REVSTAT – Statistical Journal
Volume 20, Number 5, October 2022, 587–604
https://doi.org/10.57805/revstat.v20i5.387

The Destructive Zero-Inflated Power Series Cure Rate Models for Carcinogenesis Studies

Authors: Rodrigo R. Pescim ✉
– Department of Statistics, Londrina State University, Londrina, PR, Brazil
rrpescim@uel.br

Adriano K. Suzuki
– Department of Applied Mathematics and Statistics, University of São Paulo, São Carlos, SP, Brazil
suzuki@icmc.usp.br

Gauss M. Cordeiro
– Department of Statistics, Federal University of Pernambuco, Recife, PE, Brazil
gausscordeiro@gmail.com

Edwin M. M. Ortega
– Department of Exact Science, University of São Paulo, Piracicaba, SP, Brazil
edwin@usp.br

Received: January 2020 Revised: January 2021 Accepted: January 2021

Abstract:

- In this paper, we propose a new flexible survival model called the destructive zero-inflated power series cure rate model. This new model describes a realistic interpretation for the biological mechanism of the occurrence of the event of interest in studies of carcinogenesis in the presence of the competing latent causes. The maximum likelihood method is used for estimating the model parameters. For different sample sizes, various scenarios are simulated to evaluate the precision of the estimates. The usefulness of the new cure rate survival model is illustrated by means of a cutaneous melanoma data set.

Keywords:

- *cure rate model; destructive cure model; simulation study; survival analysis; Weibull distribution; zero inflated power series distribution.*

AMS Subject Classification:

- 62N02, 62F99.

✉ Corresponding author.

1. INTRODUCTION

Cancer is the name given to a set of more than 100 diseases that have in common the disordered growth of cells, which invade tissues and organs. Dividing rapidly, these cells tend to be very aggressive and uncontrollable, determining the formation (carcinogenesis process) of malignant tumors, which can spread to other regions of the body. The carcinogenesis process (cancer formation), in general, occurs slowly and may take several years for a cancer cell proliferate and give rise to a visible tumor. That process goes through several stages (initiation of a tumor, promotion and progression) before reaching the tumor. Statistics show that cancer is one of the most important public health concern around the world and for this reason it is crucial to estimate its prevalence, incidence, and mortality/survival rates [17]. An overview of descriptive cancer data on this disease is a first step to appreciate control measures and preventive interventions in a global context of progressive cancer burden [21].

Cutaneous malignant melanoma, a type of skin cancer that originates in melanocytes (cells that produce melanin, a substance that determines skin color), is a tumor whose incidence is increasing dramatically in persons with light-colored skin in all parts of the world. As with other cancers, there are several causes of malignant melanoma formation such as environmental (imminent exposure to ultraviolet radiation), genetic and immunological factors. In most studies, the incidence doubles every 6 to 10 years. In years of potential life loss, melanoma is second to adult leukemia, as it affects younger individuals, causing a major public health problem [1]. According to World Health Organization, about 132,000 new cases of cutaneous melanoma are diagnosed worldwide each year. In particular, the American Cancer Society estimated that there will be 96,000 new cases of cutaneous melanoma in the United States and 7,000 deaths from this disease in 2020. In addition, approximately 57,000 new cases of invasive cutaneous melanoma will occur in men and 39,000 new cases in women in 2020. On the other hand, there has been a great improvement in the survival of patients with cutaneous melanoma, mainly due to the early detection of the tumor, in recent years. In general, it is taken that "cured" is related to survival beyond 5 years for patients with melanoma. This may be due to earlier diagnosis, when tumors are still at a thinner depth, as well as improved treatment and surgical techniques [8].

Survival models with a cure fraction for cutaneous melanoma data have played an important role in survival analysis in recent years. These types of survival models cover situations in which there are persons not susceptible to the occurrence of the event of interest. Consequently, a fraction (or proportion) of these individuals are not expected to experience the event of interest, that is, these individuals are considered not susceptible or "cured" in the survival analysis context. The proportion of cured individuals is denoted by the cure fraction. Cure rate models have the main purpose to include in their formulation the possibility of estimating the cure rate and they have been widely studied by several authors and used for modeling time-to-event data for various types of cancers, including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer and melanoma.

The most popular type of cure rate models are the mixture (or Berkson–Gage) cure model [2] and the promotion time cure rate model ([27] and [7]). While the Berkson–Gage cure model is based on the assumption that only one cause is responsible for the occurrence of the event of interest, that is, the unknown number of causes of the event of interest is

assumed to be a Bernoulli random variable, in the promotion time cure model the number of causes follows a Poisson distribution. In a biological context, the occurrence of the event of interest might be due to one of many competing causes [14], with the number of causes and the distribution of survival times associated with each cause [11] being unknown which leads to a latent competing causes structure. In this sense, the event of interest can be the death of a patient or a tumor recurrence, which can happen because of unknown competing causes [21]. These latent competing causes can be assigned to metastasis-component tumor cells left active after an initial treatment. A metastasis-component tumor cell is a tumor cell having the potential of metastasizing [27]. The statistical literature on distributions which accommodate different numbers of latent competitors have as the main works in the books by [19] and [16] as well as the review paper by [26] and the papers of [10], [29], [6], [24] and [4] can be mentioned as key references.

More recently, [23] extended the works of [27] and [7] by considering a cure rate model (also known as a destructive weighted Poisson cure rate model) to deal with the assumption that each initiated cell (competing cause) becomes cancerous with probability one. They argue that this development is a much more realistic alternative to the cure rate model in explaining the biological mechanism underlying the occurrence of the event in presence of a cure fraction. This is because the proposed cure rate survival model presumes that the original number of lesions, or altered cells are not repaired or eliminated after some intensive treatment, and this group (which is represented by a variable) of unrepaired cells (or latent factors) are potentially competing to give rise to a tumor, or risk of failure. Figure 1 represents the destructive model in a diagram form.

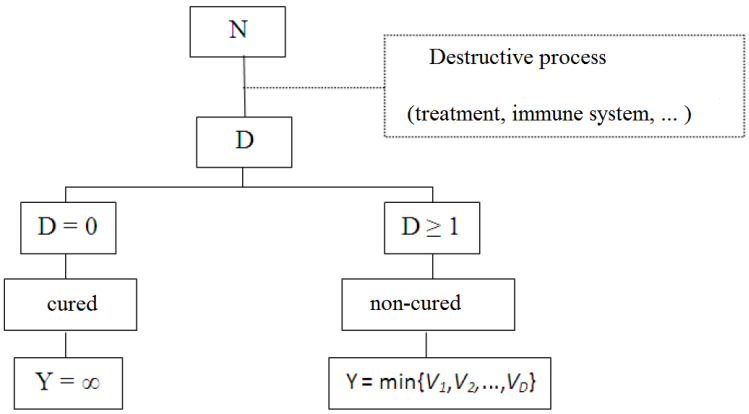

Figure 1: Representation of the proposed destructive model in a diagram form.

However, there is an amount (or proportion) of cells that have not been initiated (normal cells), which includes repaired cells, that are not being explained properly by those cure rate models that consider the number of initiated cells related to the occurrence of a tumor being a random variable that follows the power series family of distributions which has as special cases the Poisson, Bernoulli, geometric, negative binomial, etc.

In a biological context, it is noted that there is a much larger number of cells that are not initiated (normal cells) than cells that are initiated (and consequently become malignant

cells), which leads to an "excess of not initiated cells" (or "excess number of zero counts") in relation to cells which are lesioned.

In this sense, the excess of zeros (not initiated cells) can be explained in terms of zero-inflated models as follows:

1. First, there exist an amount of not initiated cells (zeros) which have never experienced any type of alterations or lesions (structural zeros).
2. On the other hand, there exist an amount of not initiated cells which have experienced alterations (or lesions), but those cells were repaired (sampling zeros).

Therefore, it is maybe desirable to construct tractable statistical models that can adequately incorporate a biological mechanism for the initiation process of carcinogenesis, and this is the main motivation for the present research work.

Here, we introduce a new cure rate survival model which extends the works of [23] and [4] by incorporating a structure to estimate the proportion of not initiated cells (those one that have never been altered/lesioned and those one that have been repaired). To create such structure, we use the concept of zero-inflated models by considering an extension of the discrete power series distributions by including an additional parameter π. Its interpretation is related to the proportion of repaired cells by means a repair system of the body. In this approach, we assume that the number of initiated cells follow the zero-inflated power series (ZIPS) [15] distribution, which is a suitable choice for modelling data sets that possesses excess of zeros and overdispersion. Furthermore, it provides a realistic interpretation related to the biological mechanism of the occurrence of the event of interest. Also, it includes a process of destruction of tumor cells after an initial treatment ([23], [3] and [22]).

The rest of the paper is outlined as follows. In Section 2, we formulate the new cure rate model. Some special models are reported in Section 3. Inference based on maximum-likelihood (ML) is discussed in Section 4. In Section 5, we perform a simulation study to verify the precision of the estimates of the model parameters. An application to a real data set on cutaneous melanoma is addressed in Section 6. Finally, Section 7 provides some concluding remarks.

2. MODEL FORMULATION

Let N be an unobservable (latent) random variable which follows the zero-inflated power series (ZIPS) distribution, denoting the initial number of initiated cells related to the occurrence (or recurrence) of a tumor for an individual in a population, with probability mass function (pmf)

$$P[N=n] = \begin{cases} \pi + (1-\pi)\,\dfrac{a_0}{g(\theta)}\,, & \text{for } n=0\,, \\ (1-\pi)\,\dfrac{a_n\theta^n}{g(\theta)}\,, & \text{for } n=1,2,3,\ldots\,, \end{cases} \tag{2.1}$$

where $0<\pi<1$, $a_n>0$ (a_n depends only on n) and $g(\theta)=\sum_{n=0}^{\infty} a_n\theta^n$ is a positive, finite and differentiable function.

Here, the parameter π is interpreted as the proportion of cells that have never experienced alterations (or modifications) in their genes, while the interpretation for the quantity $(1-\pi)$ refers to the proportion of cells which have been repaired from a body repair mechanism.

Also, we note that if $\pi = 0$, the ZIPS distribution reduces to the power series (PS) distribution proposed by [20]. Some important well-known discrete distributions belong to this family of distributions. For example, if $g(\theta) = (1+\theta)^m$ and m is positive integer, Equation (2.1) becomes the zero-inflated binomial (ZIBin) distribution. If $g(\theta) = \exp(\theta)$, it defines the zero-inflated Poisson (ZIP) distribution. Further, if $g(\theta) = (1+\theta)^{-\phi}$, $\phi > 0$ and $0 < \theta < 1$, the zero-inflated negative binomial (ZINB) distribution is obtained from Equation (2.1), among others.

For the ZIPS random variable N, the probability generating function (pgf) is

$$\mathbb{A}_N(z) = \pi + (1-\pi)\,\frac{g(\theta z)}{g(\theta)}\,, \qquad \text{for} \quad 0 \le z \le 1\,, \tag{2.2}$$

where the ratio $g(\theta z)/g(\theta)$ is the pgf of the PS distribution. For more details, see [20].

The first consequence of a prolonged treatment (destructive process) is the possible formation of precancerous lesions into the genome of the cells. These cells are denoted as malignant cells. Given $N = n$, let X_j, $j = 1, 2, ..., n$ be independent random variables (independent of N) following a Bernoulli distribution with success probability p indicating presence of the j-th lesion. The pgf of the Bernoulli random variable X_j can be expressed as

$$\mathbb{A}_{X_j}(z) = 1 - p\,(1-z)\,, \qquad \text{for} \quad 0 \le z \le 1\,. \tag{2.3}$$

The variable D, representing the total number of malignant cells among the N initial cells (competing causes) which are not eliminated by the treatment is defined as

$$D = \begin{cases} X_1 + X_2 + \cdots + X_N\,, & \text{if} \quad N > 0\,, \\ 0\,, & \text{if} \quad N = 0\,, \end{cases} \tag{2.4}$$

where $D \le N$. The idea involved in (2.4) was suggested by [28] considering that the initial N cells are primary initiated malignant cells, where X_j in (2.4) represents the number of living malignant cells that are descendants of the j-th initiated malignant cell during some time interval. In this case, D denotes the total number of living malignant cells at some specific time. The time to event for the j-th competing cause is represented by V_j, $j = 1, ..., D$. Conditional on D, the V_j's are assumed iid with cumulative distribution function $F(t)$ and survival function $S(t) = 1 - F(t)$. Also, we note that the total number of malignant cells D and the time V_j are not observable.

As pointed out by [3], in the competing causes scenario, the number of unrepaired lesions D in (2.4) and the time V taken to transform these lesions into a detectable tumor are both not observable (latent variables). In this context, we denote V a progression time. Thus, the observed time to the event of interest (the patient's death) is defined by the following random variable

$$Y = \min\{V_1, ..., V_D\} \tag{2.5}$$

for $D \ge 1$, and $Y = \infty$ if $D = 0$, which leads to a proportion p_0 of the population which is called the cured fraction.

Under this setup, [23] showed that the survival function for the population of the random variable Y in (2.5) has the form

$$S_{\text{pop}}(y) \;=\; P\big[Y \geq y\big] \;=\; \mathbb{A}_D\big(S(y)\big) \;=\; \sum_{d=0}^{\infty} P\big[D=d\big]\,\big\{S(y)\big\}^d \;=\; \mathbb{A}_N\big(\mathbb{A}_{X_j}\big(S(y)\big)\big)\,,$$

where $S(\cdot)$ is the survival function for non-cured population and $\mathbb{A}_D(\cdot)$ is the pgf for the variable D. Combining (2.2) and (2.3), the survival function of the observable lifetime of the event of interest can be expressed as

$$S_{\text{pop}}(y) \;=\; \pi + (1-\pi)\,\frac{g\big(\theta\big[1-p\,F(y)\big]\big)}{g(\theta)}\,, \tag{2.6}$$

where $F(y) = 1 - S(y)$. Hereafter, Equation (2.6) is referred to as the destructive zero inflated power series (DZIPS) cure rate model. This model includes two important special cases: For $\pi = 0$, it reduces to the destructive power series (DPS) cure rate model and if $\pi = 0$ in addition $p = 1$, it gives the power series (PS) cure rate model ([4]).

From model (2.6), the proportion p_0 of cured individuals in the population is

$$p_0 \;=\; \lim_{y\to\infty} S_{\text{pop}}(y) \;=\; \pi + (1-\pi)\,\frac{g\big(\theta(1-p)\big)}{g(\theta)}\,.$$

The density function associated with (2.6) can be expressed as

$$f_{\text{pop}}(y) \;=\; -\frac{d\,S_{\text{pop}}(y)}{dy} \;=\; -\left[(1-\pi)\,\frac{g'\big(\theta\big[1-p\,F(y)\big]\big)}{g(\theta)}\right], \tag{2.7}$$

where $g'(\cdot) = dg(\cdot)/dy$, $f(y) = dF(y)/dy$ denotes the proper density function of the time V to the event in (2.6). Note that the function $f_{\text{pop}}(y)$ is a proper function, whereas $S_{\text{pop}}(y)$ is not a proper survival function.

3. SPECIAL CASES OF THE DZIPS CURE MODEL

In this section, we present some specific models that arise from the ZIPS model formulation. Here, we consider situations where N is a random variable which follows the zero-inflated Poisson, zero-inflated binomial, zero-inflated negative binomial, and zero-inflated geometric distributions.

3.1. The destructive zero-inflated Poisson (DZIP) cure model

If we consider $a_n = \dfrac{1}{n!}$ and $g(\theta) = \exp(\theta)$ in (2.1), the number of initiated cells N follows a ZIP distribution with $\theta > 0$ and $\pi \in (0,1)$ and pmf

$$P_{\text{ZIP}}[N=n] \;=\; \begin{cases} \pi + (1-\pi)\,\mathrm{e}^{-\theta}\,, & \text{for } \; n=0\,, \\[2ex] (1-\pi)\,\dfrac{\mathrm{e}^{-\theta}\,\theta^n}{n!}\,, & \text{for } \; n=1,2,3,\ldots\,. \end{cases} \tag{3.1}$$

The corresponding survival function of the DZIP cure model is

$$S_{\text{pop}}(y) = \pi + (1-\pi)\,\mathrm{e}^{-\theta p F(y)}. \tag{3.2}$$

The cure rate is $p_0 = \pi + (1-\pi)\,\mathrm{e}^{-\theta p}$, and the corresponding density function takes the form

$$f_{\text{pop}}(y) = (1-\pi)\,\theta\,p\,f(y)\,\mathrm{e}^{-\theta p F(y)}. \tag{3.3}$$

There are some important special cases in (3.2). For $\pi = 0$, it follows the destructive Poisson cure model defined by Rodrigues *et al.* [23]. We introduce the zero-inflated Poisson cure model for $p = 1$, whereas for $\pi = 0$ in addition $p = 1$, if follows the promotion time cure model studied by [27] and [7].

3.2. The destructive zero-inflated binomial (DZIBin) cure model

If we have $a_n = \binom{m}{n}$ and $g(\theta) = (1+\theta)^m$ in (2.1), the number of initiated cells N follows a ZIBin distribution with parameters $\frac{\theta}{1+\theta}$, $\pi \in (0,1)$ (m is a positive integer) and pmf

$$P_{\text{ZIBin}}[N=n] = \begin{cases} \pi + (1-\pi)\left(\dfrac{1}{1+\theta}\right)^m, & \text{for } n = 0, \\ (1-\pi)\binom{m}{n}\left(\dfrac{\theta}{1+\theta}\right)^n\left(\dfrac{1}{1+\theta}\right)^{m-n}, & \text{for } n = 1,2,3,\ldots. \end{cases}$$

The survival function of the DZIBin cure model has the form

$$S_{\text{pop}}(y) = \pi + (1-\pi)\left[1 - \frac{\theta\,p\,F(y)}{1+\theta}\right]^m. \tag{3.4}$$

Here, the cure fraction is given by $p_0 = \pi + (1-\pi)\left[1 - \frac{\theta p}{1+\theta}\right]^m$. So, the density function of the DZIBin cure model can be expressed as

$$f_{\text{pop}}(y) = (1-\pi)\,\frac{m\,\theta\,p\,f(y)}{1+\theta}\left[1 - \frac{\theta\,p\,F(y)}{1+\theta}\right]^{m-1}. \tag{3.5}$$

The DZIBin cure model in (3.4) with $\pi = 0$ in addition to $p = m = 1$ coincides with the mixture (Berkson–Gage) cure model pioneered by [2].

3.3. The destructive zero-inflated negative binomial (DZINB) cure model

If we consider $a_n = \dfrac{\Gamma(\phi^{-1}+n)}{n!\,\Gamma(\phi^{-1})}$, $g(\theta) = (1-\theta)^{-1/\phi}$ and $\theta = \dfrac{\eta\phi}{1+\eta\phi}$ in (2.1), the number of initiated cells N follows a ZINB distribution with $\eta > 0$, $\phi \geq -1$, $\eta\phi > 0$ and $\pi \in (0,1)$, with pmf

$$P_{\text{ZINB}}[N=n] = \begin{cases} \pi + (1-\pi)\,(1+\eta\phi)^{-1/\phi}, & \text{for } n = 0, \\ (1-\pi)\,\dfrac{\Gamma(\phi^{-1}+n)}{n!\,\Gamma(\phi^{-1})}\left(\dfrac{\eta\phi}{1+\eta\phi}\right)^n(1+\eta\phi)^{-1/\phi}, & \text{for } n = 1,2,3,\ldots, \end{cases}$$

where $\Gamma(\cdot)$ denotes the gamma function.

The survival function of the DZINB cure model has the form

$$S_{\text{pop}}(y) = \pi + (1-\pi)\left[1+\eta\,\phi\,p\,F(y)\right]^{-1/\phi}, \tag{3.6}$$

the cure fraction is $p_0 = \pi + (1-\pi)\left[1+\eta\,\phi\,p\right]^{-1/\phi}$, and the associated density function becomes

$$f_{\text{pop}}(y) = (1-\pi)\,\eta\,p\,f(y)\left[1+\eta\,\phi\,p\,F(y)\right]^{-(1/\phi)-1}. \tag{3.7}$$

The DZINB cure model in (3.6) with $\pi = 0$ reduces to the destructive negative binomial model [4], whereas the negative binomial cure rate model [5] is a special case of (3.6) when $\pi = 0$ and $p = 1$.

3.4. The destructive zero-inflated geometric (DZIG) cure model

Moreover, the destructive zero-inflated geometric (DZIG) cure rate model with parameter $\theta = \eta/(1+\eta)$ is one more important special case of (3.6) when $\phi = 1$ leading to

$$S_{\text{pop}}(y) = \pi + (1-\pi)\left[1+\eta\,p\,F(y)\right]^{-1}, \tag{3.8}$$

the cure fraction is $p_0 = \pi + (1-\pi)\left[1+\eta p\right]^{-1}$ and the density function reduces to

$$f_{\text{pop}}(y) = (1-\pi)\,\eta\,p\,f(y)\left[1+\eta\,p\,F(y)\right]^{-2}. \tag{3.9}$$

4. INFERENCE AND ESTIMATION

Here, we consider the situation when the time to event of interest is not completely observed and is subject to right censoring. Let C_i denote the censoring time. We observe $t_i = \min\{Y_i, C_i\}$ and $\delta_i = 1$ if Y_i is the observed time to the event defined before and $\delta_i = 0$ if it is right censored, for $i = 1, ..., n$. Let $\boldsymbol{\gamma}$ represent the parameter vector of the distribution for the unobserved lifetime in (2.5). Here, we note that the DZIPS cure rate models in Section 3 are unidentifiable according to [18]. So, to overcome this problem, we propose to relate the model parameters p and θ (or η) to covariates $\mathbf{x}_{i1} = (x_{i11}, x_{i12}, ..., x_{i1p_1})^\top$ and $\mathbf{x}_{i2} = (x_{i21}, x_{i22}, ..., x_{i2p_2})^\top$, respectively, without common elements and $\mathbf{x}_{i2}$ without a column of intercepts. Here, the systematic components are

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_{i1}^\top\boldsymbol{\beta}_1 \quad \text{and} \quad \log(\theta_i) = \mathbf{x}_{i2}^\top\boldsymbol{\beta}_2, \tag{4.1}$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, ..., \beta_{1p_1})^\top$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, ..., \beta_{2p_2})^\top$ represent the associated parameter vectors. A critical issue is the selection of covariates to be included in the link functions in (4.1). More precisely, given a link function and a set potential covariates, the problem is to find and fit the "best" model under a "selected" subset of covariates [3]. In fact, to choose which explanatory variables will be connected to the parameters p_i and θ_i is not an easy task because it depends on several factors such as the type of cancer, the covariates available in the study, patient history, etc. It is always important to work together with the medical team to take any kind of decision. Moreover, for readers interested in this discussion, we suggest [12] and [9].

From n pairs of times and censoring indicators $(y_1, \delta_1), ..., (y_n, \delta_n)$, the observed full likelihood function under non-informative censoring can be expressed as

$$L(\boldsymbol{\nu}, \mathbf{D}) \propto \prod_{i=1}^{n} \left\{ f_{\text{pop}}(t_i; \boldsymbol{\nu}) \right\}^{\delta_i} \left\{ S_{\text{pop}}(t_i; \boldsymbol{\nu}) \right\}^{1-\delta_i}, \tag{4.2}$$

where $\boldsymbol{\nu} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\gamma}^\top)^\top$, $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{t} = (t_1, ..., t_n)$, $\mathbf{x}_1 = (\mathbf{x}_{11}, ..., \mathbf{x}_{n1})$, $\mathbf{x}_2 = (\mathbf{x}_{21}, ..., \mathbf{x}_{n2})$, and $f_{\text{pop}}(\cdot\,; \boldsymbol{\nu})$ and $S_{\text{pop}}(\cdot\,; \boldsymbol{\nu})$ are defined in Equations (2.7) and (2.6), respectively.

Next, we assume a Weibull distribution for the observed lifetime in (2.5) with cdf and pdf (for $z > 0$)

$$F(z; \boldsymbol{\gamma}) = 1 - \exp(-z^{\gamma_1} \mathrm{e}^{\gamma_2}) \qquad \text{and} \qquad f(z; \boldsymbol{\gamma}) = \gamma_1 \, z^{\gamma_1 - 1} \exp(\gamma_2 - z^{\gamma_1} \mathrm{e}^{\gamma_2}),$$

respectively, $\boldsymbol{\gamma}^\top = (\gamma_1, \gamma_2)^\top$, $\gamma_1 > 0$ and $\gamma_2 > 0$. The choice of the Weibull distribution is due to the fact that this lifetime distribution is a very popular model and it has been extensively used over the past decades for modeling data in reliability, engineering and biological studies. Also, the pdf and cdf of the Weibull distribution have closed-forms which provide simple expressions for its survival and hazard functions.

The ML estimation of the parameter vector $\boldsymbol{\nu}$ can be implemented by numerical maximization of the log-likelihood function $\ell(\boldsymbol{\nu}, \mathbf{D}) = \log L(\boldsymbol{\nu}, \mathbf{D})$ using `R` software. Further, confidence intervals and hypothesis tests can be based on the large sample normal distribution of the maximum likelihood estimator (MLE) with the variance-covariance matrix given by the inverse of the Fisher information. More specifically, under conditions that are fulfilled for the parameter vector $\boldsymbol{\nu}$ in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}\,(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu})$ is multivariate normal $N_{p_1+p_2+2}\big(0, K(\boldsymbol{\nu})^{-1}\big)$, where $K(\boldsymbol{\nu})$ is the information matrix. The asymptotic covariance matrix $K(\boldsymbol{\nu})^{-1}$ of $\widehat{\boldsymbol{\nu}}$ can be approximated by the inverse of the $(p_1 + p_2 + 2) \times (p_1 + p_2 + 2)$ observed information matrix $-\ddot{\mathbf{L}}(\boldsymbol{\nu}, \mathbf{D})$. The elements of the observed information matrix $-\ddot{\mathbf{L}}(\boldsymbol{\nu}, \mathbf{D})$ are calculated numerically. The approximate multivariate normal distribution $N_{p_1+p_2+2}\big(0, -\ddot{\mathbf{L}}(\boldsymbol{\nu}, \mathbf{D})^{-1}\big)$ for $\widehat{\boldsymbol{\nu}}$ can be used in the classical way to construct approximate confidence regions for some parameters in $\boldsymbol{\nu}$. Also, we can use the likelihood ratio (LR) statistic for comparing some special models with the DZIPS regression model.

5. SIMULATION STUDY

In this section, we conduct a simulation study in order to evaluate some properties of the MLEs. For each individual i $(i = 1, ..., n)$, the number of competing risks of the event of interest N is generated from the ZIP and ZINB distributions given in (3.1) and (3.6), respectively. We assume covariates x_{i11} and x_{i21} generated from a Bernoulli distribution with parameter 0.5 and exponential distribution with parameter one, respectively. Also, we consider the systematic components

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_{10} + \beta_{11}\, x_{i11} \qquad \text{and} \qquad \log(\psi_i) = \beta_{21}\, x_{i21}\,, \tag{5.1}$$

where ψ_i is the parameter θ_i and η_i in the DZIP and DZINB cure rate models, respectively.

We simulate from the DZIP cure fraction distribution with parameters $\pi = 0.25$, $\gamma_1 = 2$, $\gamma_2 = -0.5$, $\beta_{10} = -1$, $\beta_{11} = 0.5$ and $\beta_{21} = 1.25$; and from the DZINB distribution under two setups: the first assuming $\pi = 0.25$, $\gamma_1 = 2$, $\gamma_2 = -0.5$, $\phi = 1$, $\beta_{10} = -1.5, \beta_{11} = 0.75$ and $\beta_{21} = 1.5$ (DZIG distribution), and the second with $\pi = 0.25$, $\gamma_1 = 2$, $\gamma_2 = -0.5$, $\phi = 1.5$, $\beta_{10} = -0.5$, $\beta_{11} = 1.5$ and $\beta_{21} = 1.25$ (DZINB distribution). The censoring times are sampled from the uniform distribution in the $(0, \tau)$ interval, where τ controls the censoring proportion of the uncured population. Here, the proportions of censored observations are approximately 62%, 68% and 70%, respectively. The results are obtained from 1,000 Monte Carlo simulations where, in each replication, a random sample of size $n = 50, 100, 250, 500$ and 750 is drawn.

Table 1: Summaries of the performance of the DZIP cure model.

Sample size (n)	Parameter	Summaries of parameters			
		Mean	Bias	MSE	CP
50	π	0.2161	−0.0339	0.0265	0.930
	γ_1	2.2849	0.2849	0.2951	0.933
	γ_2	−0.5761	−0.0761	0.2439	0.930
	β_{10}	−0.8371	0.1629	5.7489	0.960
	β_{11}	1.3094	0.8094	19.3054	0.973
	β_{21}	1.4400	0.1900	0.1870	0.924
100	π	0.2302	−0.0198	0.0124	0.952
	γ_1	2.1194	0.1194	0.0970	0.931
	γ_2	−0.5096	−0.0096	0.0885	0.946
	β_{10}	−0.9897	0.0103	1.4746	0.957
	β_{11}	0.6473	0.1473	3.4111	0.966
	β_{21}	1.3148	0.0648	0.0516	0.954
250	π	0.2424	−0.0076	0.0053	0.924
	γ_1	2.0399	0.0399	0.0277	0.950
	γ_2	−0.5099	−0.0099	0.0325	0.960
	β_{10}	−1.0200	−0.0200	0.1691	0.947
	β_{11}	0.5799	0.0799	0.5453	0.953
	β_{21}	1.2507	0.0007	0.0165	0.959
500	π	0.2473	−0.0027	0.0023	0.947
	γ_1	2.0165	0.0165	0.0144	0.927
	γ_2	−0.4921	0.0079	0.0168	0.938
	β_{10}	−1.0035	−0.0035	0.0734	0.949
	β_{11}	0.5008	0.0008	0.0616	0.962
	β_{21}	1.2326	−0.0174	0.0074	0.956
750	π	0.2494	−0.0006	0.0015	0.939
	γ_1	2.0127	0.0127	0.0082	0.951
	γ_2	−0.4945	0.0055	0.0106	0.950
	β_{10}	−1.0048	−0.0048	0.0489	0.940
	β_{11}	0.5068	0.0068	0.0407	0.958
	β_{21}	1.2329	−0.0171	0.0049	0.958

Tables 1, 2 and 3 display the averages of the MLEs (mean), bias, mean square errors (MSE) and coverage probabilities (CP) for nominal 95% of the DZIP, DZIG and DZINB cure models, respectively. We conclude from these results that (for all parameters) the MSEs of the MLEs decay toward zero when the sample size increases, as expected under standard asymptotic theory. In fact, the estimates tend to be closer to the true parameter values and the CPs converge to the nominal level when the sample size n increases.

Table 2: Summaries of the performance of the DZIG cure model.

Sample size (n)	Parameter	Summaries of parameters			
		Mean	Bias	MSE	CP
50	π	0.2222	−0.0278	0.0353	0.939
	γ_1	2.3558	0.3558	0.4846	0.946
	γ_2	−0.5849	−0.0849	0.4056	0.914
	β_{10}	−1.3522	0.1478	7.4278	0.968
	β_{11}	1.9488	1.1988	27.0050	0.976
	β_{21}	1.7590	0.2590	0.5628	0.906
100	π	0.2467	−0.0033	0.0214	0.943
	γ_1	2.1990	0.1990	0.1914	0.929
	γ_2	−0.5569	−0.0569	0.1624	0.933
	β_{10}	−1.4941	0.0059	1.4062	0.962
	β_{11}	1.4290	0.6790	9.6464	0.977
	β_{21}	1.6210	0.1210	0.1626	0.937
250	π	0.2491	−0.0009	0.0086	0.940
	γ_1	2.0800	0.0800	0.0503	0.943
	γ_2	−0.5223	−0.0223	0.0568	0.945
	β_{10}	−1.4777	0.0223	0.2496	0.965
	β_{11}	0.8922	0.1422	1.0476	0.959
	β_{21}	1.5239	0.0239	0.0490	0.945
500	π	0.2540	0.0040	0.0042	0.933
	γ_1	2.0324	0.0324	0.0210	0.951
	γ_2	−0.5020	−0.0020	0.0230	0.959
	β_{10}	−1.4628	0.0372	0.1143	0.963
	β_{11}	0.7960	0.0460	0.1245	0.962
	β_{21}	1.4891	−0.0109	0.0223	0.939
750	π	0.2556	0.0056	0.0033	0.921
	γ_1	2.0276	0.0276	0.0144	0.938
	γ_2	−0.5092	−0.0092	0.0183	0.932
	β_{10}	−1.4529	0.0471	0.0842	0.944
	β_{11}	0.7859	0.0359	0.0848	0.958
	β_{21}	1.4873	−0.0127	0.0153	0.938

Table 3: Summaries of the performance of the DZINB cure model.

Sample size (n)	Parameter	Summaries of parameters			
		Mean	Bias	MSE	CP
50	π	0.0976	−0.1524	0.0436	0.995
	γ_1	2.9864	0.9864	2.1222	0.969
	γ_2	−0.8296	−0.3296	0.8123	0.931
	ϕ	2.9119	2.4119	14.9794	0.999
	β_{10}	0.1421	1.1421	23.6676	0.983
	β_{11}	3.3878	2.6378	67.0179	0.992
	β_{21}	2.1163	0.8663	2.0872	0.966
100	π	0.1392	−0.1108	0.0331	0.984
	γ_1	2.4785	0.4785	0.6232	0.970
	γ_2	−0.6812	−0.1812	0.3434	0.942
	ϕ	1.9185	1.4185	6.3233	1.000
	β_{10}	−0.7544	0.2456	6.3957	0.966
	β_{11}	2.4230	1.6730	26.6289	0.984
	β_{21}	1.6889	0.4389	0.5671	0.976
250	π	0.2005	−0.0495	0.0155	0.975
	γ_1	2.1611	0.1611	0.1189	0.977
	γ_2	−0.5978	−0.0978	0.1137	0.969
	ϕ	1.0213	0.5213	1.1649	1.000
	β_{10}	−0.9418	0.0582	0.6769	0.961
	β_{11}	1.2954	0.5454	5.1548	0.982
	β_{21}	1.3871	0.1371	0.0895	0.986
500	π	0.2362	−0.0138	0.0066	0.956
	γ_1	2.0704	0.0704	0.0408	0.970
	γ_2	−0.5519	−0.0519	0.0576	0.960
	ϕ	0.6995	0.1995	0.4031	0.996
	β_{10}	−0.9677	0.0323	0.1758	0.962
	β_{11}	0.9161	0.1661	1.0521	0.975
	β_{21}	1.3053	0.0553	0.0342	0.977
750	π	0.2468	−0.0032	0.0037	0.953
	γ_1	2.0480	0.0480	0.0261	0.964
	γ_2	−0.5188	−0.0188	0.0346	0.960
	ϕ	0.5932	0.0932	0.1934	0.978
	β_{10}	−0.9744	0.0256	0.1111	0.957
	β_{11}	0.8271	0.0771	0.2584	0.975
	β_{21}	1.2754	0.0254	0.0199	0.971

6. APPLICATION: CUTANEOUS MELANOMA DATA

In this section, we illustrate the usefulness of the DZIPS cure rate regression with an application to a real data set on cancer recurrence. The data are part of a study on cutaneous melanoma (a type of malignant cancer) extracted from [25] on 205 patients observed for the evaluation of postoperative in the period from 1962 to 1977. The cutaneous melanoma data contain information about the survival times of patients after surgery for malignant melanoma which were collected at Odense University Hospital [13].

In general, the standard treatment of cutaneous melanoma consists of broad excision of primary tumor or cicatrices at a distance of at least 5 cm down to the fascia, though not including this. On the face the tumor was removed at a distance of only 2 cm. Lymphonodectomy was only undertaken when lymph nodes were clinically suspected. The clinical data and follow-up were based on information from the case histories of the patients. For more details, see [13].

The observed survival time range, approximately from 0 to 15 years (with mean equal to 5.9 years), refers to the time until the patient's death or the censoring time. There are 72% of censoring, corresponding to the patients which had died from other causes or were still alive at the end of the study. The following variables involved in the study for each patient are: y_i: observed time (in years), x_{i11}: tumor thickness (in mm, mean = 2.92 and standard deviation = 2.96) and x_{i21}: ulceration status (absent, $n = 115$; present, $n = 90$). As we mentioned earlier, the identifiability issue is avoided if the parameter p is linked only to tumor thickness, while the parameter θ (or η) is linked to the ulceration status in the DZIP, DZINB and DZIG regressions. The survival function for these cure rate regressions are:

- **DZIP survival function**

$$S(y_i\,|\,\mathbf{x}_i) \;=\; \pi + (1-\pi)\,\exp\Big\{-\theta_i\,p_i\,\big[1-\exp\big(-y_i^{\gamma_1}\,\mathrm{e}^{\gamma_2}\big)\big]\Big\}\,,$$

  where

$$p_i = \frac{\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)}{1+\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)} \qquad \text{and} \qquad \theta_i = \exp\big(\beta_{20}+\beta_{21}\,x_{i21}\big)\,.$$

- **DZINB survival function**

$$S(y_i\,|\,\mathbf{x}_i) \;=\; \pi + (1-\pi)\,\Big\{1+\eta_i\,\phi\,p_i\,\big[1-\exp\big(-y_i^{\gamma_1}\,\mathrm{e}^{\gamma_2}\big)\big]\Big\}^{-1/\phi},$$

  where

$$p_i = \frac{\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)}{1+\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)} \qquad \text{and} \qquad \eta_i = \exp\big(\beta_{20}+\beta_{21}\,x_{i21}\big)\,.$$

- **DZIG survival function**

$$S(y_i\,|\,\mathbf{x}_i) \;=\; \pi + (1-\pi)\,\Big\{1+\eta_i\,p_i\,\big[1-\exp\big(-y_i^{\gamma_1}\,\mathrm{e}^{\gamma_2}\big)\big]\Big\}^{-1},$$

  where

$$p_i = \frac{\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)}{1+\exp\big(\beta_{10}+\beta_{11}\,x_{i11}\big)} \qquad \text{and} \qquad \eta_i = \exp\big(\beta_{20}+\beta_{21}\,x_{i21}\big)\,.$$

Figure 2(a) shows that the Kaplan–Meier survival function estimate confirms a plateau around 0.64 and this fact indicates the presence of a proportion of patients for whom the malignant melanoma will never occur again, and then, those patients can be considered as cured. Also, the empirical Kaplan–Meier curves stratified by ulceration status (upper: absent, lower: present) are displayed in Figure 2(b) and they reveal that the ulceration affects the lifetime of the patients with malignant melanoma.

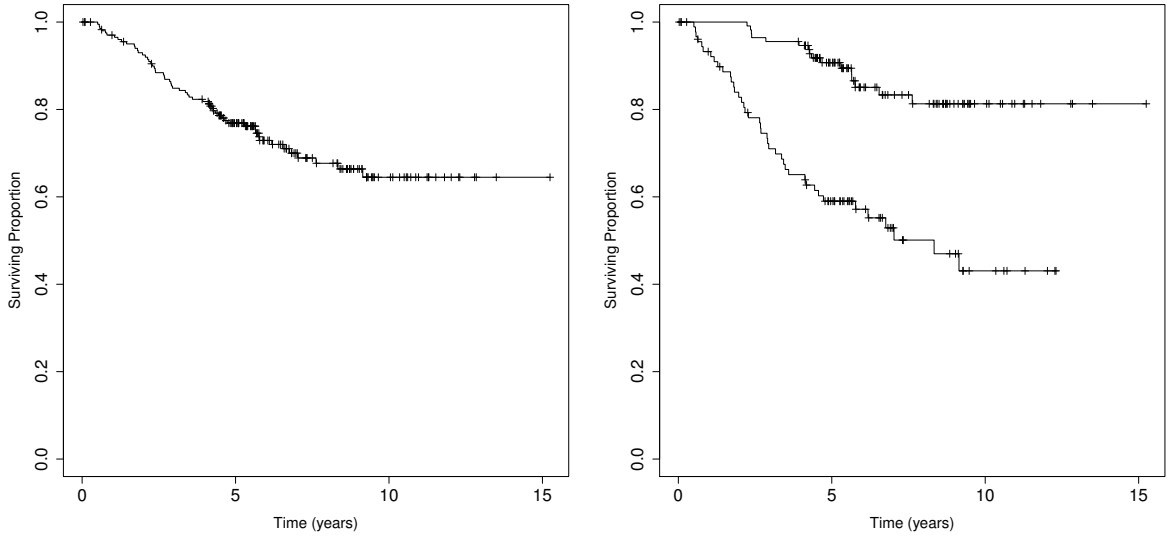

Figure 2: (a) Kaplan–Meier curve for the cutaneous melanoma data. (b) Kaplan–Meier curves stratified by ulceration status (upper: present, lower: absent).

For model comparison, we fit the DZIP, DZINB and DZIG cure models described in Section 3 to the cutaneous melanoma data. The special cases of these models were also fitted to these data, i.e., the Poisson ($\pi = 0$ and $p = 1$), the negative binomial ($\pi = 0$ and $p = 1$) and the geometric ($\pi = 0$, $p = 1$ and $\phi = 1$) models. We note that these special models belong to the PS cure models proposed by [4]. For these models, the destructive process is absent and consequently, the parameter θ (or η) is linked to both variables (ulceration status and tumor thickness). In order to compare the models, we use the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The results of the DZIPS cure models and its sub-models are reported in Table 4.

Table 4: The values of max log $L(\cdot)$, the AIC and BIC statistics for the Destructive Zero-Inflated Poisson (DZIP), Destructive Zero-Inflated Negative Binomial (DZINB), Destructive Zero-Inflated Geometric (DZIG), Poisson, negative binomial and geometric cure models.

Survival Cure Rate Model	max log $L(\cdot)$	AIC	BIC
Destructive Zero-Inflated Poisson	−201.18	416.3	439.6
Destructive Zero-Inflated Negative Binomial	−198.95	413.9	440.5
Destructive Zero-Inflated Geometric	−199.93	**413.8**	**437.1**
Poisson	−207.83	425.6	442.2
Negative Binomial	−201.52	423.0	439.7
Geometric	−205.42	420.8	437.4

According to the criteria in Table 4, the DZIG cure rate regression is the best model and so, it is selected as our working model. For this regression, we estimate the unknown parameters via ML method. All computations are performed using the R software.

The survival function for the DZIG cure rate regression is

$$S\Big(y_i;\widehat{\pi},\widehat{\gamma}_1,\widehat{\gamma}_2,\widehat{\boldsymbol{\beta}}_1,\widehat{\boldsymbol{\beta}}_2\Big) \;=\; \widehat{\pi}+(1-\widehat{\pi})\left\{1+\widehat{\eta}_i\,\widehat{p}_i\left[1-\exp\!\big(-y_i^{\widehat{\gamma}_1}\,\mathrm{e}^{\widehat{\gamma}_2}\big)\right]\right\}^{-1},$$

where

$$\widehat{p}_i=\frac{\exp\Big(\widehat{\beta}_{10}+\widehat{\beta}_{11}\,x_{i11}\Big)}{1+\exp\Big(\widehat{\beta}_{10}+\widehat{\beta}_{11}\,x_{i11}\Big)} \qquad \text{and} \qquad \widehat{\eta}_i=\exp\Big(\widehat{\beta}_{20}+\widehat{\beta}_{21}\,x_{i21}\Big)\,.$$

Here, for the cutaneous melanoma data set, the vectors $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are

$$\widehat{\boldsymbol{\beta}}_1=(\beta_{10},\beta_{11})^{\top} \qquad \text{and} \qquad \widehat{\boldsymbol{\beta}}_2=(\beta_{20},\beta_{21})^{\top}.$$

Table 5 gives the MLEs of the parameters, their standard errors and p-values from the fitted regression. We note from the fitted DZIG cure rate regression that ulceration status and tumor thickness are significant sloppy 1% and there is a significant difference for the presence or absent of ulceration status and also a difference related to the thickness of the tumor. Thus, those variables have influenced on the survival times of the patients. The estimate of the parameter π is 0.3895, and as mentioned earlier in Section 1, this indicates a proportion of those cells which never experience alterations/lesions. Consequently, the proportion of cells that were repaired by a repair system of the organism is $(1-\pi)=0.6105$ (or 61.05%).

Table 5: Results from the fitted DZIG cure rate regression.

Parameter	Estimate	Standard Error	p-value
γ_1	2.41	0.28	—
γ_2	−5.00	0.61	—
π	0.38	0.24	—
β_{10}	−4.41	0.93	<0.001
β_{11}	0.86	0.26	0.001
β_{20}	2.59	0.88	0.003
β_{21}	3.76	0.74	<0.001

Figure 3 displays the estimated survival function of the DZIG cure rate regression for patients with 0.320 mm, 1.940 mm and 4.254 mm tumor thickness, which correspond to the 5%, 50% and 80% tumor thickness quantiles. The survival rate decreases more rapidly for patients with thicker tumors in presence of ulceration. On the other hand, for patients with less thick tumor in presence of ulceration, the survival rate does not fall bellow 75% as shown in Figure 3(a).

Finally, we turn our attention to the role of the ulceration status and thickness tumor covariates on the estimation of the surviving fraction (p_0). To estimate the proportion of cured individuals, we use Equation (4.1) and the MLEs of the parameters. So, for the DZIG cure regression, the estimated cure fraction $\widehat{p}_0=\widehat{\pi}+(1-\widehat{\pi})\left[1+\widehat{\eta}\,\widehat{p}\right]^{-1}$ is 0.6450. This result is confirmed in Figure 2(a). Also, we note that the cure rate decreases when tumor thickness size increases and it is smaller for patients with presence of ulceration.

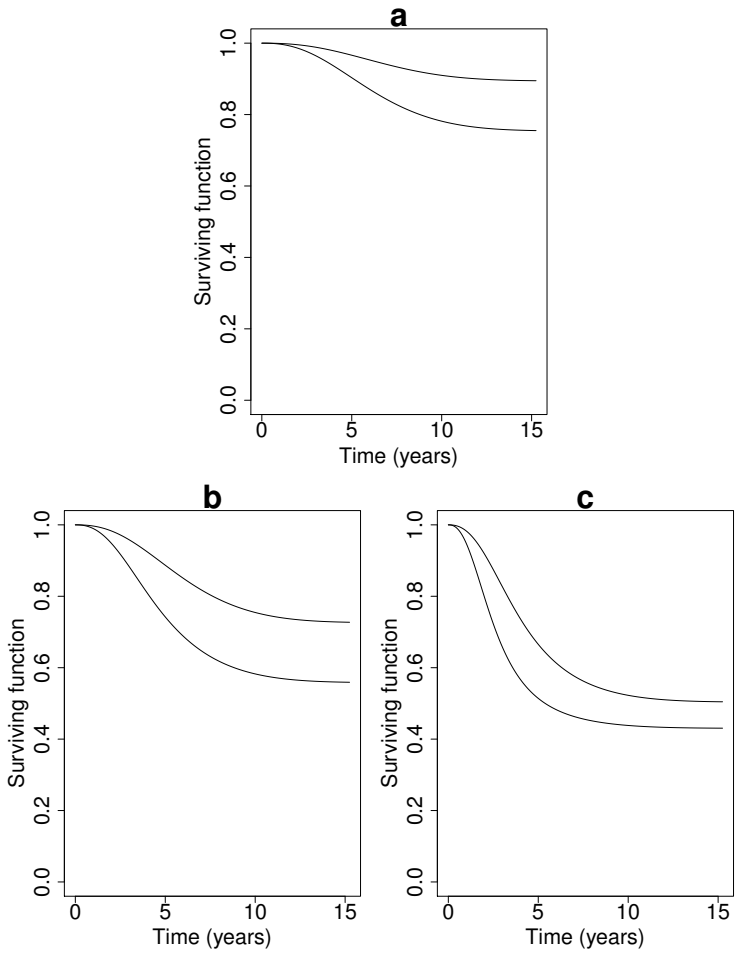

Figure 3: Estimated survival function from the DZIG cure rate regression stratified by ulceration status (upper: absent, lower: present) for patients with tumor thickness equal to: (a) 0.320 mm, (b) 1.940 mm, and (c) 4.254 mm.

7. CONCLUDING REMARKS

In this paper, we propose the destructive zero-inflated power series (DZIPS) family of cure rate models by extending the works of [23] and [4]. The DZIPS models are very flexible and contain special models such as the zero-inflated binomial (ZIBin), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated geometric (ZIG) models, among others. The proposed model allows estimation of the cure fraction by incorporating a systematic component to estimate the proportion of not initiated cells (those one that have never been altered/lesioned and those one that have been repaired). Hence, this extended family of models is very flexible in many practical situations. An application to a real cutaneous melanoma data set demonstrates that it can be used quite effectively to provide better interpretation for the underlying biological mechanism, in addition to offering a better fit than the other commonly used cure rate models.

ACKNOWLEDGMENTS

The authors are grateful to the editor and the referee for helpful comments and suggestions. We gratefully acknowledge grants from CNPq, Brazil.

REFERENCES

[1] Barral, A.M. (2001). *Immunological Studies in Malignant melanoma: importance of tnf and the thioredoxin system*, Doctorate Thesis, Linkoping University, Linkoping, Sweden.

[2] Berkson, J. and Gage, R.P. (1952). Survival curve for cancer patients following treatment, *Journal of the American Statistical Association*, **47**, 501–515.

[3] Borges, P.; Rodrigues, J. and Balakrishnan, N. (2012). Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data, *Computational Statistics and Data Analysis*, **56**, 1703–1713.

[4] Cancho, V.G.; Bandyopadhyay, D.; Louzada, F. and Yiqi, B. (2013). The destructive negative binomial cure rate model with a latent activation scheme, *Statistical Methodology*, **13**, 48–68.

[5] Cancho, V.G.; Rodrigues, J. and de Castro, M. (2011). A flexible model for survival data with a cure rate: a Bayesian approach, *Journal of Applied Statistics*, **38**, 57–70.

[6] de Castro, M.; Cancho, V.G. and Rodrigues, J. (2009). A Bayesian long-term survival model parametrized in the cured fraction, *Biometrical Journal*, **51**, 443–455.

[7] Chen, M.H.; Ibrahim, J.G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association*, **94**, 909–919.

[8] Clegg, L.X.; Feuer, E.J.; Midthune, D.N.; Fay, M.P. and Hankey, B.F. (2002). Impact of reporting delay and reporting error on cancer incidence rates and trends, *Journal of the National Cancer Institute*, **94**, 1537–1545.

[9] Collet, D. (1994). *Modelling Survival Data in Medical Research*, Chapman and Hall, London.

[10] Cooner, F.; Banerjee, S.; Carlin, B. and Sinha, D. (2007). Flexible cure rate modelling under latent activation schemes, *Journal of the American Statistical Association*, **102**, 560–572.

[11] Cox, D. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall, London.

[12] Draper, N.R. and Smith, H. (1984). *Applied Regression Analysis*, John Wiley and Sons, New York.

[13] Drzewiecki, K.T.; Ladefoged, C. and Christensen, H.E. (1980). Biopsy and prognosis for cutaneous malignant melanomas in clinical stage I, *Scandinavian Journal of Plastic and Reconstructive Surgery*, **14**, 141–144.

[14] Gordon, N.H. (1990). Application of the theory of finite mixtures for the estimation of cure rates of treated cancer patients, *Statistics in Medicine*, **9**, 397–407.

[15] Gupta, P.L.; Gupta, R.L. and Tripathi, R.C. (1995). Inflated modified power series distributions with applications, *Communications in Statistics – Theory and Methods*, **24**, 2355–2374.

[16] Ibrahim, J.G.; Chen, M.H. and Sinha, D. (2001). *Bayesian Survival Analysis*, Springer-Verlag, New York.

[17] Instituto Nacional do Câncer. *Tipos de câncer: melanoma cutâneo*, Rio de Janeiro: INCA, 2020. https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-melanoma

[18] Li, C.S.; Taylor, J. and Sy, J. (2001). Identifiability of cure models, *Statistics and Probability Letters*, **54**, 389–395.

[19] Maller, R.A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*, John Wiley and Sons, New York.

[20] Noack, A. (1950). On a class of discrete random variables, *Annals of Mathematical Statistics*, **21**, 127–132.

[21] Ortega, E.M.M.; Cordeiro, G.M.; Campelo, A.K.; Kattan, M.W. and Cancho, V.G. (2015). A power series beta Weibull regression model for predicting breast carcinoma, *Statistics in Medicine*, **44**, 1366–1388.

[22] Pescim, R.R.; Ortega, E.M.M.; Suzuki, A.K.; Cancho, V.G. and Cordeiro, G.M. (2019). A new destructive Poisson odd log-logistic generalized half-normal cure rate model, *Communication in Statistics – Theory and Methods*, **48**, 2113–2128.

[23] Rodrigues, J.; de Castro, M.; Balakrishnan, N. and Cancho, V.G. (2011). Destructive weighted Poisson cure rate models, *Lifetime Data Analysis*, **17**, 333–346.

[24] Rodrigues, J.; de Castro, M.; Cancho, V.G. and Louzada-Neto, F. (2009). On the unification of the long-term survival models, *Statistics and Probability Letters*, **79**, 753–759.

[25] Scheike, T. (2009). Timereg Package. R Package Version 3.4.0. With contributions from T. Martinussen and J. Silver. R package version 1.1-6.

[26] Tsodikov, A.D.; Ibrahim, J.G. and Yakovlev, A.Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models, *Journal of the American Statistical Association*, **98**, 1063–1078.

[27] Yakovlev, A.Y. and Tsodikov, A.D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, World Scientific, Singapore.

[28] Yang, G.L. and Chen, C.W. (1991). A stochastic two-stage carcinogenesis model: a new approach to computing the probability of observing tumor in animal bioassays, *Mathematical Biosciences*, **104**, 247–258.

[29] Zhao, Y.; Lee, A.H.; Yau, K.K.W. and Burke, V. (2009). A score test for assessing the cured proportion in the long-term survivor mixture model, *Statistics in Medicine*, **28**, 3454–3466.