Received: 14 August 2018 | Revised: 7 January 2019 | Accepted: 4 February 2019

DOI: 10.1002/bimj.201800030

RESEARCH PAPER

Biometrical Journal

Mixture and nonmixture cure fraction models assuming discrete lifetimes: Application to a pelvic sarcoma dataset

Ricardo Puziol de Oliveira[1] | André F. B. Menezes[2] | Josmar Mazucheli[2] | Jorge A. Achcar[1]

[1]Medical School, Universidade de São Paulo, Ribeirão Preto, SP, Brasil

[2]Department of Statistics, Universidade Estadual de Maringá, Paraná, PR, Brasil

Correspondence
Ricardo Puziol de Oliveira, Medical School, Universidade de São Paulo, Ribeirão Preto, SP, Brasil.
Email: rpuziol.oliveira@gmail.com

Abstract

Different cure fraction models have been used in the analysis of lifetime data in presence of cured patients. This paper considers mixture and nonmixture models based on discrete Weibull distribution to model recurrent event data in presence of a cure fraction. The novelty of this study is the use of a discrete lifetime distribution in place of usual existing continuous lifetime distributions for lifetime data in presence of cured fraction, censored data, and covariates. In the verification of the fit of the proposed model it is proposed the use of randomized quantile residuals. An extensive simulation study is considered to evaluate the properties of the estimates of the parameters related to the proposed model. As an illustration of the proposed methodology, it is considered an application considering a medical dataset related to lifetimes in a retrospective cohort study conducted by Puchner et al. (2017) that consists of 147 consecutive cases with surgical treatment of a sarcoma of the pelvis between the years of 1980 and 2012.

KEYWORDS

cure fraction, discrete Weibull distribution, long-term survivors, pelvic sarcomas, survival analysis

1 | INTRODUCTION

In many medical studies, an issue of great interest in medical research is the estimation of the fraction of patients in the studied population who never experience the event of interest. These patients are not at risk with respect to the event of interest and are considered immune, cured, nonsusceptible, or extremely long-term survivors. Standard survival analysis techniques, as for example, the Cox proportional hazards (Cox, 1972) model, provide no direct estimation for the cure fraction that is a motivation for the use of mixture and nonmixture cure fraction models.

According to Vahidpour (2016), in the literature it is presented different models to be fitted by data in presence of cure fraction with great emphasis on the mixture cure fraction models, also known as standard cure fraction models (see, e.g., De Angelis, Capocaccia, Hakulinen, Soderman, & Verdecchia, 1999; Lambert, Thompson, Weston, & Dickman, 2006), which have been widely used for modeling survival data in presence of cure fraction and the nonmixture cure fraction models that are not very popular (see Achcar, Coelho-Barros, & Mazucheli, 2012; Tsodikov, Ibrahim, & Yakovlev, 2003; Vahidpour, 2016). Different approaches have been presented in the literature to model cure fraction for univariate lifetime data: (see, e.g., Achcar et al., 2012; Cancho & Bolfarine, 2001; De Angelis et al., 1999; Farewell, 1982; Lambert et al., 2006; Lu, 2010; Othus et al., 2012; Price & Manatunga, 2001; Yin & Ibrahim, 2005; Yu, Tiwari, Cronin, & Feuer, 2004, among many other studies).

Let us denote by T a positive random variable related to the the time until the event. Following Maller and Zhou (1996), the standard fraction model (or mixture cure fraction model) assumes that the probability of the time-to-event to be greater than a specified time t (the survival function) is given by

$$S(t) = \rho + (1 - \rho)S_0(t), \tag{1}$$

where $\rho \in (0, 1)$ is the mixing parameter that represents the proportion of "long-term survivors," "nonsusceptible," or "cured patients," and $S_0(t)$ denotes a proper survival function for the noncured or susceptible group in the population. Observe that if $t \to \infty$, then $S(t) \to \rho$, that is, the survival function has an asymptote at the cure fraction ρ. The probability density and the hazard functions corresponding to (1) are given, respectively, by

$$f(t) = (1 - \rho)f_0(t) \quad \text{and} \quad h(t) = \frac{(1 - \rho)f_0(t)}{\rho + (1 - \rho)S_0(t)}. \tag{2}$$

As an alternative for the mixture cure fraction model, a nonmixture model has been proposed in the literature that defines an asymptote for the cumulative hazard and thus for the cure fraction (see Tsodikov et al., 2003). In this case, the survival function for the nonmixture cure fraction model is given by

$$S(t) = \rho^{F_0(t)} = \exp\{\ln(\rho)F_0(t)\}, \tag{3}$$

where $\rho \in (0, 1)$ is the probability of cured patients and $F_0(t) = 1 - S_0(t)$ denotes a proper distribution function for the noncured or susceptible group in the population. The probability density and the hazard functions corresponding to (3) are given, respectively, by

$$f(t) = -\ln(\rho)f_0(t)\exp\{\ln(\rho)F_0(t)\} \quad \text{and} \quad h(t) = -\ln(\rho)f_0(t). \tag{4}$$

The main goal of this paper is to introduce discrete Weibull cure fraction models to investigate long-term lifetimes and risk factors with a special application related to the lifetimes of patients receiving a treatment for pelvic sarcomas. The main reason for the use of the discrete Weibull distribution, is that similarly to the continuous case of the Weibull probability distribution, possibly the most used lifetime distribution in medical or engineering applications due to the great flexibility of fit, the discrete case of the Weibull distribution is also a good option for the modeling of lifetime discrete data. Besides the great flexibility of fit, this model only has two parameters that implies in great simplicity to get the inferences of interest. It is important to point out that other flexible parametric distributions like the Burr XII or the Lindley also could be assumed following the same approach introduced in this paper. According to Sugarbaker (2001), an important feature characteristic of the natural history of retroperitoneal and pelvic sarcomas is how it differs significantly from the more common abdominal and pelvic adenocarcinomas and from visceral sarcoma. These differences are important when planning a treatment since the treatment of pelvic sarcomas are surgically difficult due to the anatomic proximity of the pelvis to many neurovascular structures and the urinary and intestinal tracts, with poor oncological outcomes and high complication rates.

In this way, it is considered a retrospective cohort study conducted by Puchner et al. (2017) that consists of 147 consecutive cases with surgical treatment of a sarcoma of the pelvis between the years of 1980 and 2012. The records included 68 males (46%) and 79 females (54%) with an average age of 38 ± 20 years at time of surgery. The diagnosis was based on conclusive clinical and imaging findings and it was always confirmed by biopsy and histological analysis. Some prognostic and oncological factors were also reported associated to the lifetimes of the patients: patient gender, tumor grade, radiotherapy, chemotherapy, among many others.

The paper is organized as follows: in Section 2 it is presented the proposed discrete Weibull models using the standard mixture and the nonmixture cure fraction models. The inference methods and residuals procedures are introduced in Section 3. Section 4 presents a simulation study carried out to evaluate the model properties considering a hypothetic cure fraction. Finally, Section 5 presents the analysis of pelvic sarcoma under the proposed methodology and Section 6 closes the paper with some concluding remarks and directions for future research.

2 | DISCRETE WEIBULL CURE FRACTION MODELS

In this section, we assume that the distribution of the lifetimes for the susceptible populations follows a discrete Weibull (DW) distribution introduced by Nakagawa and Osaki (1975), which can be considered as a discrete analog of the continuous Weibull distribution. The probability mass function (p.m.f.) of a DW distribution is defined by

$$\Pr(T = t \mid \phi, \beta) = \phi^{t^\beta} - \phi^{(t+1)^\beta}, \qquad t \in \mathbb{N}_0 = \{0, 1, 2, \ldots\} \tag{5}$$

and its corresponding survival function is given by

$$S(t \mid \phi, \beta) = \Pr(T > t \mid \phi, \beta) = \phi^{(t+1)^{\beta}}, \tag{6}$$

where $\beta > 0$ and $0 < \phi < 1$. Note that, when $\beta = 1$, the UW distribution reduces to the geometric distribution and when $\beta = 2$, it reduces to the Rayleigh distribution introduced by Roy (2004). This model has been applied to many areas, including competing risks, extreme values, failure times, regional analyses of precipitation, and reliability (see, e.g., Almalki & Nadarajah, 2014; Brunello & Nakano, 2015; Englehardt & Li, 2011; Khan, Khalique, & Abouammoh, 1989; Kulasekera, 1994; Murthy, Xie, & Jiang, 2004; Roy, 2002).

From Klakattawi et al. (2018), the τth ($0 < \tau < 1$) quantile of DW distribution is given by

$$M = \left\lfloor \left(\frac{\log(1-\tau)}{\log(\phi)} \right)^{\frac{1}{\beta}} - 1 \right\rfloor, \qquad \tau \geq 1 - \phi, \tag{7}$$

where $\lfloor x \rfloor$ denotes the integer less than or equal to x.

Remark 1. If the random variable X follows a Weibull distribution with scale μ and shape β parameters, denoted by $\mathrm{W}(\beta, \mu)$. Thus for $\phi = \mathrm{e}^{-\frac{1}{\mu}}$ we have that

$$T = \lfloor X \rfloor . \tag{8}$$

From (8) we can easily generate random samples from a $\mathrm{DW}(\beta, \phi)$. More specifically, we first generate a random sample X from a $\mathrm{W}(\beta, \mu)$ distribution, and then by considering $T = \lfloor X \rfloor$, we obtain a generated sample from $\mathrm{DW}(\beta, \phi)$.

Using the proposed methodology, the mixture and nonmixture cure fraction model for the lifetime T assuming the DW distribution are given, respectively, by

$$S(t \mid \phi, \beta, \rho) = \Pr(T > t \mid \phi, \beta, \rho) = \rho + (1-\rho)\, \phi^{(t+1)^{\beta}}, \tag{9}$$

and

$$S(t \mid \phi, \beta, \rho) = \Pr(T > t \mid \phi, \beta, \rho) = \exp\left\{ \ln(\rho) \left[1 - \phi^{(t+1)^{\beta}} \right] \right\}, \tag{10}$$

where $\rho \in (0, 1)$ is the cure fraction parameter.

Remark 2. Since the DW model has no closed form for the expected value and variance, the subsequently mixture and nonmixture cure fraction models also have no closed form for their expected values and variances. However, the expected values and the variances could be obtained using numerical methods directly from the definition of the r-th moment given by

$$\mathbb{E}\left(T^{r}\right) = \sum_{k=0}^{\infty} k^{r} \left\{ \phi^{\log\left[1+\left(\frac{k}{\theta}\right)^{\alpha}\right]} - \phi^{\log\left[1+\left(\frac{k+1}{\theta}\right)^{\alpha}\right]} \right\},$$

where, in particular, for $r = 1$, we have $\mathbb{E}(T) = \sum_{k=1}^{\infty} \phi^{\log[1+(\frac{k}{\theta})^{\alpha}]}$ and for $r = 2$ we have $\mathbb{E}(T^2) = \sum_{k=1}^{\infty} (2k-1) \phi^{\log[1+(\frac{k}{\theta})^{\alpha}]}$.

3 | INFERENCE AND RESIDUALS

3.1 | Maximum likelihood method

Let us consider the situation when the lifetime, T_i, is not completely observed and may be subject to right censoring. Let C_i be the censoring time for the ith individual. From a sample of size n , we observe $T_i = \min\{T_i, C_i\}$ and $\delta_i = I(T_i < C_i)$, where $\delta_i = 1$ if T_i is an observed lifetime and $\delta_i = 0$ if it is right censored lifetime. In this case, the log-likelihood function considering the DW distribution with p.m.f. defined in (5), can be written as

$$\ell(\beta, \phi \mid \boldsymbol{t}, \boldsymbol{\delta}) = \sum_{i=1}^{n} \delta_i \, \log\left[\phi^{t^{\beta}} - \phi^{(t+1)^{\beta}} \right] + \sum_{i=1}^{n} (1 - \delta_i) \, \log\left[\phi^{(t+1)^{\beta}} \right], \tag{11}$$

where $\boldsymbol{t} = (t_1, \dots, t_n)^{\top}$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^{\top}$.

Assuming the DW mixture model the log-likelihood function can be expressed as

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{t}, \boldsymbol{\delta}) = r \,\ln(1-\rho) + \sum_{i=1}^{n} \delta_i \,\ln\left[\phi^{t_i^{\beta}} - \phi^{(t_i+1)^{\beta}}\right] + \sum_{i=1}^{n} (1-\delta_i)\,\ln\left[\rho + (1-\rho)\,\phi^{(t_i+1)^{\beta}}\right], \tag{12}$$

where $\boldsymbol{\theta} = (\beta, \phi, \rho)^{\top}$ and $r = \sum_{i=1}^{n} \delta_i$ is the number of uncensored observations.

Additionally, considering the discrete Weibull nonmixture model, the likelihood can be written as

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{t}, \boldsymbol{\delta}) = r \,\ln(-\ln\rho) + \sum_{i=1}^{n} \delta_i \,\ln\left[\phi^{t_i^{\beta}} - \phi^{(t_i+1)^{\beta}}\right] + (\ln\rho) \sum_{i=1}^{n} 1 - \phi^{(t_i+1)^{\beta}}. \tag{13}$$

The maximum likelihood estimates (MLEs) $\widehat{\boldsymbol{\theta}}$ for the unknown parameters in the vector parameter $\boldsymbol{\theta}$ are obtained by maximizing the log-likelihood functions defined in Equations (11), (12), and (13) using standard optimization methods, such as Newton-Raphson and quasi-Newton. In this study, the MLEs were obtained by the quasi-Newton method available in the `SAS/NLMIXED` procedure (SAS, 2010). Under suitable regularity conditions (see, Lehmann & Casella, 1998, pp. 461–463), the asymptotic distribution of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ is a multivariate Normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$, which can be consistently estimated by the inverse of the observed Fisher information matrix given by

$$\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\theta}}\right) = \left[-\frac{\partial\, \ell(\boldsymbol{\theta} \mid \boldsymbol{t}, \boldsymbol{\delta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^{\top}}\right]^{-1} \tag{14}$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. The required second derivatives are computed numerically using the `SAS/NLMIXED` procedure.

We further propose to relate the parameters ϕ and ρ of the mixture and non-mixture models to a vector of explanatory variables $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, respectively. Thus, we assume the following link functions

$$\log\left(-\log(\phi_i)\right) = \boldsymbol{x}_i^{\top}\,\boldsymbol{\alpha} \qquad \text{and} \qquad \log\left(\frac{\rho_i}{1-\rho_i}\right) = \boldsymbol{z}_i^{\top}\,\boldsymbol{\delta}, \tag{15}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ denote the vectors of unknown parameters. It is noteworthy, that the log-log link function in ϕ is motivated by the analytical formula for the quantile function (7), which facilitates the interpretation of the coefficients. According to Klakattawi et al. (2018) the regression parameters $\boldsymbol{\alpha}$ can be interpreted in relation to the logarithm of the median.

To discriminate the proposed models we can consider the Akaike Information Criterion (AIC) (Akaike, 1974) that is given by AIC $= -2\,\ell(\widehat{\boldsymbol{\theta}}) + 2\,p$, where p is the number of model parameters. Among all considered models those one with the smallest value is commonly take as the favored model to describe the data (Rohde, 2014).

3.2 | Randomized quantile residuals

Given that the response is discrete, in the evaluation and study of departures from the model assumptions we propose the use of the randomized quantile residuals introduced by Dunn and Smyth (1996), which are defined as follows,

$$\widehat{r}_i = \Phi^{-1}(u_i), \qquad i = 1, \ldots, n, \tag{16}$$

where $\Phi(\cdot)$ is the standard normal distribution function and u_i is a random value from the uniform distribution on the interval

$$u_i = \begin{cases} \left[F(t_i - 1 \mid \widehat{\boldsymbol{\theta}}), F(t_i \mid \widehat{\boldsymbol{\theta}})\right], & \text{for } \delta_i = 1 \\ \left[F(t_i \mid \widehat{\boldsymbol{\theta}}), 1\right], & \text{for } \delta_i = 0, \end{cases} \tag{17}$$

where $F(t_i \mid \widehat{\boldsymbol{\theta}})$ is the cumulative distribution function of mixture and nonmixture DW models. Apart from the variability due to the estimates of the parameters these residuals have standard normal distribution if the proposed model is is correctly specified (Dunn & Smyth, 1996).

Hence, to check if the model assumption is adequate we can examine the half-normal plots with simulated envelope proposed by Atkinson (1981). The simulated envelope can be construct as follows

- **(i)** fit the model and generate sample set of n independent observations using the estimated parameters of the fitted model;
- **(ii)** fit the model from the generated sample, calculate the absolute values of the residuals and arrange them in order;
- **(iii)** repeat steps (i) and (ii) B number of times;
- **(iv)** consider the n sets of the B ordered statistics of the residuals, then for each set calculate the quantile $\gamma/2$, the median and the quantile $1-\gamma/2$;
- **(v)** plot these values and the ordered residuals of the original sample set versus the expected order statistics of a half-normal distribution, which is approximated as

$$\Phi^{-1}\left(\frac{i+n-0.125}{2n+0.5}\right).$$

According to Atkinson (1981) if the model was correctly specified then no more than $\gamma \times 100\%$ of the observations are expected to appear outside the envelope bands. Additionally, if a large proportion of the observations lies outside the envelope, thus one have evidence against the adequacy of the fitted model.

4 | SIMULATION STUDY

In this section, we present results of a simulation study to assess the performance of the maximum likelihood estimators of the parameters of DW mixture and non-mixture models considering different sample sizes, different fixed parameter values and different fixed proportion of cure fractions. In order to simulate a random sample from DW distribution containing observed times, censored times, and cure fraction we consider the following algorithm:

1. Fix the parameter values, that is, β and ϕ of the DW distribution as well as the value of cure fraction ρ and the proportion of censored observations λ;
2. Find the value of τ to achieve the desired proportion of censored times λ. To that end we need to find the root of the following equation:

$$\frac{1}{\tau}\int_0^{\tau} S_y(c)\,\mathrm{d}t\,\mathrm{d}c - \lambda = 0, \tag{18}$$

   where S_y denotes the survival function of the observed times;

   Remark 3. In this step the package LindleyR (Mazucheli, Fernandes, & de Oliveira, 2016) is used to solve (18);

3. Generate $Z_i \sim \mathrm{Bernoulli}(\rho)$ for the cure fraction, $Y_i \sim \mathrm{DW}(\beta, \phi)$ for the observed times and $C_i \sim \mathrm{Uniform}(\tau)$ for the censored times;
4. If $y_i > c_i$ or $z_i = 1$ set $\delta_i = 0$, otherwise set $\delta_i = 1$, where δ_i is the censoring indicator.

In all scenarios, we considered random samples of size $n = 100, 300$, and 500. For each sample size we performed 10,000 simulations and calculated the bias and the root of mean squared error. We fixed $\beta = 0.9$ and considering that $\phi_i = \exp(-\exp(\alpha_0 + \alpha_1\, x_i))$, the true values for the parameters were fixed as $\alpha_0 = -2.0$ and $\alpha_1 = -1.0$. The covariate x_i is simulated from $\mathrm{Bernoulli}(0.5)$ distribution and their values were remained constant during the simulations. The proportion of censored times was fixed at $\lambda = 0.1$. Tables 1 and 2 present the results of the Monte Carlo simulations.

In the first experiment, it was considered that the cure fraction parameter, ρ, was not affected by the covariate, taking the following values: $\rho = 0.1, 0.2$, and 0.3. From the results presented in Table 1 it is observed that:

- **(i)** The mixture model has lower biases than the nonmixture model, mainly for the α_1, ρ, and β parameters;
- **(ii)** For both models and scenarios the parameter α_0 presented moderate bias;
- **(iii)** As expected, the RMSE for both models decrease as the sample size increase, conforming the consistency of the maximum likelihood estimates.

TABLE 1 Estimated bias and root mean squared errors for the parameters of the DW mixture and nonmixture models — Scenario I

			Bias				RMSE			
Model	ρ	n	α_0	α_1	ρ	β	α_0	α_1	ρ	β
1	0.1	100	0.167	0.083	−0.051	0.018	0.091	0.058	0.122	0.007
		300	0.186	0.096	−0.010	0.006	0.054	0.027	0.038	0.002
		500	0.192	0.094	−0.009	0.005	0.048	0.019	0.021	0.001
	0.2	100	0.165	0.078	−0.024	0.021	0.099	0.064	0.067	0.008
		300	0.183	0.097	−0.003	0.007	0.055	0.028	0.021	0.002
		500	0.191	0.094	−0.005	0.005	0.049	0.020	0.012	0.001
	0.3	100	0.159	0.074	−0.004	0.024	0.108	0.073	0.049	0.009
		300	0.179	0.095	−0.001	0.009	0.058	0.030	0.015	0.003
		500	0.191	0.093	−0.002	0.005	0.051	0.022	0.010	0.002
2	0.1	100	−0.900	−0.028	0.054	0.167	0.894	0.069	0.111	0.036
		300	−0.870	−0.020	0.090	0.156	0.784	0.023	0.042	0.027
		500	−0.866	−0.023	0.092	0.155	0.766	0.014	0.027	0.025
	0.2	100	−0.628	−0.025	0.040	0.149	0.484	0.074	0.062	0.031
		300	−0.605	−0.007	0.056	0.135	0.394	0.024	0.023	0.021
		500	−0.597	−0.011	0.055	0.133	0.373	0.015	0.014	0.019
	0.3	100	−0.451	−0.013	0.038	0.131	0.302	0.082	0.048	0.027
		300	−0.429	0.007	0.039	0.115	0.216	0.026	0.016	0.016
		500	−0.418	0.005	0.038	0.111	0.193	0.016	0.011	0.014

1: mixture model; 2: nonmixture model.

The second Monte Carlo experiment was carried out considering that the cure fraction parameter is affected by the covariate x_i through the following regression structure:

$$\rho(x_i) = \frac{\exp\left(\delta_0 + \delta_1\, x_i\right)}{1 + \exp\left(\delta_0 + \delta_1\, x_i\right)},$$

where the true values of the parameters were taken as $\delta_0 \approx (-2.19, -1.38, -0.84)$ and $\delta_1 = -1.0$, so that the cure fraction for the two levels of the covariate are $\rho(x_i = 0) = (0.1, 0.2, 0.3)$ and $\rho(x_i = 1) \approx (0.0392, 0.0842, 0.1361)$. Some points are very clear from the results presented in Table 2.

(i) The mixture model has lower biases than the nonmixture model, mainly for α_1, δ_0, δ_1, and β;

(ii) The parameter α_0 presented a moderate biased for both models. Specially, it is positively biased for the mixture model, whereas it is negatively biased for the nonmixture model;

(iii) As expected, the RMSE for both models decrease as the sample size increase, conforming the showing consistency of the maximum likelihood estimates;

(iv) The parameter α_0 in nonmixture model presented a moderate RMSE.

5 | STATISTICAL ANALYSIS OF THE PELVIC SARCOMA (PS) DATASET

For all patients with pelvic sarcoma in the study introduced by Puchner et al. (2017), three lifetimes lifetime were investigated: the overall survival to death, the time to infection and the time to metastasis. In addition, the prognostic factors of age, gender, tumor grade, radiotherapy, chemotherapy, and tumor volume were also recorded. One of the main goals of this study, was to check the influence of potential prognostic factors on the three responses. To that end, the mixture and nonmixture discrete Weibull models were assumed in the data analysis.

TABLE 2 Estimated bias and root mean squared errors for the parameters of the DW mixture and nonmixture models — Scenario II

			Bias					RMSE				
Model	$\rho(x_i = 0)$	n	α_0	α_1	δ_0	δ_1	β	α_0	α_1	δ_0	δ_1	β
1	0.1	100	0.167	0.079	−0.161	−1.329	0.018	0.090	0.056	1.246	1.208	0.007
		300	0.187	0.096	−0.026	−0.080	0.006	0.054	0.026	0.080	0.700	0.002
		500	0.192	0.094	−0.026	−0.022	0.004	0.048	0.019	0.045	0.177	0.001
	0.2	100	0.168	0.078	−0.034	−0.216	0.019	0.098	0.061	0.137	2.345	0.008
		300	0.184	0.097	−0.010	−0.018	0.007	0.055	0.027	0.043	0.137	0.002
		500	0.191	0.093	−0.011	−0.002	0.005	0.049	0.019	0.025	0.083	0.001
	0.3	100	0.165	0.075	−0.001	−0.083	0.022	0.106	0.067	0.104	0.381	0.008
		300	0.182	0.096	−0.006	−0.012	0.008	0.057	0.029	0.031	0.091	0.002
		500	0.192	0.092	−0.008	−0.002	0.005	0.051	0.020	0.020	0.056	0.001
2	0.1	100	−0.899	−0.621	−0.399	−2.509	0.160	0.938	1.815	0.090	0.876	0.034
		300	−0.866	−0.340	0.106	−0.121	0.156	0.780	0.168	0.081	0.746	0.027
		500	−0.864	−0.331	0.105	−0.059	0.156	0.763	0.137	0.050	0.180	0.026
	0.2	100	−0.648	−0.361	0.052	−1.925	0.158	0.512	0.345	0.124	0.217	0.034
		300	−0.628	−0.299	0.067	−0.017	0.147	0.424	0.123	0.043	0.127	0.024
		500	−0.621	−0.300	0.066	−0.003	0.145	0.403	0.110	0.026	0.077	0.023
	0.3	100	−0.489	−0.299	0.055	−0.071	0.148	0.338	0.185	0.099	0.271	0.031
		300	−0.471	−0.260	0.045	−0.002	0.133	0.252	0.098	0.031	0.084	0.020
		500	−0.460	−0.262	0.043	0.005	0.130	0.229	0.088	0.020	0.051	0.018

1: mixture model; 2: nonmixture model.

TABLE 3 Inference results for the three lifetimes

		Survival Time			Infection Time			Metastasis Time		
Model	**Parameter**	**MLE**	**95% C.I.**	**AIC**	**MLE**	**95% C.I.**	**AIC**	**MLE**	**95% C.I.**	**AIC**
Weibull	μ	19.6126	(8.5066, 30.7186)	761.5373	26.0097	(7.8812, 44.1382)	311.0863	37.1204	(8.8439, 65.3968)	423.8983
	β	0.6181	(0.4921, 0.7442)		0.4562	(0.3025, 0.6098)		0.6229	(0.4521, 0.7936)	
Mixture	μ	23.9595	(6.7375, 41.1815)	751.9783	7.7367	(1.4697, 14.0038)	309.0108	31.7240	(1.0969, 62.3511)	411.5414
	β	0.9199	(0.7088, 1.1310)		0.6420	(0.4156, 0.8683)		1.0576	(0.7822, 1.3330)	
	ρ	0.3663	(0.2593, 0.4734)		0.7256	(0.6009, 0.8503)		0.6261	(0.5156, 0.7367)	
Nonmixture Weibull	μ	41.3519	(10.2851, 72.4186)	754.2051	9.1000	(1.3374, 16.8625)	309.0300	43.4083	(-0.4776, 87.2941)	411.3725
	β	0.9714	(0.7394, 1.2035)		0.6609	(0.4277, 0.8941)		1.1071	(0.8196, 1.3946)	
	ρ	0.3632	(0.2455, 0.4809)		0.7255	(0.5993, 0.8517)		0.6246	(0.5133, 0.7359)	

5.1 | Statistical analysis not considering the presence of covariates

We firstly fitted models for the PS data based on the mixture and nonmixture discrete Weibull without considering the presence of covariates. The corresponding inference results are given in Table 3. From these results, it is observed that all models, on computational aspects, did not show instability and the estimation method converged successful. As expected, the smallest AIC values were obtained assuming the mixture and nonmixture models.

Moreover, the estimated proportion of cure fraction for the survival times, that is, the proportions of people in population who will not die due to sarcoma pelvic, are given by 36.63 and 36.32% considering the mixture and nonmixture models, respectively, while from the Kaplan–Meier estimate it is approximately 37.40%. Regarding the time to infection it was verified that the estimated proportions of cure fraction are given by 72.56 and 72.55% for the mixture and nonmixture models, respectively, whereas from the Kaplan–Meier estimate the cure fraction is 72.80%. Finally, for the time to metastasis the estimated proportions of cure fraction are given by 62.61 and 62.46% assuming the mixture and nonmixture models, respectively, and it is 62.20% from the Kaplan–Meier estimate.

Figure 1 displays the plots of theoretical estimated survival functions based on the maximum likelihood estimates along with the empirical survival functions in presence of censored data using the Kaplan–Meier method. It is observed that the DW

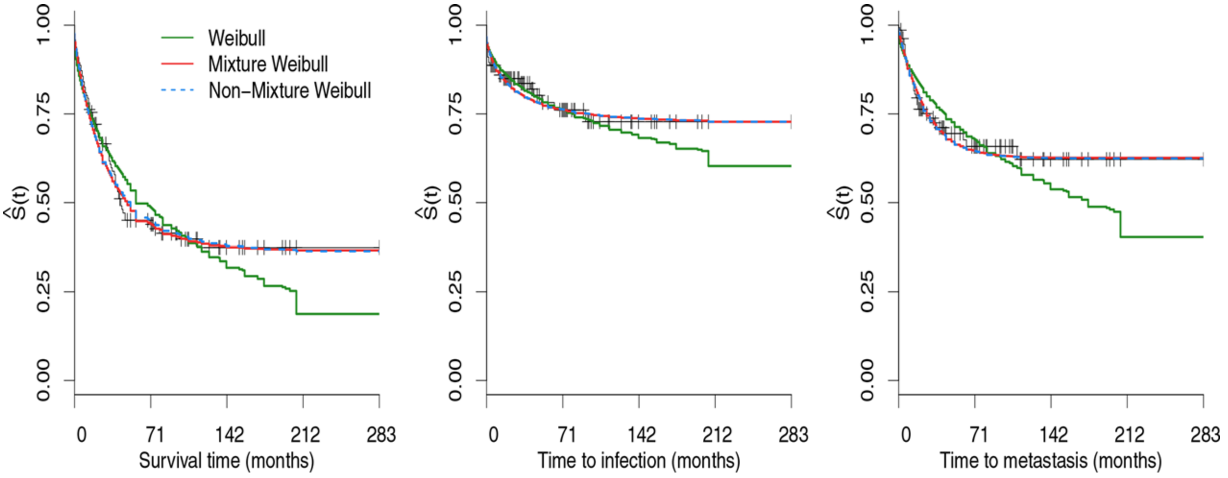

FIGURE 1 Kaplan–Meier curves and the estimated survival functions for the three times

mixture and nonmixture models have better fit than the DW distribution with no cure rate. We can conclude that the mixture and nonmixture models captured the cure rate with great accuracy.

Finally, it is compared in Figure 2 the empirical estimates based on the Kaplan–Meier versus the corresponding predicted values obtained from the parametric models. Clearly, we can observe from these plots that the predicted values obtained from the mixture and nonmixture models are closer to the empirical values than those obtained from the standard discrete Weibull. Nevertheless, for the time to infection it seems that no one of the models have a satisfactory fit.

5.2 | Statistical analysis in the presence of covariates

In what follows let us consider the following covariates associated to each patient:

- age40_i: age of the patients at start of follow-up, classified as less than 40 years ($\text{age40}_i = 0$) versus greater or equal to 40 years ($\text{age40}_i = 1$);
- sex_i: patient gender, classified as male ($\text{sex}_i = 0$) versus female ($\text{sex}_i = 0$);
- chemo_i: whether the patient received chemotherapy ($\text{chemo}_i = 1$) or not ($\text{chemo}_i = 0$);
- radio_i: whether the patient received radiotherapy ($\text{radio}_i = 1$) or not ($\text{radio}_i = 0$);
- grade_i: tumor grade of the patient classified as G3 ($\text{grade}_i = 1$) versus G2 ($\text{grade}_i = 0$);
- log2volume_i: tumor volume (log2-transformed).

In our analysis considering regression models, it was assumed that the covariates affect the probability of being cured and the parameter ϕ, considering the following regression structures:

$$\log\left(\frac{\rho_i}{1-\rho_i}\right) = \delta_0 + \delta_1\,\text{age40}_i + \delta_2\,\text{sex}_i + \delta_3\,\text{chemo}_i + \delta_4\,\text{radio}_i + \delta_5\,\text{grade}_i + \delta_6\,\text{log2volume}_i \tag{19}$$

and

$$\log\left(-\log(\phi_i)\right) = \alpha_0 + \alpha_1\,\text{age40}_i + \alpha_2\,\text{sex}_i + \alpha_3\,\text{chemo}_i + \alpha_4\,\text{radio}_i + \alpha_5\,\text{grade}_i + \alpha_6\,\text{log2volume}_i \tag{20}$$

for $i = 1, \ldots, 147$.

The inference results for the overall survival to death, time to infection, and time to metastasis are reported in Tables 4, 5, and 6, respectively.

From the obtained results of Table 4, it was observed that for both models (mixture and nonmixture) the covariates grade (δ_5) and tumor volume (δ_6) affect the overall survival to death since the zero value is not included in the 95% confidence intervals. The significance of these covariates indicate that the probability of being "cured," that is, not dying due to sarcoma pelvic, depends on person's tumor grade and volume. Patients who have tumor grade G3 seem to have lower chance of not dying compared to patients with tumor grade G2. In addition, we can observe that as the tumor volume increases the odds of not dying due to pelvic

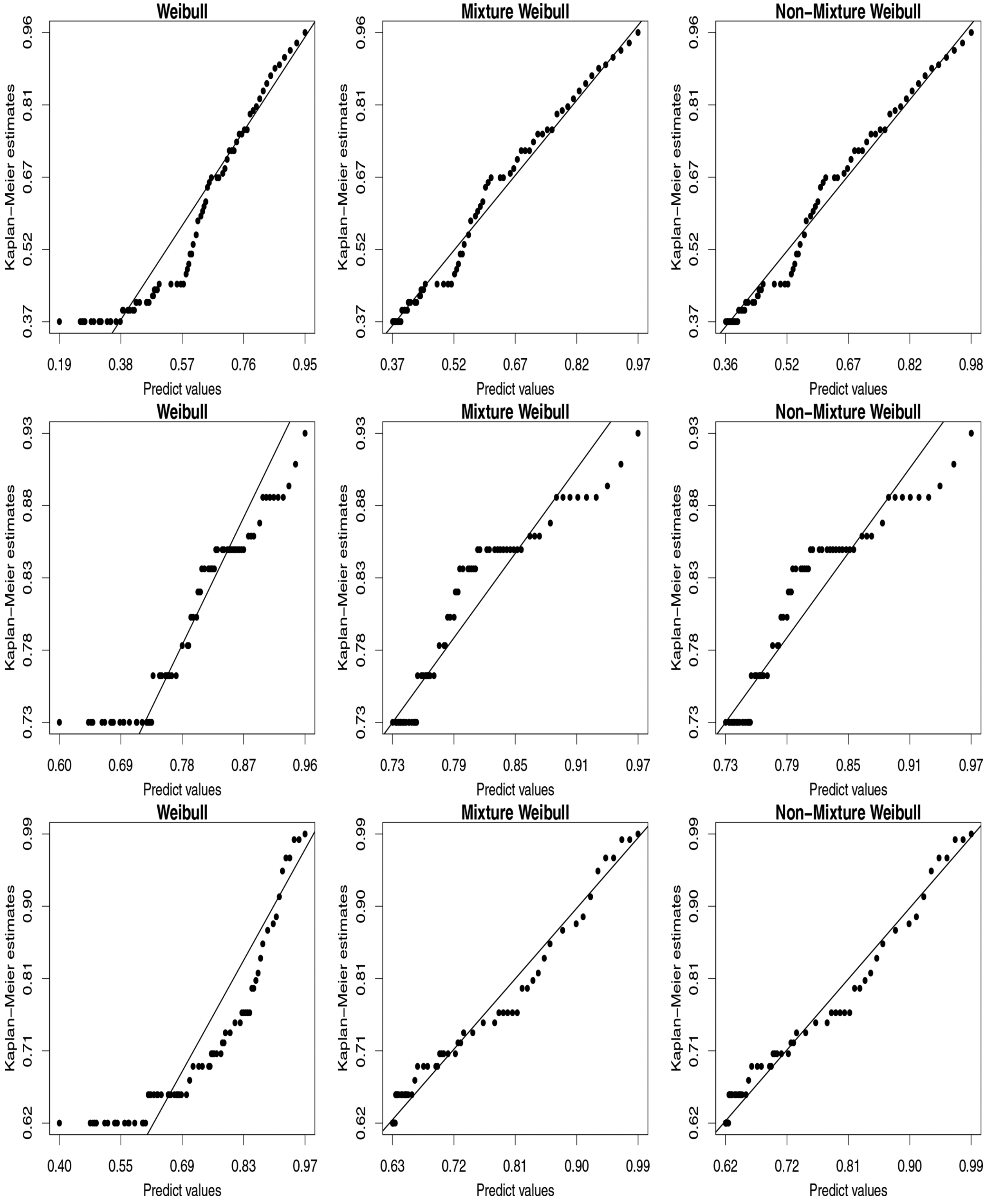

FIGURE 2 Plots of the Kaplan–Meier estimates for the survival function versus the respective predicted values obtained from the parametric models for the three lifetimes. (Upper Panel: Survival time. Middle Panel: Infection time. Lower Panel: Metastasis time.)

sarcoma decreases, since the parameter δ_6 associated to the tumor volume has a negative sign. A difference exist regarding the significance of the covariate age40 (δ_1) and radio (δ_4), because they are statistically not significants in the mixture model. For the DW mixture model only the covariate tumor volume (α_6) affects the median of the survival time, whereas in the nonmixture model the covariates age40 (α_1) and radio (α_4) are also significant at the 5% significance level. From the obtained results of Table 4, it is observed that the inferences for the regression parameters $\delta_0, \delta_1, \ldots, \delta_6$ are similar since the regression parameters

TABLE 4 Inference results for the discrete Weibull cure rate regression models — Survival time

	Mixture			Nonmixture		
Parameter	MLE	SE	95% C.I.	MLE	SE	95% C.I.
δ_0	6.2878	2.1468	(2.0431, 10.5325)	7.0469	1.8754	(3.3389, 10.7549)
δ_1	−1.6036	0.8449	(−3.2741, 0.0669)	−1.8573	0.7275	(−3.2956, −0.4189)
δ_2	−0.3822	0.4918	(−1.3546, 0.5901)	−0.2539	0.5196	(−1.2813, 0.7734)
δ_3	0.1498	0.6348	(−1.1054, 1.4050)	−0.0420	0.6864	(−1.3991, 1.3151)
δ_4	−1.4408	0.7506	(−2.9248, 0.0433)	−1.5720	0.6847	(−2.9258, −0.2181)
δ_5	−2.2784	0.7974	(−3.8551, −0.7018)	−2.4059	0.7795	(−3.9471, −0.8648)
δ_6	−0.4533	0.1567	(−0.7632, −0.1434)	−0.5113	0.1483	(−0.8046, −0.2181)
α_0	−0.4324	2.1326	(−4.6489, 3.7841)	1.3765	1.3332	(−1.2594, 4.0124)
α_1	−0.7533	0.7378	(−2.2121, 0.7054)	−1.4037	0.5125	(−2.4171, −0.3904)
α_2	−0.5068	0.3380	(−1.1750, 0.1614)	−0.5165	0.3768	(−1.2615, 0.2286)
α_3	−0.6700	0.4044	(−1.4695, 0.1296)	−0.9122	0.5152	(−1.9309, 0.1065)
α_4	−0.7493	0.6431	(−2.0208, 0.5222)	−1.2694	0.4968	(−2.2516, −0.2872)
α_5	0.2029	0.6441	(−1.0705, 1.4764)	−0.4409	0.5908	(−1.6090, 0.7272)
α_6	−0.2206	0.1108	(−0.4397, −0.0015)	−0.3697	0.1065	(−0.5802, −0.1591)
β	1.1115	0.1793	(0.7569, 1.4661)	1.1596	0.1375	(0.8877, 1.4315)

MLE, maximum likelihood estimates; SE, standard error; 95% C.I., 95% confidence interval.

TABLE 5 Inference results for the discrete Weibull cure rate regression models — Time to infection

	Mixture			Nonmixture		
Parameter	MLE	SE	95% C.I.	MLE	SE	95% C.I.
δ_0	−16.4294	11.3838	(−38.9385, 6.0798)	−9.5600	7.3180	(−23.9030, 4.7830)
δ_1	−3.6810	1.9659	(−7.5682, 0.2062)	−3.2156	1.6364	(−6.4229, −0.0082)
δ_2	8.0530	4.6309	(−1.1036, 17.2096)	4.7289	2.6811	(−0.5260, 9.9837)
δ_3	7.6546	4.7980	(−1.8324, 17.1417)	4.7702	2.9249	(−0.9625, 10.5030)
δ_4	−3.9429	2.3855	(−8.6598, 0.7740)	−3.1499	1.7926	(−6.6632, 0.3635)
δ_5	−3.9574	2.7146	(−9.3250, 1.4102)	−2.1437	1.8444	(−5.7586, 1.4712)
δ_6	1.1875	0.7627	(−0.3206, 2.6956)	0.7942	0.5411	(−0.2664, 1.8548)
α_0	−8.5747	1.9547	(−12.4398, -4.7095)	−11.9181	3.0884	(−17.9713, −5.8650)
α_1	−0.9091	0.6660	(−2.2261, 0.4079)	−2.2399	1.0368	(−4.2719, −0.2079)
α_2	1.9000	0.8354	(0.2483, 3.5518)	3.2018	0.9450	(1.3496, 5.0540)
α_3	2.3371	0.8708	(0.6154, 4.0589)	3.6894	1.0830	(1.5668, 5.8120)
α_4	−1.4132	0.7004	(−2.7981, −0.0284)	−2.5335	1.0909	(−4.6717, −0.3954)
α_5	−0.7874	0.7070	(−2.1855, 0.6106)	−1.2127	1.0060	(−3.1844, 0.7590)
α_6	0.6456	0.1777	(0.2942, 0.9969)	0.9284	0.2783	(0.3830, 1.4738)
β	0.3737	0.0895	(0.1968, 0.5507)	0.4387	0.1170	(0.2095, 0.6680)

MLE, maximum likelihood estimates; SE, standard error; 95% C.I., 95% confidence interval.

are related in both cure models to the same cure parameter ρ. In other way, the obtained inferences for the regression parameters $\alpha_0, \alpha_1, \dots, \alpha_6$ related to the parameter ϕ are in general different since they are associated to different scales given by the survival functions (9) and (10).

Figure 3 presents the half-normal with simulated envelope considering the randomized quantile residuals. It is observed, for both models, that all points lie inside the envelopes, suggesting that there is no serious violation of the model assumptions. Additionally, it is noteworthy that the nonmixture model fits the data better than the mixture, since the observed residuals are closer to the median line.

In respect to the time to infection the results from Table 5 showed that all covariates do not affect the probability of being cured assuming the mixture model, since the 95% confidence intervals include the zero value. On the other hand, the covariate

TABLE 6 Inference results for the discrete Weibull cure rate regression models — Time to metastasis

Parameter	Mixture			Nonmixture		
	MLE	SE	95% C.I.	MLE	SE	95% C.I.
δ_0	7.1233	1.9217	(3.3570, 10.8897)	6.9469	1.8971	(3.2288, 10.6651)
δ_1	−0.3116	0.6474	(−1.5805, 0.9573)	−0.2062	0.6569	(−1.4938, 1.0813)
δ_2	−0.6215	0.5210	(−1.6425, 0.3996)	−0.6769	0.5173	(−1.6908, 0.3370)
δ_3	−1.8591	0.7426	(−3.3146, −0.4036)	−1.8487	0.7351	(−3.2894, −0.4079)
δ_4	0.3352	0.6439	(−0.9269, 1.5973)	0.4549	0.6303	(−0.7805, 1.6904)
δ_5	−1.1753	0.9068	(−2.9526, 0.6021)	−1.1484	0.8935	(−2.8996, 0.6028)
δ_6	−0.4700	0.1669	(−0.7971, −0.1429)	−0.4570	0.1645	(−0.7795, −0.1345)
α_0	−4.8403	1.5214	(−7.8222, −1.8584)	−4.2341	1.7083	(−7.5823, −0.8858)
α_1	−0.2126	0.6624	(−1.5109, 1.0858)	−0.0527	0.7592	(−1.5407, 1.4353)
α_2	−0.3770	0.4503	(−1.2596, 0.5057)	−0.6333	0.5045	(−1.6221, 0.3554)
α_3	−2.5830	0.6866	(−3.9287, −1.2373)	−2.9090	0.7692	(−4.4166, −1.4014)
α_4	0.9750	0.5980	(−0.1970, 2.1471)	1.2403	0.6482	(−0.0301, 2.5108)
α_5	3.6399	0.9562	(1.7659, 5.5140)	3.6976	1.0326	(1.6736, 5.7215)
α_6	−0.1645	0.1219	(−0.4035, 0.0745)	−0.2751	0.1368	(−0.5433, −0.0069)
β	1.7503	0.2417	(1.2766, 2.2239)	1.8718	0.2582	(1.3658, 2.3778)

MLE, maximum likelihood estimates; SE, standard error; 95% C.I., 95% confidence interval.

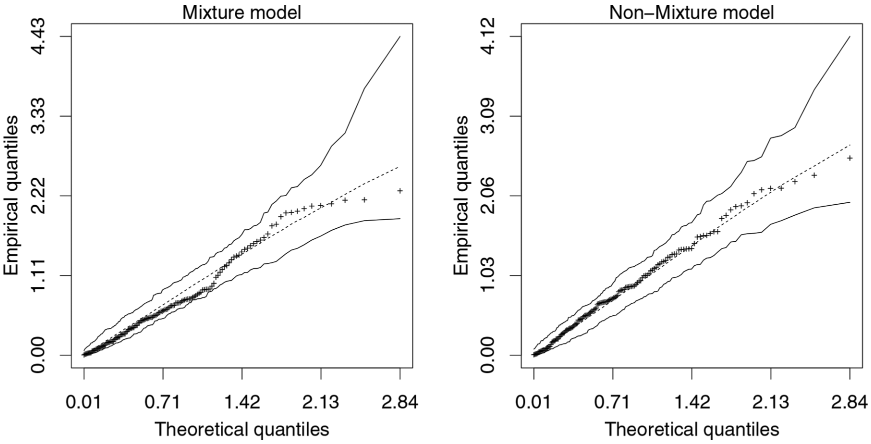

FIGURE 3 Half-normal plot with simulated envelope for the randomized quantile residuals — Survival time

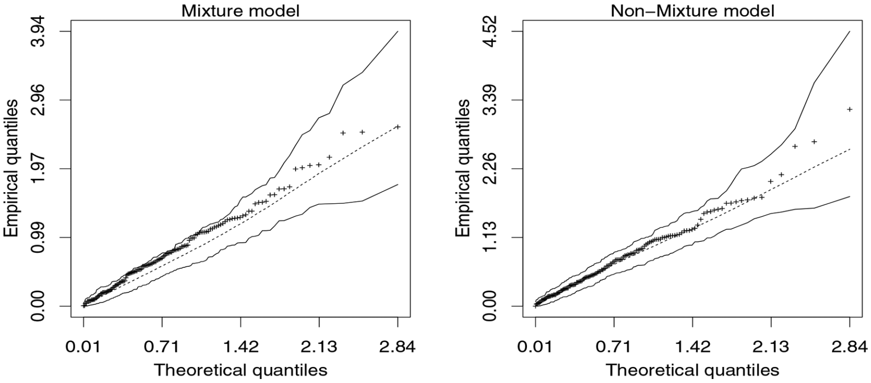

FIGURE 4 Half-normal plot with simulated envelope for the randomized quantile residuals — Time to infection

age40 (δ_1) was only statistically significant under the nonmixture model, although the upper limit of the 95% confidence interval for δ_1 assuming the nonmixture model is very close to zero, indicating some significance of the covariate age40.

A visual inspection of the half-normal plots given in Figure 4 suggests that although all points lie inside the envelope, indicating that there is no serious violation of the model assumptions, it is also observed that the DW mixture model does not fit the

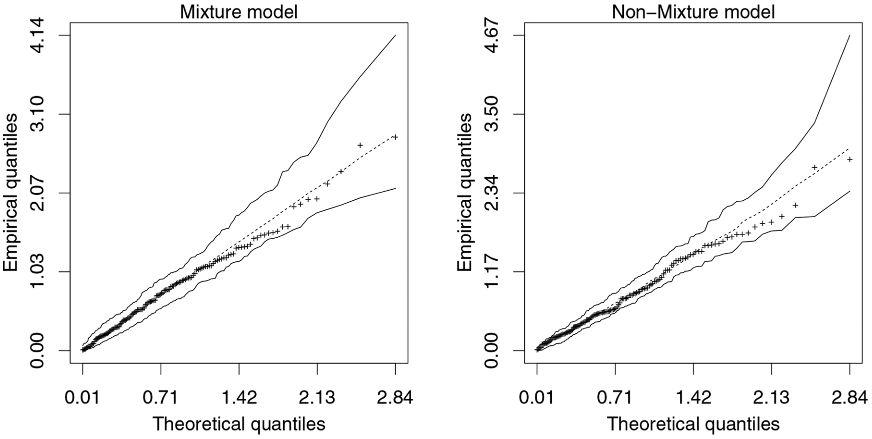

FIGURE 5 Half-normal plot with simulated envelope for the quantile residuals — Time to metastasis

TABLE 7 Values of model discrimination criteria (AIC) for mixture and non-mixture models

	Mixture		**Nonmixture**	
Time	**Without covariates**	**With covariates**	**Without covariates**	**With covariates**
Survival	751.9783	694.6915	754.2051	697.4294
Infection	309.0108	267.6405	309.0300	270.0043
Metastasis	411.5414	370.2743	411.3725	370.9714

data as well as the DW nonmixture model, since most of the observed residuals of the mixture model are near to the boundary of the envelope.

The results of Table 6 related to the metastasis times indicate that the covariate chemotherapy (δ_3) and tumor volume are statistically significant for both the mixture and nonmixture models. The significance of theses covariates reveal that the probability of being “cured,” depends if the patients received or not the chemotherapy, which means that individuals who received the chemotherapy seem to have lower chance of not presented metastasis compared to individuals who do not received chemotherapy. Additionally, we expected that as the tumor volume increases the odds of not presented metastasis decreases, since the parameter δ_6 has negative sign.

The simulated envelope plots of the mixture and nonmixture models corresponding to the metastasis time are shown in Figure 5. For both models, we can see that most of the observed randomized quantile residuals are within the simulated envelope, showing no evidence of violated model assumptions.

Table 7 reports the AIC values considering the regression model and the model not including the effects of the covariates. From these results, it is observed that the mixture and nonmixture models have the similar values of AIC. Regardless the response variable it is verified that the model with covariates presented the smallest values of AIC. However, for the time to infection the AIC values indicated no improvement on the fit to the data, as expected, since for this time the covariates were not significant.

6 | CONCLUDING REMARKS

The main goal of this paper was the introduction of mixture and nonmixture cure fraction models assuming a discrete Weibull distribution in place of standard existing continuous lifetime distributions, with special application to the statistical analysis of a dataset related to long-term oncological treatment outcomes of resection of pelvic sarcomas. The obtained results of this study show many advantages for the use of discrete cure fraction models in terms of great accuracy for the obtained point and interval inferences, great computational simplicity to get the inferences of interest under the classical approach and with simple interpretations for the parameters of the models that is an important point in medical applications. In this way, the results of this study could be of great interest for the search of appropriate univariate lifetime distributions in presence of cure fraction, censored data and covariates. It was also observed that using regression models, the identification of important covariates was easily obtained with good accuracy assuming the DW model. These results are implication of the best simplicity of the likelihood function assuming the DW model when compared to standard continuous Weibull cure fraction models commonly used in the analysis of lifetime data in presence of cure fraction, censored data and covariates.

ACKNOWLEDGMENT

The authors would like to thank the Editor, Associate Editor, and two referees for the valuable comments that have improved considerably the manuscript.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

ORCID

Ricardo Puziol de Oliveira https://orcid.org/0000-0001-6134-5975
André F. B. Menezes https://orcid.org/0000-0002-3320-9834
Josmar Mazucheli https://orcid.org/0000-0001-6740-0445
Jorge A. Achcar https://orcid.org/0000-0002-9868-9453

REFERENCES

Achcar, J. A., Coelho-Barros, E. A., & Mazucheli, J. (2012). Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications*, *51*(1), 1–9.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Almalki, S. J., & Nadarajah, S. (2014). Modifications of the weibull distribution: A review. *Reliability Engineering & System Safety*, *124*, 32–55.

Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, *68*(1), 13–20.

Brunello, G., & Nakano, E. (2015). Inference Bayesian in the discrete Weibull model in data with presence of censorship (in portuguese). *TEMA (São Carlos)*, *16*(2), 97–110.

Cancho, V. G., & Bolfarine, H. (2001). Modeling the presence of immunes by using the exponentiated-Weibull model. *Journal of Applied Statistics*, *28*(6), 659–671.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, *34*, 187–220.

De Angelis, R., Capocaccia, R., Hakulinen, T., Soderman, B., & Verdecchia, A. (1999). Mixture models for cancer survival analysis: Application to population-based data with covariates. *Statistics in Medicine*, *18*(4), 441–454.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, *5*(3), 236–244.

Englehardt, J. D., & Li, R. (2011). The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis*, *31*(3), 370–381.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, *38*(4), 1041–1046.

Khan, M. A., Khalique, A., & Abouammoh, A. (1989). On estimating parameters in a discrete Weibull distribution. *Reliability, IEEE Transactions on*, *38*(3), 348–350.

Klakattawi, H. S., Vinciotti, V., & Yu, K. (2018). A simple and adaptive dispersion regression model for count data. *Entropy*, *20*(2), 1–15.

Kulasekera, K. B. (1994). Approximate MLE's of the parameters of a discrete Weibull distribution with type I censored data. *Microelectronics Reliability*, *34*(7), 1185–1188.

Lambert, P. C., Thompson, J. R., Weston, C. L., & Dickman, P. W. (2006). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, *8*(3), 576–594.

Lehmann, E. J., & Casella, G. (1998). *Theory of point estimation*. Berlin: Springer Verlag.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*, *20*(2), 661–674.

Maller, R. A., & Zhou, X. (1996). *Survival analysis with long-term survivors*. New York: Wiley.

Mazucheli, J., Fernandes, L. B., & de Oliveira, R. P. (2016). *LindleyR: The Lindley Distribution and Its Modifications*. R package version 1.1.0.

Murthy, D. P., Xie, M., & Jiang, R. (2004). *Weibull models* (vol. 505). Hoboken, New Jersey: John Wiley & Sons.

Nakagawa, T., & Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, *24*(5), 300–301.

Othus, M., Barlogie, B., LeBlanc, M. L., & Crowley, J. J. (2012). Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, *18*(14), 3731–3736.

Price, D. L., & Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, *20*(9-10), 1515–1527.

Puchner, S. E., Funovics, P. T., Böhler, C., Kaider, A., Stihsen, C., Hobusch, G. M., Panotopoulos, J., & Windhager, R. (2017). Oncological and surgical outcome after treatment of pelvic sarcomas. *PloS One*, *12*(2), e0172203.

Rohde, C. A. (2014). *Introductory statistical inference with the likelihood function*. New York: Springer-Verlag.

Roy, D. (2002). Discretization of continuous distributions with an application to stress-strength reliability. *Bulletin of the Calcutta Statistical Association*, *52*(205–208), 295–314.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, *53*(2), 255–260.

SAS (2010). *The NLMIXED Procedure, SAS/STAT® User's Guide, Version 9.4*. Cary, NC: SAS Institute Inc.

Sugarbaker, P. (2001). *Management of abdominopelvic sarcoma* (pp. 147–163). Dordrecht: Springer Netherlands.

Tsodikov, A., Ibrahim, J., & Yakovlev, A. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, *98*(464), 1063–1078.

Vahidpour, M. (2016). *Cure rate models*. PhD thesis, École Polytechnique de Montréal.

Yin, G., & Ibrahim, J. G. (2005). Cure rate models: A unified approach. *Canadian Journal of Statistics*, *33*(4), 559–570.

Yu, B., Tiwari, R. C., Cronin, K. A., & Feuer, E. J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, *23*(11), 1733–1747.

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: de Oliveira RP, Menezes AFB, Mazucheli J, Achcar JA. Mixture and nonmixture cure fraction models assuming discrete lifetimes: Application to a pelvic sarcoma dataset. *Biometrical Journal*. 2019;1–14. https://doi.org/10.1002/bimj.201800030