Scilight

Journal of Modern Applied Statistical Methods
https://jmasm.com

Article

The Analysis of Labor Time and Healthcare Management via Generalized Gamma Zero-Inflated Cure-Rate Regression Model

Gleici da Silva Castro Perdoná [1,*], Hayala Cristina Cavenague de Souza [1], Francisco Louzada [2], Adesina Akintan [3], Hadiza Absulaziz Idris [4], Josaphat Byamugisha [5] and Lawal Olawale Oyeneyin Oyeneying [6]

[1] Department of Social Medicine, Ribeirão Preto School of Medicine, University of São Paulo, São Paulo 14040-900, Brazil
[2] Institute of Mathematical Science and Computing, University of São Paulo, São Paulo 13566-590, Brazil
[3] Department of Obstetrics and Gynecology, Mother and Child Hospital, Akure 340001, Nigeria
[4] Nyanya General Hospital, Abuja 900101, Nigeria
[5] Department of Gynecology and Obstetrics School of Medicine, Makerere University College of Health Sciences, Kampala P.O. Box 7072, Uganda
[6] Department of Obstetrics and Gynecology, University of Medical Sciences Teaching Hospital Medical Village, Akure 351101, Nigeria
* Correspondence: pgleici@fmrp.usp.br

How To Cite: Perdoná, G.d.S.C.; de Souza, H.C.C.; Louzada, F.; et al. The Analysis of Labor Time and Healthcare Management via Generalized Gamma Zero-Inflated Cure-Rate Regression Model. *Journal of Modern Applied Statistical Methods* **2026**, *25*(1), 6. https://doi.org/10.53941/jmasm.2026.100006

Abstract: Standard survival models assume that time is greater than zero. However, there are several practical examples where a proportion of the data presents times equal to zero. Based on this issue, survival models including this proportion have been proposed and are called zero-inflated survival models. It is important to consider flexible distributions in this context in order to model more complex patterns. In this paper, we propose a Generalized Gamma zero-inflated cure-rate survival model, motivated by the pattern observed in the labor progression times. Based on simulation study, we show that estimation methods (maximum likelihood estimates and asymptotic confidence intervals) present a good performance even for small sample sizes. Concerning the model selection, we verified that Likelihood Ratio Test showed the best results. The proposed model was considered to analyze the labor time of pregnant women from sub-Saharan Africa. For diagnostic analysis, we used Cox-Snell residuals and Local Influence methods. In general, the model showed as an adequate tool to describe the labor and we conclude that this model can be a tool in the study of childbirth times supporting management in obstetrical healthcare. We acknowledge the World Health Organization for granting us permission to use the data set.

Keywords: cure-rate; duration of labor; flexible distributions; survival analysis; zero-inflation

1. Introduction

Survival analysis is an area of statistics that mainly studies the time until the occurrence of a particular event of interest, with the advantage that the incomplete (censored) times can be included in the analysis. The usual survival models consider that the time should be greater than zero, but there are diverse situations already described in literature in which we can find proportion of individuals with times equal to zero, which we can call zero-inflation. The zero-inflation is a property that occurs when the studied variable presents more zero values than expected and cannot be fitted by usual probability distributions. This property has been extensively studied, especially in the context of count data, for which we can find proposed distributions since the 1960s [1–3] that were extended by many other authors [4–6]. Several practical applications for these models are found in the medical field [7–9]. Concerning the zero-inflation for continuous variables, there is also a large number of situations for which specific

models have been proposed [10–14].

Particularly for survival analysis, there is a lack of proposals that allow the variable time to assume values equal to zero. Braekers and Grouwels [15] present an extension of Cox Proportional Hazards models applied to mice laboratory studies and Garcia [16] presents a Discrete Weibull model in the context of survival after heart surgery. Recently, authors [17–19] proposed extensions for the mixture cure model [20] and promotion cure model [21], respectively, considering Weibull distribution in the financial area. Finally, Calsavara et al. [22] proposed a defective model to deal with the presence of times equal to zero in two medical data sets. Besides including the proportion of zero-inflation, the last three models also include the long-term proportion (or cure-rate), which occurs when a group of subjects do not present the event of interest, although after a long follow-up, and has become very common in survival data sets of several areas [23].

In terms of risk function, which describes the instantaneous risk of event occurrence, the Weibull distribution considered in proposed zero-inflated cure-rate (ZICR) models [17, 18, 24] is restrictive because it allows only increasing, decreasing or constant functions. Thus, it is important to consider more flexible distributions for the baseline model in order to describe a bigger range of patterns that can be seen in the data.

1.1. Motivation Problem: The Study of Duration of Labor for Supporting Obstetric Healthcare

The health of pregnant women and their children is one of the pillars of public health in the world [25]. The poor quality of care becomes one of the greatest obstacles in reducing preventable deaths [26]. This perception leads to a global shift towards investment in quality of care during labor as the most impactful and cost-effective strategy to save millions of lives in the coming years [27]. The insufficient supply of beds is one of the points that contribute to a poor quality of care. Hospital overcrowding can lead to delays in patient admission, especially in emergencies. Such delays can intensify the occurrence of adverse outcomes during periods of overcrowding [28]. In such cases, good healthcare management is essential and to study the time between hospital admission and birth can help in improve this management.

In this context, the Word Health Organization (WHO) developed a project called Better Outcomes in Labor Difficulty (BOLD), which aimed to improve the quality of intrapartum care by developing a cohort study. BOLD researchers collected data from women giving birth in 9 and 4 hospitals in Nigeria and Uganda, respectively, from 2014 to 2015. They considered eligible women admitted for spontaneous or induced vaginal delivery, with a single foetus, during the first stage of labour, with cervical dilatation less than 7 cm [29–31]. In terms of the duration of labor (i.e., the time between hospital admission and vaginal birth), three different women groups can be observed:

(1) Women who arrive at the hospital already having had a stillbirth (fetal death);
(2) Women whose natural time to child birth is accelerated by an intervention (augmentation of labor) or who may not undergo vaginal birth due to a cesarean section.
(3) Women that undergo vaginal birth without fetal death or intervention.

Then, as it is not possible to observe the duration of labor, the time for women in Group 1 is equal to zero, while the time for women in the second group is positive, but incomplete (censored), because the intervention (cesarean or augmentation) decreases the natural labor pattern and women in Group 3 present the complete duration of labor. Given the characteristics observed in the three groups of BOLD data set, we believe that the class of the ZICR models its a important tool to describe the duration of labor. In this case, the cure proportion is related to those women in whom censorship (intervention) occurs after a long time of labor. It is important to mention that this study excluded women with elective or pre-labour cesarean section. Thus, the decision about cesarean section, as all other interventions, occurs during the labor period due clinical needs.

Therefore, the main objective of this paper is to present a version of the ZICR model [17], considering a more flexible baseline distribution: the Generalized Gamma and to evaluate the properties of the inference methods for this model. As described above, monitoring hospitals is of great importance in health management. So, an additional objective of this paper is to present a real application in the BOLD dataset, showing how the proposed model can be a useful tool in this context, particularly when comparing different facilities. All analyses were performed using the R software [32].

2. Materials and Methods

2.1. The Generalized Gamma Zero-Inflated Cure Rate Model

The GG-ZICR model is obtained when the Generalized Gamma (GG) distribution is associated as the distribution for the subjects susceptible to failure, which means that GG distribution represents the survival behavior of the non-negative random variable T. Following the parametrization suggested by Meeker [33], the

survival function (Equation (1)) and probability density function (pdf) (Equation (2)) of the GG distribution are, respectively, given by

$$S_{GG}(t) = 1 - \Gamma_1\left\{\lambda^{-2}\exp(\lambda w); \lambda^{-2}\right\} \tag{1}$$

and

$$f_{GG}(t) = \frac{\lambda}{\sigma t}\frac{\exp\left[\lambda^{-2}\left\{\lambda w + \log(\lambda^{-2})\right\} - \exp\left\{\lambda w + \log(\lambda^{-2})\right\}\right]}{\Gamma(\lambda^{-2})}, \tag{2}$$

where, $-\infty < \mu < \infty$, $\sigma > 0$ and $\lambda > 0$, $t > 0$, $w = \frac{\log(t)-\mu}{\sigma}$ and $\Gamma_1((\lambda, w); \lambda)$ is the incomplete gamma function. As an example of the flexibility of GG distribution, Figure 1a presents examples of its hazard function for different parameter values.

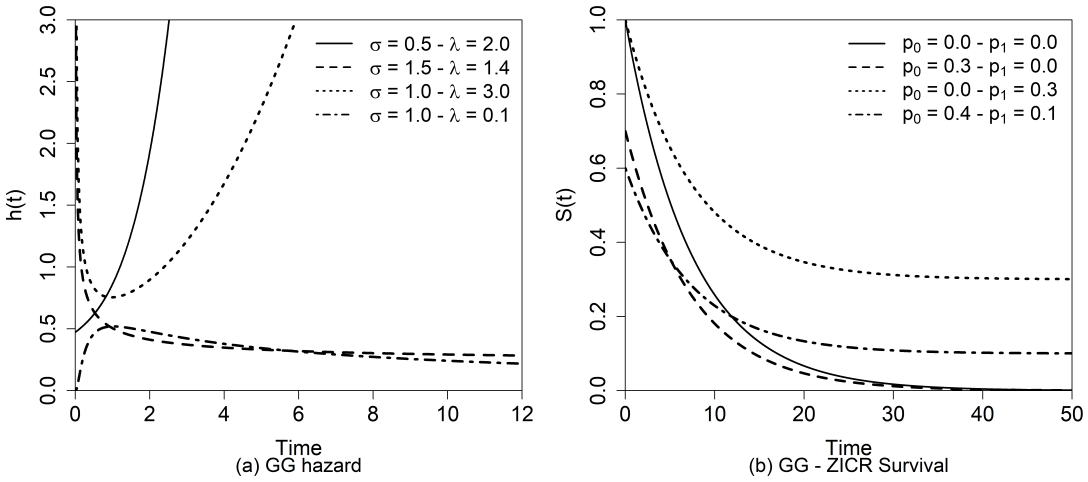

Figure 1. Examples of GG hazard shapes for different values of σ and λ (**a**) and GG-ZICR survival shapes for different values of p_0 and p_1 (**b**).

Then, based on the model proposed by Louzada et al. [17], the survival and probability sub-density function GG-ZICR model are given by

$$S(t) = p_1 + (1 - p_0 - p_1)\left[1 - \Gamma_1\left\{\lambda^{-2}\exp(\lambda w); \lambda^{-2}\right\}\right] \tag{3}$$

and

$$f(t) = p_0 I_{\{0\}}(t) + (1 - p_1 - p_0)\frac{\lambda}{\sigma t}\frac{\exp\left[\lambda^{-2}\left\{\lambda w + \log(\lambda^{-2})\right\} - \exp\left\{\lambda w + \log(\lambda^{-2})\right\}\right]}{\Gamma(\lambda^{-2})} I_{\{\mathbb{R}^+\}}(t), \tag{4}$$

respectively, where $p_1 \in (0, 1)$ is the proportion of subjects immune to the event and, $p_0 \in (0, 1 - p_1)$ is the proportion of individuals with time equal to zero.

Depending on the values of the parameters $\mu, \sigma, \lambda, p_0$ and p_1, the GG-ZICR distribution reduces to other proposed distributions, such as the Weibull ZICR [17] ($\lambda = 1$), Log Normal ZICR ($\lambda \to 0$) and the Gamma ZICR ($\lambda = \sigma$), as well as the particular cases without cure-rate or zero-inflation, if $p_1 = 0$ or $p_0 = 0$, respectively. This is an important advantage, because it is possible to fit the GG-ZICR model and then evaluate if a particular case presents a better fit by evaluating the parameter estimates and using hypothesis testing or comparison methods.

Figure 1b presents survival shape examples of Model (3) for different values of p_0 and p_1. It can be observed that when $p_0 > 0$, $S(t = 0) < 1$ and, when $p_1 > 0$, $S(t \to \infty) > 0$, which characterize an improper survival function [23].

2.1.1. Likelihood Function

The point estimation is maximum likelihood based. The value of likelihood contribution depends on the group of the subject, as follows:

- p_0, if $t_i = 0$
- $p_1 + (1 - p_0 - p_1)S_{GG}(t_i)$ if $t_i > 0$ and censored
- $(1 - p_0 - p_1)f_{GG}(t_i)$ if $t_i > 0$ and not censored.

For $t > 0$, we assumed the independent random censoring process, the observed data is given by the pair

(t_i, δ_i), where $\delta_i = 0$, if the ith subjetct is censored and $\delta_i = 1$, otherwise. Then, the Likelihood function for $\boldsymbol{\vartheta} = (p_0, p_1, \mu, \sigma, \lambda)$, corresponding to the observed data is given by Equation (5)

$$\begin{aligned} L(\boldsymbol{\vartheta}, t_i, \delta_i) &= \prod_{i:\, t_i=0} p_0 \prod_{i:\, t_i>0} \left[\{(1-p_0-p_1) f_{GG}(t_i)\}^{\delta_i} \{p_1 + (1-p_0-p_1) S_{GG}(t_i)\}^{1-\delta_i}\right] \\ &= p_0{}^m \prod_{i:\, t_i>0} \left[\{(1-p_0-p_1) f_{GG}(t_i)\}^{\delta_i} \{p_1 + (1-p_0-p_1) S_{GG}(t_i)\}^{1-\delta_i}\right] \end{aligned} \tag{5}$$

where $m(< n)$ is the number of subjects with $t = 0$ and the logarithmic likelihood is given by

$$\begin{aligned} l(\boldsymbol{\vartheta}, t_i, \delta_i) &= \log\{L(\boldsymbol{\vartheta}, t_i, \delta_i)\} \\ &= m \log(p_0) + \sum_{i:\, t_i>0} \delta_i \log(1-p_0-p_1) \\ &\quad + \sum_{i:\, t_i>0} \delta_i \log\left(\frac{\lambda}{\sigma t} \frac{\exp\left[\lambda^{-2}\left\{\lambda w + \log(\lambda^{-2})\right\} - \exp\left\{\lambda w + \log(\lambda^{-2})\right\}\right]}{\Gamma(\lambda^{-2})}\right) \\ &\quad + \sum_{i:\, t_i>0} (1-\delta_i) \log\left(p_1 + [1-p_0-p_1]\left[1 - \Gamma_1\left\{\lambda^{-2}\exp(\lambda w_i); \lambda^{-2}\right\}\right]\right) \end{aligned} \tag{6}$$

where $w_i = \frac{\log(t_i)-\mu}{\sigma}$, $\delta_i = 0$ if $T_i < C_i$ (censored time) and $\delta_i = 1$, otherwise.

Maximum Likelihood Estimates (MLE) are obtained by solving the non-linear system of equations $U(\boldsymbol{\vartheta} = (p_0, p_1, \mu, \sigma, \lambda)) = \left(\frac{\partial l(\boldsymbol{\vartheta})}{\partial p_0}, \frac{\partial l(\boldsymbol{\vartheta})}{\partial p_1}, \frac{\partial l(\boldsymbol{\vartheta})}{\partial \mu}, \frac{\partial l(\boldsymbol{\vartheta})}{\partial \sigma}, \frac{\partial l(\boldsymbol{\vartheta})}{\partial \lambda}\right) = 0$. Due to the difficulty in solving this system arithmetically, iterative techniques can be used. In this paper, the *optim()* routine with method = "BFGS", a quasi-Newton method [34] , available on the statistical software R [32], was considered.

As noted above, the GG-ZICR model extends to a variety of ZICR models. Thus, when the parameters of the GG ZICR model are estimated, it can be evaluated if any particular case fits better to the data set by using model selection methods, as hypothesis tests and comparison criteria. Thus, in this paper, we will consider the Likelihood Ratio Test (LRT) and the following comparison methods: Akaike Information Criterion (AIC) [35], Consistent AIC (cAIC) [36,37] and Bayesian Information Criterion (BIC) [38].

The LRT consists of comparing the maximum logarithmic likelihood value under the null hypothesis, $l(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{H0}})$, with the maximum likelihood value under the alternative hypothesis, $l(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{H1}})$. The test statistic is the deviance, given by $d = -2\left\{l(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{H0}}) - l(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{H1}})\right\}$ and its compared with a Chi-squared distribution, where the degrees of freedom are equal to the dimension of the parameter to be tested. Concerning the comparison methods, AIC, BIC and cAIC for a specific model M are given by: $AIC = 2k - 2\log\left\{L(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{M}})\right\}$, $BIC = k\log(n) - 2\log\left\{L(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{M}})\right\}$ and $cAIC = k[log(n)+1] - 2\log\left\{L(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{M}})\right\}$, where $L(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{M}})$ is the maximum likelihood, k is the number of parameters for the model M and n is the sample size.

2.1.2. The Regression Model

Regression models can be used when the aim is to access how independent variables are associated with the final outcome and are an important tool in practice as it helps researchers to verify the impact of one or more individual features in the survival, supporting decision-making. In this Section, we propose to link all the parameter of model (3) with a set of $k(< n)$ independent variables, $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k)$, the p_0 and p_1 were modeled by logistic link for all fits. This approach allows for the incorporation of individual-specific covariates into the model while still retaining the flexibility of the model. In our application, demographic characteristics of the patient were considered to motivate the investigation of how these factors influence the cure fraction and the probability of instantaneous events, zero inflation. Then, the regression version GG-ZICR model is given by the survival function

$$S_i(t) = p_{1i} + (1-p_{0i}-p_{1i})\left(1 - \Gamma_1\left[\lambda_i{}^{-2}\exp\left\{\lambda_i \frac{\log(t)-\mu_i}{\sigma_i}\right\}; \lambda_i{}^{-2}\right]\right). \tag{7}$$

The link functions are defined according the parametric space of each parameter and are given by

$$\begin{cases} \mu_i & = \eta_{1i}, \\ \log(\sigma_i) & = \eta_{2i}, \\ \log(\lambda_i) & = \eta_{3i}, \\ \left(\log\left(\frac{p_{0i}}{1-p_{0i}-p_{1i}}\right), \log\left(\frac{p_{1i}}{1-p_{0i}-p_{1i}}\right)\right) & = (\eta_{4i}, \eta_{5i}), \end{cases} \tag{8}$$

where the linear predictors for the ith individual are given by

$$\eta_{ji} = \beta_{j0} + \beta_{j1}x_{1i} + \beta_{j1}x_{2i} + ... + \beta_{jk}x_{ki} = \beta_{j0} + \sum_{c=1}^{k} \beta_{jc}x_{ci}, \tag{9}$$

for p_0 and p_1

$$p_{0i} = \frac{exp(\beta_{40} + \sum_{c=1}^{k} \beta_{4c}x_{ci})}{[1 + exp(\beta_{40} + \sum_{c=1}^{k} \beta_{4c}x_{ci}) + exp(\beta_{50} + \sum_{c=1}^{k} \beta_{5c}x_{ci})]} \tag{10}$$

$$p_{1i} = \frac{exp(\beta_{50} + \sum_{c=1}^{k} \beta_{5c}x_{ci})}{[1 + exp(\beta_{40} + \sum_{c=1}^{k} \beta_{4c}x_{ci}) + exp(\beta_{50} + \sum_{c=1}^{k} \beta_{5c}x_{ci})]} \tag{11}$$

where x_{ci} is the value of the variable c-th for the individual i-th, $j = 1, 2, 3, 4, 5$, $c = 1, 2, ..., k$ and $i = 1, 2, ..., n$. The linear coefficients β_{jc} are also based on maximum likelihood. To obtain interval estimates of β_{jc}, we consider the standard asymptotic confidence interval, i.e, the approximate $(1 - \upsilon)$ 100% confidence interval (CI) for β_{jc} is given by $\hat{\beta_{jc}} \pm z_{1-\frac{\upsilon}{2}}\sqrt{Var(\hat{\beta_{jc}})}$, where $z_{1-\frac{\upsilon}{2}}$ represents the $(1 - \frac{\upsilon}{2})$% standard normal quantile and Var($\hat{\beta_{jc}}$) is obtained from the observed information matrix.

2.2. Diagnostic Analysis

Diagnostic analysis is a very important step in the modeling process. At this stage, it is possible to assess whether the fit is consistent with prior assumptions, through residual analysis, and whether there are observations that influence the estimates disproportionately by influence analysis. Concerning classical survival models, there are several diagnostic methods already presented [39–41], but, particularly for cure-rate models, there is still limited research on this topic and, to the best of our knowledge, no methods have yet been proposed in the context of zero-inflated survival models.

In this section, we present details about the diagnostic tools that we propose to use in the context of ZICR models: Cox-Snell residuals and Local influence methods. We also provide a justification for why Martingale residuals are not an adequate tool for the proposed models. Second, ref. [41], page 365, model diagnostic tools for the overall mixture cure model and for the latency, including three types of model checking: To check for the functional form of covariates and to diagnose the presence of outliers, to evaluate the fit of the model and to evaluate the departure from the PH assumption for the latency. The authors noted that, for the martingale residuals, they are bounded from below and insensitive to covariate effects on incidence. Cox-Snell residuals [39] are useful to evaluate the overall fit [40,41] and are obtained from the values of the estimated cumulative hazard function, as follows: $r_i = \hat{H} = -\log\{\hat{S}(t_i|x_i)\}$. Usually, Cox-Snell residuals are compared with a unit exponential distribution to evaluate if standard survival models are properly fitted, but authors [41] showed that for mixture cure-rate models, these residuals are not a sample of the unit exponential distribution even though the model is correctly specified. Even so, they argue that it is still correct to check the fit by comparing the cumulative hazard rate (or survival) of the pairs (r_i, δ_i) against the cumulative hazard (or survival) function of the unit exponential distribution. Figure 2 presents an example of the obtained residuals when different models are fitted for a GG-ZICR simulated sample. It was observed that when the model is correct, the plot of Cox-Snell residuals survival against the unit exponential survival was closer to the identity line than when the model is incorrect. This indicates that Cox-Snell residuals are a suitable way to verify if the baseline distribution is correctly specified.

An widely used alternative to Cox-Snell residual are the Martigale residuals, which are given by $rm_i = \delta_i - r_i$. When obtained for classical survival models, rm_i can assume values between $(-\infty, 1)$, but authors [41] showed that, when obtained for cure-rate mixture models, Martingale residuals have a lower bound equal to $\log(p_1)$ that may limit the use of this method.In the context of ZICR models, we can show that martingale residuals are also bounded and has a limited use. When $t_i = \infty$ and $\delta_i = 0$, we obtain the lower bound, as follows:

$$\begin{aligned} rm_i &= \delta_i + log(S(t_i|x_i)) \\ &= 0 + log(p_1 + (1 - p_0 - p_1)S_0(t_i)) \\ &= 0 + log(p_1 + (1 - p_0 - p_1) \times 0) \\ &= log(p_1), \end{aligned}$$

which is the same presented for cure-rate model. And, when $t_i = 0$ and $\delta_i = 1$, we obtain the upper bound:

$$\begin{aligned} rm_i &= \delta_i + log(S(t_i|x_i)) \\ &= 1 + log(p_1 + (1 - p_0 - p_1)S_0(t_i)) \\ &= 1 + log(p_1 + (1 - p_0 - p_1) \times 1) \\ &= = 1 + log(1 - p_0) < 1. \end{aligned}$$

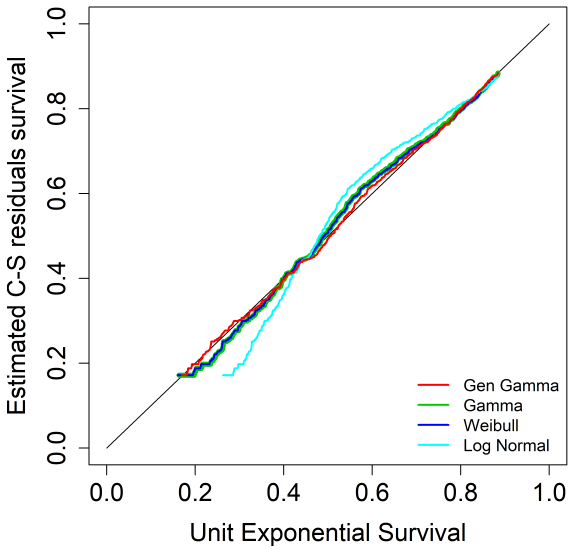

Figure 2. Example of Cox-Snell residuals when different models are fitted for a GG-ZICR simulated sample.

Concerning the methods for assessing the influence of atypical points when adjusting statistical models, let $l(\boldsymbol{\vartheta}) = \log\{L(\boldsymbol{\vartheta}, t_i, \delta_i)\}$ be the loglikelihood and $\boldsymbol{\omega}$ the vector of perturbations, which is a subset of $\Omega \in \mathbb{R}^n$. To simplify notation, we will denote the perturbed loglikelihood $l(\boldsymbol{\vartheta}|\boldsymbol{\omega})$ by $l(\boldsymbol{\vartheta}_\omega)$. The likelihood displacement is a measure useful to evaluate the difference between the maximum likelihood estimates in the perturbed and non-perturbed case and is given by $LD(\omega) = 2\{l(\hat{\boldsymbol{\vartheta}}) - l(\hat{\boldsymbol{\vartheta}}_\omega)\}$. A very used perturbation scheme is the case-deletion. In this scheme, $LD_i = 2\{l(\hat{\boldsymbol{\vartheta}}) - l(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\omega i}})\}$, were $\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\omega i}}$ is MLE obtained after deleting the ith subject. The perturbation scheme based on case deletion it is useful in identifying large deviations in fit caused by a single case. In order to have a more complete understanding of the influence of single and multiple cases, ref. [42] proposed the local influence methods based on normal curvature. The basic idea was to study the behavior of $LD(\omega)$ around the null perturbation vector, $\boldsymbol{\omega_0}$.

Let Δ be a perturbation matrix given by $\Delta_{ij} = \frac{\partial l(\boldsymbol{\vartheta}|\omega)}{\partial \vartheta_i \partial \omega_i}$. The normal curvature for $\hat{\vartheta}$ in the direction vector $\boldsymbol{d}$ ($||d|| = 1$) is given by $C_d = 2|\boldsymbol{d}^T\Delta^T\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1}\Delta\boldsymbol{d}|$, where $\boldsymbol{\Sigma}(\boldsymbol{\vartheta})$ is the covariance matrix. The main diagnostic measure in this approach is $\boldsymbol{d}_{max}$, which can be obtained by the eingenvector corresponding to largest eigenvalue of $\boldsymbol{B}(\boldsymbol{\vartheta}) = \Delta^T\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1}\Delta$ evaluated at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega_0}$. Another measure is $C_i = 2|b_{ii}|$, where b_{ii} is the i-th diagonal element of the matrix $\boldsymbol{B}(\boldsymbol{\vartheta})$. Index plots of $\boldsymbol{d}_{max}$ and C_i can be used to visualize the potential influential measures under small perturbations. In this paper, we will consider two schemes of perturbation: case-weight perturbation and response perturbation.

Let $\boldsymbol{\omega} = \{\omega_1, \omega_2, \ldots, \omega_n\}^T$ be the perturbation vector. The perturbed log-likelihood for the case-weight scheme is given by

$$l(\boldsymbol{\vartheta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \times l(\boldsymbol{\vartheta}, t_i, d_i),$$

where $l(\boldsymbol{\vartheta}, t_i, d_i)$ is presented in Equation (6). In the response perturbation scheme, we address the perturbation in the response value, t_i, as $t_i(\omega_i) = t_i + S_t \times \omega_i$. Then perturbed log-likelihood is given by

$$l(\boldsymbol{\vartheta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} l(\boldsymbol{\vartheta}, t_i + S_t \times \omega_i, d_i).$$

In both cases, the perturbation matrix will be numerically obtained using the *numDeriv* package available on software R.

3. Results

3.1. Simulation Studies

In order to assess the behavior of the MLEs and 95% CIs for increasing sample size, we performed a Monte Carlo simulation study, examining the absolute bias and Root Mean Square Error (RMSE) of point estimates and Coverage Probabilities (CP) of interval estimates. To assess if an estimated CP converges the CI fixed level, $(1-\upsilon)\%$, we set the nominal coverage, given by the bounded values: $(1-\upsilon) \pm 1.96\sqrt{(\upsilon)(1-\upsilon)/B}$, where B is the number of samples generated to obtain the CP [43]. The GG-ZICR samples were generated considering the CDF inversion method [44], with a random censure process and a Uniform distribution for the censored times [17].

The simulation study is based on $B = 1000$ sample replications, where the sample size varies as $n = 50$, 100, 250, 500, 750 and 1000. To generated the samples (y) of the simulation, was generated by two uniforme in function of parameters, see Appendix A. We considered four different combinations of β_{jc}, which we call Scenarios I, II, III and IV. In the first three scenarios, we assume x as a single binary covariate with values drawn from a Bernoulli distribution with parameter 0.5 and that this covariate is associated with all the parameters of the GG-ZICR regression model (7), following the systematic component described in (8). Therefore, the first three scenarios have a total of 10 linear coefficients ($\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, \beta_{40}, \beta_{41}, \beta_{50}, \beta_{51}$). In order to assess how the behavior of the estimation methods is impacted by the increase in the number of covariates, we consider an additional situation (Scenario IV) in which two binary covariates are inserted in each parameter, thus totaling 15 linear coefficients. It is important to note that the linear coefficients of scenarios II and IV are the same, except for the values of $\beta_{12}, \beta_{22}, \beta_{32}, \beta_{42}$ and β_{52}, which facilitates comparison between cases with fewer and more covariates. The values of β_{jc} in each scenario are described in Table 1. Figure 3 shows the survival function shapes obtained for each parameter scenario considered.

Table 1. Parameter values fixed in each scenario (I, II, III,IV).

		Scenario I	Scenario II	Scenario III	Scenario IV
μ	β_{10}	1.10	2.00	1.65	2.00
	β_{11}	1.75	0.90	−1.80	0.90
	β_{12}				0.30
σ	β_{20}	−0.15	0.25	−0.60	0.25
	β_{21}	−1.05	−1.15	1.80	−1.15
	β_{22}				-0.75
λ	β_{30}	0.35	0.15	−0.20	0.15
	β_{31}	−0.55	−0.25	0.55	−0.25
	β_{32}				−0.50
p_0	β_{40}	−3.00	−2.00	−0.50	−2.00
	β_{41}	1.00	2.00	0.75	2.00
	β_{42}				1.00
p_1	β_{50}	−2.50	−1.50	−0.35	−1.50
	β_{51}	0.30	1.50	1.75	1.50
	β_{52}				1.00

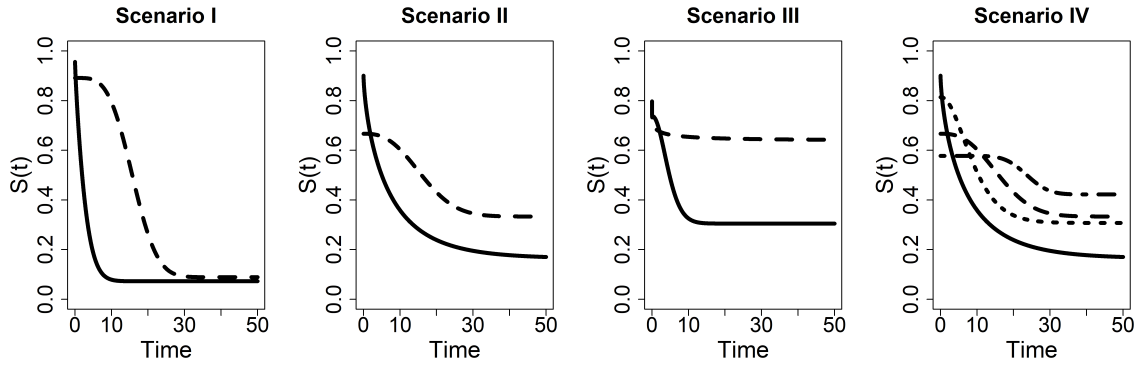

Figure 3. Survival curves of scenarios I, II, III and IV.

Figures 4–6 present the Bias, RMSE and CP of MLE of zero-inflated cure-rate regression model linear coefficients β_{10}, β_{11}, β_{20}, β_{21}, β_{30}, β_{31}, β_{40}, β_{41}, β_{50}, β_{51}, considering the four simulated scenarios. The main results are described above:

- In general, as the sample size increases, the biases and Root Mean Square Error (RMSE) are closer to zero and CPs are closer to 0.95.
- In most cases, the biases of the parameters related to μ and σ (β_{10}, β_{11}, β_{20}, β_{21}) are closer to zero (< 0.5) even for small sample sizes. This pattern does not occur in the third and fourth scenarios, where the parameters are close to zero when $n \geq 100$ (β_{11} and β_{21}) e $n \geq 500$ (β_{21}).
- The RMSE is less than 1 for β_{10} and β_{20} when $n \geq 100$. In contrast, the RMSE for β_{11} and β_{21} needs larger sample sizes to approach zero, especially in scenario III e IV.
- The CP of β_{10}, β_{11}, β_{20} and β_{21} in scenarios III and IV needs a bigger sample size to achieve the nominal coverage.
- The parameter related to λ (β_{30}, β_{31}) needs a bigger sample size to approximate Bias and RMSE to zero in the four scenarios, more evident in scenario III e IV.
- The parameters related to zero-inflation and cure-rate proportions (β_{40}, β_{41}, β_{50}, β_{51}) need a sample size of 250 to achieve Bias and RMSE near zero, more evident for β_{50}, β_{51} in all scenarios and β_{40}, β_{41} in Scenario I.
- The majority of CP for the parameters related to zero-inflation and cure-rate are close to 0.95, even if $n = 50$.

In summary, these results show that the parameters related to μ, σ and λ present better results in the first and second scenarios, which are less censored than scenario III. In relation to the increase in the number of covariates, comparing scenario IV, it is observed that the impact in general is small, with emphasis on the lower CP for the parameters related to μ, σ e λ.

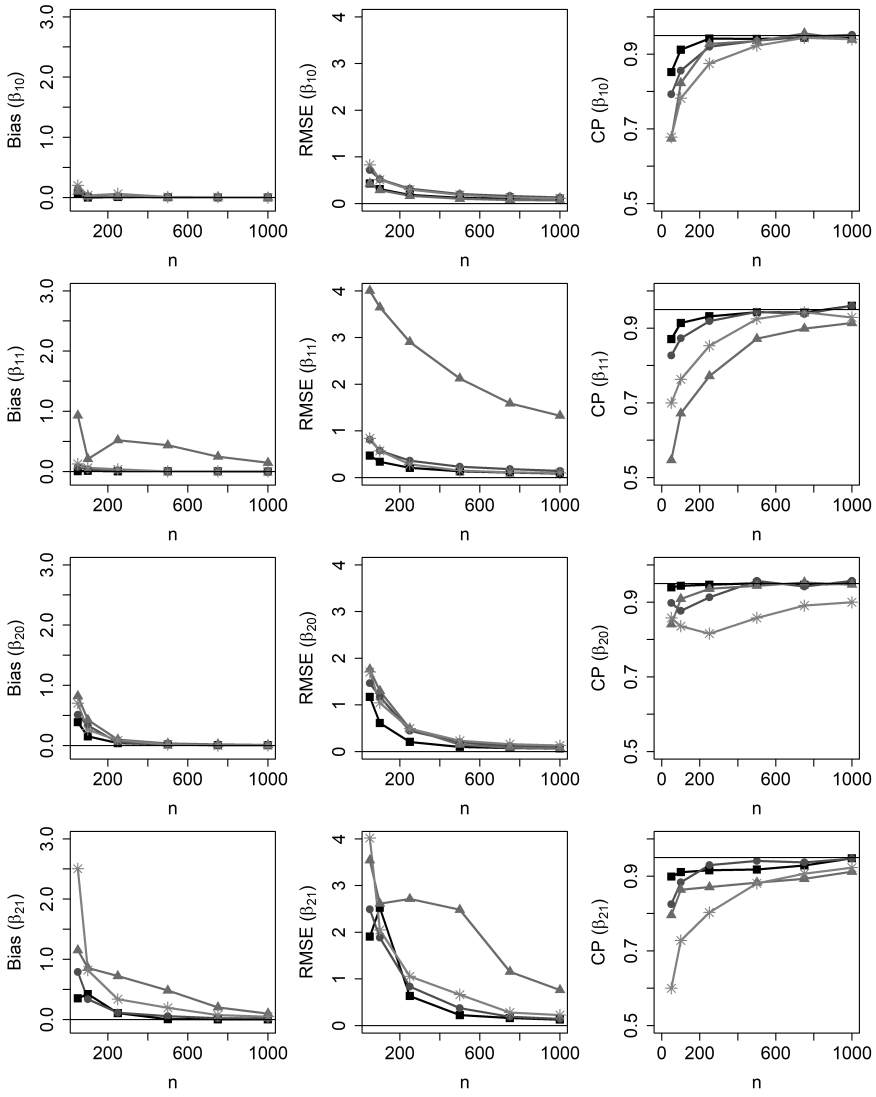

Figure 4. Absolute bias, RMSE and CP of MLE for the parameters ($\widehat{\beta}_{10}$, $\widehat{\beta}_{11}$, $\widehat{\beta}_{20}$, $\widehat{\beta}_{21}$) of GG-ZICR regression model considering four different scenarios: I (square), II (circle), III (triangle) and IV(asterisk).

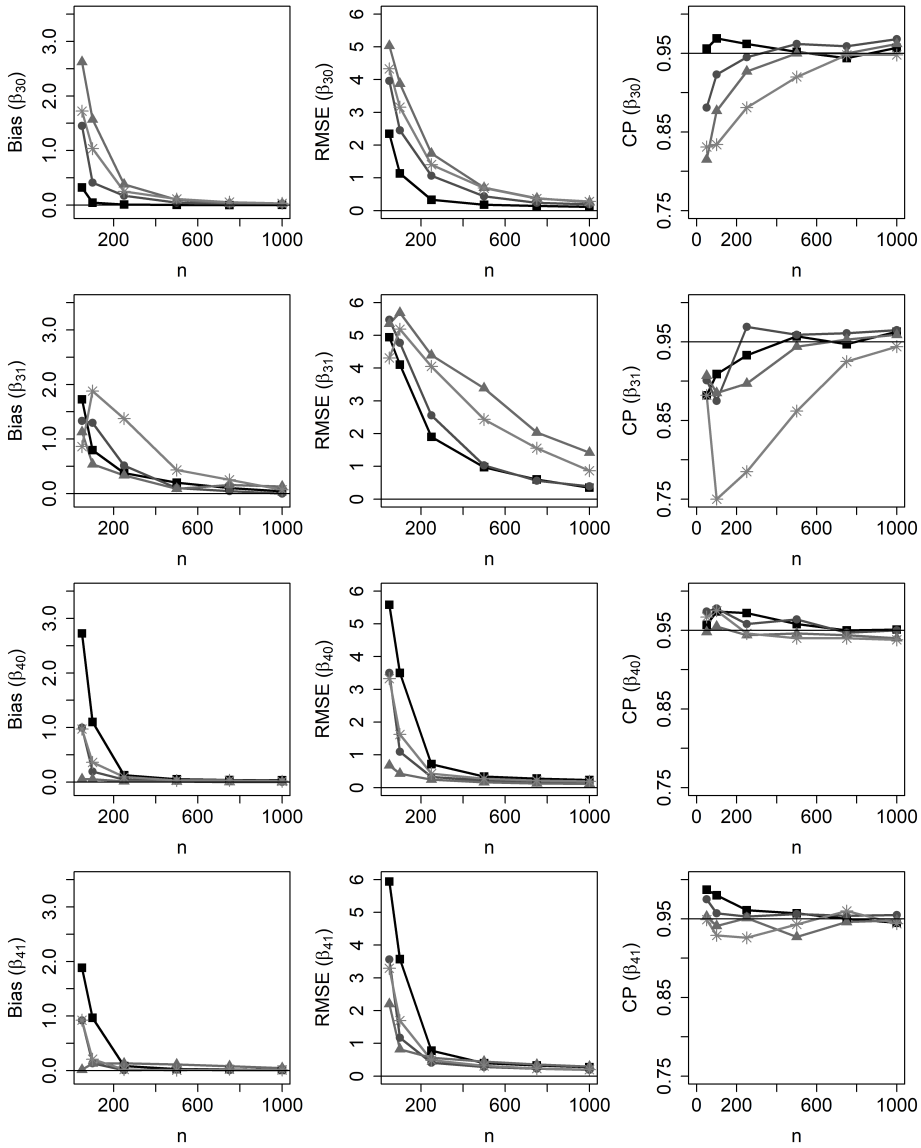

Figure 5. Absolute bias, RMSE and CP of MLE for the parameters $(\widehat{\beta}_{30}, \widehat{\beta}_{31}, \widehat{\beta}_{40}, \widehat{\beta}_{41})$ of GG-ZICR regression model considering four different scenarios: I (square), II (circle), III (triangle) and IV(asterisk).

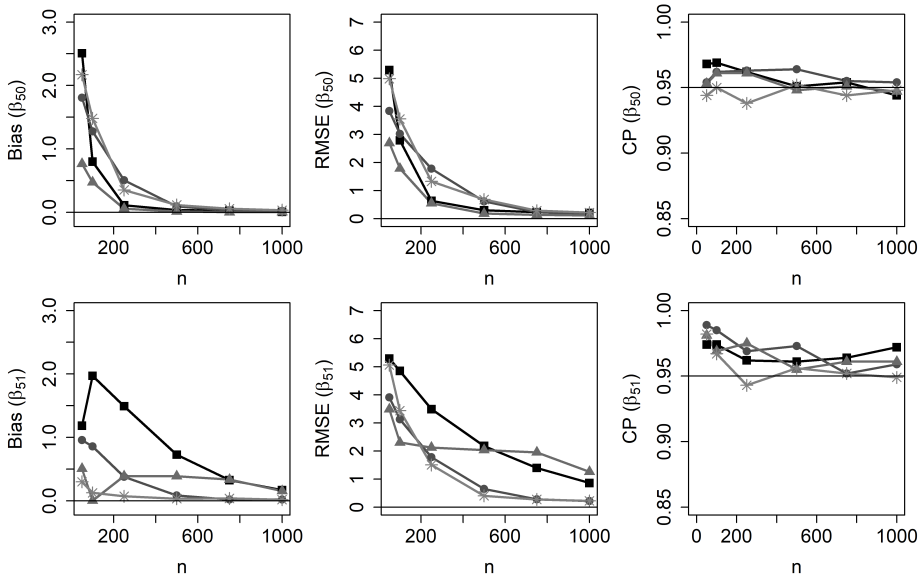

Figure 6. Absolute bias, RMSE and CP of MLE for the parameters $(\widehat{\beta}_{50}, \widehat{\beta}_{51})$ of GG-ZICR regression model considering four different scenarios: I (square), II (circle), III (triangle) and IV(asterisk).

3.2. Comparison Methods

In order to evaluate the performance of comparison methods, we generate $B = 1000$ samples with size equal to $n = 50, 100, 250, 500, 750$ and 1000 and the following fixed parameter values: $\mu = 2$, $\sigma = 1.5$, $p_0 = 10\%$, $p_1 = 16.5\%$ and $\lambda = (0, 0.5, 1, 1.5)$. For each sample, the comparison methods were obtained for testing three different null hypothesis:

- $H_0 : \lambda = 0$
- $H_0 : \lambda = 1$
- $H_0 : \lambda = \sigma = 1.5$

which are related with the three main particular distributions: Log-Normal, Weibull and Gamma, respectively. The percentage frequencies of samples in which each method indicates the GG-ZICR model are presented in Table 2 and reflect the rate of rejection of null hypothesis. The percentages under the null hypothesis are highlighted in bold.

When the samples are generated under the alternative hypothesis, it is observed that as the value of λ distances from the null hypothesis and the sample size increases, the rate of rejection of null hypothesis increases. Under the null hypothesis related to Log-Normal distribution ($\lambda = 0$), all the criteria present rates of rejection closer to zero. In contrast, for the hypothesis $\lambda = 1$ and $\lambda = \sigma$, it is observed that AIC present high rates of rejection of the null hypothesis even for the samples generated under the null hypothesis. LRT presents better results, since the rejection rate is close to 5% when the null hypothesis is true and the sample size is greater than 100. However, it is still necessary to evaluate comparison methods with a better performance when the sample size is small and the true value of λ is near to the null hypothesis value.

Table 2. Percentage frequency of samples which each comparison method (LRT, AIC, cAIC, BIC) indicates the GG model for different values of λ.

		$H_0 : \lambda = 0$				$H_0 : \lambda = 1$				$H_0 : \lambda = 1.5$			
	N	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$
LRT	50	0.0	10.6	25.5	48.2	31.9	10.8	10.3	20.6	43.6	19.6	12.0	20.8
	100	0.0	19.1	48.3	71.4	70.8	22.4	8.6	17.3	80.4	38.9	14.2	11.9
	250	0.2	44.8	86.6	96.8	96.6	57.7	6.0	30.8	98.4	84.3	21.9	7.0
	500	0.5	78.5	99.5	100.0	99.8	89.6	6.7	62.5	100.0	99.0	43.4	10.1
	750	0.5	93.4	99.8	100.0	100.0	97.8	7.4	78.6	100.0	99.9	56.1	4.9
	1000	0.8	97.9	100.0	100.0	100.0	99.4	5.0	91.2	100.0	100.0	76.9	3.2
AIC	50	0.0	21.3	41.0	59.9	56.6	32.0	20.2	0.1	63.5	40.7	25.0	30.0
	100	0.0	33.3	64.8	85.3	82.9	42.5	20.7	31.6	87.8	60.6	31.7	23.6
	250	0.2	63.8	94.4	98.7	97.7	77.7	16.5	50.5	98.8	92.9	44.2	19.7
	500	0.5	90.3	100.0	100.0	99.9	96.8	16.3	79.5	100.0	99.9	65.1	21.8
	750	0.5	98.0	100.0	100.0	100.0	99.3	19.2	90.5	100.0	99.9	75.7	15.5
	1000	0.8	99.2	100.0	100.0	100.0	99.7	15.7	96.8	100.0	100.0	90.0	14.8
cAIC	50	0.0	7.3	20.5	41.9	21.0	5.9	7.8	18.6	33.8	12.2	9.7	18.1
	100	0.0	9.8	36.2	60.4	55.0	12.1	3.3	11.2	70.7	25.9	7.6	7.7
	250	0.2	26.7	73.9	92.4	92.6	31.8	1.7	14.1	97.1	69.6	9.4	2.2
	500	0.5	57.4	96.6	100.0	99.8	70.7	1.6	33.4	99.9	95.9	18.6	5.5
	750	0.5	78.7	99.7	100.0	100.0	88.9	3.3	52.0	100.0	99.5	28.9	0.9
	1000	0.8	92.9	100.0	100.0	100.0	96.8	0.9	68.8	100.0	100.0	47.1	0.3
BIC	50	0.0	10.2	25.0	47.9	31.1	10.5	10.2	20.5	42.6	19.1	11.9	20.6
	100	0.0	14.3	42.5	65.4	63.5	16.6	6.0	14.0	76.5	31.3	10.3	9.9
	250	0.2	32.5	79.3	94.1	94.1	40.9	2.8	18.7	97.4	74.9	12.7	3.3
	500	0.5	62.7	97.5	100.0	99.8	75.6	2.1	39.6	99.9	97.6	25.0	6.1
	750	0.5	83.7	99.7	100.0	100.0	91.9	3.6	58.2	100.0	99.6	36.4	1.2
	1000	0.8	94.5	100.0	100.0	100.0	97.9	1.3	75.5	100.0	100.0	52.7	0.5

4. Data Application: Sub-Saharan African Pregnant Women

In order to show how the GG-ZICR model can be applied to study survival data, in this Section an application of the GG-ZICR is presented. As described in the Introduction Section, the data set was collected by WHO as a step of the BOLD project and consists of data on women in labor from Nigeria and Uganda. Here, we considered the labor time of 7062 which were selected according to clinical characteristics of interest. These characteristics are described in the Flowchart presented in Appendix B.

Fitting the GG-ZICR model without covariates, the following parameter values were obtained: $\hat{\mu} = 2.16$, $\hat{\sigma} = 1.03$, $\hat{\lambda} = 0.0002$, $\hat{p_0} = 0.045$ and $\hat{p_1} = 0.040$. These results enable us to infer that the particular case when $\lambda = 0$, the Log-Normal ZICR, is also adequate to the data, since the value of $\hat{\lambda}$ is very close to zero. Additionally,

the values of p_0 and p_1 indicate that 4.5% of women in this sample arrive to the health facility already presenting fetal death and 4.0% do not present vaginal delivery at the end of the follow-up.

Table 3 presents the values of comparison methods obtained for the GG-ZICR model and its main particular cases. The LRT tests rejected the null hypothesis that $\lambda = 1$ (Weibull) or $\lambda = \sigma$ (Gamma) and does not reject the hypothesis that $\lambda = 0$ (Log Normal). The Log-Normal and Generalized Gamma present the smaller AIC, BIC and cAIC. The Cox-Snell residuals (Figure 7b) also indicate Log-Normal and Generalized Gamma as being the most correctly fit for this data. Figure 7a presents the fitted and KM Survival curves. The fitted survival curves represent an estimate of the proportion of woman who still have the possibility of vaginal delivery over time after admission. Based on Log Normal, for example, after 6 h of admission, 62.6% is the percentage of women can still have a vaginal delivery. In order to evaluate the local influence, in Figure 7c, we present the index plot of $\boldsymbol{d}_{max}$ considering the case-weight perturbation scheme. The result indicates that observation 462 is a potential influence case. This observation present a non-censored time of 0.02 h, which is the smaller time between the non-zero values. The withdrawal of this observation does not change to a large extent the estimates nor the residuals, so we chose to keep the observation in the analysis.

Table 3. AIC, cAIC, BIC and Log-Likelihood (LL) of ZICR estimated models considering Generalized Gamma, Gamma, Weibull and Log Normal as baseline distributions. (*) Models which were rejected by LRT.

Baseline Distribution	AIC	cAIC	BIC	LL
Generalized Gamma	30,747.75	30,787.02	30,782.02	−15,368.87
Gamma	31,062.35	31,093.77	31,089.77	−15,527.18 *
Weibull	31,195.06	31,226.48	31,222.48	−15,593.53 *
Log Normal	30,745.73	30,777.14	30,773.14	−15,368.86

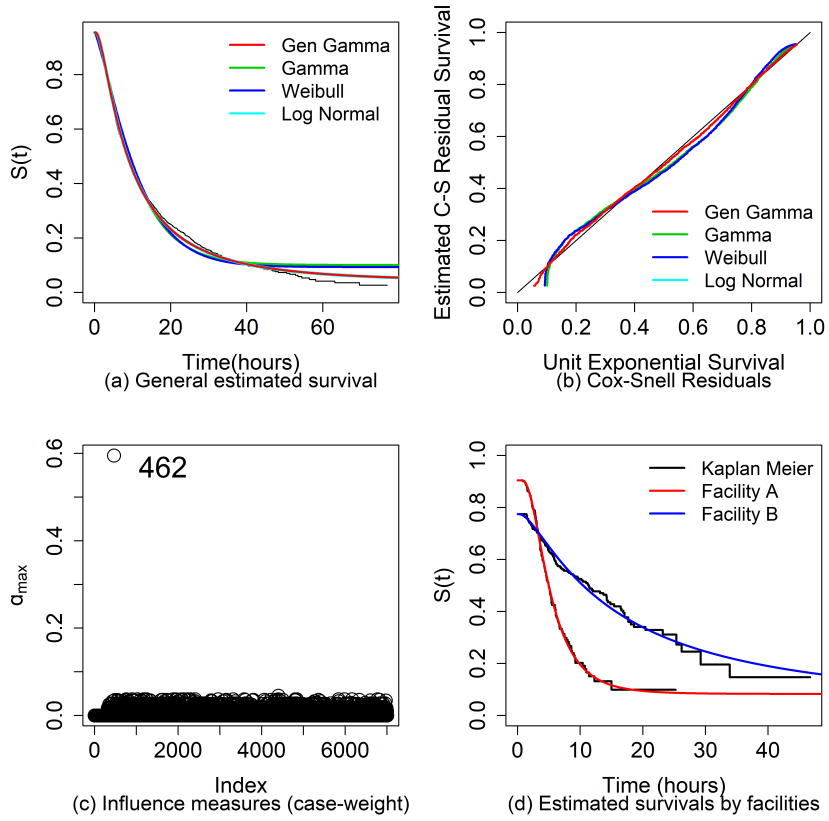

Figure 7. Survival curves (**a**), Cox-Snell residuals (**b**) and influence measures (**c**) for estimated ZICR models with Generalized Gamma, Gamma, Weibull and Log-Normal as baseline distributions. Survival curves estimated for facilities A and B (**d**).

The analysis without covariate allow us to understand the pattern of childbirth admission and development in the whole population. However, when it comes to health management, it may be interesting to evaluate the characteristics of the different facilities separately. In this context, we can use can use the regression model. To

exemplify this, we present the analysis considering two facilities from Nigeria, which we will call Facility A and Facility B.

The data set is composed of 272 and 279 observations for Facility A and B, respectively. Table 4 presents descriptive results of socio-demographic variables for each facility. The age and marital status are similar in the facilities. The proportion of multiparous and woman with one or more previous abortions is higher in Facility B and the proportion of woman with high level of education is higher in Facility A. Yoruba and Other are the ethnic groups more frequent in facility B and A, respectively.

Table 4. Socio-demographic data.

	Facility A	**Facility B**
Age (mean (SD))	28.2 (4.8)	28.62 (5.6)
Married (%)	271 (99.6)	275 (98.6)
Parity > 0 (%)	128 (47.1)	159 (57.0)
Previous abortions > 0 (%)	72 (26.5)	87 (31.2)
Ethnic groups (%)		
Ibo	89 (32.7)	35 (12.5)
Yoruba	28 (10.3)	233 (83.5)
Other	155 (57.0)	11 (4.0)
Education: High (%)	249 (91.5)	227 (81.4)

Table 5 present the parameter estimates for GG-ZICR considering the Facility as a binary covariate, which assumes 0 for Facility A and 1 for Facility B. Figure 7d presents the fitted curves and Kaplan-Meier curves for each facility. The confidence interval for β_{31} is strictly negative, indicating that the Facility B has, significantly, a small value for λ. On the other hand, the confidence intervals for β_{11}, β_{21} and β_{41} are strictly positive, indicating that the Facility B has, significantly, higher values for parameters μ, σ and p_0 than Facility A. The estimate for p_0 is 9.56% in Facility A and 22.58% in Facilty B. In the practical point of view, we can infer that Facility B presents a higher level of fetal death in admission and a faster time of childbirth than Facility A.

Table 5. MLE, Standard Deviation (SD), Lower and Upper limit (CI 95%) of GG-ZICR regression model parameters with the Facility as a covariate.

		MLE	**SD**	**Lower**	**Upper**
μ	β_{10}	1.62	0.07	1.49	1.75
hline	β_{11}	1.11	0.17	0.78	1.44
σ	β_{20}	−0.46	0.08	−0.61	−0.32
hline	β_{21}	0.62	0.10	0.42	0.81
λ	β_{30}	−7.85	0.21	−8.26	−7.44
hline	β_{31}	−3.61	0.21	−4.02	−3.20
p_0	β_{40}	−2.15	0.21	−2.57	−1.74
hline	β_{41}	0.97	0.28	0.43	1.52
p_1	β_{50}	−2.30	0.55	−3.38	−1.22
	β_{51}	−0.60	2.18	−4.87	3.67

5. Discussion

The GG-ZICR model enabled us to deal with the possibility of zero-inflation in the context of survival analysis, which characterizes an innovation for this statistical area, since there is a lack of models presented in this issue. The proposed model is based on a very flexible distribution, the Generalized Gamma [33], which extends four popular distributions: Log-Normal, Weibull, Gamma and Exponential. Finally, the model is an extension of the Berkson and Gage model [20], and therefore it presents a parameter that includes a proportion of subjects that do not present the outcome even after a long follow-up, called long-term proportion or cure rate.

The regression structure presented here enables us to link covariates with all the model parameters, and therefore the users can evaluate how a covariate is associated with the zero inflation, the cure-rate and other GG parameters, separately. The point estimation is likelihood-based and we considered the asymptotic confidence intervals.

The properties of inference methods were evaluated by a Monte Carlo simulation study. It can be observed that the inference methods for the parameters related with the GG distribution (μ, σ, λ) present good performance even for small sample sizes when the zero-inflation and cure-rate are medium or small. In contrast, the estimation of the parameters p_0 an p_1 are less biased the bigger the zero-inflation and cure-rate are. In general, the bias and RMSE are closer to zero and the CP converges to the nominal as the sample size increase. One important consequence of non-identifiability is the possible instability of parameter estimates, as noted by [45]. However, the simulation results presented here do not indicate evidence of such instability in the estimates, as reported by the results of [46]. Nevertheless, this issue should be carefully evaluated for the specific model under consideration.

We also evaluate the properties of comparison methods: AIC, BIC, consistent AIC and LRT. It can be observed that the LRT presents a better performance and the AIC has a limited application, as it tends to reject the null hypothesis, even though the null hypothesis is true. Thus, it is still necessary to evaluate alternative comparison methods with a better performance, particularly when the sample size is small.

Finally, in the application on the BOLD data set, which consists of data about African pregnant women in labor, the GG-ZICR model was considered to study the time between hospital admission and vaginal birth. Analyzing the whole sample, it was estimated that 4.5% of the considered sample present fetal death on the arrival of hospital and 4.0% do not present vaginal delivery at the end of the follow-up. When two specific facilities were evaluated, it was possible to observe that the facility with higher level of education and smaller proportion of previous abortions, also presented the smaller proportion of fetal death at admission. The main contribution of the GG-ZICR model to the analysis of these data is the greater flexibility to identify patterns of occurrence of childbirth over time and to infer the possibility of adjusting more parsimonious models (particular cases). Thus, in this application, we identify that Log-Normal distribution also has a good fit to the data by observing the estimated value of parameter λ and the comparison method results.

6. Conclusions

The GG-ZICR model provides an innovative and flexible framework for survival analysis in the presence of zero-inflation and long-term survivors. By allowing covariates to be linked with all model parameters, it enables a detailed understanding of how explanatory variables influence different aspects of the process, such as zero-inflation, cure rate, and the underlying distribution parameters. Simulation studies demonstrated that the proposed inference methods perform well under various scenarios, particularly as sample size increases, though further investigation of comparison criteria remains necessary. The application to the BOLD dataset illustrates the model's practical utility in capturing complex patterns of childbirth outcomes, highlighting its ability to accommodate specific cases such as the Log-Normal distribution. Overall, the GG-ZICR model broadens the methodological tools available for survival analysis, offering both theoretical contributions and practical insights for real-world data. Recent references such as [46] broaden the possibilities for deriving the baseline function and may bring new insights to the problem of timing of delivery.

Author Contributions

G.d.S.C.P., F.L. and H.C.C.d.S.: conceptualization, methodology, software, data curation, writing—original draft preparation, visualization, investigation; G.d.S.C.P.: supervision; A.A., H.A.I., J.B. and L.O.O.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by CAPES, CEPID-ICMC/USP/ FAPESP grant number 2013/07375-0, FAEPA, CNPq.

Institutional Review Board Statement

Project approved in Ethics Statement, ID protocol: A65879, title: Development of a Simplified, Effective, Labour, Monitoring Study, in 10/09/2014.

Informed Consent Statement

Data collection was conducted in an observational study that did not expose participants to any risk. All potential participants were approached by research assistants at the time of admission to the maternity ward, preferably in the early stages of labor. Research assistants determined whether or not women were able to provide consent and were trained to ensure the voluntariness of consent. At the time of data collection, all potential participants received information about the research in their language of choice, in accordance with ethical requirements for research involving human beings. Those who agreed to participate in the study were invited to sign the Informed Consent Form (ICF).

Data Availability Statement

This manuscript reports on a secondary analysis of the World Health Organization BOLD Project database. This project was implemented by WHO to further understand the patterns of labor progression in sub-Saharan African women and to contribute to the development of evidence-based guidelines to improve women's experience with labor and childbirth. The UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), a cosponsored program executed by the World Health Organization (WHO) provided access to the BOLD data set for this analysis. The named authors alone are responsible for the views expressed in this publication. The authors would like to thank João Paulo de Sousa for allowing us to participate in the project.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the author G.d.S.C.P. used "Chatgpt" to revise the references. After using this "Chatgpt", the author G.d.S.C.P. reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Appendix A: R Code for Generated Sample of Zero-Cure Gamma Generalized Model

```
model <- function(N, varp, x){
  v1 <- matrix(varp[1:2], ncol=1)
  v2 <- matrix(varp[3:4], ncol=1)
  v3 <- matrix(varp[5:6], ncol=1)

  mu  <- as.vector(x%*%v1)
  sig <- exp(as.vector(x%*%v2))
  Qe  <- exp(as.vector(x%*%v3))

  v4 <- matrix(varp[7:8], ncol=1)
  v5 <- matrix(varp[9:10], ncol=1)

  p14 <- as.vector(x%*%v4)
  p15 <- as.vector(x%*%v5)

  denom = 1+exp(p14)+exp(p15)

  gamm_zer = exp(p14)/denom
  gamm_inf = exp(p15)/denom

  U = runif(N)
  V = runif(N, gamm_zer, 1-gamm_inf)
  y = numeric(N)
  ind0 = U <= gamm_zer
  ind1 = U > gamm_zer & U <= 1-gamm_inf
  ind2 = U > 1-gamm_inf
  y[ind0] = 0
  y[ind1] = qgengamma(
    (V[ind1]-gamm_zer[ind1])/
    (1-gamm_zer[ind1]-gamm_inf[ind1]),
    mu = mu[ind1],
    sigma = sig[ind1],
    Q = Qe[ind1]
  )
```

```
y[ind2] = Inf
tM = max(y[y<Inf])
Z = runif(N,0,tM)
time = pmin(y,Z)
delta = ifelse(y <= Z,1,0)
data.frame(time,delta)
```

Appendix B

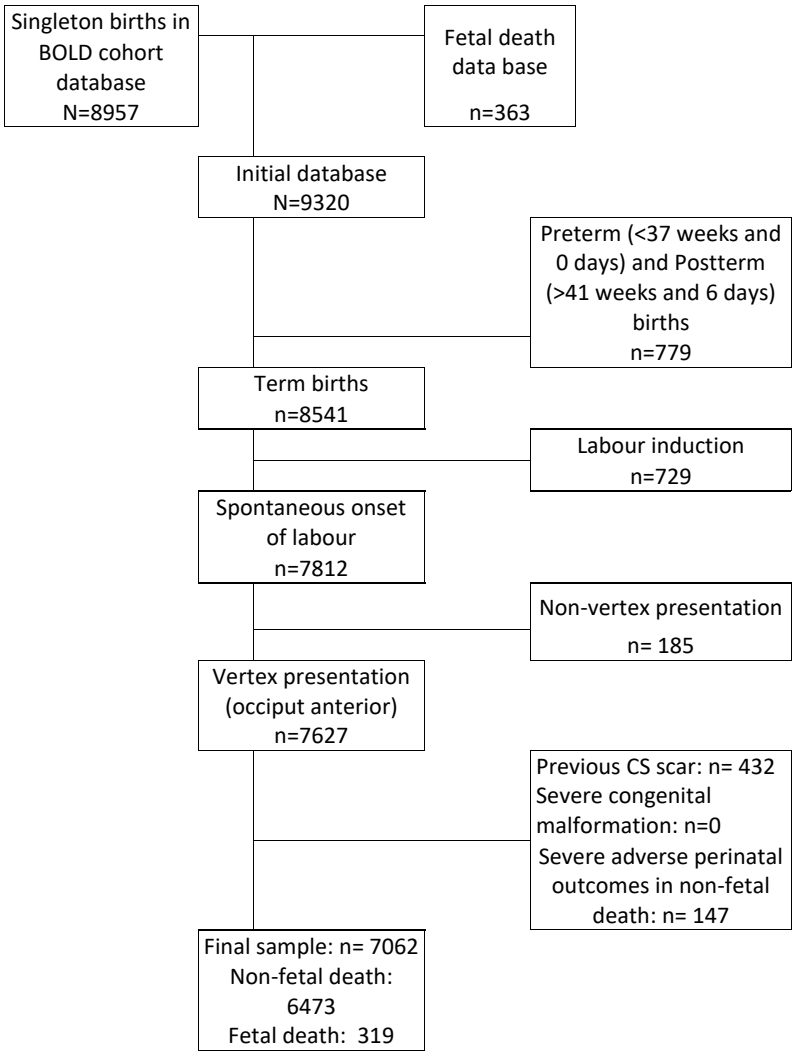

Figure B1. Flowchart of patient selection.

References

1. Cohen, A.C. An extension of a truncated Poisson distribution. *Biometrics* **1960**, *16*, 446–450.
2. Johnson, N.L.; Kotz, S.; Kemp, A. *Discrete Distributions*; Wiley: New York, NY, USA, 1969.
3. Umbach, D. On inference for a mixture of a Poisson and a degenerate distribution. *Commun. Stat. Theory Methods* **1981**, *10*, 299–306.
4. Alghamdi, F.M.; Ahsan-ul-Haq, M.; Hussain, M.N.S.; et al. Discrete Poisson Quasi-XLindley distribution with mathematical properties, regression model, and data analysis. *J. Radiat. Res. Appl. Sci.* **2024**, *17*, 100874.
5. Alghamdi, F.M.; Gemeay, A.M.; Abd-Elmougod, G.A.; et al. A bias-reduced estimator for generalized Poisson regression with application to carbon dioxide emission in Canada. *Sci. Rep.* **2025**, *15*, 39224.
6. Yousif, H.M.; Algamal, Z. Improving geographically weighted Poisson regression model based on metaheuristic algorithms: Application to cancer rate data. *Comput. J. Math. Stat. Sci.* **2025**, *4*, 407–423.
7. Workie, M.S.; Lakew, A.M. Bayesian count regression analysis for determinants of antenatal care service visits among pregnant women in Amhara regional state, Ethiopia. *J. Big Data* **2018**, *5*, 7.

8. Bertoli, W.; Conceição, K.S.; Andrade, M.G.; et al. On the zero-modified Poisson–Shanker regression model and its application to fetal deaths notification data. *Comput. Stat.* **2018**, *33*, 807–836.
9. Smits-Seemann, R.R.; Pettit, J.; Li, H.; et al. Infection-related mortality in Hispanic and non-Hispanic children with cancer. *Pediatr. Blood Cancer* **2017**, *64*, e26502.
10. Aitchison, J. On the distribution of a positive random variable having a discrete probability mass at the origin. *J. Am. Stat. Assoc.* **1955**, *50*, 901–908.
11. Tong, E.N.; Mues, C.; Thomas, L. A zero-adjusted gamma model for mortgage loan loss given default. *Int. J. Forecast.* **2013**, *29*, 548–562.
12. Tooze, J.A.; Grunwald, G.K.; Jones, R.H. Analysis of repeated measures data with clumping at zero. *Stat. Methods Med. Res.* **2002**, *11*, 341–355.
13. Liu, L.; Strawderman, R.L.; Johnson, B.A.; et al. Analyzing repeated measures semi-continuous data, with application to an alcohol dependence study. *Stat. Methods Med. Res.* **2016**, *25*, 133–152.
14. Ospina, R.; Ferrari, S.L.P. A general class of zero-or-one inflated beta regression models. *Comput. Stat. Data Anal.* **2012**, *56*, 1609–1623.
15. Braekers, R.; Grouwels, Y. A semi-parametric Cox regression model for zero-inflated left-censored time to event data. *Commun. Stat. Theory Methods* **2016**, *45*, 1969–1988.
16. Garcia, P.N.d.A. Modelo Weibull Discreto com Fração de Cura e Excessos de Zeros. PhD Thesis, Universidade de Brasília, Brasília, Brazil, 2016.
17. Louzada, F.; Moreira, F.F.; de Oliveira, M.R. A zero-inflated non default rate regression model for credit scoring data. *Commun. Stat. Theory Methods* **2018**, *47*, 3002–3021.
18. de Oliveira, M.R.; Moreira, F.; Louzada, F. The zero-inflated promotion cure rate model applied to financial data on time-to-default. *Cogent Econ. Financ.* **2017**, *5*, 1395950.
19. Ocloo, S.K.; Brew, L.; Nasiru, S.; et al. On the extension of the Burr XII distribution: Applications and regression. *Comput. J. Math. Stat. Sci.* **2023**, *2*, 1–30.
20. Berkson, J.; Gage, R.P. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* **1952**, *47*, 501–515.
21. Chen, M.H.; Ibrahim, J.G.; Sinha, D. A new Bayesian model for survival data with a surviving fraction. *J. Am. Stat. Assoc.* **1999**, *94*, 909–919.
22. Calsavara, V.F.; Rodrigues, A.S.; Rocha, R.; et al. Zero-adjusted defective regression models for modeling lifetime data. *J. Appl. Stat.* **2019**, *46*, 2434–2459.
23. Maller, R.; Zhou, X. *Survival Analysis with Long-term Survivors*; Wiley: New York, NY, USA, 1996.
24. de Souza, H.C.C.; Louzada, F.; Ramos, P.L.; et al. A Bayesian approach for the zero-inflated cure model: An application in a Brazilian invasive cervical cancer database. *J. Appl. Stat.* **2022**, *49*, 3178–3194.
25. Gyimah, S.O.; Takyi, B.K.; Addai, I. Challenges to the reproductive-health needs of African women. *Soc. Sci. Med.* **2006**, *62*, 2930–2944.
26. Lawn, J.E.; Blencowe, H.; Pattinson, R.; et al. Stillbirths: Where? When? Why? How to make the data count? *Lancet* **2011**, *377*, 1448–1463.
27. Bhutta, Z.A.; Das, J.K.; Bahl, R. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet* **2014**, *384*, 347–370.
28. Hoot, N.R.; Aronsky, D. Systematic review of emergency department crowding: Causes, effects, and solutions. *Ann. Emerg. Med.* **2008**, *52*, 126–136.
29. Oladapo, O.T.; Souza, J.P.; Bohren, M.A.; et al. WHO Better Outcomes in Labour Difficulty (BOLD) project: Innovating to improve quality of care around the time of childbirth. *Reprod. Health* **2015**, *12*, 48.
30. Souza, J.P.; Oladapo, O.T.; Bohren, M.A.; et al. The development of a simplified, effective, labour monitoring-to-action (SELMA) tool for better outcomes in labour difficulty (BOLD): Study protocol. *Reprod. Health* **2015**, *12*, 49.
31. Oladapo, O.T.; Souza, J.P.; Fawole, B. Progression of the first stage of spontaneous labour. *PLoS Med.* **2018**, *15*, e1002492.
32. R Core Team. The R Project for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 3 October 2025).
33. Meeker, W.Q.; Escobar, L.A. *Statistical Methods for Reliability Data*; Wiley: New York, NY, USA, 1998.
34. Ram, B.; Mishra, S.K.; Lai, K.K.; et al. Quantum Broyden Fletcher Goldfarb Shanno Method. In *Unconstrained Optimization and Quantum Calculus*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 101–123.
35. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723.
36. Acquah, H. On the Comparison of Akaike Information Criterion and Consistent Akaike Information Criterion in Selection of an Asymmetric Price Relationship: Bootstrap Simulation Results. *AGRIS Line Pap. Econ. Inform.* **2013**, *5*, 3.
37. Bozdogan, H. Model selection and Akaike's information criterion (AIC). *Psychometrika* **1987**, *52*, 345–370.
38. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464.
39. Cox, D.R.; Snell, E.J. A general definition of residuals. *J. R. Stat. Soc. Ser. B* **1968**, 248–275.
40. Colosimo, E.A.; Giolo, S.R. *Análise de Sobrevivência Aplicada*; Survival Analysis Applied; Edgard Blücher: São Paulo, Brazil, 2006.

41. Peng, Y.; Taylor, J.M. Residual-based model diagnosis methods for mixture cure models. *Biometrics* **2017**, *73*, 495–505.
42. Cook, R.D. Influence assessment. *J. Appl. Stat.* **1987**, *14*, 117–131.
43. Louzada-Neto, F. Extended hazard regression model for reliability and survival analysis. *Lifetime Data Anal.* **1997**, *3*, 367–381.
44. Ross, G.J. *Nonlinear Estimation*; Springer: Berlin/Heidelberg, Germany, 1990.
45. Hanin, L.; Huang, L.S. Identifiability of cure models revisited. *J. Multivar. Anal.* **2014**, *130*, 261–274.
46. Louzada, F.; Ramos, P.L.; de Souza, H.C.; et al. On the unification of zero-adjusted cure survival models. *Commun. Stat. Simul. Comput.* **2025**. https://doi.org/10.1080/03610918.2025.2496297.