihood	−408.20		−411.22		−462.28		−409.09		−411.22	
AIC	828.40		834.44		934.56		828.18		832.44	
BIC	852.47		858.50		954.62		848.23		852.49	

Table 4. Hypothesis tests for different cases in the Fcr_0 and Fcr_{-1} models in melanoma data set 1.

Model	Model to be tested in H_0	Hypothesis	LR test statistic	LR test p-value	SW test statistic	SW test p-value
Fcr_0	Gcr	$H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$	1.158	0.282	1.779	0.075
	PTcr	$H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$	6.036	0.014	1.970	0.049
Fcr_{-1}	Mcr	$H_0 : \gamma = -1$ versus $H_1 : \gamma \neq -1$	102.120	<0.001	1.113	0.266
	PTcr	$H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$	<0.000	>0.999	0.000	>0.999

Table 5. Estimates and standard error (s.e.) for different cure rate models parameterized in the cure rate term in melanoma data set 2.

	Fcr_0		Fcr_{-1}		Mcr		Gcr		PTcr	
parameters	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.
β_0	1.4978	0.3060	1.8345	0.4617	1.8802	0.4266	1.8021	0.4194	1.8473	0.4244
β_1	−1.1714	0.3071	−1.5018	0.4097	−1.3620	0.4729	−1.4806	0.3570	−1.4973	0.4027
β_2	−0.1090	0.0342	−0.2001	0.0762	−0.1862	0.1246	−0.1784	0.0538	−0.1989	0.0740
α_1	−13.7623	3.0936	−1.4558	3.8693	−0.9618	3.3729	−3.2227	5.2449	−1.6886	3.3856
α_2	6.4145	1.4524	1.0740	1.6539	0.8998	1.4225	1.7906	2.4037	1.1739	1.4915
λ	0.4561	0.1502	1.9691	4.4481	2.5609	6.4886	1.0867	1.8329	1.7250	3.1634
γ	6.5619	2.8931	−0.0006	0.9346	-		-		-	
log-likelihood	−201.68		−207.45		−221.08		−205.38		−207.45	
AIC	417.37		428.90		454.15		422.75		426.90	
BIC	440.63		452.16		474.09		442.69		446.84	

5.2. Melanoma data set 2

We reviewed the classical melanoma data set provided in the timereg [21] package of [13], which contains data related to the survival of patients after operation for malignant melanoma collected at Odense University Hospital. The information pertains to 205 patients with melanoma, among whom 54.6% were censored. The data consider two covariates:, ulceration (presence or absence of ulceration, 115 yes and 90 no) and thickness (tumor thickness, in cm; mean and standard deviation 2.92 and 2.96, respectively), included in the analysis as $p_i = (1 + \exp(\beta_0 + \beta_1 \times \text{ulceration}_i + \beta_2 \times \text{thickness}_i))^{-1}$, $i = 1, \ldots, 205$. In this case, Table 5, the regression coefficients are not so close, but they are consistent (all have the same sign). In addition, the value for the log-likelihood function of Fcr_0 is considerably lower than the other values. This suggests that Fcr_0 is preferable to the Mcr, Gcr, and PTcr models. This is confirmed by the hypothesis tests in Table 6. Figure 2 presents the survival function (SF) for two different profiles based on the Fcr_0, Gcr, and PTcr models. Note that the Gcr and PTcr models yield more similar results, as their estimates, particularly the regression coefficients associated with the cure rate, are closely aligned. However, for longer time horizons (e.g., greater than 10 years), both models tend to underestimate the survival function for individuals without ulceration and overestimate it for those with ulceration, when compared to the Fcr_0 model. Finally, Figure 3 presents the estimated

cure rates for patients, depending on their ulceration status and tumor thickness. For instance, patients with a tumor thickness of 4 cm have a cure rate estimated at 0.74 and 0.47 for ulceration absent and present, respectively. Seen from another point of view, according to the Fcr_0, $\exp(-\beta_1) \approx 0.3$ and $\exp(-\beta_2) \approx 0.9$. Therefore, the odds ratio for patients with ulceration is decreased by 70% compared to patients without ulceration, whereas the odds ratio decreases by 10% for each millimeter increase in tumor thickness.

Table 6. Hypothesis tests for different cases in the Fcr_0 and Fcr_{-1} models in melanoma data set 2.

Model	Model to be tested in H_0	Hypothesis	LRT statistic	LRT p-value	Wald statistic	Wald p-value
Fcr_0	Gcr	$H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$	7.385	0.007	−3.621	<0.001
	PTcr	$H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$	11.531	0.001	3.037	0.002
Fcr_{-1}	Mcr	$H_0 : \gamma = -1$ versus $H_1 : \gamma \neq -1$	27.256	<0.001	0.668	0.504
	PTcr	$H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$	0.005	0.946	0.443	0.658

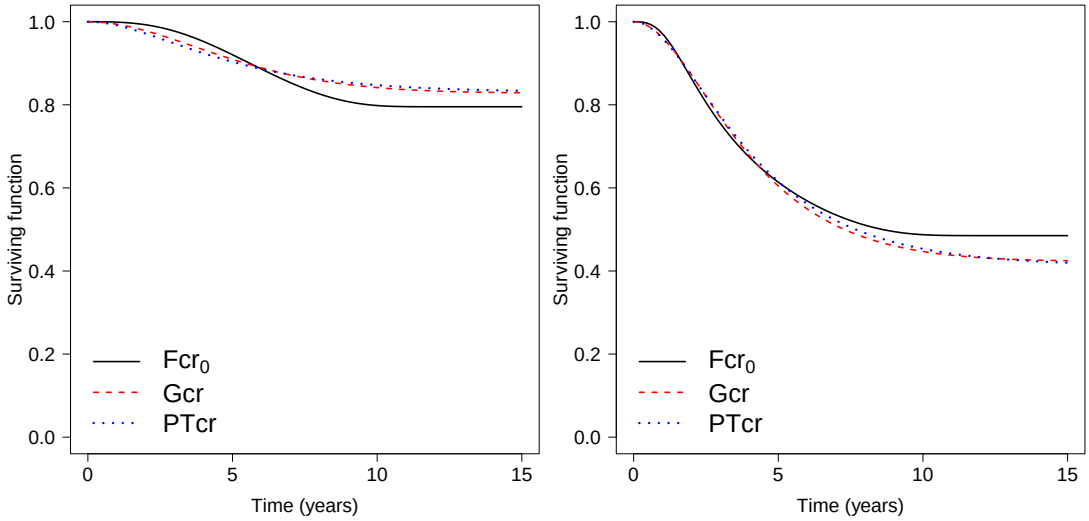

Figure 2. Surviving function under the Fcr_0, Gcr, and PTcr models for different profiles: ulceration status absent and tumor thickness 1.29 mm (left panel) and ulceration status present and tumor thickness 3.54 mm.

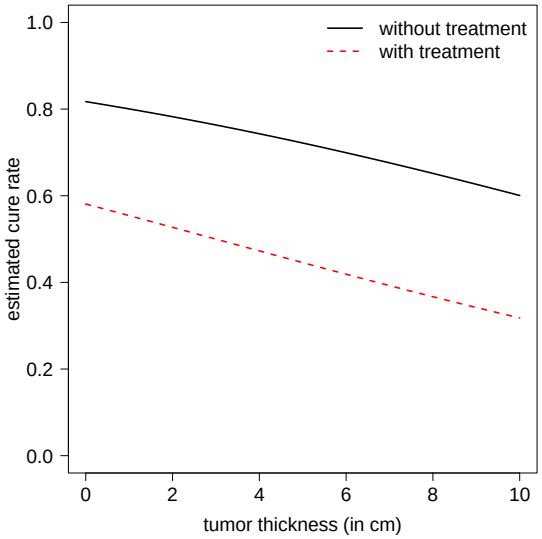

Figure 3. Estimated cure rate for different profiles of patients in melanoma data set 2.

6. Discussion

In this paper, we emphasize the importance of reparameterizing cure rate models to enhance interpretability and identifiability, particularly when incorporating covariates. The model proposed by [1] provides a flexible framework, but lacks direct interpretability of parameters and presents identifiability issues in practical applications. First, the parameter θ does not have any functional interpretation except for the PTcr model when $\gamma \to 0$. As a result, it is challenging to derive meaningful interpretations of the regression coefficients. This lack of interpretability makes it difficult to compare these coefficients with those of other cure models discussed in the literature. Second, furthermore, for the original model [1], the author discusses that the model defined in (2.2) is identifiable if at least one covariate with three or more levels is included as $\theta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$, which is also a disadvantage.

We have highlighted the critical importance of reparameterizing in the context of cure rate models. Through our proposed reparameterization, we have demonstrated how careful consideration of parameter significance can improve understanding of clinical outcomes, particularly in oncology. By reparameterizing the model in terms of the cure rate using the Lambert W function, we obtain models that retain flexibility while allowing for a meaningful interpretation of the regression coefficients. For the Fcr_k model, $k = 0$ or $k = -1$ is parameterized directly in terms of the cure rate, and the identifiability of the model follows in a simple way. In this case, in order to take into account covariates, we consider $p_i = p(\mathbf{x}_i, \boldsymbol{\beta}) = g(\mathbf{x}_i^\top \boldsymbol{\beta})$, where $g : \mathbb{R} \to (0, 1)$ is a monotone function. Here, we consider $g(u) = \log(u/(1-u))$, i.e., the logit function; however, the result is valid for any monotone function g. For instance, for the usual probit, log-log, and complementary log-log links.

The findings suggest that a more interpretable parameterization not only facilitates more transparent communication of results to a medical audience, but also facilitates the practical application of these models in clinical settings. By linking model parameters to clinically relevant quantities, such as the fraction of patients who have been cured, practitioners can make more informed decisions about treatment efficacy. Furthermore, our simulation studies and real data

applications support the assertion that well-defined parameters lead to improved model performance and inference. We encourage researchers to prioritize interpretability in their modeling efforts, which can significantly enhance the utility of statistical models in addressing real-world medical challenges.

Author contributions

Diego I. Gallardo: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft; Yolanda M. Gómez: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft; Marcelo Bourguignon: Conceptualization, Formal analysis, Methodology, Software; Héctor J. Gómez: Funding acquisition, Software, Writing – review & editing. All authors have read and approved the final version of the manuscript for publication

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was partially supported by a grant from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) 11230397.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. F. S. Milienos, On a reparameterization of a flexible family of cure models, *Stat. Med.*, **41** (2022), 4091–4111. https://doi.org/10.1002/sim.9498
2. J. W. Boag, Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy, *J. R. Stat. Soc. B*, **11** (1949), 15–44. https://doi.org/10.1111/j.2517-6161.1949.tb00020.x
3. A. Y. Yakovlev, A. D. Tsodikov, *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, Singapore: World Scientific, 1996. https://doi.org/10.1142/2420
4. A. D. Tsodikov, J. G. Ibrahim, A. Y. Yakovlev, Estimating Cure Rates from Survival Data: An Alternative to Two-Component Mixture Models, *J. Am. Stat. Assoc.* **98** (2003), 1063–1078. https://doi.org/10.1198/016214503000000001007
5. J. Rodrigues , M. De Castro, V. G. Cancho, N. Balakrishnan, COM–Poisson cure rate survival models and an application to a cutaneous melanoma data, *J. Stat. Plan. Infer.*, **139** (2009), 3605–3611. https://doi.org/10.1016/j.jspi.2009.04.014
6. V. G. Cancho, F. Louzada, E. M. Ortega , The power series cure rate model: An application to a cutaneous melanoma data, *Commun. Stat. Simul. Comput.*, **42** (2013), 586–602. https://doi.org/10.1080/03610918.2011.639971

7. D. I. Gallardo, J. S. Romeo, R. Meyer, A simplified estimation procedure based on the EM algorithm for the power series cure rate model, *Commun. Stat. Simul. Comput.*, **46** (2017), 6342–6359. https://doi.org/10.1080/03610918.2016.1202276
8. D. I. Gallardo, Y. M. Gómez, M. De Castro, A flexible cure rate model based on the polylogarithm distribution, *J. Stat. Comput. Simul.*, **88** (2018), 2137–2149. https://doi.org/10.1080/00949655.2018.1451850
9. J. Rodrigues, V. G. Cancho, M. De Castro, F. Louzada-Neto, On the unification of long-term survival models, *Stat. Probab. Lett.*, **79** (2009), 753–759. https://doi.org/10.1016/j.spl.2008.10.029
10. R. Corless, G. Gonnet, D. Hare, D. Jeffrey, D. Knuth, On the Lambert W function, *Adv. Comput. Math.*, **5** (1996), 329–359. https://doi.org/10.1007/BF02124750
11. Avraham Adler, lamW: Lambert-W Function, R package version 2.1.0, 2015. Available from: https://cran.r-project.org/web/packages/lamW/.
12. H. W. Borchers, pracma: Practical Numerical Math Functions, R package version 2.4.2, 2022. Available from: https://CRAN.R-project.org/package=pracma.
13. The R Foundation, R: A Language and Environment for Statistical Computing, 2025. Available from: https://www.R-project.org/.
14. M. De Castro, V. G. Cancho, J. Rodrigues, A Bayesian Long-term Survival Model Parametrized in the Cured Fraction, *Biometrical J.*, **51** (2009), 443–455. https://doi.org/10.1002/bimj.200800199
15. R. C. Mittelhammer, G. G. Judge, D. J. Miller, *Econometric Foundations*, New York: Cambridge University Press, 2000.
16. D. R. Cox, C. V. Hinkley, *Theoretical Statistics*, New York: Chapman and Hall/CRC, 1974. https://doi.org/10.1201/b14832
17. E. L. Lehmann, The Power of Rank Tests, *Ann. Math. Stat.*, **24** (1953), 23–43. https://doi.org/10.1214/aoms/1177729080
18. L. Hanin, L. Huang, Identifiability of cure models revisited, *J. Multivar. Anal.*, **130** (2014), 261–274. https://doi.org/10.1016/j.jmva.2014.06.002
19. H. Akaike, A new look at the statistical model identification, *IEEE T. Automat. Contr.*, **19** (1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705
20. G. Schwarz, Estimating the dimension of a model, *Ann. Stat.*, **6** (1978), 461–464.
21. T. H. Scheike, M. Zhang, Analyzing Competing Risk Data Using the R timereg Package, *J. Stat. Software*, **38** (2011), 1–15. https://doi.org/10.18637/jss.v038.i02

AIMS Press