stats

MDPI

Article

Defective Gamma–G Family for Cure Fraction Models: Novel Survival Methods with Applications to Cancer Data

Cynthia A. V. Tojeiro [1,2,*], Vera L. D. Tomazella [3], Agatha S. Rodrigues [2] and Pedro R. D. Marinho [2]

[1] Institute of Mathematics and Statistics, Federal University of Goiás (UFG), Goiânia 74690-900, GO, Brazil
[2] Department of Statistics, Federal University of Paraíba (UFPB), João Pessoa 58051-900, PB, Brazil; agatha@ccen.ufp.br (A.S.R.); pedro@de.ufpb.br (P.R.D.M.)
[3] Department of Statistics, Federal University of São Carlos (UFSCar), São Carlos 13565-905, SP, Brazil; vera@ufscar.br
* Correspondence: cynthiatojeiro@ufg.br

Abstract

In this paper, we propose two novel defective survival models within the Gamma–G family: the defective Gamma–Gompertz and the defective Gamma–Dagum distributions. In contrast to the corresponding Gamma–G mixture cure formulation, in which the Gamma–G distributional parameters are combined with an explicit cure fraction mixing parameter, the proposed defective formulation induces the cure fraction through the limiting behavior of the survival function. Thus, within the same Gamma–G baseline structure, the model avoids introducing an additional cure fraction parameter. The motivation for these new models lies in the limited set of defective distributions currently available, despite the increasing demand for flexible cure rate models in biomedical applications. By extending the defective property to the Gamma–G construction, our approach fills this methodological gap while providing models that are both interpretable and computationally efficient. We show that the Gamma–G construction preserves defectiveness whenever the baseline distribution is defective, thus establishing a coherent theoretical foundation. Both models allow covariate effects through regression structures on shape, scale, and, in the case of the Gamma–Dagum distribution, on the cure fraction parameter, resulting in flexible and interpretable specifications. Parameters are estimated via maximum likelihood, and an extensive Monte Carlo study confirms estimator consistency and accurate coverage in finite samples. The practical relevance of the models is illustrated with two large clinical datasets on melanoma and cervical cancer from the São Paulo Cancer Registry. Results reveal that the proposed models provide competitive goodness-of-fit and offer useful insights into long-term survival compared to traditional cure rate approaches. Overall, this work introduces a unifying and flexible framework for defective survival models, extending their applicability and delivering practical improvements over existing cure models.

Keywords: cure fraction; defective distribution; defective Gamma–G distribution; regression model; survival analysis

Check for updates

Academic Editors: Bogdan Oancea, Adrian Pană and Cătălina Liliana Andrei

Received: 1 May 2026
Revised: 30 May 2026
Accepted: 10 June 2026
Published: 17 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In many biomedical and epidemiological studies, it is increasingly common to observe that a proportion of individuals never experience the event of interest, even after long-term follow-up [1–4]. This subpopulation, referred to as the cure fraction, poses a challenge to classical survival models, which assume that all individuals are susceptible and will eventually fail. Ignoring this cured portion can lead to biased estimates, poor risk

stratification, and misleading conclusions, especially in cancer research and clinical studies where long-term remission or immunity is possible.

To address such scenarios, Berkson and Gage [2], building on the work of Boag [1], proposed the standard mixture model for the cure fraction. The survival function is adjusted to $S(t) = p + (1 - p)S_0(t)$ such that $S_0(t)$ is a standard survival function. Thus, it is apparent that $S(t)$ converges to p as time increases. In Berkson and Gage [2] an analysis is carried out on patients with stomach cancer, and from there, various other cure fraction studies have been proposed in the literature, focusing on this standard mixture model. The most common choices for $S_0(t)$ are the Weibull, Log-logistic, and Lognormal models. Recently, different models have been proposed for this purpose, as seen in Martinez et al. [5], Shi et al. [6], Wang and Pal [7], de Souza et al. [8], Mota et al. [9], and Cancho et al. [10].

As an alternative to mixture models, the literature has introduced approaches based on defective distributions. These distributions are characterized by probability density functions whose integrals over the entire domain yield a value $p \in (0, 1)$ rather than 1 [11]. Consequently, the corresponding cumulative distribution functions converge not to 1, but to p_0, leading the survival function to asymptotically approach $1 - p_0$. Models exhibiting this behavior are commonly referred to as defective models. It is important to emphasize that defective distributions are inherently improper. When applied in the context of cure rate modeling, the proportion of immune or long-term surviving individuals in the population can be directly obtained by evaluating the limiting value of the survival function based on the estimated parameters.

In defective models, the fraction of immunized or long-term surviving individuals is induced directly by the limiting value of the survival function, avoiding the need for an explicit cure-fraction mixing parameter. In contrast, mixture cure models estimate the cured proportion jointly with the remaining model parameters through likelihood optimization. In the specific context of a Gamma–G mixture cure formulation, this would require combining the Gamma–G distributional parameters with an additional cure fraction parameter. The proposed defective Gamma–G formulation instead incorporates the cure fraction directly into the survival structure. This distinction should not be interpreted as a universal reduction in dimensionality relative to all classical cure models, but rather as an integrated parametrization within the defective Gamma–G framework. In the literature, three defective distributions commonly employed for this purpose are the Gompertz, inverse Gaussian, and Dagum distributions [12,13].

The Gompertz distribution holds significant relevance in various real-world applications, including reliability analysis, human mortality modeling, and actuarial science. It is particularly versatile, as it can accommodate both right- and left-skewed data. Recognized as one of the classical distributions used to model survival functions under mortality laws, it has been extensively studied in diverse contexts by numerous researchers [14]. This distribution is characterized by two positive parameters, one for location and another for shape, both positive. Notably, when the shape parameter takes on negative values, the distribution becomes defective. For example, a Gompertz model was fitted to breast cancer data in [15] and later adapted in a modified form for pediatric cancer data in [16]. In ref. [14], the model is further extended to incorporate covariates.

The Dagum distribution is another notable example of a model that can assume a defective form, making it particularly relevant in the context of survival analysis with cure fraction modeling. Originally proposed by Dagum [17] in the context of modeling personal income distribution, the Dagum distribution has since been recognized for its flexibility in handling heavy-tailed data and various hazard rate shapes. Its applications have extended to fields such as reliability analysis, demography, and economics. More

recently, ref. [18] introduced a defective version of the Dagum distribution, where the cumulative distribution function converges to a value less than one, allowing the survival function to stabilize at a positive level representing the cure fraction. A notable feature of this defective model is that the cure fraction is directly expressed as a function of a single shape parameter, simplifying interpretation and estimation. The model accommodates both increasing and decreasing hazard functions and has shown strong performance in applications involving long-term survivors in medical studies.

Complementing this line of research, Calsavara et al. [19] applied defective Gompertz and inverse Gaussian distributions to interval-censored survival data, demonstrating their effectiveness in capturing long-term survivor behavior. In a subsequent study, ref. [12] extended these models to handle zero-inflated data structures. New defective distributions have been constructed using generator families such as Marshall–Olkin [20] and Kumaraswamy [21], further broadening the methodological tools available for modeling survival data with cure fractions. These developments reflect a growing interest in flexible defective alternatives for cure rate analysis, particularly in settings where the long-term survival plateau can be represented directly through the limiting behavior of the survival function.

Motivated by the need for flexible models that can capture both the presence of long-term survivors and complex hazard structures, we propose in this paper a novel class of defective distributions derived from the Gamma–G family. Unlike traditional defective models, which are based directly on baseline distributions, our approach incorporates the Gamma–G construction to enhance distributional flexibility while preserving the defective property. This allows the cured fraction to emerge from the limiting behavior of the survival function, providing an integrated parametrization for long-term survival within the proposed defective framework.

In recent decades, several generator-based models have been developed to increase the flexibility of classical distributions in survival analysis. Among them, the Gamma–G family first introduced by Zografos and Balakrishnan [22] has gained considerable attention due to its ability to generate flexible models by composing the cumulative distribution function of the Gamma distribution with the survival function of a given baseline distribution. Since its introduction, numerous variants have been explored in the literature, such as the Gamma–Dagum [23], Gamma–log-logistic [24], Gamma–exponentiated Weibull, Gamma–extended Fréchet [25], Gamma–half normal [26], Gamma–inverse Weibull [27], Gamma–linear failure rate [28], and Gamma–logistic [29] and [30], among others. These constructions generally arise from transformations of well-established baseline distributions, and in many cases, linear combinations or mixtures involving a distribution G lead to new probabilistic models in which G appears as a special case.

The Gamma–G family has not yet been explored in the context of defective survival distributions. This gap provides the main motivation for the present work. Our proposal builds upon and extends the existing line of research on defective models, such as those introduced by [20,21,31] by incorporating the defective property into the Gamma–G construction. In doing so, we establish a coherent and unifying framework that embeds the cure fraction directly into the survival function while preserving the desirable flexibility of the Gamma–G family. Beyond this theoretical contribution, it is equally important to highlight the practical relevance of enlarging the toolbox of defective survival models for biomedical applications, since survival datasets often exhibit clear evidence of a cured fraction, typically seen as a plateau in the Kaplan–Meier curve.

Within this framework, we introduce two novel defective distributions: the defective Gamma–Gompertz and the defective Gamma–Dagum. These models provide flexible parametric alternatives for estimating survival functions in the presence of a cured fraction

in the population. A key feature of the proposed formulation is that the cure fraction is induced by the limiting value of the survival function, rather than introduced through an explicit mixing parameter. More specifically, when compared with the corresponding Gamma–G mixture cure formulation, in which the Gamma–G parameters are combined with an additional mixing parameter for the cure fraction, the proposed defective Gamma–G models avoid the inclusion of this explicit cure fraction parameter. Thus, the long-term survivor proportion is expressed as a function of the model parameters under the defective structure. Beyond their theoretical appeal, the proposed models also accommodate covariates through regression structures, providing flexible and interpretable tools for analyzing survival data with a cure fraction.

Motivating Example: Melanoma and Cervical Cancer in the State of São Paulo

Melanoma and cervical cancer (CC) remain significant public health concerns in both Brazil and the global context. Although melanoma accounts for a smaller proportion of skin cancer cases compared to non-melanoma types, it is considerably more aggressive and associated with a higher mortality rate. In Brazil, estimates indicate over 8400 new melanoma cases annually during the 2023–2025 period, with a higher incidence in the southern regions of the country [32]. Melanoma progression and prognosis are closely linked to clinical staging at diagnosis, highlighting the importance of accurate early detection and stratified survival analysis.

Cervical cancer remains a significant global health concern among women, primarily caused by persistent infection with oncogenic types of the human papillomavirus (HPV). While HPV infection is common and often asymptomatic, a small proportion of cases may progress to precancerous lesions and eventually develop into invasive cervical cancer over time. According to GLOBOCAN 2022 estimates, cervical cancer ranked as the fourth most frequent cancer in the female population worldwide, with approximately 662,300 new cases and 348,900 deaths, accounting for around 3.1% of all cancer cases in women [33]. In Brazil, projections for the period 2023–2025 indicate an average of 17,010 new cases annually, corresponding to an age-standardized incidence rate of approximately 15.38 cases per 100,000 women [32]. The disease continues to show marked regional disparities, with the North and Northeast regions presenting the highest incidence rates. Despite being largely preventable through HPV vaccination and regular cytological screening, cervical cancer remains one of the leading causes of cancer-related mortality among Brazilian women. In 2022, the estimated national mortality was close to 9000 deaths, corresponding to an age-standardized rate of approximately 8.5 deaths per 100,000 women [34].

The datasets used in this study were provided by the *Fundação Oncocentro de São Paulo* (FOSP), a public institution linked to the São Paulo State Health Department. FOSP is responsible for coordinating the Hospital Cancer Registry system in the state, collecting, organizing, and evaluating clinical and epidemiological information related to cancer cases treated in hospitals across São Paulo. In addition to monitoring cancer incidence and outcomes, FOSP supports the planning and implementation of public health policies aimed at prevention, early diagnosis, and improved oncological care. The datasets made available include detailed information on tumor staging, vital status, and follow-up time, allowing for survival modeling in scenarios that account for the presence of a potentially cured fraction.

In this study, we analyze two independent datasets from the state of São Paulo, with a focus on modeling the cured fraction. The first dataset comprises 7823 patients diagnosed with melanoma between 2000 and 2014, with follow-up until 2018. The event of interest was death due to melanoma, with right-censoring applied to patients who did not die from this cause during the study period. The maximum observation time was approximately

18 years, and the median follow-up time was 5.1 years. According to the American Joint Committee on Cancer (AJCC) staging system, early clinical stages (I or II) are associated with a favorable prognosis and are typically treated with surgery. In contrast, patients diagnosed at advanced stages (III or IV) face significantly worse outcomes, with melanoma-specific survival at 10 years varying from 24% to 88% [35]. In this sample, 68.15% of the patients were classified as stage I or II, and the Kaplan–Meier survival curve for this subgroup reveals a visible plateau, suggesting the presence of long-term survivors and motivating the application of cure models.

The second dataset includes 12,030 women diagnosed with cervical cancer (C53) between 2012 and 2022. The event of interest was death due to cervical cancer, with censoring applied to all other cases. Survival analysis stratified by clinical stage revealed marked differences in risk dynamics. Among patients diagnosed at early stages (Stage I/II), only 15.8% experienced the event, while 54.5% of those diagnosed at advanced stages (Stage III/IV) died from the disease. Moreover, the survival curve for early-stage patients exhibits a clear plateau, consistent with the existence of a cured subgroup, while the curve for late-stage patients trends toward zero.

These findings underscore the importance of cure rate models in survival analysis, as traditional models assuming that all individuals remain at risk indefinitely may fail to capture the full complexity of cancer progression and prognosis. Modeling a cured fraction allows for more realistic inference on long-term outcomes and the effect of clinical staging at diagnosis on patient survival.

The article is organized as follows. In Section 2, we review classical defective models, with a focus on the Gompertz and Dagum distributions, and discuss their properties in the context of survival data with a cure fraction. Section 3 introduces the proposed defective Gamma–G models, including the Gamma–Gompertz and Gamma–Dagum distributions, and presents a formal proof that the defective property is preserved under the Gamma–G construction. In Section 4, we discuss the estimation of parameters using the log-likelihood function and numerical methods. Section 5 presents a Monte Carlo simulation study to evaluate the asymptotic performance of the maximum likelihood estimators. In Section 6, we apply the proposed models to two real-world datasets on cancer patients provided by the Fundação Oncocentro de São Paulo (FOSP). Finally, in Section 7, we summarize the main findings and provide concluding remarks.

2. Methodology

2.1. Defective Model

A distribution is called defective if the integral of its density function does not result in 1, but in a value $p_0 \in (0,1)$, when the domain of the parameters are changed. Figure 1 illustrates the cumulative function of a defective distribution where $p_0 = 0.75$. Obviously, the defective distribution is not proper.

In this context, the interpretation of p is the proportion of individuals cured or not susceptible to the event of interest, representing a model that accommodates a cure rate or long-term model. The defective models have the advantage of not needing the assumption of the presence of immune individuals in the data set. In order to use the defective distributions theory in a competitive way, we need a larger variety of these distributions.

Defective distributions are distinguished by the fact that their probability density functions integrate to less than one over the entire support of the random variable when at least one of their parameters deviates from the standard configuration [36]. This property makes such distributions particularly suitable for modeling survival data in the presence of a cured fraction, as the survival function inherently accounts for the immune subgroup without the need for an explicit mixture formulation.

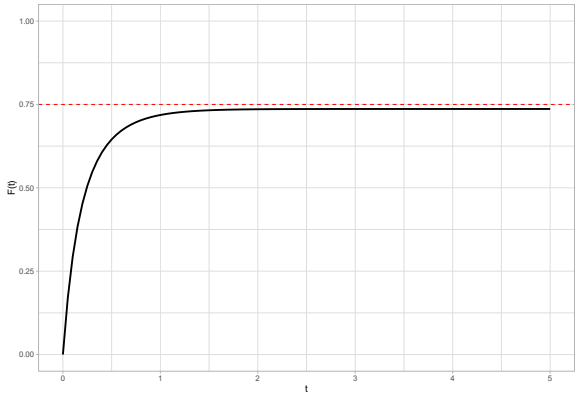

Figure 1. Example of a cumulative function of a defective distribution. The different coloured lines represent different parameter configurations, illustrating cumulative distribution functions that converge to limiting values below one.

2.1.1. Defective Gompertz Distribution

Consider the Gompertz distribution, which depends on two parameters, typically one for location and another for shape. When the shape parameter assumes negative values, the distribution becomes defective, thereby extending its parametric domain and enabling its application in scenarios involving long-term survivors.

The Gompertz distribution, in its appropriate formulation, is defined by two positive parameters, $\alpha > 0$ and $\beta > 0$, where α represents the scale parameter, and β denotes the shape parameter.

The probability density function of the Gompertz distribution for the random variable representing the time to the event of interest, denoted as t, and the corresponding survival function are, respectively, as follows:

$$f_{\text{Go}}(t \mid \alpha, \beta) = \beta e^{\alpha t} e^{-\frac{\beta}{\alpha}(e^{\alpha t}-1)}, \tag{1}$$

and

$$S_{\text{Go}}(t \mid \alpha, \beta) = e^{-\frac{\beta}{\alpha}(e^{\alpha t}-1)}. \tag{2}$$

We have that $S_{\text{Go}}(t) \to 0$ as $t \to \infty$. Therefore, the distribution presented in this way is not suitable for data with a cure fraction.

The distribution of the Gompertz model becomes improper or defective when values less than zero are considered for the parameter α. This probability distribution defined for positive continuous values has been reported by several authors, such as [21,31].

The proportion of immune individuals within the population, denoted by p_0, is obtained by evaluating the limit of the survival function as time t approaches infinity, under the condition $\alpha < 0$:

$$p_0 = \lim_{t\to\infty} S_{\text{Go}}(t \mid \alpha, \beta) = \lim_{t\to\infty} e^{-\frac{\beta}{\alpha}(e^{\alpha t}-1)} = e^{\frac{\beta}{\alpha}}, \tag{3}$$

with $p_0 \in (0,1)$. Once the model parameters are estimated, the cure fraction p_0 can be directly computed from this closed-form expression.

2.1.2. Defective Dagum Distribution

Another illustrative example is the Dagum distribution, which depends on three parameters—typically two shape parameters and one scale parameter. In its classical form, the type I Dagum distribution [17] has a survival function that tends to zero as time progresses, thus being unsuitable for modeling data with a cure fraction. To address this

limitation, a defective version of the Dagum distribution was proposed by modifying the parameter structure, resulting in a proper long-term survival model.

The standard survival function of the type I Dagum distribution is given by

$$S(t \mid \kappa, \alpha, \beta, \gamma) = 1 - \left[\kappa + \left(\frac{t}{\beta}\right)^{-\alpha}\right]^{-\gamma}, \quad t > 0,$$

where $\kappa = 1$, $\alpha > 0$, $\gamma > 0$ are shape parameters, and $\beta > 0$ is a scale parameter.

Based on this formulation, [18] introduced a defective version of the Dagum distribution by setting $\gamma = 1$ and $\kappa = \theta^{-1}$, with $0 < \theta \leq 1$. Under this reparametrization, the probability density function (PDF) and survival function become

$$f_{\mathrm{DD}}(t \mid \alpha, \beta, \theta) = \frac{\alpha\beta\theta^2\, t^{-(\alpha+1)}}{(\beta + \theta t^{-\alpha})^2}, \tag{4}$$

and

$$S_{\mathrm{DD}}(t \mid \alpha, \beta, \theta) = \frac{\beta + \theta t^{-\alpha} - \theta\beta}{\beta + \theta t^{-\alpha}}, \quad t > 0. \tag{5}$$

The survival function converges to a positive constant as $t \to \infty$, implying the existence of a cure fraction:

$$p_0 = \lim_{t\to\infty} S_{\mathrm{DD}}(t \mid \alpha, \beta, \theta) = 1 - \theta. \tag{6}$$

Thus, the cure fraction depends only on the parameter $0 < \theta < 1$, which provides a great advantage for the defective Dagum (DD) distribution in relation to other defective distributions found in the literature, which typically depend on all distribution parameters. When $\theta = 1$, the model reduces to the standard (proper) Dagum distribution with no cured individuals. The flexibility and interpretability of the defective Dagum model make it a useful alternative in survival analysis with long-term survivors.

3. Defective Gamma–G Family

The Gamma–G (GG) family of distributions, introduced by Zografos and Balakrishnan [22], is a flexible class of continuous probability distributions obtained by compounding the Gamma distribution with a given baseline distribution. The central idea is to construct new models by transforming a random variable G (with cumulative distribution function $G(t)$) through the density of the incomplete Gamma distribution. More formally, if T is a continuous random variable with cumulative distribution function $G(t)$, then the probability density function (pdf) and cumulative distribution function (cdf) of the corresponding Gamma–G distribution are defined as

$$g^*(t) = \Gamma(\mu)^{-1} g(t)[-\log(1 - G(t))]^{\mu-1} \tag{7}$$

and

$$G^*(t) = Q[\mu, -\log(1 - G(t))], \tag{8}$$

for t in the domain of g and $\mu > 0$, the shape parameter, where

$$Q(\mu, x) = \Gamma(\mu)^{-1} \int_0^x t^{\mu-1} e^{-t}\, dt$$

is the regularized lower incomplete gamma function, and

$$\Gamma(\mu) = \int_0^\infty t^{\mu-1} e^{-t}\, dt$$

is the gamma function.

The main result of this section is that if a given distribution G is defective, then its extension within the *Gamma–G* class is also defective.

Proposition 1. *Let $S(t)$ be the survival function of a defective model and $S^*(t) = S_{\mathrm{GG}}(t)$ (or $S^*(t) = S_{\mathrm{GD}}(t)$) be its continuous Gamma–G extension. Then $S^*(t)$ is also a defective survival function.*

Proof. Let $S(t) = 1 - G(t)$ denote the survival function of a defective baseline distribution. Assume that

$$\lim_{t\to\infty} S(t) = p_0, \qquad 0 < p_0 < 1,$$

and let $\mu > 0$. The survival function of the corresponding Gamma–G extension can be written as

$$S^*(t) = 1 - G^*(t) = 1 - Q(\mu, -\log(S(t))),$$

where $Q(\mu, \cdot)$ denotes the regularized lower incomplete gamma function.

Since $S(t) \to p_0 \in (0,1)$ as $t \to \infty$, and since the function $x \mapsto -\log(x)$ is continuous on $(0,1)$, we have

$$-\log(S(t)) \longrightarrow -\log(p_0) > 0.$$

Moreover, for fixed $\mu > 0$, the function $z \mapsto Q(\mu, z)$ is continuous for $z > 0$. Therefore,

$$\begin{aligned} \lim_{t\to\infty} S^*(t) &= \lim_{t\to\infty} [1 - Q(\mu, -\log(S(t)))] \\ &= 1 - Q(\mu, -\log(p_0)) \\ &= 1 - \frac{1}{\Gamma(\mu)} \int_0^{-\log(p_0)} u^{\mu-1} e^{-u}\, du =: p_0^*. \end{aligned} \tag{9}$$

Given that $p_0 \in (0,1)$, it follows that $-\log(p_0) > 0$. Hence,

$$0 < Q(\mu, -\log(p_0)) < 1,$$

and consequently,

$$0 < p_0^* = 1 - Q(\mu, -\log(p_0)) < 1.$$

Thus, the Gamma–G extension has a positive limiting survival probability strictly between 0 and 1. Therefore, the Gamma–G construction preserves the defective property under the stated conditions. □

To generate a broader class of defective distributions, we propose in this work two new defective models, named *Gamma–Gompertz* and *Gamma–Dagum*.

3.1. Defective Gamma–Gompertz Distribution

The defective Gamma–Gompertz distribution is a novel extension within the Gamma–G family of distributions, designed to handle survival data where a cure fraction is present. This distribution arises by compounding the Gompertz distribution with the family Gamma–G (GG) distribution.

By applying Equations (7) and (8) together with the density function defined in Equation (1) and the survival function given in Equation (2) of the Gompertz model, one can derive the probability density function and the survival function of the Gamma–Gompertz (GGo) model. These expressions are given, respectively, by

$$g_{GGo}(x) = \Gamma(\mu)^{-1} \beta e^{\alpha t} e^{-\frac{\beta}{\alpha}(e^{\alpha t}-1)} \left[\frac{\beta}{\alpha}(e^{\alpha t} - 1)\right]^{\mu-1} \tag{10}$$

and

$$S_{GGo}(t) = 1 - \frac{1}{\Gamma(\mu)} \int_0^{-log(S_G(t))} t^{\mu-1} e^{(-t)} dt. \tag{11}$$

Similarly to the Gompertz model, when $\alpha < 0$, the Gamma–Gompertz model is also classified as defective. Then, the cure fraction of the Gamma–Gompertz defective model is given by

$$p_0 = 1 - Q\left(\mu, -\log\left(\exp\left(\frac{\beta}{\alpha}\right)\right)\right) = 1 - Q\left(\mu, -\frac{\beta}{\alpha}\right), \tag{12}$$

where

$$Q(\mu, z) = \frac{1}{\Gamma(\mu)} \int_0^z t^{\mu-1} e^{-t}\, dt$$

is the regularized incomplete gamma function, and $\mu > 0, \beta > 0$ and $\alpha < 0$.

3.2. Defective Gamma–Dagum Distribution

By combining Equations (7) and (8) with the density function defined in Equation (4) and the survival function presented in Equation (5) for the Dagum model, one can derive the probability density function and the survival function of the Gamma–Dagum (GD) model. These expressions are given, respectively, by

$$g_{\mathrm{GD}}(t) = \Gamma(\mu)^{-1} \cdot f_{\mathrm{DD}}(t \mid \alpha, \beta, \theta) \cdot [-\log(S_{\mathrm{DD}}(t \mid \alpha, \beta, \theta))]^{\mu-1}, \tag{13}$$

$$S_{\mathrm{GD}}(t) = 1 - \frac{1}{\Gamma(\mu)} \int_0^{-\log(S_{\mathrm{DD}}(t|\alpha,\beta,\theta))} u^{\mu-1} e^{-u}\, du, \tag{14}$$

where $f_{\mathrm{DD}}(t \mid \alpha, \beta, \theta)$ and $S_{\mathrm{DD}}(t \mid \alpha, \beta, \theta)$ are defined in Equations (4) and (5), respectively.

In these expressions, $\alpha > 0$, $\beta > 0$, $\theta \in (0, 1)$ and $\mu > 0$ is the shape parameter associated with the gamma compounding in the Gamma–G framework.

The cure fraction of the Gamma–Dagum defective model is given by

$$p_0 = \lim_{t\to\infty} S_{\mathrm{GD}}(t) = 1 - Q(\mu, -\log(1-\theta)), \tag{15}$$

where $Q(\mu, z) = \frac{1}{\Gamma(\mu)} \int_0^z t^{\mu-1} e^{-t} dt$ is the regularized incomplete gamma function, and $\mu > 0, \alpha > 0, \beta > 0$ and $\theta \in (0, 1)$.

For fixed $\mu > 0$, Equation (15) establishes a bijection between $\theta \in (0, 1)$ and $p_0(\mu, \theta) \in (0, 1)$. In particular,

$$p_0 = 1 - Q(\mu, -\log(1-\theta)) \quad \Longleftrightarrow \quad \theta = 1 - \exp\Big\{-Q_\mu^{-1}(1-p_0)\Big\},$$

where $Q_\mu^{-1}(\cdot)$ denotes the inverse of the regularized incomplete gamma function $Q(\mu, \cdot)$ with respect to its second argument. Therefore,

$$\theta(p_0, \mu) = 1 - \exp\Big\{-Q_\mu^{-1}(1-p_0)\Big\}. \tag{16}$$

By continuity, $\theta \to 0^+ \iff p_0 \to 1$ and $\theta \to 1^- \iff p_0 \to 0$. This reparameterization allows us to work directly with the cure fraction p_0 while preserving the structural properties of the GD model.

Figure 2 presents several scenarios depicting the behavior of the density, survival, and hazard functions associated with the defective Gamma–Gompertz and defective Gamma–Dagum.

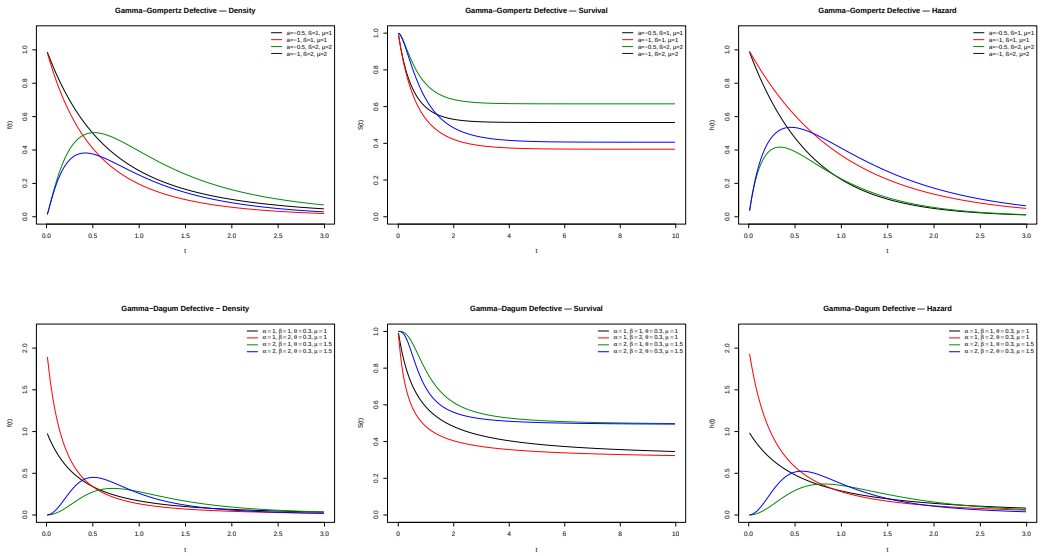

Figure 2. The density, survival, and hazard functions of the defective Gamma–Gompertz and Gamma–Dagum distributions are shown for different parameter configurations. The labels in each panel indicate the parameter values used to generate the curves.

4. Inference

In this section, we describe the inferential procedure, which is based on the maximum likelihood approach and asymptotic large-sample theory. Let $T_i \geq 0$ be the random variable representing the time until the occurrence of the event of interest, and let C_i denote the censoring time, for $i = 1, \ldots, n$. We observe $t_i = \min(T_i, C_i)$ and the censoring indicator δ_i, where $\delta_i = 1$ if the event is observed and $\delta_i = 0$ otherwise. Thus, the observed data are denoted by $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta})$, with $\mathbf{t} = (t_1, \ldots, t_n)^T$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$.

Suppose that the data are independent and identically distributed (i.i.d.) from a distribution with density and survival functions denoted by $f(\cdot, \boldsymbol{\vartheta})$ and $S(\cdot, \boldsymbol{\vartheta})$, respectively, where $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_q)^T$ is a vector of parameters.

The likelihood function of $\boldsymbol{\vartheta}$ can be written as (see [37])

$$L(\boldsymbol{\vartheta}, \mathbf{D}) \propto \prod_{i=1}^{n} \left[f(t_i, \boldsymbol{\vartheta})^{\delta_i} S(t_i, \boldsymbol{\vartheta})^{1-\delta_i} \right].$$

The corresponding log-likelihood function is

$$\log L(\boldsymbol{\vartheta}, \mathbf{D}) = \text{const} + \sum_{i=1}^{n} \delta_i \log f(t_i, \boldsymbol{\vartheta}) + \sum_{i=1}^{n} (1 - \delta_i) \log S(t_i, \boldsymbol{\vartheta}).$$

For the Gamma–Gompertz distribution given by Equations (10) and (11), the log-likelihood is

$$\begin{aligned} \log L(\boldsymbol{\vartheta}, \mathbf{D}) = \text{const} + \sum_{i=1}^{n} \delta_i \Bigg\{ & \log(\beta) + \alpha t_i - \frac{\beta}{\alpha} \left(e^{\alpha t_i} - 1 \right) - \log \Gamma(\mu) \\ & + (\mu - 1) \log \left[\frac{\beta}{\alpha} \left(e^{\alpha t_i} - 1 \right) \right] \Bigg\} \\ & + \sum_{i=1}^{n} (1 - \delta_i) \log \left[1 - \frac{1}{\Gamma(\mu)} \int_0^{\frac{\beta}{\alpha}(e^{\alpha t_i} - 1)} u^{\mu - 1} e^{-u} \, du \right], \end{aligned} \tag{17}$$

where $\vartheta = (\mu, \alpha, \beta)$, with $\mu > 0$, $\beta > 0$, and $\alpha < 0$ for the defective case, and δ_i is the censoring indicator.

For the Gamma–Dagum distribution defined by Equations (13) and (14), the log-likelihood is

$$\begin{aligned}\log L(\boldsymbol{\vartheta}, \mathbf{D}) = \text{const} + \sum_{i=1}^{n} \delta_i \Bigg[& \log\left(\frac{\alpha\beta\theta^2}{\Gamma(\mu)}\right) - (\alpha+1)\log t_i - 2\log\left(\beta + \theta t_i^{-\alpha}\right) \\ & + (\mu-1)\log(-\log S_{\mathrm{DD}}(t_i)) - \log S_{\mathrm{DD}}(t_i)\Bigg] \\ & + \sum_{i=1}^{n}(1-\delta_i)\log\left[1 - \frac{1}{\Gamma(\mu)}\int_0^{-\log S_{\mathrm{DD}}(t_i)} u^{\mu-1}e^{-u}\,du\right],\end{aligned} \tag{18}$$

where $\vartheta = (\mu, \alpha, \beta, \theta)$, and δ_i is the censoring indicator.

The maximum likelihood estimator $\hat{\vartheta}$ is obtained by solving

$$U(\vartheta) = \frac{\partial l(\vartheta)}{\partial \vartheta} = 0,$$

where $l(\vartheta) = \log L(\vartheta)$ is the log-likelihood.

The estimates were computed numerically via the `optim` routine in R [38]. Numerical maximization was performed by minimizing the negative log-likelihood function using the `optim` routine in R. To ensure that the estimates remain in the admissible parameter space, constrained parameters were handled through transformations. For the defective Gamma–Gompertz model, we used $\mu = \exp(\eta_\mu)$, $\beta = \exp(\eta_\beta)$, and $\alpha = -\exp(\eta_\alpha)$, which guarantee $\mu > 0$, $\beta > 0$, and $\alpha < 0$. For the defective Gamma–Dagum model, positivity constraints for μ, α, and β were imposed through exponential transformations, whereas parameters constrained to the interval $(0, 1)$, such as the cure fraction p_0, were handled through a logistic transformation. To reduce sensitivity to starting values, the optimization was initialized from multiple admissible initial points. Convergence was assessed using the convergence code returned by `optim`, inspection of the final log-likelihood value, and numerical stability of the observed information matrix. Replications or fits leading to non-finite likelihood values, non-invertible Hessian matrices, or estimates outside the admissible parameter space were treated as numerical failures and excluded from summary measures when appropriate.

Under standard regularity conditions, $\hat{\vartheta}$ is asymptotically normal,

$$\hat{\vartheta} \sim N\big(\vartheta, \Sigma(\hat{\vartheta})\big),$$

with variance–covariance matrix estimated by

$$\hat{\Sigma}(\hat{\vartheta}) = \left(-\frac{\partial^2 l(\vartheta)}{\partial\vartheta\,\partial\vartheta^T}\bigg|_{\vartheta=\hat{\vartheta}} \right)^{-1}.$$

In addition to standard errors, we also computed the estimated correlation matrix of the maximum likelihood estimators, obtained from $\hat{\Sigma}(\hat{\vartheta})$. This diagnostic was used to assess possible strong dependence among parameter estimates and to investigate finite-sample identifiability and numerical stability in the real-data applications.

A $100(1-\alpha)\%$ confidence interval for ϑ_i is then

$$\left(\hat{\vartheta}_i - z_{\alpha/2}\sqrt{\Sigma_{ii}},\ \hat{\vartheta}_i + z_{\alpha/2}\sqrt{\Sigma_{ii}}\right),$$

where Σ_{ii} is the i-th diagonal element of $\hat{\Sigma}$.

As the cure fraction p_0 depends on the parameters, its confidence interval is obtained using the delta method [39], which gives

$$\widehat{p}_0 \sim N\Big(p_0,\, g'(\hat{\vartheta})^T \hat{\Sigma}(\hat{\vartheta}) g'(\hat{\vartheta})\Big),$$

where $g'(\hat{\vartheta})$ is the gradient of p_0 with respect to ϑ.

In Section 5, we perform a Monte Carlo simulation study to evaluate the asymptotic properties of the maximum likelihood estimators. Simulation studies are widely used to assess estimator performance, particularly when analytical derivations are infeasible.

Defective Gamma–Gompertz and Gamma–Dagum Regression Models

In this section, we present a general strategy for incorporating covariate information into the framework of defective Gamma–Gompertz and Gamma–Dagum distributions. In many practical applications, covariates influence survival time by accelerating or delaying the occurrence of the event of interest. Therefore, it is essential to account for such factors when performing survival analysis.

To incorporate these effects, we consider parametrizing the distribution parameters (e.g., shape, scale, or cure fraction) as functions of covariates through appropriate link functions. The regression structures for the baseline parameters differ according to the defective models proposed.

Let $\mathbf{w}_i^\top = (1, w_{i1}, \ldots, w_{iq})$ and $\mathbf{x}_i^\top = (1, x_{i1}, \ldots, x_{ip})$ be vectors of covariates for individual i. The regression structures for the Gamma–Gompertz parameters are specified as

$$\alpha_i = \mathbf{w}_i^\top \boldsymbol{\zeta}, \qquad \beta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\eta}),$$

where $\boldsymbol{\zeta} = (\zeta_0, \ldots, \zeta_q)^\top$ and $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_p)^\top$ are unknown regression coefficient vectors.

For the Gamma–Gompertz model, the identity link for α allows both positive values (no cure fraction) and negative values (defective case, presence of a cure fraction). The cure fraction is computed from the estimated parameters μ, α, and β as

$$p_{0_i}^* = 1 - \frac{1}{\Gamma(\mu)} \int_0^{-\log(p_0)} u^{\mu-1} e^{-u}\, du,$$

where $p_0 = \lim_{t\to\infty} S_{\mathrm{Go}}(t \mid \alpha_i, \beta_i)$ and $S_{\mathrm{Go}}(t \mid \alpha_i, \beta_i)$ is given by (2). Here, p_0 denotes the limiting survival probability of the defective Gompertz baseline distribution, whereas $p_{0_i}^*$ denotes the induced cure fraction of the corresponding Gamma–Gompertz defective model.

For the defective Gamma–Dagum (GD) model, the existence of a cure fraction depends on the parameter $\theta \in (0, 1)$. To ensure positivity of the baseline parameters we set

$$\alpha_i = \exp(\mathbf{w}_i^\top \boldsymbol{\zeta}), \qquad \beta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\eta}),$$

The long-term survival probability is parameterized as

$$p_{0i} = \frac{\exp(\mathbf{z}_i^\top \boldsymbol{\nu})}{1 + \exp(\mathbf{z}_i^\top \boldsymbol{\nu})},$$

where $\boldsymbol{\nu} = (\nu_0, \ldots, \nu_r)^\top$ is the regression coefficient vector associated with the cure fraction; we then recover the compounding parameter via the bijection in (16),

$$\theta_i = \theta(p_{0i}, \mu) = 1 - \exp\Big\{ - Q_\mu^{-1}(1 - p_{0i}) \Big\},$$

so that, under this cure-parameterization, θ is not estimated as a free parameter but is deterministically linked to (p_{0i}, μ), while covariates act on both the baseline parameters (α_i, β_i) and the cure fraction p_{0i}.

Thus, the complete parameter vectors for each model are

$$\boldsymbol{\Psi}_{\mathrm{GG}} = (\boldsymbol{\zeta}^\top, \boldsymbol{\eta}^\top, \mu)^\top, \qquad \boldsymbol{\Psi}_{\mathrm{GD}} = (\boldsymbol{\zeta}^\top, \boldsymbol{\eta}^\top, \boldsymbol{\nu}^\top, \mu)^\top,$$

This regression specification offers several advantages. First, in the Gamma–Gompertz model, the cure fraction is induced by the limiting value of the survival function. The sign of α_i determines whether a given covariate profile corresponds to a proper or defective survival distribution: When $\alpha_i < 0$, the model is defective, and a positive cure fraction is obtained; when $\alpha_i \geq 0$, no cure fraction is induced. Thus, relative to the corresponding Gamma–Gompertz mixture cure formulation, the defective specification avoids the inclusion of an explicit additional mixing parameter for the cure fraction. Second, the Gamma–Dagum model allows a direct assessment of covariate effects on long-term survival through $\boldsymbol{\nu}$, with θ_i deterministically linked to (p_{0i}, μ) through (16). Finally, the proposed regression structures allow different survival patterns across covariate profiles through the model parameters.

5. Simulation Studies

To assess the performance of the maximum likelihood estimators (MLEs), we conducted Monte Carlo simulation studies. Two models were considered, each designed to reflect one of the real-data applications analyzed in this article. We investigated the performance of several information criteria for selecting between the Gamma–Gompertz defective and Gompertz defective models, and between the Gamma–Dagum defective and Dagum defective models.

5.1. Gamma–Gompertz Model

In this simulation setting for the Gamma–Gompertz model, we consider a single binary covariate $x \sim \mathrm{Bernoulli}(0.5)$. The true parameter values used in the data generation were based on the estimates obtained from the application to malignant neoplasms of the skin (Section 6.1). The following values were used: $\mu = 1.0,\ 1.3$; $\zeta_0 = -0.13$, $\zeta_1 = -0.23$; and $\eta_0 = -2.7$, $\eta_1 = 2.2$. These parameter values lead to the following true cure fractions under the defective Gamma–Gompertz distribution with $\mu = 1.3$: $p_{00}^* \approx 0.726$ for $x = 0$, and $p_{01}^* \approx 0.275$ for $x = 1$. These values, reported as the true cure fraction values in Table 1, are obtained from $p_0^* = 1 - Q(\mu, -\log(p_0))$, where $p_0 = \exp(\beta/\alpha)$ denotes the limiting survival probability of the defective Gompertz baseline distribution. The sample sizes considered were $n = 100,\ 500,\ 1000,\ 2500,\ 5000$. Simulations were performed under the defective Gamma–Gompertz regression model to evaluate the bias, root mean squared error (RMSE), and coverage probability of the maximum likelihood estimators. For numerical summaries, we report the results for the data-driven defective Gamma–Gompertz setting with $\mu = 1.3$. Figures additionally include the special case $\mu = 1.0$, which corresponds to the defective baseline model within the same family, to provide a visual reference.

To introduce random censoring, censoring times were drawn from a uniform distribution on the interval $(0, \tau)$, with τ chosen to control the right-censoring proportion. The observed times and censoring indicators were generated following the steps of Algorithm 1.

Table 1 reports Monte Carlo results (bias, RMSE, and coverage probability) for the MLEs at $\mu = 1.3$. Overall, bias and RMSE decline with sample size, and coverage approaches the nominal 95% level. The cure–fraction parameters (p_{00}, p_{01}) are particularly stable across n, while $\zeta_0, \zeta_1, \eta_0, \eta_1$ show higher small-sample variability that diminishes as n grows.

Algorithm 1 Data generation for the defective Gamma–Gompertz model

1. Fix the true parameters: $\mu, \zeta_0, \zeta_1, \eta_0, \eta_1$.
2. Define the desired censoring proportion p_c.
3. Estimate the appropriate censoring distribution parameter τ to achieve p_c.
4. For each subject $i = 1, \ldots, n$:
   (a) Draw $x_i \sim \text{Bernoulli}(0.5)$.
   (b) Compute:
   $$\alpha_i = \zeta_0 + \zeta_1 x_i, \quad \beta_i = \exp(\eta_0 + \eta_1 x_i).$$
   (c) Compute the Gompertz baseline limit and the Gamma–Gompertz cure fraction:
   $$p0_i = \exp\left(\frac{\beta_i}{\alpha_i}\right), \quad p0_i^* = 1 - \text{pgamma}(-\log(p0_i), \mu).$$
   (d) Draw $u_i \sim \mathcal{U}(0,1)$. If $u_i < p0_i^*$, set $w_i = \infty$; otherwise, draw $u_i' \sim \mathcal{U}(0, 1 - p0_i^*)$ and compute w_i such that:
   $$\text{pgamma}(-\log(S_{\text{Go}}(w_i)), \mu) = u_i',$$
   where $S_{\text{Go}}(t)$ denotes the survival function of the Gompertz distribution with parameters α_i, β_i.
   (e) Draw a censoring time $c_i \sim \mathcal{U}(0, \tau_{x_i})$.
   (f) Determine $t_i = \min\{w_i, c_i\}$. If $t_i = w_i$, set $\delta_i = 1$; otherwise, $\delta_i = 0$.
5. After completing steps (a)–(g) for all $i = 1, \ldots, n$, the final dataset is:
   $$\{(t_i, x_i, \delta_i) : i = 1, \ldots, n\}.$$

Table 1. Results of Monte Carlo simulation for the defective Gamma–Gompertz model. True value, mean of point estimation, bias, RMSE, and coverage probability for the MLEs.

Parameter	True	*n*	Mean	Bias	RMSE	CP
μ	1.300	100	1.506	0.206	0.481	0.927
		500	1.354	0.054	0.161	0.941
		1000	1.346	0.046	0.118	0.959
		5000	1.328	0.028	0.055	0.981
ζ_0	−0.130	100	−0.154	−0.024	0.078	0.946
		500	−0.135	−0.005	0.026	0.948
		1000	−0.134	−0.004	0.019	0.957
		5000	−0.132	−0.002	0.008	0.961
ζ_1	−0.230	100	−0.248	−0.018	0.158	0.954
		500	−0.233	−0.003	0.062	0.950
		1000	−0.235	−0.005	0.044	0.963
		5000	−0.233	−0.003	0.020	0.962
η_0	−2.700	100	−2.469	0.231	0.757	0.908
		500	−2.623	0.077	0.326	0.932
		1000	−2.632	0.068	0.238	0.935
		5000	−2.650	0.050	0.112	0.919
η_1	2.200	100	2.132	−0.068	0.469	0.936
		500	2.175	−0.025	0.212	0.952
		1000	2.181	−0.019	0.145	0.946
		5000	2.182	−0.018	0.069	0.941
p_{00}	0.726	100	0.718	−0.008	0.083	0.946
		500	0.724	−0.001	0.034	0.948
		1000	0.726	0.000	0.024	0.955
		5000	0.726	0.000	0.011	0.947
p_{01}	0.275	100	0.274	−0.001	0.096	0.945
		500	0.271	−0.003	0.040	0.944
		1000	0.275	0.000	0.029	0.960
		5000	0.276	0.001	0.013	0.956

Figure 3 displays the bias (points) and root mean square error (RMSE; vertical bars) of the MLEs across sample sizes. We focus on the data-driven setting with $\mu = 0.8$ and include $\mu = 1.0$ (defective Gompertz limit within the same family within the same family) as a reference. Across both settings, bias shrinks toward zero and RMSE declines as n increases. The largest small-sample variability appears at $n = 100$, most visibly for η_0 and η_1, and becomes negligible for $n \geq 1000$. The cure fraction parameters p_{00} and p_{01} are particularly stable, with near-zero bias and small RMSE even for moderate samples ($n \approx 500$). These patterns indicate well-behaved MLEs in the data-driven scenario considered here. Moreover, the empirical coverage probabilities for all parameters approach the nominal level as the sample size increases (Figure 4), regardless of the specification within the defective Gamma–Gompertz family (including the $\mu = 1.3$ and $\mu = 1$ cases). The average interval lengths shrink with larger n, reflecting the expected gains in precision.

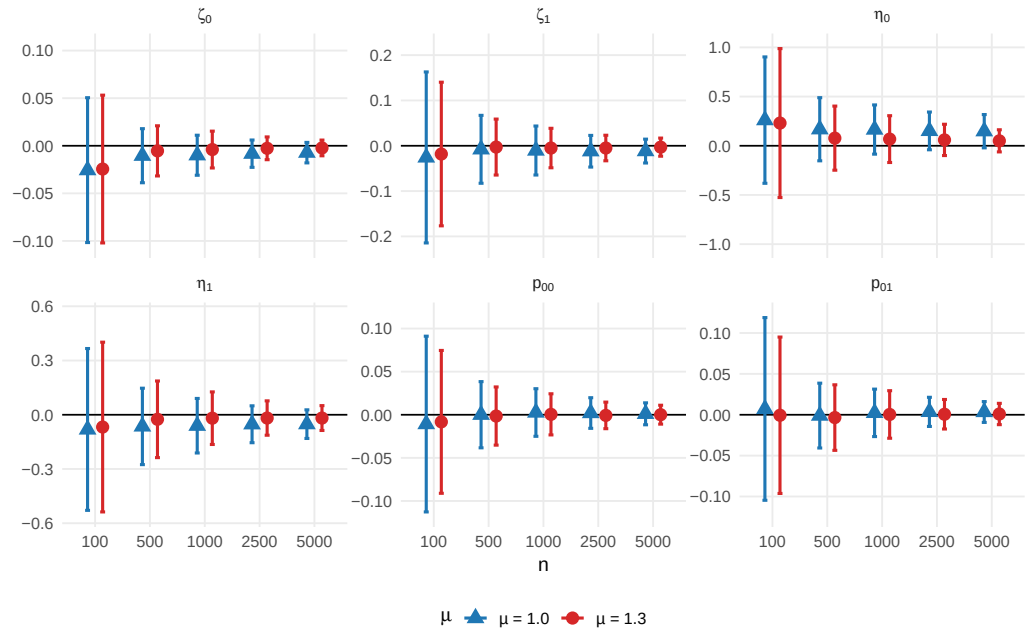

Figure 3. Bias (points) and RMSE (bars) of the MLEs as a function of the sample size under two data-generating settings: $\mu = 1.3$ (application-driven) and $\mu = 1$ (defective Gompertz limit).

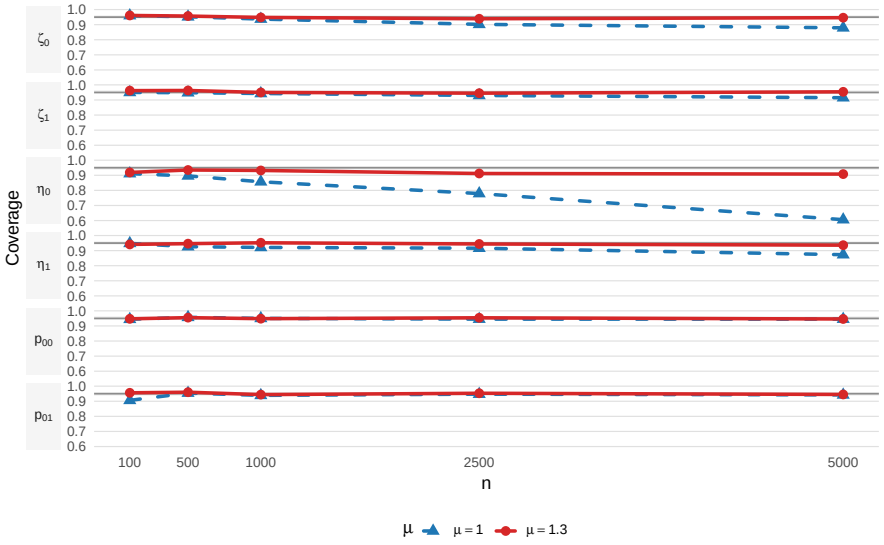

Figure 4. Empirical coverage probabilities for 95% confidence intervals of the parameters versus sample size for simulated data from the defective Gamma–Gompertz distribution.

We considered AIC, AICc, BIC, HQIC, and CAIC, defined by AIC $= -2\ell + 2k$; AICc $=$ AIC $+ \dfrac{2k(k+1)}{n-k-1}$; BIC $= -2\ell + k\log n$; HQIC $= -2\ell + 2k\log\log n$; CAIC $= -2\ell + k(\log n + 1)$, where ℓ is the maximized log-likelihood, k the number of parameters, and n the sample size. All criteria were computed in every Monte Carlo replication for both model families (Gamma–Gompertz and Gamma–Dagum). Within each family, we compared two specifications: the Gamma–Gompertz defective model and the baseline Gompertz defective model. Figure 5 presents the summaries of the model-selection criteria and the corresponding selection frequencies across scenarios and sample sizes for the defective Gamma–Gompertz model compared with the defective Gompertz model.

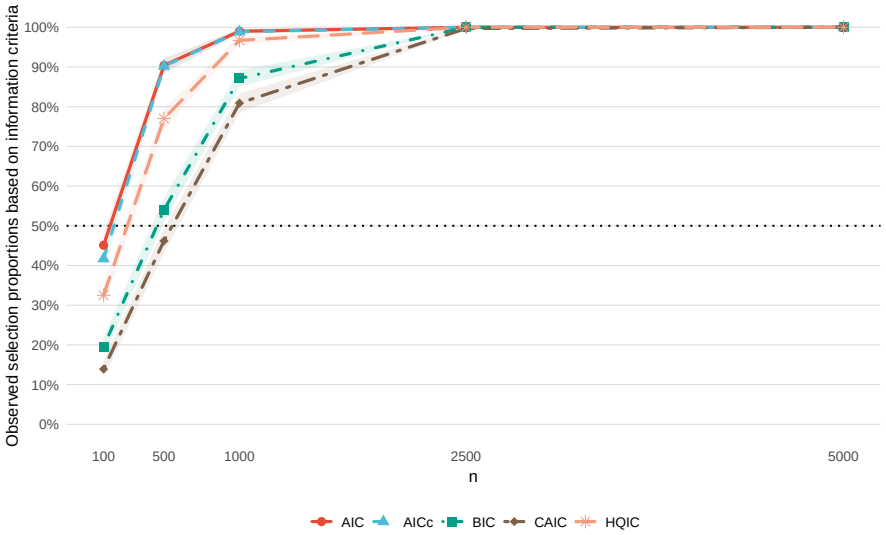

Figure 5. Selection proportion favoring the defective Gamma–Gompertz over defective Gompertz across n for $\mu = 1.3$.

Figure 5 shows the proportion of datasets in which the Gamma–Gompertz model achieves the lowest information criterion compared to the Gompertz baseline, across sample sizes for $\mu = 1.3$. The selection proportion increases with n and approaches one for larger samples, indicating that standard information criteria consistently favor the Gamma–Gompertz model in this scenario and effectively distinguish it from the Gompertz baseline.

Overall, the results show that the proposed modeling framework yields consistent and accurate MLEs: Bias and RMSE decrease and empirical coverage approaches the nominal level as the sample size increases, while the cure fractions and covariate effects are reliably recovered in the scenario considered.

5.2. Gamma–Dagum Model

In this simulation setting for the Gamma–Dagum model, we consider a single binary covariate $x \sim \text{Bernoulli}(0.5)$. The true parameter values used in the data generation are based on the estimates obtained from the application to cervical cancer data (Section 6.2), specifically for malignant neoplasms of the cervix uteri. The following values were used: $\mu = 0.8,\ 1.0$; $\zeta_0 = 0.65$, $\zeta_1 = -0.15$; $\eta_0 = -15$, $\eta_1 = 4.3$; $\nu_0 = 1.6$, and $\nu_1 = -2.3$. These lead to cure fractions of $p_0 = 0.744$ for $x = 0$, and $p_0 = 0.251$ for $x = 1$. The sample sizes considered were $n = 50,\ 100,\ 500,\ 1000,\ 2500,\ 5000$. Simulations were performed under the defective Gamma–Dagum regression model to evaluate the bias, root mean squared error (RMSE), and coverage probability of the maximum likelihood estimators. For numerical summaries, we report results at the data-driven setting only Gamma–Dagum with $\mu = 0.8$. Figures additionally include the special case $\mu = 1.0$, which corresponds to the defective baseline model within the same family, to provide a visual reference.

To introduce random censoring, censoring times were assumed to follow a uniform distribution over the interval $(0, \tau)$. The value of τ was chosen to control the desired proportion of right-censoring. The observed times and censoring indicators were generated following the steps described in Algorithm 2.

Algorithm 2 Data generation for the defective Gamma–Dagum model

1. Fix the true parameters: $\mu, \zeta_0, \zeta_1, \eta_0, \eta_1, \nu_0, \nu_1$.
2. Define the desired censoring proportion p_c.
3. Estimate the appropriate censoring distribution parameter τ to achieve p_c.
4. For each subject $i = 1, \ldots, n$:
   (a) Draw $x_i \sim \text{Bernoulli}(0.5)$.
   (b) Compute:
   $$\alpha_i = \exp(\zeta_0 + \zeta_1 x_i), \quad \beta_i = \exp(\eta_0 + \eta_1 x_i), \quad p0_i = \frac{\exp(\nu_0 + \nu_1 x_i)}{1 + \exp(\nu_0 + \nu_1 x_i)}.$$
   (c) Compute the cure fraction under the Gamma–Dagum model:
   $$p0_i^* = 1 - \text{pgamma}(-\log(p0_i), \mu).$$
   (d) Draw $u_i \sim \mathcal{U}(0,1)$. If $u_i < p0_i^*$, set $w_i = \infty$; otherwise, draw $u_i' \sim \mathcal{U}(0, 1 - p0_i^*)$ and compute w_i such that:
   $$\text{pgamma}(-\log(S_{\text{DD}}(w_i)), \mu) = u_i'.$$
   (e) Draw a censoring time $c_i \sim \mathcal{U}(0, \tau_{x_i})$.
   (f) Determine $t_i = \min\{w_i, c_i\}$. If $t_i = w_i$, set $\delta_i = 1$; otherwise, $\delta_i = 0$.
5. After completing steps (a)–(g) for all $i = 1, \ldots, n$, the final dataset is:
   $$\{(t_i, x_i, \delta_i) : i = 1, \ldots, n\}.$$

Table 2 presents the results of the Monte Carlo simulation in terms of bias, root mean square error (RMSE), and coverage probability (CP) for the maximum likelihood estimators (MLEs) under the data–driven configuration $\mu = 0.8$ and different sample sizes. In general, as the sample size increases, both bias and RMSE decrease for most parameters, confirming the consistency of the estimators. The estimators associated with the cure fraction parameters (ν_0, ν_1, p_{00}, and p_{01}) exhibit low bias and RMSE with coverage close to 95% even at moderate sample sizes.

Table 2. Results of Monte Carlo simulation for the defective Gamma–Dagum model. True value, mean of point estimation, bias, RMSE, and coverage probability for the MLEs.

Parameter	True	*n*	Mean	Bias	RMSE	CP
μ	0.800	100	3.521	2.721	6.079	0.923
		500	0.978	0.178	1.221	0.931
		1000	0.817	0.017	0.130	0.936
		5000	0.805	0.005	0.061	0.944
ζ_0	0.650	100	0.721	0.071	0.449	0.933
		500	0.666	0.016	0.249	0.937
		1000	0.652	0.002	0.106	0.942
		5000	0.649	−0.001	0.053	0.947
ζ_1	−0.150	100	−0.114	0.036	0.417	0.902
		500	−0.157	−0.007	0.182	0.935
		1000	−0.153	−0.003	0.075	0.954
		5000	−0.150	0.000	0.038	0.948

Table 2. *Cont.*

Parameter	True	n	Mean	Bias	RMSE	CP
η_0	-15.000	100	-14.759	0.241	9.140	0.824
		500	-15.507	-0.507	4.664	0.933
		1000	-15.070	-0.070	1.838	0.939
		5000	-14.991	0.009	0.910	0.948
η_1	4.300	100	4.805	0.505	5.577	0.916
		500	4.617	0.317	2.564	0.928
		1000	4.360	0.060	1.002	0.954
		5000	4.298	-0.002	0.501	0.952
ν_0	1.600	100	-1.138	-2.738	6.240	0.970
		500	1.448	-0.152	1.529	0.976
		1000	1.571	-0.029	0.412	0.941
		5000	1.579	-0.021	0.197	0.952
ν_1	-2.300	100	-3.303	-1.003	2.272	0.969
		500	-2.439	-0.139	0.505	0.967
		1000	-2.324	-0.024	0.159	0.953
		5000	-2.312	-0.012	0.078	0.954
p_{00}	0.744	100	0.718	-0.026	0.109	0.957
		500	0.738	-0.007	0.035	0.958
		1000	0.741	-0.003	0.016	0.943
		5000	0.742	-0.003	0.008	0.950
p_{01}	0.251	100	0.230	-0.020	0.105	0.925
		500	0.240	-0.011	0.048	0.958
		1000	0.245	-0.005	0.020	0.949
		5000	0.246	-0.005	0.011	0.937

Figure 6 displays the bias (points) and root mean square error (RMSE; vertical bars) of the MLEs across sample sizes. We focus on the data-driven setting with $\mu = 0.8$ and include $\mu = 1.0$ (the defective Dagum limit within the same family) as a reference. In both settings, biases contract toward zero and RMSE declines as n increases. The largest small-sample variability is observed at $n = 100$—most visibly for η_0, η_1, and also ν_0, ν_1—and becomes negligible from $n \geq 1000$. The cure fraction parameters p_{00} and p_{01} are especially stable, with near-zero bias and small RMSE even for moderate samples. Figure 7 reports empirical coverage probabilities (CP) by parameter and sample size. CPs approach the nominal 95% level as n increases for both $\mu = 0.8$ and $\mu = 1.0$; the interval bands tighten with larger n, reflecting the expected gains in precision. Moreover, Figure 8 shows the proportion of datasets in which the Gamma–Dagum model achieves the lowest information criterion compared to the Dagum baseline, across sample sizes for $\mu = 0.8$. The selection proportion increases with n and approaches one for larger samples, indicating that standard information criteria consistently favor the Gamma–Dagum specification in this setting and effectively distinguish it from the Dagum baseline.

Overall, the proposed defective Gamma–Dagum model exhibits robust estimation properties, particularly for low to moderate values of μ, providing reliable inference for both the cure fraction and the time-to-event components.

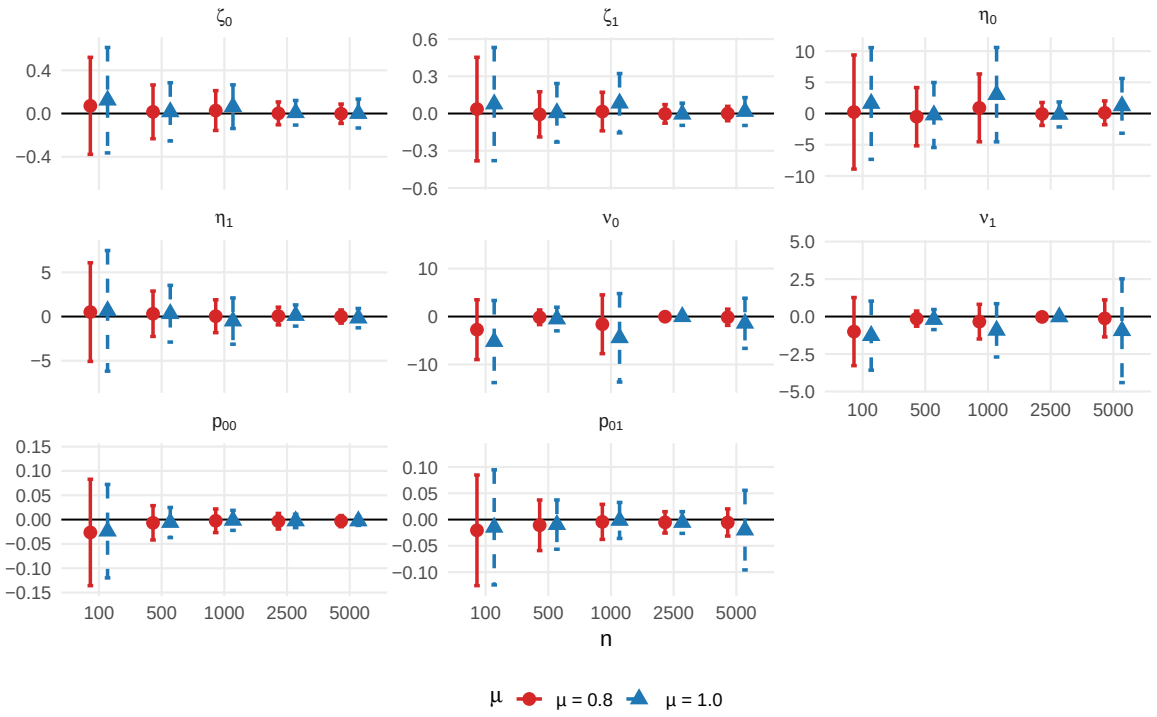

Figure 6. Bias (points) and RMSE (bars) of the MLEs as a function of the sample size under two data-generating settings: $\mu = 0.8$ (application-driven) and $\mu = 1$ (defective Dagum limit).

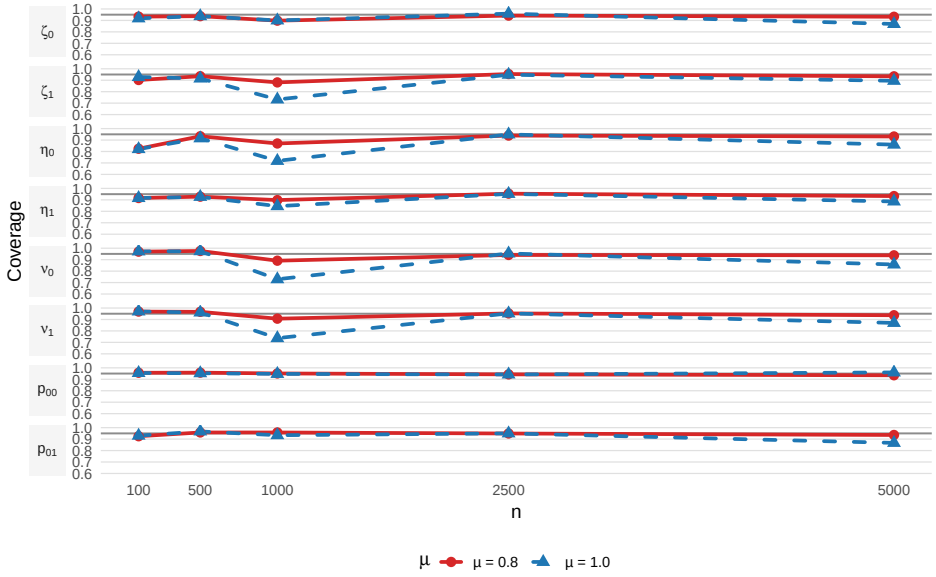

Figure 7. Empirical coverage probabilities for 95% confidence intervals of the parameters versus sample size for simulated data from the defective Gamma–Dagum distribution.

Although the simulation design was intentionally based on the empirical configurations observed in the two real-data applications, we acknowledge that it considers a simplified covariate structure and a non-informative censoring mechanism. More complex scenarios involving multiple or continuous covariates, time-dependent effects, informative censoring, and alternative censoring distributions constitute important directions for future research and are discussed in the Section 6.3.

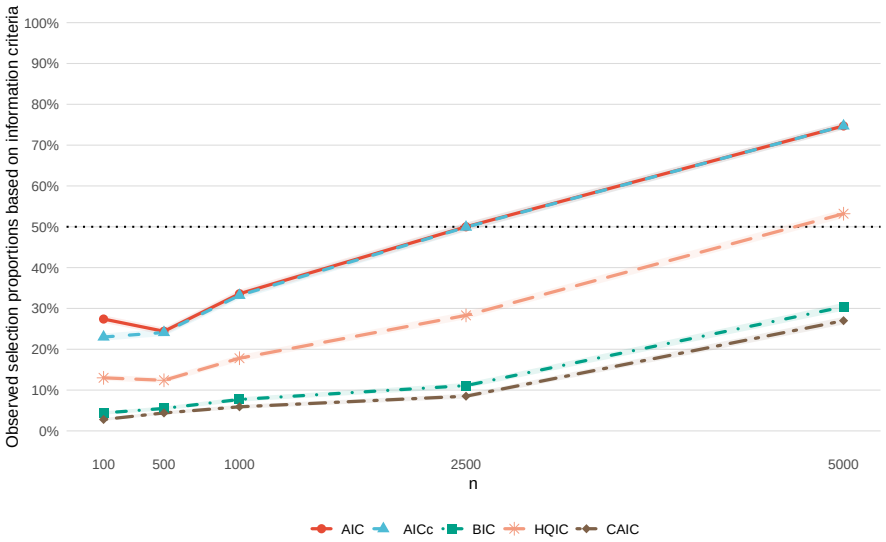

Figure 8. Selection proportion favoring the defective Gamma–Dagum over defective Dagum across n for $\mu = 0.8$.

6. Application

As previously described in Section Motivating Example: Melanoma and Cervical Cancer in the State of São Paulo, this study involves two real-world datasets provided by the *Fundação Oncocentro de São Paulo* (FOSP): one on melanoma and another on cervical cancer. FOSP is a public institution under the São Paulo State Health Department, responsible for managing the Hospital Cancer Registry system (RHC/FOSP), which began in 2000 with the goal of registering cancer cases treated across the state. All records were made available by FOSP through the RHC/SP database and can be freely downloaded from https://fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/rhc-registro-hospitalar-de-cancer/download-de-arquivos/ (accessed on 9 June 2026).

In what follows, we assess the performance of the proposed defective Gamma–Gompertz and Gamma–Dagum cure rate models using these datasets. We also compare the estimated survival curves with those obtained via the Kaplan–Meier estimator [40] and report maximum likelihood estimates, standard errors, 95% confidence intervals, and AIC values [41] for each fitted model.

Both datasets are anonymized and comply with the ethical requirements established by the Brazilian National Research Ethics Commission (CONEP), which does not require prior approval for studies based on secondary public data. All statistical analyses were conducted using the free software `R` [38].

6.1. Malignant Neoplasms of the Skin

As introduced in Section Motivating Example: Melanoma and Cervical Cancer in the State of São Paulo, this application focuses on melanoma, a malignant tumor of melanocytes that, despite being less common than other skin cancers, is associated with high mortality and aggressive progression, particularly in advanced clinical stages.

The dataset comprises 7823 patients diagnosed between 2000 and 2014 in the state of São Paulo, Brazil. All patients had a minimum follow-up of two months, with data available until 2018. The event of interest was cancer-specific mortality. Patients who did not die from this cause during follow-up were treated as right-censored.

This data set was initially studied by Calsavara et al. [35], where the effect of surgery was evaluated using a non-proportional hazards frailty model. In the present study, we focus exclusively on distinguishing between metastatic (stage III–IV) and non-metastatic

(stage I–II) cases. In the sample, 68.15% of the patients were classified as non-metastatic, while 31.85% had metastatic disease. Kaplan–Meier survival curves (Figure 9) reveal a clear contrast between these groups: Early-stage patients exhibit long-term survival with the curve stabilizing around 75%, while advanced-stage patients show poorer outcomes with a plateau near 26%. These findings motivate the use of cure models that can explicitly capture the presence of long-term survivors.

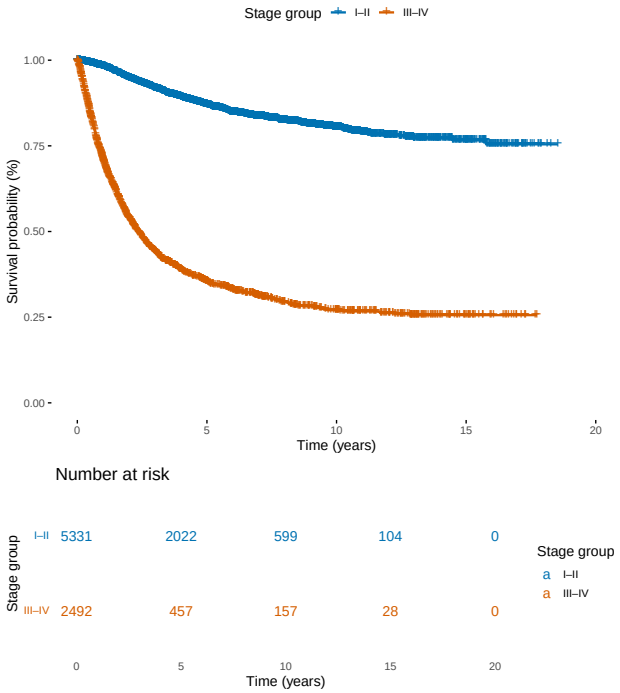

Figure 9. Kaplan–Meier estimates for survival curves of melanoma patients, stratified by clinical stage.

Our aim is to estimate survival functions and cure fractions for this population. The analysis compares the fitted defective Gamma–Gompertz cure rate model with the Kaplan–Meier estimator. We report maximum likelihood estimates, confidence intervals, and AIC values to assess model fit. Additionally, the estimated hazard functions highlight distinct patterns across clinical stages: a low and decreasing risk over time for early-stage patients, and a unimodal risk pattern for those diagnosed at advanced stages.

Table 3 presents the maximum likelihood estimates and 95% confidence intervals for the Gompertz defective and Gamma–Gompertz defective models. In both cases, the regression coefficients for the shape parameter are negative, which guarantees the defective property and confirms the existence of a cure fraction. For the Gamma–Gompertz defective model, the estimated cure fractions are substantially higher for patients at stages I–II ($p^*_{0_{I-II}} = 0.720$) than for those at stages III–IV ($p^*_{0_{III-IV}} = 0.263$).

To assess finite-sample estimation stability in the melanoma application, we examined the estimated correlation matrix of the maximum likelihood estimators for the defective Gamma–Gompertz model. As shown in Table 4, moderate to high correlations were observed among some parameters, particularly between μ and η_0 and between η_0 and η_1. Nevertheless, no estimated correlation was close to ± 1, the estimated covariance matrix was positive definite, and all confidence intervals were finite and well defined. These results suggest that, although estimator dependence is present, there is no evidence of severe non-identifiability in this application.

High correlations among some estimators are expected in this class of models, since the cure fraction is not represented by an isolated mixing parameter but is induced by combinations of distributional parameters. Therefore, these correlations do not necessarily

imply lack of theoretical identifiability. Rather, they indicate finite-sample dependence among the estimators and reinforce the need for numerical diagnostics and cautious interpretation of individual parameters.

Table 3. Estimated parameters for the melanoma dataset under the Gompertz defective and Gamma–Gompertz defective models.

Parameter	Gompertz Defective		Gamma–Gompertz Defective	
	MLE	**95% CI**	**MLE**	**95% CI**
μ	–	–	1.295	(1.208, 1.382)
		$\boldsymbol{\alpha}$ (shape)		
ζ_0	−0.065	(−0.092, −0.038)	−0.131	(−0.162, −0.099)
ζ_1	−0.215	(−0.255, −0.176)	−0.226	(−0.266, −0.186)
		$\boldsymbol{\beta}$ (scale)		
η_0	−3.499	(−3.618, −3.380)	−2.680	(−2.901, −2.459)
η_1	2.555	(2.417, 2.692)	2.197	(2.056, 2.339)
		$\boldsymbol{p_0}$ (cure fraction)		
$p_{0_{I-II}}$	**0.629**	(0.531, 0.727)	**0.720**	(0.676, 0.763)
$p_{0_{III-IV}}$	**0.250**	(0.224, 0.276)	**0.263**	(0.239, 0.287)
log-likelihood	−6439.41	–	−6411.62	–

Table 4. Estimated correlation matrix of the maximum likelihood estimators for the defective Gamma–Gompertz model fitted to the melanoma dataset.

	μ	ζ_0	ζ_1	η_0	η_1
μ	1.000	−0.528	−0.101	0.892	−0.566
ζ_0	−0.528	1.000	−0.515	−0.766	0.757
ζ_1	−0.101	−0.515	1.000	0.142	−0.526
η_0	0.892	−0.766	0.142	1.000	−0.822
η_1	−0.566	0.757	−0.526	−0.822	1.000

A comparison of the defective Gamma–Gompertz and defective Gompertz models with mixture cure specifications based on Gompertz, exponential, Weibull, and Weibull-exponential distributions was conducted in terms of AIC and BIC (Table 5). The defective Gamma–Gompertz model achieved the lowest AIC and BIC among all candidate models, indicating the best overall fit. It also outperformed its defective Gompertz baseline, supporting the added flexibility provided by the Gamma–G construction.

Table 5. AIC and BIC statistics, with the lowest values highlighted in bold—Malignant melanoma of skin.

Model	AIC	BIC
Defective Gamma–Gompertz	**12,833.2**	**12,868.1**
Defective Gompertz	12,886.8	12,916.7
Gompertz Mixture	12,888.0	12,930.0
Exponential Mixture	12,897.0	12,925.0
Weibull Mixture	12,895.0	12,930.0
Weibull–Exponential Mixture	12,888.0	12,937.0

The Gamma–Gompertz defective model proved to be well suited for capturing the survival dynamics observed in the data. The Kaplan–Meier curves (Figure 10) indicate a

substantial cured fraction among early-stage patients, as evidenced by the plateau observed during long-term follow-up. In contrast, the curve for advanced-stage patients stabilizes at a considerably lower level. These findings support the use of defective survival models that explicitly account for the presence of long-term survivors through a flexible hazard structure. Moreover, the estimated hazard functions (Figure 11) exhibit distinct patterns across clinical stages: A unimodal hazard is observed for advanced-stage patients (III–IV), while early-stage patients (I–II) show a decreasing hazard, reflecting a low-risk profile over time. This behavior highlights the flexibility of the Gamma–Gompertz framework and its suitability for modeling heterogeneous survival trajectories across subgroups.

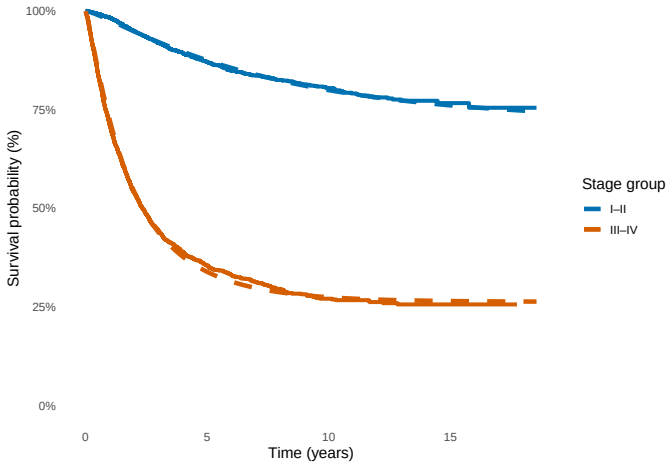

Figure 10. Kaplan–Meier estimates (solid curves) and Gamma–Gompertz defective model estimates (dashed curves) for the survival of melanoma patients, stratified by clinical stage.

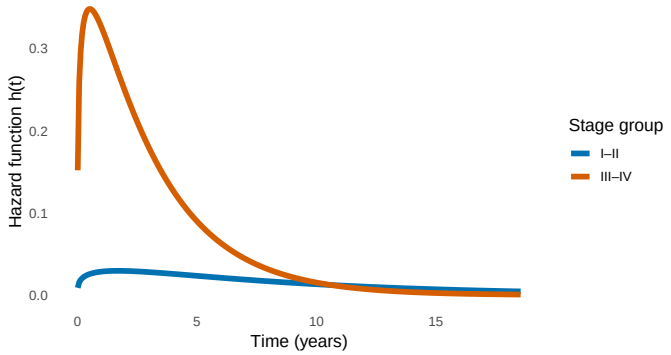

Figure 11. Estimated hazard functions under the Gamma–Gompertz defective model for melanoma patients, stratified by clinical stage.

6.2. Malignant Neoplasms of Cervix Uteri

As discussed in Section Motivating Example: Melanoma and Cervical Cancer in the State of São Paulo, cervical cancer (ICD-10 code C53) remains a major public health challenge in Brazil, particularly due to disparities in access to prevention and treatment. In this application, we analyze a retrospective cohort of 12,030 women diagnosed with cervical cancer in the state of São Paulo between 2012 and 2022.

The dataset analyzed comprises 12,030 women diagnosed with cervical cancer in the state of São Paulo, Brazil. Among them, 6147 were classified as Stage I–II and 5883 as Stage III–IV at diagnosis. A total of 4177 patients (34.7%) experienced the event of interest (death due to cancer), while 7853 (65.3%) were right-censored. In the early-stage group (I–II), only 15.8% (969 patients) experienced the event, whereas this proportion rose to 54.5% (3208 patients) in the advanced-stage group (III–IV).

The event of interest was cancer-specific death, and individuals who did not experience the event during follow-up were treated as right-censored. The primary goal of this analysis is to evaluate the relationship between clinical staging at diagnosis (Stage I–II vs. Stage III–IV) and long-term survival.

We fitted the defective Gamma–Dagum model to the data, allowing the parameters to vary according to the staging group. This model accommodates a cure fraction and captures different hazard shapes between groups through its flexible structure. Kaplan–Meier curves were used for visual comparison, as shown in Figure 12. We also report maximum likelihood estimates, standard errors, confidence intervals, and AIC values to assess model fit and parameter stability.

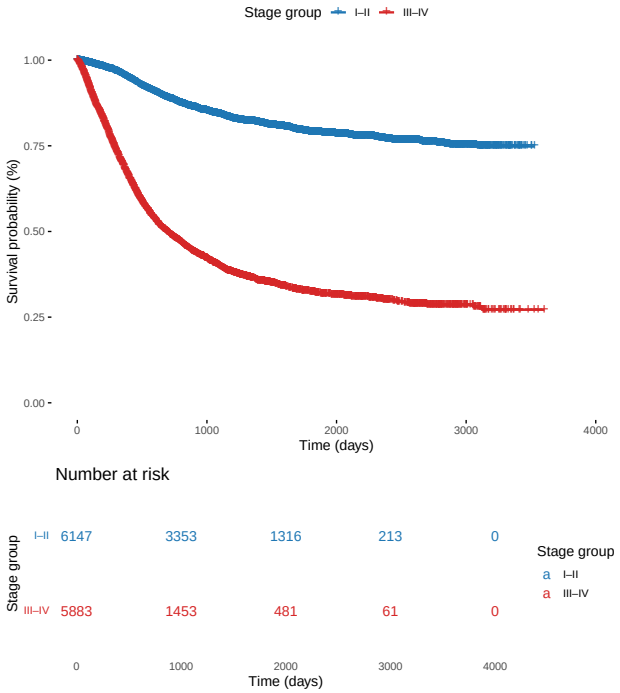

Figure 12. Kaplan–Meier estimates for survival curves of cervical cancer patients, stratified by clinical stage.

In this analysis, we explore Kaplan–Meier survival curves stratified by clinical staging. The results reveal a marked difference in survival dynamics: Patients diagnosed at early stages (Stage I–II) exhibit evidence of long-term survival and a potential cure fraction, whereas those diagnosed at advanced stages (Stage III–IV) show significantly lower survival probabilities, with no apparent plateau in the tail of the survival curve.

Among the women analyzed, a substantial proportion presented censored survival times, indicating they were either alive or had died from other causes during the follow-up period. Notably, patients in the advanced-stage group accounted for the majority of observed cancer-specific deaths, highlighting the critical importance of early detection and timely intervention in improving survival outcomes for this population.

Table 6 presents the maximum likelihood estimates and 95% confidence intervals for the Dagum defective and Gamma–Dagum defective models. The estimated cure fractions are considerably higher for stages I–II ($p^*_{0_{I-II}} = 0.732$) than for stages III–IV ($p^*_{0_{III-IV}} = 0.250$).

For the cervical cancer application, we also examined the estimated correlation matrix of the maximum likelihood estimators for the defective Gamma–Dagum model. As shown in Table 7, stronger correlations were observed, particularly between ζ_0 and η_0, between μ and ν_0, and between η_0 and ν_0. This indicates substantial finite-sample dependence among

some parameter estimates and suggests that individual parameter interpretation should be made with caution. Nevertheless, the estimated covariance matrix was positive definite, all estimates were finite, and the confidence intervals were well defined, indicating that the model remained numerically estimable in this application.

Table 6. Parameter estimates for the cervical cancer dataset under the defective Dagum and defective Gamma–Dagum models.

Parameter	Dagum Defective		Gamma–Dagum Defective	
	MLE	**95% CI**	**MLE**	**95% CI**
μ	–	–	0.758	(0.642, 0.875)
		$\boldsymbol{\alpha}$ (shape)		
ζ_0	0.466	(0.389, 0.542)	0.657	(0.528, 0.786)
ζ_1	−0.121	(−0.206, −0.035)	−0.146	(−0.227, −0.064)
		$\boldsymbol{\beta}$ (scale)		
η_0	−12.215	(−12.979, −11.452)	−15.358	(−17.483, −13.233)
η_1	3.279	(2.455, 4.104)	4.282	(3.143, 5.422)
		$\boldsymbol{\nu}$ (cure fraction, logit scale)		
ν_0	0.913	(0.787, 1.039)	1.658	(1.209, 2.106)
ν_1	−2.158	(−2.337, −1.979)	−2.271	(−2.465, −2.076)
		$\boldsymbol{p_0}$ (cure fraction)		
$p_{0_{I-II}}$	**0.714**	(0.688, 0.739)	**0.732**	(0.709, 0.755)
$p_{0_{III-IV}}$	**0.224**	(0.202, 0.246)	**0.250**	(0.227, 0.274)
log-likelihood	−35427.41	–	−35421.56	–

Table 7. Estimated correlation matrix of the maximum likelihood estimators for the defective Gamma–Dagum model fitted to the cervical cancer dataset.

	μ	ζ_0	ζ_1	η_0	η_1	ν_0	ν_1
μ	1.000	−0.831	0.141	0.909	−0.541	−0.959	0.526
ζ_0	−0.831	1.000	−0.607	−0.985	0.878	0.875	−0.618
ζ_1	0.141	−0.607	1.000	0.491	−0.896	−0.260	0.472
η_0	0.909	−0.985	0.491	1.000	−0.816	−0.930	0.612
η_1	−0.541	0.878	−0.896	−0.816	1.000	0.628	−0.611
ν_0	−0.959	0.875	−0.260	−0.930	0.628	1.000	−0.689
ν_1	0.526	−0.618	0.472	0.612	−0.611	−0.689	1.000

The Kaplan–Meier survival curves displayed in Figure 13 suggest the presence of a substantial cure fraction among patients diagnosed at early stages, as evidenced by a clear plateau in the curve after a certain follow-up period. Conversely, this pattern is not observed for advanced-stage patients, indicating a lack of long-term survivors in this group. These findings support the use of defective survival models, which explicitly incorporate the possibility of cure through a flexible hazard structure. Moreover, the nonparametric hazard function shown in Figure 14 presents a unimodal shape, reinforcing the adequacy of the proposed Gamma–Dagum defective modeling approach.

Table 8 presents the maximum likelihood estimates and corresponding 95% confidence intervals for the parameter θ under the Dagum and Gamma–Dagum models. In all cases, the estimated values of θ lie strictly within the interval $(0, 1)$, as required for the distribution to be defective. This confirms the theoretical adequacy of the defective formulation in modeling long-term survival.

A comparison of the Gamma–Dagum and Dagum defective models with mixture cure models (exponential, Weibull, Weibull–exponential and Log-normal) was conducted in terms of AIC and BIC (Table 9). The defective Gamma–Dagum model achieved the lowest AIC and BIC among all candidate models. It also clearly outperformed its defective Dagum baseline, indicating the best overall fit.

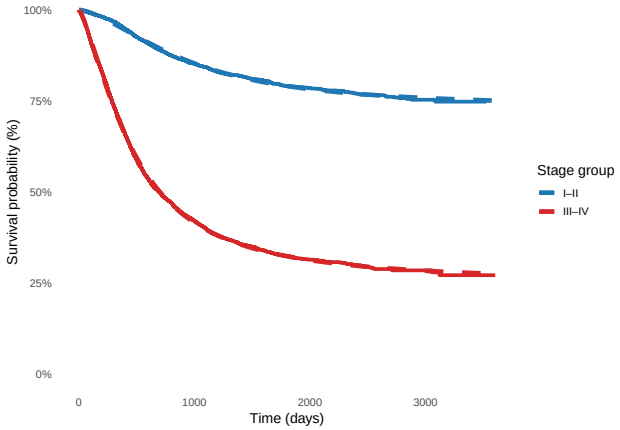

Figure 13. Kaplan–Meier estimates (solid curves) and defective Gamma–Dagum model estimates (dashed curves) for survival curves of cervical cancer patients, stratified by clinical stage.

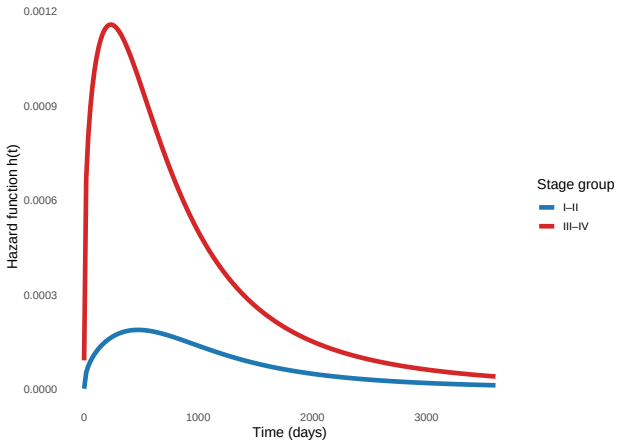

Figure 14. Estimated hazard functions under the defective Gamma–Dagum model for cervical cancer patients, stratified by clinical stage.

Table 8. θ estimates for the cervical cancer dataset.

Model	Stage	MLE	CI (95%)
Dagum	I + II	0.286	(0.261, 0.312)
	III + IV	0.776	(0.754, 0.798)
Gamma–Dagum	I + II	0.268	(0.245, 0.291)
	III + IV	0.750	(0.726, 0.773)

Moreover, the estimated θ values differ substantially between staging groups, with lower values observed for early-stage patients (Stage I–II), indicating a lighter tail and higher long-term survival, while higher values in the advanced-stage group (Stage III–IV) are consistent with increased mortality risk and reduced cure probability. These findings support the use of the Gamma–Dagum defective model for characterizing heterogeneous survival patterns across clinical stages.

Table 9. AIC and BIC statistics, with the lowest values highlighted in bold—malignant neoplasms of cervix uteri.

Model	AIC	BIC
Defective Gamma–Dagum	**70,857.1**	**70,908.9**
Defective Dagum	70,866.8	70,911.2
Dagum Mixture	70,864.3	70,923.4
Exponential Mixture	71,123.2	71,145.3
Weibull Mixture	71,092.8	71,122.4
Weibull–Exponential Mixture	70,947.5	70,991.9

6.3. Limitations and Future Work

Some limitations should be acknowledged. First, the simulation study was designed to reflect the empirical configurations of the two real-data applications and considered a simplified covariate structure with non-informative censoring. More complex scenarios involving multiple continuous covariates, time-dependent effects, informative censoring mechanisms, and alternative censoring distributions should be investigated in future studies. Second, the real-data applications focused mainly on clinical stage, which is a relevant prognostic factor consistently available in both datasets. However, other variables, such as age, sex, treatment information, biomarkers, and socioeconomic indicators, may also affect long-term survival and should be incorporated when available with sufficient quality and completeness. Third, the estimated correlation matrices indicated moderate to high dependence among some maximum likelihood estimators, especially in the Gamma–Dagum application. Although the covariance matrices were positive definite and the confidence intervals were well defined, individual parameter interpretation should be made with caution. This is particularly relevant for the Gamma–G parameter μ, which should be interpreted primarily as a structural flexibility parameter rather than as a directly clinical quantity. Its inclusion should therefore be supported by model comparison criteria, numerical stability diagnostics, and substantive plausibility in order to reduce the risk of overfitting. Future work should further investigate identifiability, predictive validation, and comparisons with semiparametric and more flexible survival frameworks.

7. Conclusions

This study introduced two defective survival models within the Gamma–G family, namely the defective Gamma–Gompertz and defective Gamma–Dagum distributions. The proposed models provide flexible parametric alternatives for survival data in the presence of long-term survivors, with the cure fraction induced by the limiting value of the survival function rather than introduced through an explicit mixture component.

The theoretical development showed that the Gamma–G construction preserves defectiveness when applied to defective baseline distributions, providing a coherent framework for extending classical defective models. The proposed formulations also allow covariate effects to be incorporated into model parameters, enabling regression-based analysis of heterogeneous survival patterns and long-term survival probabilities.

The Monte Carlo simulation study indicated that the maximum likelihood estimators generally performed well in the scenarios considered, with decreasing bias and RMSE as the sample size increased and empirical coverage probabilities approaching the nominal level. These results support the finite-sample adequacy of the estimation procedure under the data-generating configurations analyzed in this work.

The applications to melanoma and cervical cancer data from the São Paulo Cancer Registry illustrated the practical relevance of the proposed models. For the melanoma dataset, the defective Gamma–Gompertz model captured distinct survival patterns across

clinical stages and provided cure-fraction estimates consistent with the observed Kaplan–Meier plateaus. For the cervical cancer dataset, the defective Gamma–Dagum model provided a flexible fit to the survival data and reflected the marked difference in long-term survival between early and advanced clinical stages.

Overall, the defective Gamma–Gompertz and defective Gamma–Dagum models expand the class of defective distributions available for cure-rate survival analysis. Although the results are promising, the observed dependence among some parameter estimates highlights the importance of numerical diagnostics, careful interpretation, and further investigation of identifiability in flexible defective models. Future research may extend this framework to competing risks, frailty structures, Bayesian inference, predictive validation, and applications involving richer clinical covariate information.

Author Contributions: Conceptualization, methodology, formal analysis, investigation, software, validation, data curation, visualization, writing—original draft preparation, and writing—review and editing, C.A.V.T., V.L.D.T., A.S.R. and P.R.D.M.; supervision, C.A.V.T., V.L.D.T., A.S.R. and P.R.D.M.; project administration and funding acquisition, C.A.V.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CAPES through a postdoctoral fellowship under the PDPG-FAPIII program, process number 88887.235722/2025-00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in this study are publicly available from Fundação Oncocentro de São Paulo (FOSP) at: https://fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/rhc-registro-hospitalar-de-cancer/download-de-arquivos/ (accessed on 9 June 2026). The real cancer registry datasets are not redistributed by the authors. The R code used to fit the proposed models, perform the Monte Carlo simulation studies, compute the numerical analyses, obtain the estimator correlation matrices, and reproduce the main figures and tables is available at: https://github.com/CynthiaTojeiro552/Defective-Gamma-G-Cure-Models (accessed on 9 June 2026). Instructions for preparing the required data files are provided in the repository.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Boag, J.W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc. Ser. B Methodol.* **1949**, *11*, 15–53. [CrossRef]
2. Berkson, J.; Gage, R.P. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* **1952**, *47*, 501–515. [CrossRef]
3. Farewell, V.T. The Use of Mixture Models for the Analysis of Survival Data with Long-Term 806 Survivors. *Biometrics* **1982**, *38*, 1041–1046. [CrossRef]
4. Maller, R.A.; Zhou, X. *Survival Analysis with Long-Term Survivors*; Wiley: New York, NY, USA, 1996.
5. Martinez, E.Z.; Achcar, J.A.; Jácome, A.A.; Santos, J.S. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput. Methods Programs Biomed.* **2013**, *112*, 343–355. [CrossRef] [PubMed]
6. Shi, H.; Ma, D.; Faisal Beg, M.; Cao, J. A functional proportional hazard cure rate model for interval-censored data. *Stat. Methods Med. Res.* **2022**, *31*, 154–168. [PubMed]
7. Wang, P.; Pal, S. A two-way flexible generalized gamma transformation cure rate model. *Stat. Med.* **2022**, *41*, 2427–2447. [CrossRef] [PubMed]
8. de Souza, H.C.C.; Louzada, F.; Ramos, P.L.; de Oliveira Júnior, M.R.; Perdoná, G.d.S.C. A 816 Bayesian approach for the zero-inflated cure model: An application in a Brazilian invasive 817 cervical cancer database. *J. Appl. Stat.* **2022**, *49*, 3178–3194. [PubMed]
9. Mota, A.; Milani, E.A.; Leão, J.; Ramos, P.L.; Ferreira, P.H.; Junior, O.G.; Tomazella, V.L.D.; Louzada, F. A new cure rate frailty regression model based on a weighted Lindley distribution applied to stomach cancer data. *Stat. Methods Appl.* **2023**, *32*, 883–909. [CrossRef]

10. Cancho, V.G.; Bedia, E.C.; Cordeiro, G.M.; Prataviera, F.; Ortega, E.M.; Santo, A.P. A survival regression with cure fraction applied to cervical cancer. *Comput. Stat.* **2023**, *38*, 403–418.
11. Balka, J.; Desmond, A.F.; McNicholas, P.D. Bayesian and likelihood inference for cure rates based on defective inverse Gaussian regression models. *J. Appl. Stat.* **2011**, *38*, 127–144.
12. Calsavara, V.F.; Rodrigues, A.S.; Rocha, R.; Louzada, F.; Tomazella, V.; Souza, A.C.; Costa, R.A.; Francisco, R.P. Zero-adjusted defective regression models for modeling lifetime data. *J. Appl. Stat.* **2019**, *46*, 2434–2459. [CrossRef]
13. Rodrigues, A.S.; Borges, P. Long-term Dagum-power variance function frailty regression model: Application in health studies. *Stat. Methods Med. Res.* **2025**, *34*, 407–439. [CrossRef] [PubMed]
14. Gieser, P.W.; Chang, M.N.; Rao, P.; Shuster, J.J.; Pullen, J. Modelling cure rates using the Gompertz model with covariate information. *Stat. Med.* **1998**, *17*, 831–839. [CrossRef]
15. Haybittle, J. The estimation of the proportion of patients cured after treatment for cancer of the breast. *Br. J. Radiol.* **1959**, *32*, 725–733. [CrossRef] [PubMed]
16. Cantor, A.B.; Shuster, J.J. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Stat. Med.* **1992**, *11*, 931–937. [CrossRef] [PubMed]
17. Dagum, C. A new model of personal income distribution: Specification and estimation. *Econ. Appliquée* **1977**, *30*, 413–437. [CrossRef]
18. Martinez, E.; Achcar, J. A new straightforward defective distribution for survival analysis in the presence of a cure fraction. *J. Stat. Theory Pract.* **2018**, *12*, 688–703. [CrossRef]
19. Calsavara, V.F.; Rodrigues, A.S.; Rocha, R.; Tomazella, V.; Louzada, F. Defective regression models for cure rate modeling with interval-censored data. *Biom. J.* **2019**, *61*, 841–859. [CrossRef] [PubMed]
20. Rocha, R.; Nadarajah, S.; Tomazella, V.; Louzada, F. Two new defective distributions based on the Marshall–Olkin extension. *Lifetime Data Anal.* **2016**, *22*, 216–240. [PubMed]
21. Rocha, R.; Nadarajah, S.; Tomazella, V.; Louzada, F.; Eudes, A. New defective models based on the Kumaraswamy family of distributions with application to cancer data sets. *Stat. Methods Med. Res.* **2017**, *26*, 1737–1755. [PubMed]
22. Zografos, K.; Balakrishnan, N. On families of beta- and generalized gamma-generated distributions and associated inference. *Stat. Methodol.* **2009**, *6*, 344–362. [CrossRef]
23. Oluyede, B.O.; Huang, S.; Pararai, M. A New Class of Generalized Dagum Distribution with Applications to Income and Lifetime Data. *J. Stat. Econom. Methods* **2014**, *3*, 125.
24. Ramos, M.; Cordeiro, G.; Marinho, P.R.; Cícero, R.; Dias, C.; Hamedani, G. The Zografos-Balakrishnan Log-Logistic Distribution: Properties and Applications. *J. Stat. Theory Appl.* **2013**, *12*, 225–244. [CrossRef]
25. da Silva, R.V.; de Andrade, T.A.; Maciel, D.B.; Campos, R.P.; Cordeiro, G.M. A New Lifetime Model: The Gamma Extended Fréchet Distribution. *J. Stat. Theory Appl.* **2013**, *12*, 39–54. [CrossRef]
26. Alzaatreh, A.; Knight, K. On The Gamma-Half Normal Distribution and Its Applications. *J. Mod. Appl. Stat. Methods* **2013**, *12*, 103–119. [CrossRef]
27. Oluyede, B.; Pararai, M.; Warahena-Liyanage, G. A New Class of Generalized Inverse Weibull Distribution with Applications. *J. Appl. Math. Bioinform.* **2014**, *4*, 17–35.
28. Cordeiro, G.; Otrega, E.; Popović, B. The gamma-linear failure rate distribution: Theory and applications. *J. Stat. Comput. Simul.* **2013**, *84*, 2408–2426. [CrossRef]
29. Castellares, F.; Santos, M.; Montenegro, L.; Cordeiro, G. A Gamma-Generated Logistic Distribution: Properties and Inference. *Am. J. Math. Manag. Sci.* **2015**, *34*, 14–39. [CrossRef]
30. Salinari, G.; De Santis, G. One or more rates of ageing? The extended gamma-Gompertz model (EGG). *Stat. Methods Appl.* **2020**, *29*, 211–236. [CrossRef]
31. Martinez, E.; Achcar, J. The defective generalized Gompertz distribution and its use in the analysis of lifetime data in presence of cure fraction, censored data and covariates. *Electron. J. Appl. Stat. Anal.* **2017**, *10*, 463–484. [CrossRef]
32. Instituto Nacional de Câncer (INCA). Estimativa 2023: Incidência de Câncer no Brasil. 2022. Available online: https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil (accessed on 9 June 2026).
33. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
34. Ministério da Saúde—DATASUS. Informações de Saúde (TABNET)—Mortalidade por Câncer. 2023. Available online: https://datasus.saude.gov.br/ (accessed on 9 June 2026).
35. Calsavara, V.F.; Milani, E.A.; Bertolli, E.; Tomazella, V. Long-term frailty modeling using a non-proportional hazards model: Application with a melanoma dataset. *Stat. Methods Med. Res.* **2020**, *29*, 2100–2118. [PubMed]
36. Balka, J.; Desmond, A.F.; McNicholas, P.D. Review and implementation of cure models based on first hitting times for Wiener processes. *Lifetime Data Anal.* **2009**, *15*, 147–176. [CrossRef] [PubMed]

37. Klein, J.; Moeschberger, M. *Survival Analysis: Statistical Methods for Censored and Truncated Data*; Springer: New York, NY, USA, 2003.
38. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
39. Olive, D.J. *Statistical Theory and Inference*; Springer: Berlin/Heidelberg, Germany, 2014.
40. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [CrossRef]
41. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.