# Alternative methods for modeling of the cure rate in survival studies

**Vera Tomazella**

Uniiversidade Federal de são Carlos-São Carlos-SP
veratomazella@gmail.com

VIII Encontro dos Alunos de Pós-graduação em Estatística e Experimentação Agronômica

# Survival Analysis (SA) x Long-Term Survival Analysis (LTSA)

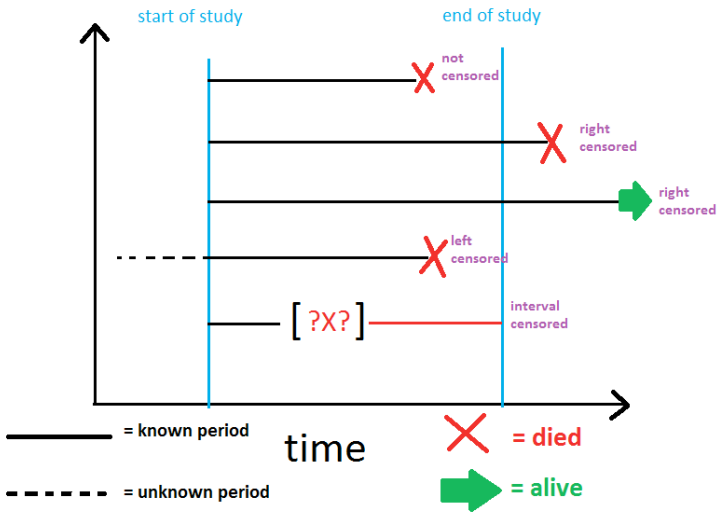
- **SA**: It is assumed that all experimental units ("individuals") present the event of interest.

- **Long-Term Survival Analysis (LTSA)**:
  - In survival analysis studies in which there are a cure fraction are common.
  - With the fast development of medical treatments, the data in the population generally reveal that a proportion of patients can be cured
  - The cure fraction is the proportion of the observed individuals which, for some reason, are not susceptible to the event of interest.
  - These data sets may be applied in different areas such as in

1. Medicine - recurrence of a cancer
2. Social area { - occurrence of divorces
   - time until the birth of the first child

# Specifying Censorship

- Features which are typically encountered in analysis of survival data:
  - individuals do not all enter the study at the same time
  - when the study ends, some individuals still haven't had the event yet
  - other individuals drop out or get lost in the middle of the study, and all we know about
  - them is the last time they were still 'free' of the event



# Specifying Survival Time

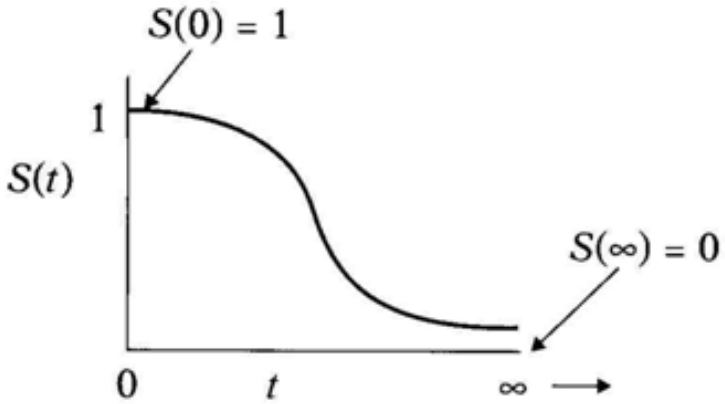
Let T$\in \Re^+$ a random variable denoting survival time. The $T$ distribution function can be written as::

$$F(t) = P(T \leq t) = \int_0^t f(u)du$$

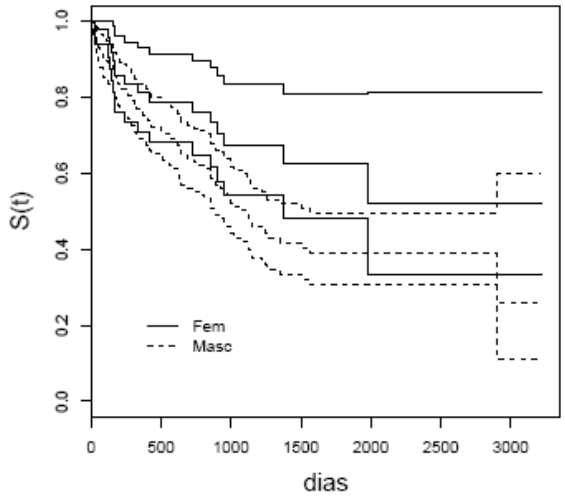
where $f$ is the $f.d.p$ of $T$.
We define the Survival Function, $S(t)$, as the probability of an individual surviving a time greater than $t$, that is,

$$S(t) = 1 - F(t)$$

# Shape of the Survival Function




When $\lim_{t\to\infty} S(t) \neq 0$



# Improper Survival Function - ISF

$S_{pop}(t) \equiv$ Population Survival Function (ISF)

$$S_{pop}(t) = 1 - \gamma + \int_t^{\infty} f(u)du \quad , \quad \gamma \leq 1$$

Properties:

1. If $\gamma = 1 \Rightarrow S_{pop}(t) = S(t)$, that is, this class contains the usual FS of Survival Analysis,
2. $S_{pop}(0) = 1$;
3. $S_{pop}(t) \downarrow t$;
4. $lim_{t\to\infty} S_p(t) = 1 - \gamma = p_0 \equiv$ Cure Fraction. .

# Application with Breast cancer data set

- The study came from a real-world medical data set collected at a hospital in Brazil from Feb/2011 to Oct/2013. These data contain information from 78 patients diagnosed with triple-negative breast cancer and treated with neoadjuvant chemotherapy.

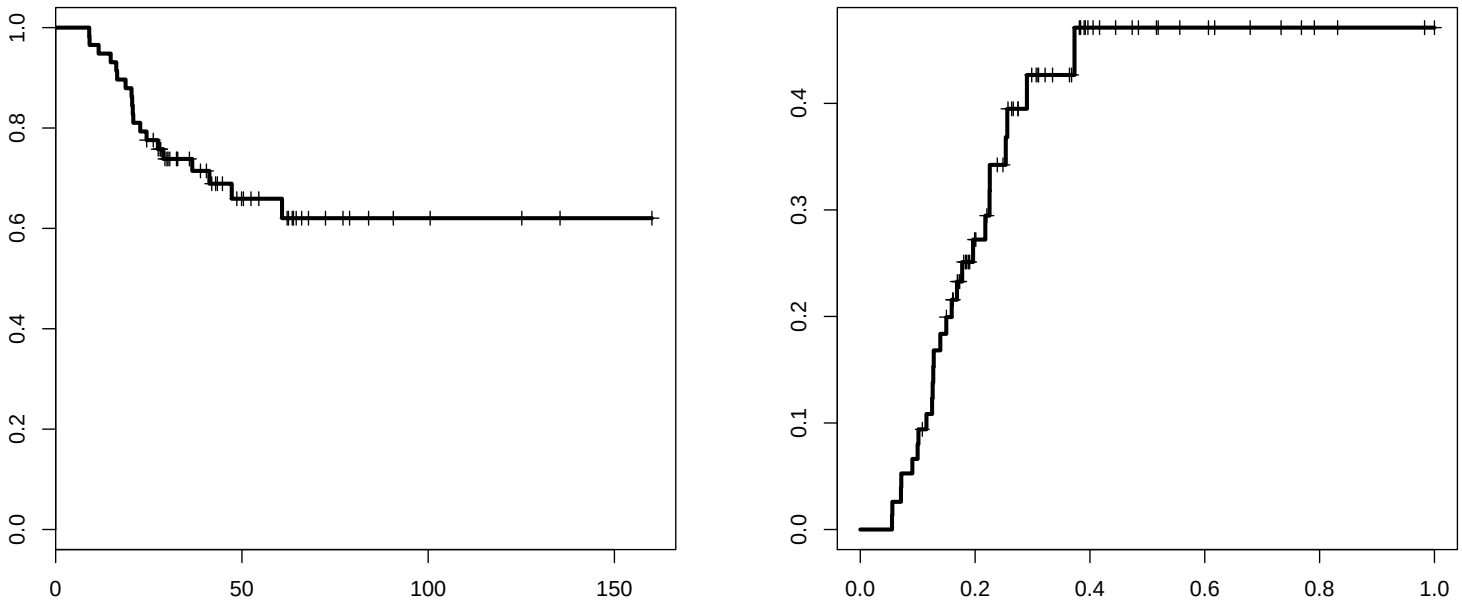



Figure 1: Kaplan-Meier estimated survival curve and cumulative hazard function.

# Cutaneous Melanoma data set

The data set was collected by Eastern Cooperative Oncology Group from 1991 to 1995 on cutaneous melanoma to evaluate the postoperative treatment performance with a high dose of interferon alpha-2b to prevent the recurrence.

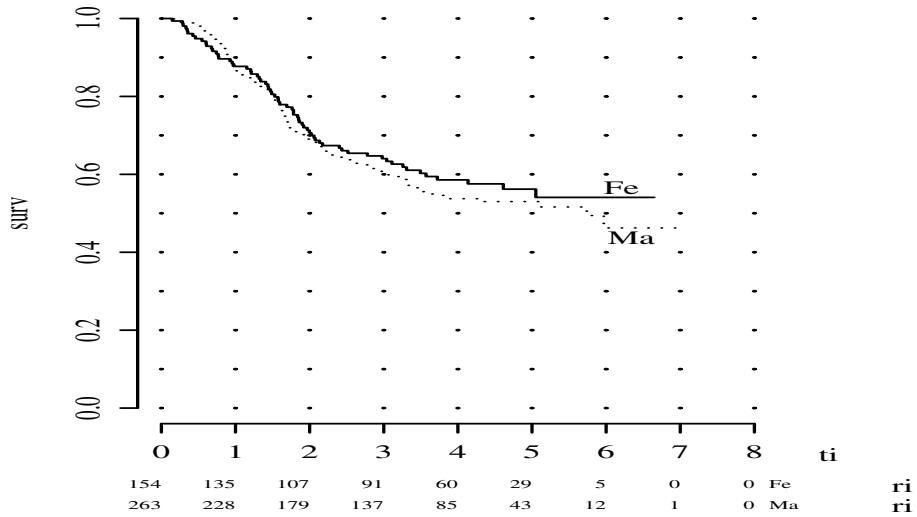


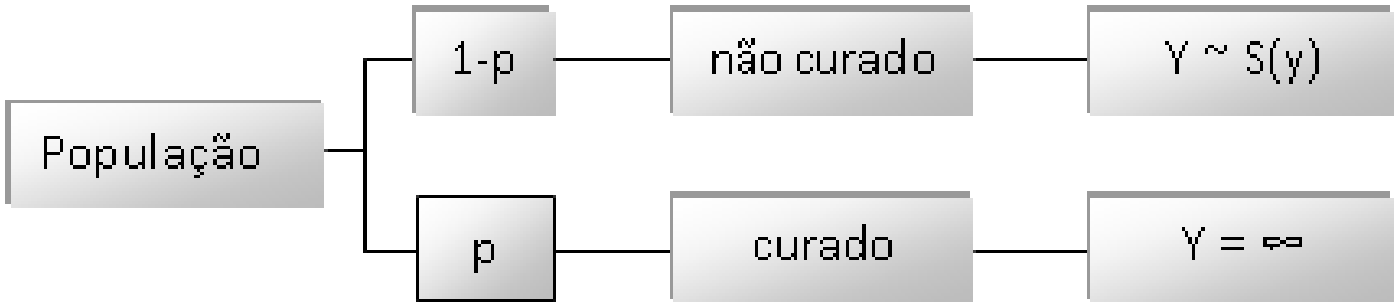

Figure 2: Kaplan-Meier estimated survival curve for data stratified by patient's gender with the number of patients at risk.

# Characteristics of the survival curve of long-term

- At the survival curve an asymptote is clearly reached
- There are Individuals NOT susceptible to the event of interest.
- High censoring rates.
- When $\lim_{t\to\infty} S(t) \neq 0$

# Standard mixture models

- The pioneering work was presented by Boag (1949) and Berkson & Gage (1952);




- The survival function for the population ($S_{pop}(y)$) is given by

$$S_{pop}(y) = p + (1-p)S(y)$$

$S(y)$: Usual survival function (group of uncured)

- Se $p = 1$, então $S_{pop}(t) = S(t)$;
- $S_{pop}(0) = 1$;
- $S_{pop}(t)$ é decrescente;
- $\lim_{t\to\infty} S_{pop}(t) = 1 - p$ (imprópria).

# Non-mixture model:Unified approach

- Unified models have been proposed by Tsodikov et al. (2003) and Rodrigues et al. (2009).
  - $N$ number of causes for the event of interest (latent) with $p_n = P[N = n]$ and $q_n = P[N > n]$, with $n = 1, 2, ...,$ and $T = min\{Z_1, ..., Z_N\}$ where $T = \infty$ if $N = 0$ and $Z_k$, $k = 1, ..., n$ represent the time of occurrence of the event of interest due to the $k$ -th cause.
  - The population survival function is given by

$$\begin{aligned} S_{pop}(t) &= P[N = 0] + P[Z_1 > t, Z_2 > t, ..., Z_N > t, N \geq 1] \\ &= P[N = 0] + \sum_{n=1}^{\infty} P[N = n] P[Z_1 > t, Z_2 > t, ..., Z_N > t] \\ &= p_0 + \sum_{n=1}^{\infty} p_n S(t)^n \\ &= A[S(t)], \end{aligned} \tag{1}$$

  $A(.)$ is the generating function of the sequence $p_n$.

# Cure rate models: Unified approach

The density and risk functions associated with the long-term survival function are given, respectively, by

$$f_{pop}(t) = f(t)\frac{dA(s)}{ds}|_{s=S(t)} \tag{2}$$

and

$$h_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)} = \frac{f(t)\frac{dA(s)}{ds}|_{s=S(t)}}{S_{pop}(t)}. \tag{3}$$

Some distributions are widely used for probability-generating functions, such as Bernoulli, Binomial, Poisson, Negative Binomial and Geometry.

# $N$ has a Negative Binomial (NB) distribution

- We assume that the unobserved (latent) random variable (RV) $N$ has a Negative Binomial (NB) distribution with probability mass function expressed as

$$\mathrm{P}(N = n) = \frac{\Gamma(n + \alpha^{-1})}{n!\Gamma(\alpha^{-1})} \left(\frac{\alpha\theta}{1 + \alpha\theta}\right)^{n} (1 + \alpha\theta)^{-1/\alpha}, \tag{4}$$

  where $n = 0, 1, \ldots$, $\theta > 0$, $\alpha \geq -1$, $1 + \alpha\theta > 0$.

- The long-term SF for the RV $T$ is given by

$$S_p(t; \theta, \alpha) = (1 + \alpha\theta(1 - S_T(t)))^{-1/\alpha}, \quad t > 0, \tag{5}$$

  where $S_T(t)$ is the proper SF.

- We calculate the cure fraction $p$ as

$$p = \lim_{t \to \infty} S_p(t; \theta, \alpha) = (1 + \alpha\theta)^{-1/\alpha}$$

# Defective Distributions

- A distribution is called defective if the integral of its density function does not result in 1 but it results in a value $p \in (0, 1)$, when we change its domain

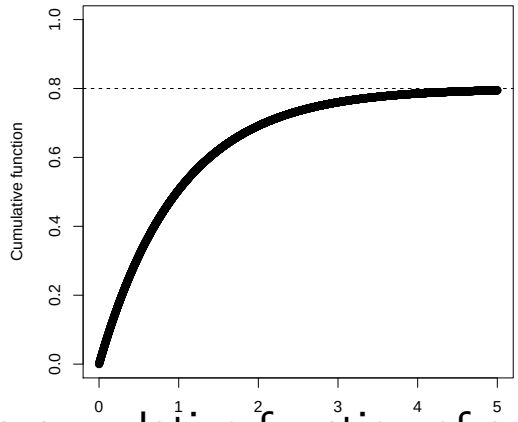



Figure 3: Example of a cumulative function of a defective distribution

- In a defective model it is possible to estimate a cure rate with the use of a natural improper distribution.
- Some defective distributions
  - Gompertz Defective distribution ( rocha2014)
  - Inverse Gaussian Defective distribution (balka2009)
  - Marshall-Olkin family of defective model ( rocha2017)
  - Kumaraswamy Family of defective model (rocha2015b)

# Inverse Gaussian Defective Distribution

- The probability density, survival and hazard functions of the inverse Gaussian model are given by

$$g_0(t) = \frac{1}{\sqrt{2b\pi t^3}} \exp\left\{-\frac{1}{2bt}(1-at)^2\right\}, \tag{6}$$

$$S_0(t) = 1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{2a/b}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right], \tag{7}$$

$$h_0(t) = \frac{\frac{1}{\sqrt{2b\pi t^3}} \exp\left\{-\frac{1}{2bt}(1-at)^2\right\}}{1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{2a/b}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right]}. \tag{8}$$

  where $a > 0$, $b > 0$ and $t > 0$. $\Phi(\cdot)$ represents the cumulative distribution of the standard normal.
- The defective inverse Gaussian distribution is the inverse Gaussian distribution that allows negative values of a. When $a < 0$ the cure rate is calculated by

$$p = \lim_{t\to\infty} S_0(t) = \lim_{t\to\infty} 1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{2a/b}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right] = (1 - e^{2a/b}) \in (0,1).$$

# Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data


RESEARCH ARTICLE | Full Access

## Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data

Jeremias Leão, Víctor Leiva ✉, Helton Saulo, Vera Tomazella

First published: 14 August 2018 | https://doi.org/10.1002/sim.7929 | Cited by: 1

# Frailty models

- The frailty model is characterized by using a random effect, that is a non observable random variable, and it represents a generalization of the Cox model and it may be incorporated in the baseline hazard rate (HR) multiplicatively
- The conditional Harzard and survival functions are given by

$$h_{T|U=u_i}(t;\boldsymbol{x}) = u_i h_0(t)\exp(\boldsymbol{x}_i{}^\top\boldsymbol{\varphi}), \quad (9)$$

$$S_{T|U=u_i}(t;\boldsymbol{x}) = \exp(-u_i H_0(t)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})) \quad i=1,\ldots,n, \quad (10)$$

where $u$ represents the frailty variable and $h_0(.)$ and $H_0(.)$ are baseline hazard rate and cumulative hazard rate respectively and $\boldsymbol{x}$ is observed variable

- To get the unconditional SF, we need to integrate out the frailty component as

$$S_T(t) = \int_0^\infty \exp(-uH_0(t)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi}))f_U(u)\,\mathrm{d}u = Q(H_0(t))e^{(\boldsymbol{x}^\top\boldsymbol{\varphi})}. \quad (11)$$

where $Q(\cdot)$ denotes the Laplace transform.

- An important point in Frailty models is the choice of the distribution for the Frailty variable. In this work we considered Birnbaum and Saunders distribu-

# A parameterized version of the BS distribution

- Birnbaum and Saunders (1969), introduced a distribution to fatigue life data model.
- A RV $U$ is BS distributed, $U \sim \mathrm{BS}(\alpha, \beta)$ and the PDF is given by

$$f_U(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2}\left[\frac{u}{\beta} + \frac{\beta}{u} - 2\right]\right) \frac{u^{-3/2}[u+\beta]}{2\alpha\beta^{1/2}}, \quad u > 0, \tag{12}$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters respectively

- Santos-Neto et al. (2012) proposed BS distribution parameterized by its mean and precision. The shape parameters and precision parameters are given by

$$\delta = 2/\alpha^2 > 0 \tag{13}$$

and

$$\mu = \beta[1 + \alpha^2/2] > 0 \tag{14}$$

# A parameterized version of the BS distribution

- The PDF of the BS distribution parameterized $U \sim \mathrm{RBS}(\mu, \delta)$ is given by

$$f_U(u) = \frac{\exp(\delta/2)\sqrt{\delta+1}}{4\sqrt{\pi\mu}u^{3/2}}\left[u + \frac{\delta\mu}{\delta+1}\right]\exp\left(-\frac{\delta}{4}\left[\frac{u\{\delta+1\}}{\delta\mu} + \frac{\delta\mu}{u\{\delta+1\}}\right]\right). \quad (15)$$

- $\mathrm{E}[U] = \mu$ and $\mathrm{Var}[U] = \dfrac{\mu^2}{(\delta+1)^2(2\delta+5)}$

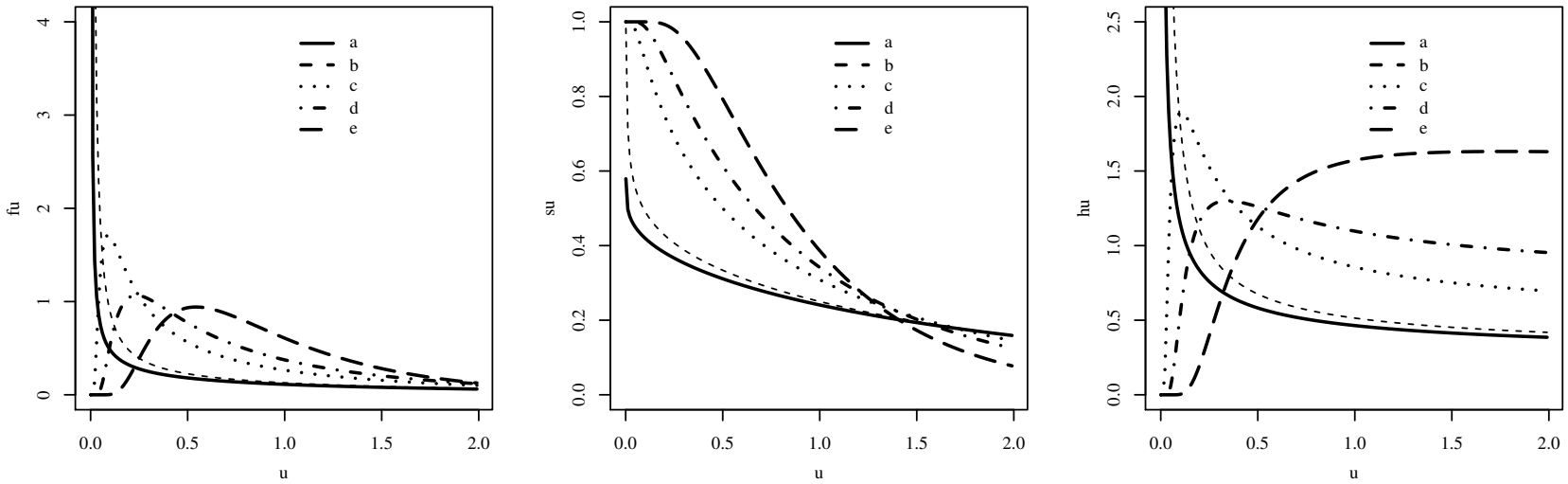



Figure 4: Plots of PDF, SF and HR of the BS distribution $U \sim \mathrm{RBS}(\mu = 1, \delta)$.

# A RBS frailty model for survival data

- We assume that the frailty $U$ has a RBS distribution with parameters $\mu = 1$ and $\delta$, where $\mathrm{E}[U] = 1$ and $\mathrm{Var}[U] = (2\delta + 5)/(\delta + 1)^2$. The variance quantifies the amount of heterogeneity among patients.
- The Laplace transform for the RBS distribution is given by

$$Q(s) = \frac{\exp\left(\frac{\delta}{2}\left(1 - \frac{\sqrt{\delta+4s+1}}{\sqrt{\delta+1}}\right)\right)\left(\sqrt{\delta+4s+1} + \sqrt{\delta+1}\right)}{2\sqrt{\delta+4s+1}}. \qquad (16)$$

- Evaluating (16) at $s = H_0(t)$, we get the unconditional SF and HR under the RBS frailty as

$$S_T(t;\delta) = \frac{\exp\left(\frac{\delta}{2}\left(1 - \sqrt{\delta+4H_0(t)+1}/\sqrt{\delta+1}\right)\right)\left(\sqrt{\delta+4H_0(t)+1} + \sqrt{\delta+1}\right)}{2\sqrt{\delta+4H_0(t)+1}}. \qquad (17)$$

$$h_T(t;\delta) = h_0(t)\left(\frac{\delta(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t)+1}+4H_0(t)+3)+2}{(\delta+4H_0(t)+1)(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t)+1}+1)}\right). \qquad (18)$$

# Analyzing cutaneous melanoma data set

- In this application we assumed the regression structure.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}, \quad i = 1, \ldots, 426.$$

Table 1: ML estimates (with SE and $p$-value) of the indicated parameter for the NBCrBSF model with the melanoma data.

| Parameter/Covariate name | Parameter | Estimate | SE | $p$-value |
|---|---|---|---|---|
| Dispersion/competing causes (NB) | $\alpha$ | 8.0498 | 2.8105 | – |
| Shape/baseline HR (Weibull) | $\kappa$ | 3.2591 | 0.6153 | – |
| Scale/baseline HR (Weibull) | $\gamma$ | 6.1634 | 3.5862 | – |
| Precision/frailty (BS) | $\delta$ | 0.7918 | 3.5633 | – |
| Constant | $\beta_0$ | 0.4727 | 0.5931 | 0.4254 |
| Treatment | $\beta_1$ | 0.2232 | 0.1076 | 0.0381 |
| Age | $\beta_2$ | −0.0057 | 0.0037 | 0.1268 |
| Gender | $\beta_3$ | −0.1060 | 0.1136 | 0.3505 |
| Nodule category | $\beta_4$ | −0.6233 | 0.1891 | 0.0010 |

- The estimate frailty variance

$$\widehat{\mathrm{Var}}(U) = 2\widehat{\delta} + 5/(\widehat{\delta}+1)^2 = 2.0506$$

- To calculate the cure rate parameter for each patient we consider the logistic regression model defined as

$$p_{0_i} = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}, \quad i = 1, \ldots, m.$$

Table 2: ML estimates (with estimated asymptotic SE and 95% confidence interval) of the cure fraction stratified by treatment, nodule category and patient's gender for the NBCrBSF model with the melanoma data.

| Treatment | Nodule category | Gender | Estimate | SE | 95% confidence interval |
|---|---|---|---|---|---|
| 0 | Absent | Female | 0.6160 | 0.1678 | (0.6001, 0.6319) |
| | | Male | 0.5907 | 0.2359 | (0.5683, 0.6131) |
| | Present | Female | 0.4624 | 0.1956 | (0.4438, 0.4810) |
| | | Male | 0.4362 | 0.2240 | (0.4149, 0.4575) |
| 1 | Absent | Female | 0.6673 | 0.2508 | (0.6435, 0.6911) |
| | | Male | 0.6433 | 0.2966 | (0.6151, 0.6715) |
| | Present | Female | 0.5181 | 0.2978 | (0.4898, 0.5464) |
| | | Male | 0.4917 | 0.3357 | (0.4598, 0.5236) |

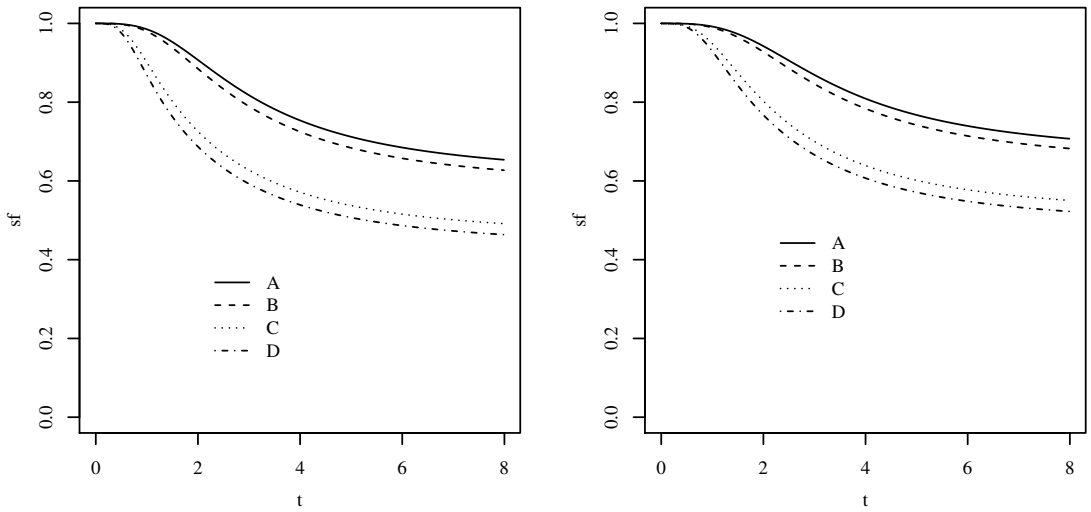



Figure 5: Overall SF fitted with the NBCrBSF model stratified by nodule category and patient's gender (A: absent and female, B: present and female, C: absent and male, D: present and male) for patients with no treatment = 0 (left) and with treatment = 1 (right), using melanoma data.

- We observed for example that the patient who did not receive treatment with absence of nodules and female has a cure rate equal to 0.61

# Diagnostic Analysis

- We carried out a diagnostic analysis based on local influence. note that cases #255, #290, #279 and #341 are detected as potentially influential observations under the considered perturbation schemes.

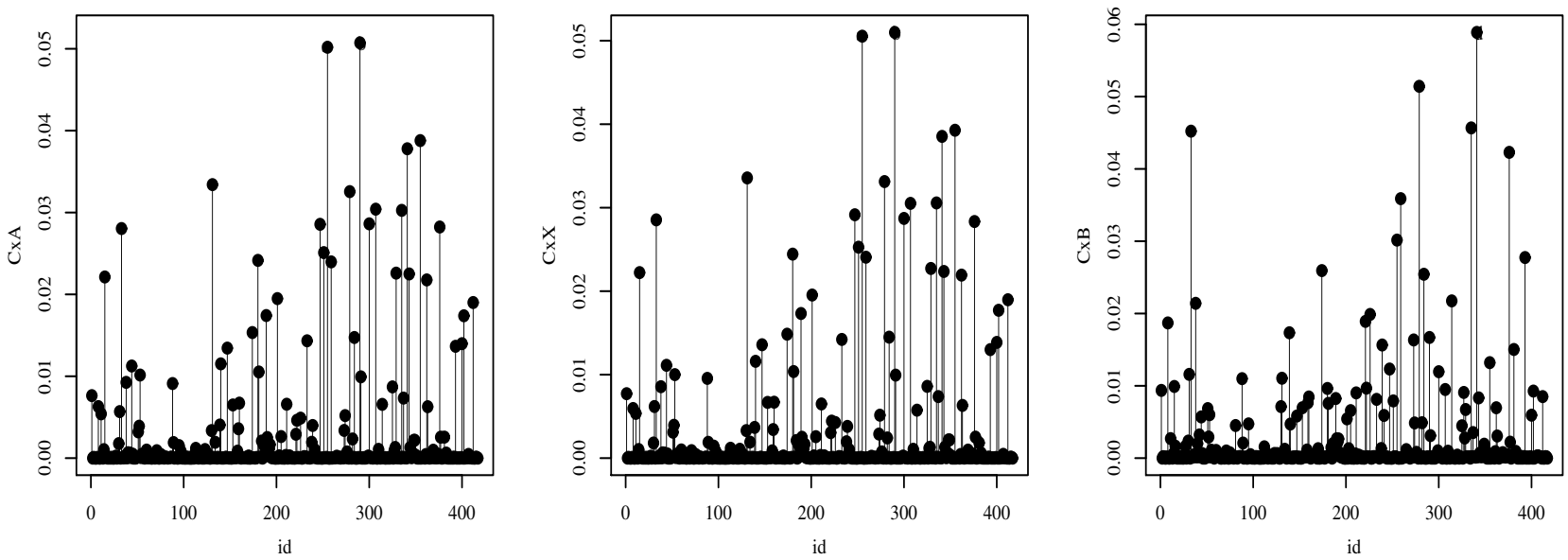



Figure 6: Index plots of $C_i$ for $\alpha$ (left), $\boldsymbol{\xi} = (\delta, \gamma, \kappa)^\top$ (center) and $\boldsymbol{\beta}$ (right) with case-weight perturbation and melanoma data.

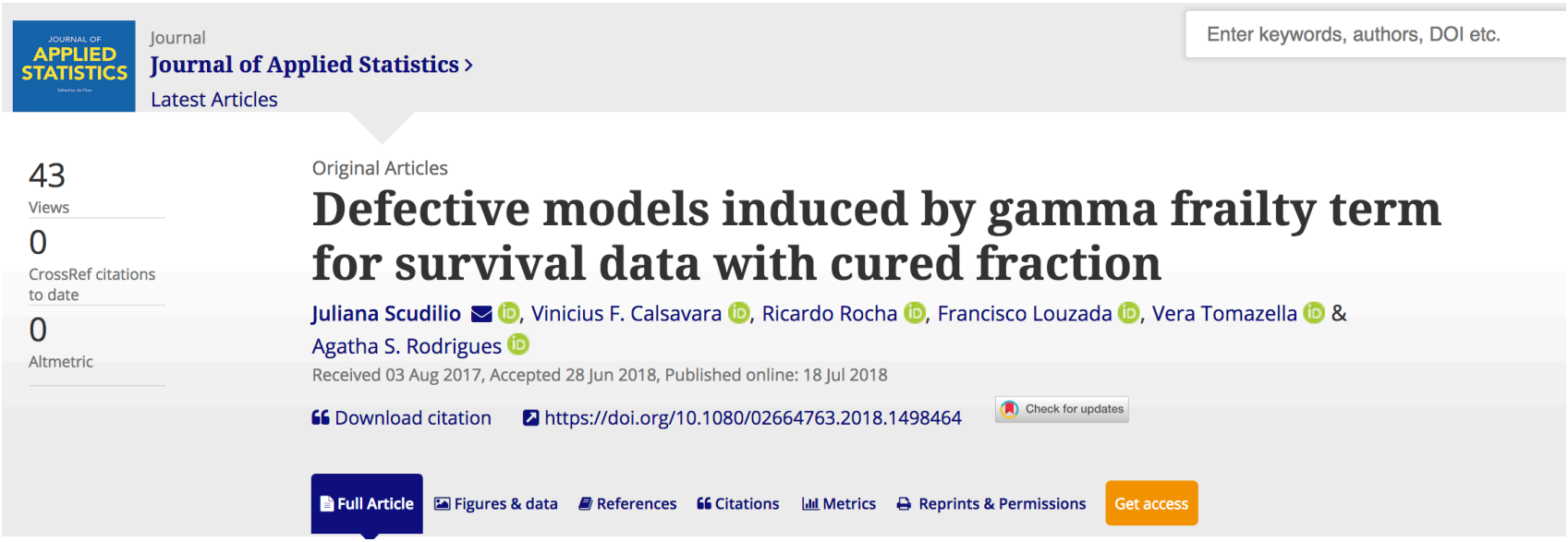
# Conclusion

- We proposed a new methodology based on a cure rate model with frailty described by the reparamerized Birnbaum-Saunders distribution.
- The proposed methodology encompassed estimation and inference about the model parameters, as well as local influence diagnostics under different perturbation schemes.
- We illustrated the methodology with data of malignant melanoma. The empirical results showed the potentiality of this methodology

# Main references


- Berkson J, Gage RP. Survival curve for cancer patients following treatment. Journal of the American Statistical Association 1952; 47: 501-515.
- Rodrigues J, Cancho VG, de Castro M, Louzada F. On the unification of long-term survival models. Statistics and Probability Letters 2009; 79: 753-759. Birnbaum, ZW, Saunders, SC (1969) A new family of life distributions. J. Appl. Prob., 6, 319-327.
- Leao J, Leiva V, Saulo H, Tomazella V. Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. Biom J. 2017; 59:291-314.
- Leiva V, Santos-Neto M, Cysneiros FJA, Barros M. Birnbaum-Saunders statistical modelling: A new approach. Statistical Modelling 2014; 14: 21-48.
- Santos-Neto M, Cysneiros FJA, Leiva V, Barros M. Reparameterized Birnbaum-Saunders regression models with varying precision. Electron J Stat. 2016; 10:2825-2855.

# Defective models induced by gamma frailty term for survival data with cure fraction



# Defective models induced by frailty term

- Let $V \sim Gamma(1/\theta, 1/\theta)$, with $\mathbb{E}(V) = 1$ and $\text{Var}(V) = \theta$ (**?**). The Laplace transform of the gamma frailty distribution is expressed by

$$\mathcal{L}_g(s) = (1 + \theta s)^{-1/\theta}. \tag{19}$$

- The unconditional survival, density and hazard functions in the gamma frailty model are given by

$$S(t) = \left[1 - \theta \log S_0(t)\right]^{-1/\theta}, \tag{20}$$

$$f(t) = h_0(t) \left[1 - \theta \log S_0(t)\right]^{-1-1/\theta}, \tag{21}$$

and

$$h(t) = h_0(t) \left\{1 - \theta \log S_0(t)\right\}^{-1}, \tag{22}$$

where $S_0(t)$ can be either proper or not proper survival function

# Defective gamma-inverse Gaussian model

- Let $S_0(t)$ be the Survival function of the defective inverse Gaussian model with parameters $a < 0$ and $b > 0$ and gamma frailty term
- The survival and hazard functions of the defective gamma-inverse Gaussian model are given by

$$\begin{aligned} S(t) &= [1 - \theta \log S_0(t)]^{-1/\theta} \\ &= \left\{1 - \theta \log \left\{1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{\frac{2a}{b}}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right]\right\}\right\}^{-1/\theta} \end{aligned} \quad (23)$$

$$h(t) = \frac{\frac{1}{\sqrt{2b\pi t^3}} \exp\left\{-\frac{1}{2bt}(1-at)^2\right\}}{1-\left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right)+e^{2a/b}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right]} \left\{1 - \theta \log\left[1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{2a/b}\Phi\left(\frac{-1-at}{\sqrt{bt}}\right)\right]\right]\right\}^{-1} \quad (24)$$

- We calculate the cure fraction $p$ for the defective gamma-inverse Gaussian model as
- $p = \lim_{t\to\infty} S(t) = \lim_{t\to\infty} [1 - \theta \log S_0(t)]^{-1/\theta} = \left[1 - \theta \log\left(1 - e^{2a/b}\right)\right]^{-1/\theta}$

# Analyzing the breast cancer data set

- In this application we assumed the regression structure.

$$h(t|V,\boldsymbol{x}) = V\, h_0(t) e^{\boldsymbol{x}^\top \boldsymbol{\beta}},$$

We considered the model only with the covariate $N$ (tumor location), $N = 0$ ( neigboring lymph nodes do not have cancer )and $N = 1$ (neigboring lymph nodes have cancer )

Table 3: Maximum likelihood estimates of the gamma-inverse Gaussian model with the covariate Location of Tumor $N = 0$ and $N = 1$

| Parameter | Estimate | Std. Error | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|
| $a$ | -5.1892 | 2.588 | -10.2616 | -0.1168 |
| $b$ | 1.9289 | 0.6592 | 0.6369 | 3.2209 |
| $\theta$ | -0.801 | 4.1804 | -8.9945 | 7.3925 |
| $\beta_0$ | 3.6569 | 2.5964 | -1.4319 | 8.7456 |
| $\beta_1$ | 1.0042 | 0.9295 | -0.8177 | 2.8261 |
| $p_0$ | 0.8245 | 0.0951 | 0.6380 | 0.9999 |
| $p_1$ | 0.5384 | 0.0837 | 0.3743 | 0.7025 |

Note that $a = -5.1892$. In this case we have a defective gamma-inverse Gaussian model.

- The cure fraction is $p = \left\{1 - \theta e^{x\beta} \log\left(1 - e^{\frac{2a}{b}}\right)\right\}^{-\frac{1}{\theta}}$
- The proportion of cured individuals was estimated in $p_0 = 0.82$ for the group $N = 0$ (red line) and $p_1 = 0.53$, for the group N = 1 (green line)

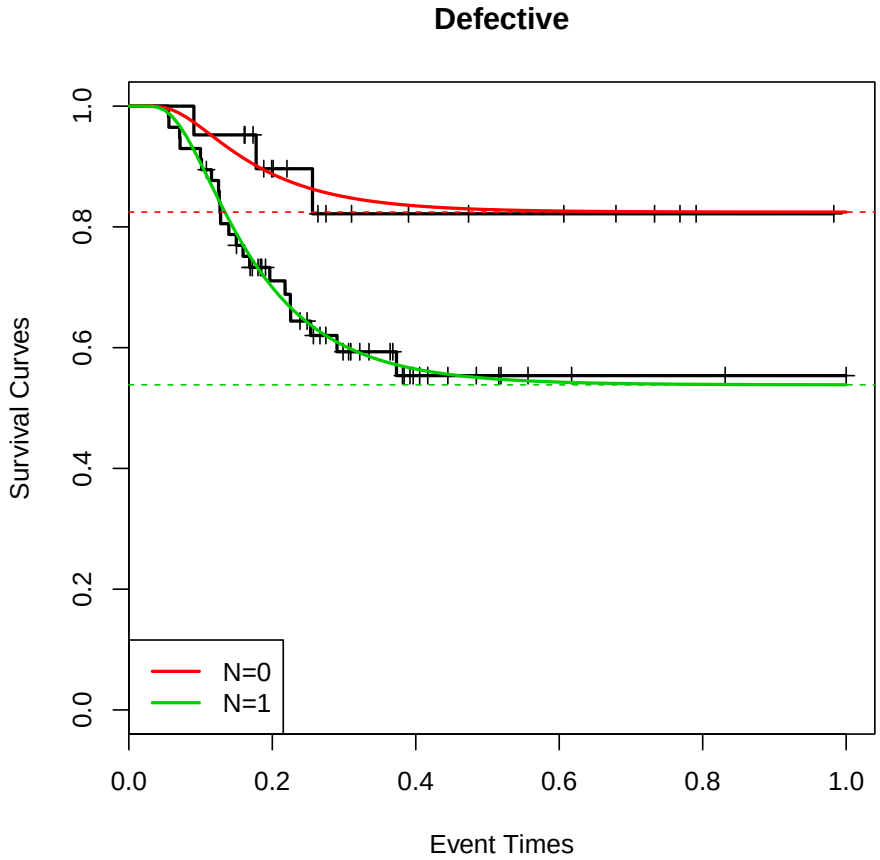



Figure 7: Survival curves of the gamma-inverse Gaussian model ($N = 0$ and $N = 1$)

# Conclusion

- Once you have a defective model, it will lead to a cure fraction when the estimation procedure presents a value out of the usual range of parameters.
- We showed that when $a > 0$, we have the frailty models, gamma-inverse Gaussian..
- When we have $a < 0$, we have the defective inverse Gaussian induced by the frailty gamma.
- We showed that we can induce new defective distributions when using the gamma frailty term,

- We illustrated the methodology with data of breast cancer. The empirical results showed the potentiality of this methodology

# Main references


- Balka, J., Desmond, A. F. & McNicholas, P. D. (2009). Review and implementation of cure models based on first hitting times for wiener processes. Lifetime Data Analysis, 15(2), 147-176.
- Rocha, R., Tomazella, V. & Louzada, F. (2014). Inferência clássica e Bayesiana para o modelo de fração de cura Gompertz defeituoso. Revista Brasileira de Biometria, 32(1), 104-114.
- Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F. & Eudes, A. (2015). New defective models based on the Kumaraswamy family of distributions with application to cancer data sets. Statistical methods in medical research, pages 1-15.
- Rocha, R., Nadarajah, S., Tomazella, V. & Louzada, F. (2017). A new class of defective models based on the Marshall-Olkin family of distributions for cure rate modeling. Computational Statistics & Data Analysis, 107, 48-63.
- Wienke, A. (2003). Frailty models. Wiley StatsRef: Statistics Reference Online.

# THANK YOU!