

The Poisson Generalized Birnbaum-Saunders Cure Model and Application in Breast Cancer Data

Mojtaba Meshkat, Ahmad Reza Baghestani and Farid Zayeri*

Department of Biostatistics, School of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

The cure rate survival models are generally used to model lifetime data with long term survivors. We assumes the number of competing causes of the event of interest has the Poisson distribution and the time to the event of interest follows the Generalized Birnbaum-Saunders distribution. The Poisson GB-S distribution has been defined and usefull representations for its density function have been presented which facilitates obtaining some mathematical properties. The parameters of the model with cure rate have been estimated using the maximum likelihood method. For different sample sizes and censoring percentages, several simulations have been performed and a real data set from the medical area has been analyzed.

Keywords: Cure fraction models; Generalized Birnbaum-Saunders distribution; Poisson distribution; Lifetime data

Introduction

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest [1]. The time to event or survival time can be measured in days, weeks, years, etc. In survival analysis, subjects are usually followed over a specified time period and the focus is on the time at which the event of interest occurs. A number of models are available to analyze the relationship of a set of predictor variables with the survival time. Methods include parametric, nonparametric and semiparametric approaches. Parametric methods assume that the underlying distribution of the survival times follows certain known probability distributions. Popular ones include the exponential, Weibull, and lognormal distributions.

The Birnbaum-Saunders (BS) distribution has a close relation to the normal distribution and has been applied in a wide variety of fields and distribution is a non-negative support that has attracted considerable interest [2]. This distribution is positively skewed, unimodal and has two parameters [2,3]. Generalized Birnbaum-Saunders (GB-S) distribution is a flexible lifetime model that admits different degrees of kurtosis and asymmetry while possessing unimodality and bimodality proposed by Díaz-García et al. , Owen and Leiva [4-8]. Also, GB-S distributions produce models whose parameter estimates are usually non-sensitive to outliers and robust to irregular data [8-11]. This interest in the GB-S distribution is due to its physical and theoretical arguments, relationship with the normal distribution and other attractive properties.

For $t > 0$, the survival function, cumulative distribution function and probability density function of the GB-S distribution are [6,7]:

$$S_{GB-S} = \Phi \left(-\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right) \quad (1)$$

$$F_{GB-S} = 1 - S_{GB-S} = \Phi \left(\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right) \quad (2)$$

$$f_{GB-S} = \frac{\lambda}{\alpha t \sqrt{2\pi}} \left(\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right) e^{\left\{ -\frac{1}{2\alpha^2} \left[\left(\frac{t}{\beta} \right)^{2\lambda} + \left(\frac{\beta}{t} \right)^{2\lambda} - 2 \right] \right\}} \quad (3)$$

Where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\alpha > 0$, $\beta > 0$ and $\lambda > 0$ are shape scale and location parameters, respectively.

In survival analysis, there are situation which the studied population is a mixture of uncured (susceptible individuals - who may experience the event of interest), and cured (non-susceptible individuals - who will never experience the event), the standard survival models are usually not appropriate because they do not account for the possibility of cure. Many patients with disease like cancer can be long-term survivors, and thus cure models can be a useful tool to analyze and describe their survival data.

Cure rate models contain situations where subjects not sensitive to the occurrence of the event of interest there are situations where a fraction of individuals are not expected to experience the event of interest ; that is, those individuals are cured. These models have become very popular due to the significant progress in treatment therapies leading to enhanced cure rates. In the medical studies, many patients with disease like cancer can be long-term survivors, and thus cure models can be a useful tool to analyze and describe their survival data.

The most popular type of cure rate model is the mixture cure model (MCM) which was introduced by Boag to study cases where there was a proportion of cured patients among treatment receivers for mouth cancer [12]. This model was developed later by Berkson et al. [13-16]. In this distribution, it is assumed that a specific number of the patients, say p_0 , are cured, in the sense that they do not present the event of interest during a long period of time and can be observed as to be immune to the cause of death under study or cured. In this model, the survivor function for the entire population is given by

*Corresponding author: Zayeri F, Department of Biostatistics, School of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran, Tel: +989125277488; E-mail: fzayeri@yahoo.com

Received November 28, 2017; Accepted January 23, 2018; Published January 29, 2018

Citation: Meshkat M, Baghestani AR, Zayeri F (2018) The Poisson Generalized Birnbaum-Saunders Cure Model and Application in Breast Cancer Data. J Biom Biostat 9: 389. doi: 10.4172/2155-6180.1000389

Copyright: © 2018 Meshkat M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

$$S_{POP}(t) = p_0 + (1 - p_0)S(t) \quad t > 0 \quad (4)$$

where $S(t)$ is survival function and p_0 is the cured fraction.

The other type of cure rate model is the Non-Mixture Cure Model (NMC), which was first proposed by Yakovlev et al. and was further extended by some other authors [17-23]. The literature on distributions which accommodate other different latent competing causes is rich and growing rapidly.

Cancho et al. introduced the Geometric Birnbaum-Saunders regression model with cure rate with an application to a cutaneous melanoma data. It was demonstrated that it can provide better fitting in comparison with the MB-S distribution [24]. Hashimoto et al. and Cordeiro et al. presented the Poisson Birnbaum-Saunders model with long-term survivors in breast cancer data and Negative binomial Birnbaum-Saunders model with long-term survivors in melanoma data [25,26].

In this paper, we have proposed a new distribution family here, the Poisson generalized Birnbaum-Saunders cure rate model (PGB-S), which is conceived inside a latent competing causes scenario with cure fraction where there is no information regarding the exact cause of the individual death or tumor recurrence and only the minimum lifetime value among all of the risks is observed and a fraction of the population is not susceptible to the event of interest.

This paper is organized as follows: in Section 2 the PGB-S model is formulated by defining the density, cumulative distribution and hazard rate functions of the Poisson generalized Birnbaum-Saunders (PGB-S) distribution and some of its properties are also studied. In Section 3 we explain the maximum-likelihood estimation procedure and parameter inference. In Section 4 the performance of the parameter estimates using Monte Carlo simulation is evaluated. Then, an application to a real data set on breast cancer is given in Section 5. Finally, in Section 6 some conclusions are drawn.

Materials and Methods

Model formulation

We formulate the model within a biological context. Let M define the random variable unobservable that denotes the number of causes (risk factors) of the event under study for an individual in the population. We assume that M follows a Poisson distribution with parameter φ and mass probability

$$P(M = m) = \frac{e^{-\varphi} \varphi^m}{m!} \quad m = 0, 1, \dots \quad (5)$$

The promotion time for the j th cause (for example tumor cell) to produce the event of interest is defined by Z_j , $j=1, \dots, M$. We assume that, conditional on M , the Z_j s are i.i.d. random variables having the generalized Birnbaum-Saunders cumulative function given by eqn. (2). Also, we assume that Z_1, Z_2, \dots are independent of M . The observable time to event of interest is defined by the random variable $T = \min(Z_1, \dots, Z_M)$, and $T = \infty$ if $M=0$ with $P(T = \infty | M=0) = 1$. Under this setup and used the eqn. (5), the survival function for the population is

$$S_{POP}(t) = P(M=0) + P(Z_1 > t, \dots, Z_M > t | M \geq 1) = e^{-\varphi} + \sum_{m=1}^{\infty} [S_{GB-S}]^m P(M=m) \\ = e^{-\varphi} + \sum_{m=1}^{\infty} \left\{ \Phi \left[-\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right]^m \frac{e^{-\varphi} \varphi^m}{m!} \right\} = \exp \left\{ -\varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right\} \quad (6)$$

The cure function is defined by $p_0 = S_{POP}(\infty) = \exp(-\varphi)$ and the corresponding density function becomes:

$$f_{POP}(t) = \varphi f_{GB-S}(t) \exp \left\{ -\varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right\} \quad (7)$$

where $f_{GB-S}(t)$ is the probability density function of the distribution given in eqn. (3). In addition, the population hazard function is given by:

$$h_{POP}(t) = \varphi f_{GB-S} = \frac{\varphi \lambda}{\alpha t \sqrt{2\pi}} \left(\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right) e^{\left\{ -\frac{1}{2\alpha^2} \left[\left(\frac{t}{\beta} \right)^{2\lambda} + \left(\frac{\beta}{t} \right)^{2\lambda} - 2 \right] \right\}} \quad (8)$$

The function $h_{POP}(t)$ is multiplicative in φ and $f_{GB-S}(t)$ thus, it has the proportional hazard structure when the covariates are modeled through φ .

The distribution in eqn. (6) can be written as a mixture distribution [13]:

$$S_{POP}(t) = [\exp(-\varphi)] + [1 - \exp(-\varphi)] \left\{ \frac{\exp \left[-\varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right] - \exp(-\varphi)}{1 - \exp(-\varphi)} \right\} \quad (9)$$

Then, from eqns. (4) and (9), the survival function for the non-cured population from the Poisson Generalized Birnbaum Saunders (PGB-S) survival function is hereafter given by

$$S_{PGB-S}(t) = P(T > t | M \geq 1) = \frac{\exp \left[-\varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right] - \exp(-\varphi)}{1 - \exp(-\varphi)} \quad (10)$$

It should be noted that $S_{PGB-S}(0) = 1$ and $S_{PGB-S}(\infty) = 0$, so that it is a proper survival function. The density function of the PGB-S distribution is given by:

$$f_{PGB-S}(t) = \frac{\varphi \lambda t^{-(\lambda+1)} (t^{2\lambda} + \beta^{2\lambda}) \exp \left\{ -\frac{1}{2\alpha^2} \left[\left(\frac{t}{\beta} \right)^{2\lambda} + \left(\frac{\beta}{t} \right)^{2\lambda} - 2 \right] - \varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right\}}{\sqrt{2\pi\alpha\beta^\lambda} (1 - e^{-\varphi})} \quad (11)$$

From eqn. (11), the parameter β controls the scale of the density distribution while the parameters α , λ and φ control its shape. As φ approaches zero, the PGB-S distribution converges to a GB-S distribution. In Figure 1, the PGB-S density function has been plotted for selected values of the parameters. The PGB-S distribution can be used to model data in survival analysis.

From eqns. (10) and (11), it is easy to verify that the hazard rate function of the non-cured population is

$$h_{PGB-S}(t) = \frac{\varphi \lambda t^{-(\lambda+1)} (t^{2\lambda} + \beta^{2\lambda}) \exp \left\{ -\frac{1}{2\alpha^2} \left[\left(\frac{t}{\beta} \right)^{2\lambda} + \left(\frac{\beta}{t} \right)^{2\lambda} - 2 \right] - \varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right\}}{\sqrt{2\pi\alpha\beta^\lambda} \left\{ \exp \left[-\varphi \Phi \left[\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right] \right] - \exp(-\varphi) \right\}} \quad (12)$$

Numerical examples allow us to determine that the hazard rate function eqn. (12) is either increasing or unimodal (Figure 2).

The first moment [27] of the PGB-S distribution is given by

$$\mu_r = E(T^r) = \int_0^{\infty} t^r f_{PGB-S}(t) dt \quad (13)$$

where $f_{PGB-S}(t)$ is the probability density function of the distribution given in eqn. (11).

The skewness and kurtosis of the distribution can be calculated from the ordinary moments given in (13) using numerically obtainable in common statistical software such as SAS university. Figure 3 shows graphical representations of kurtosis and skewness where it can be

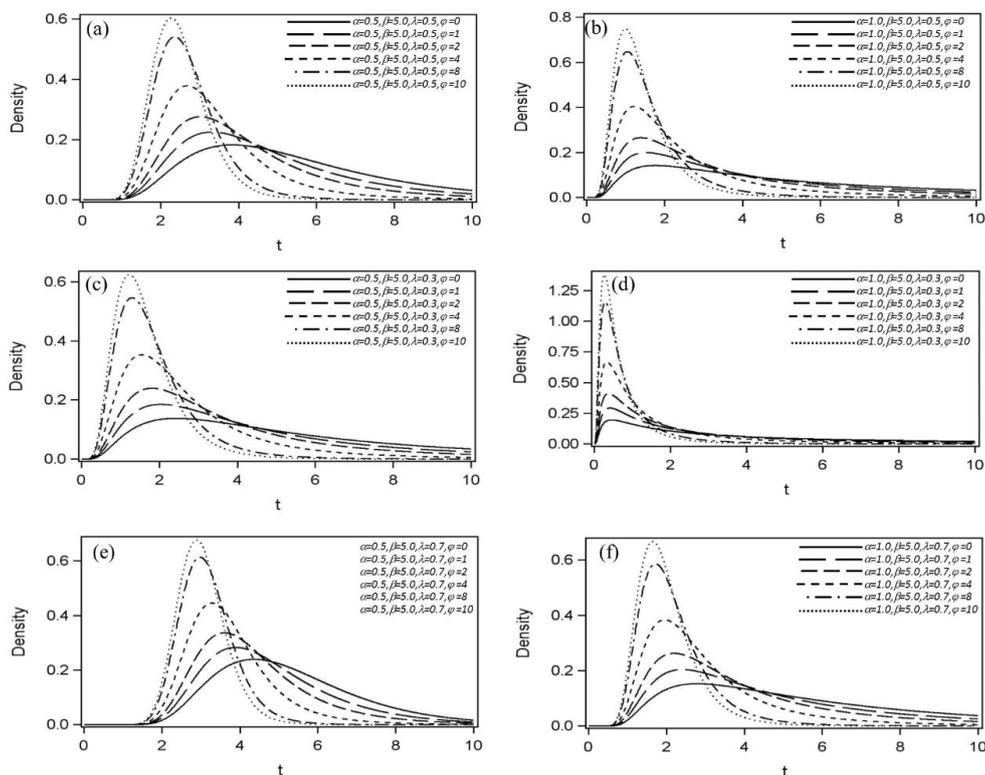


Figure 1: Probability density function of the PGB-S distribution. The parameters were fixed $\varphi=0$ 1 2 4 8 10 and $\alpha=0.5, \beta=5.0, \lambda=0.5$ (a), $\alpha=1.0, \beta=5.0, \lambda=0.5$ (b), $\alpha=0.5, \beta=5.0, \lambda=0.3$ (c), $\alpha=1.0, \beta=5.0, \lambda=0.3$ (d), $\alpha=0.5, \beta=5.0, \lambda=0.7$ (e), $\alpha=1.0, \beta=5.0, \lambda=0.7$ (f).

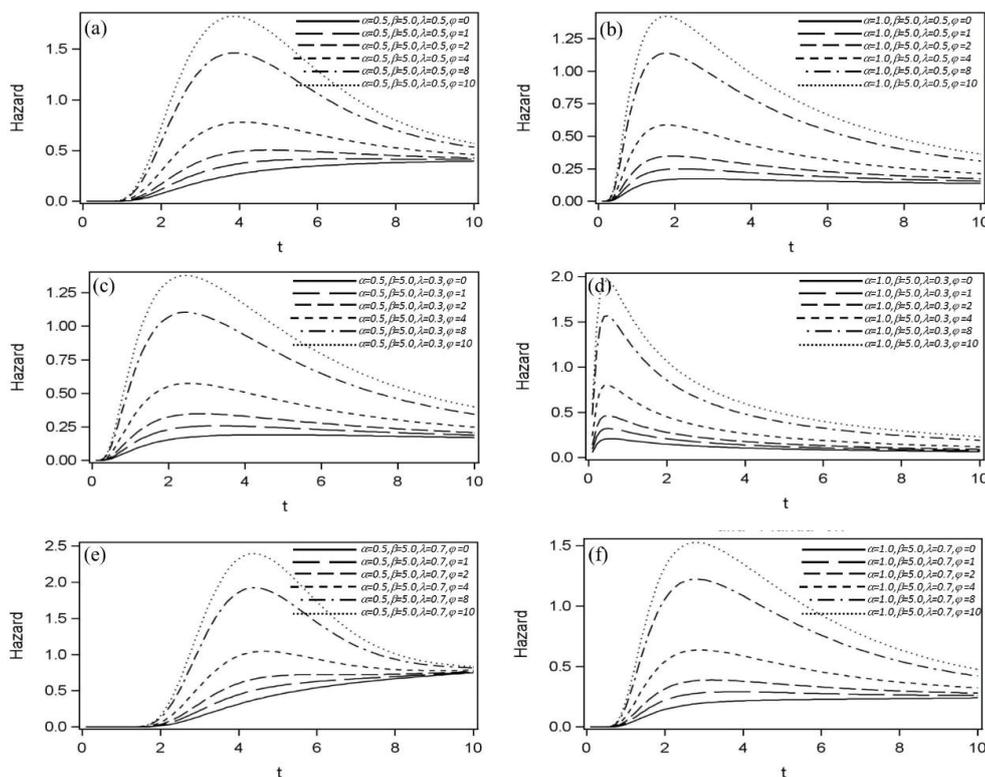


Figure 2: Hazard function of the PGB-S distribution. The parameters were fixed $\varphi=0$ 1 2 4 8 10 and $\alpha=0.5, \beta=5.0, \lambda=0.5$ (a), $\alpha=1.0, \beta=5.0, \lambda=0.5$ (b), $\alpha=0.5, \beta=5.0, \lambda=0.3$ (c), $\alpha=1.0, \beta=5.0, \lambda=0.3$ (d), $\alpha=0.5, \beta=5.0, \lambda=0.7$ (e), $\alpha=1.0, \beta=5.0, \lambda=0.7$ (f).

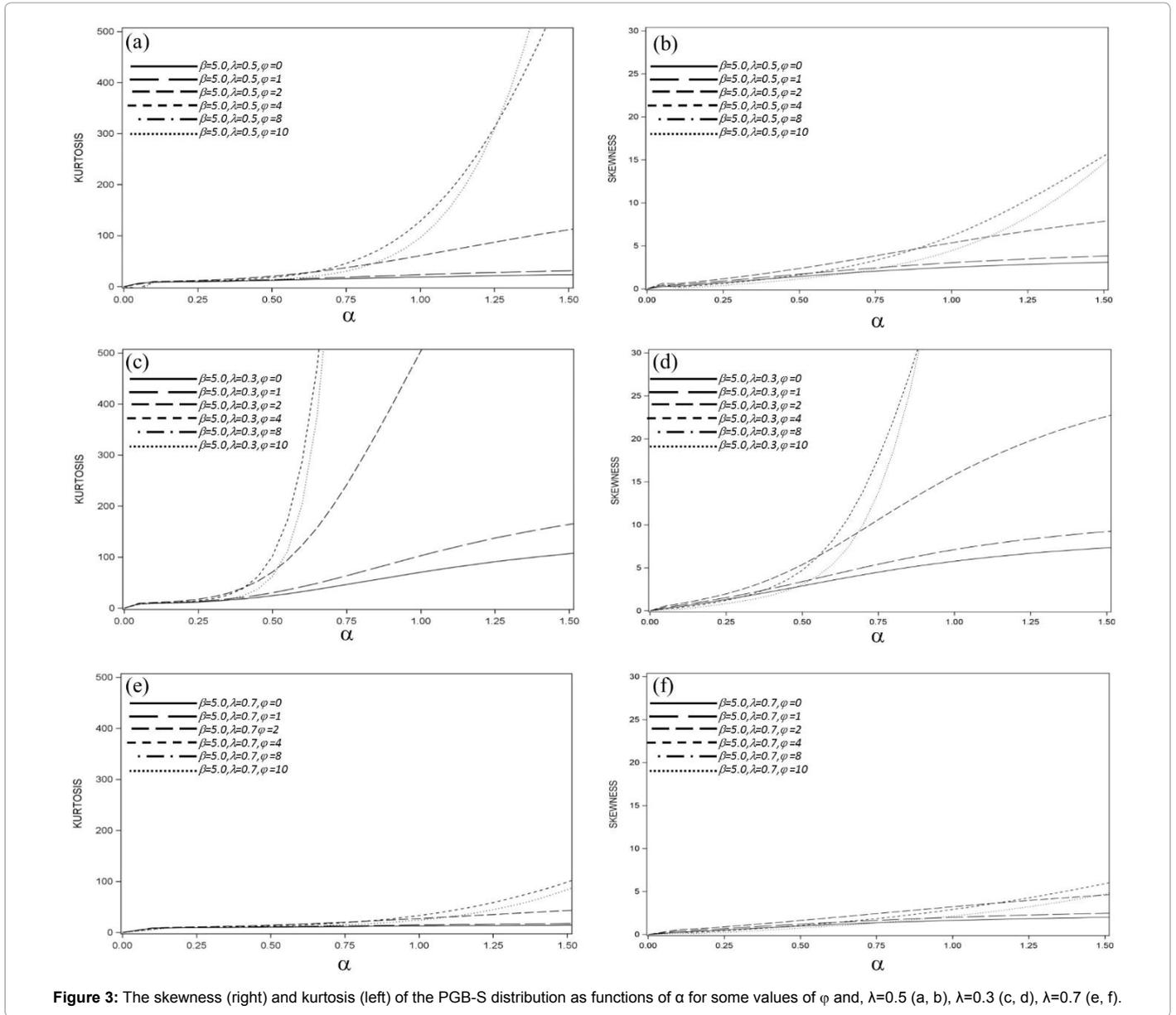


Figure 3: The skewness (right) and kurtosis (left) of the PGB-S distribution as functions of α for some values of ϕ and, $\lambda=0.5$ (a, b), $\lambda=0.3$ (c, d), $\lambda=0.7$ (e, f).

observed that both skewness and kurtosis are increasing functions of α , ϕ and λ .

Inference

The lifetimes of all the sampling units may not be observable, due to censoring that could be present in the data. We consider the case that the minimum lifetime of all the competing causes is not completely observed and is subject to right censoring. If we consider C_i denote the censoring time, Then we observe $\delta_i = I(T_i \leq C_i)$ and $t_i \min\{T_i, C_i\}$, where $\delta_i = 1$ if T_i is the observed time to the event defined before and $\delta_i = 0$ if it is right censored, for $i = 1, \dots, n$. Hereafter, we Consider $\gamma = (\alpha, \beta, \lambda)^T$ as the parameter vector of the distribution function of the time-to-event, from n pairs of times and censoring indicators $(t_1, \delta_1), \dots, (t_n, \delta_n)$, the observed likelihood function under non-informative censoring [28] is

$$L(\phi, \gamma) \propto \prod_{i=1}^n [f_{POP}(t_i; \phi, \gamma)]^{\delta_i} [S_{POP}(t_i; \phi, \gamma)]^{1-\delta_i} \quad (14)$$

where $S_{POP}(t_i)$ and $f_{POP}(t_i)$, respectively, have been defined in eqns. (6) and (7). In much medical research, the lifetimes are usually affected by explanatory variables such as the sex, age, and many others. The parameter ϕ in eqn. (6) is now linked to a vector X_i of explanatory variables by assuming a log link, $\log(\phi) = X_i^T \eta$ $i = 1, \dots, n$, that is

$$p_{0i} = \exp(-\phi) = \exp(-\exp(X_i^T \eta)) \quad (15)$$

where $\eta = (\eta_1, \dots, \eta_p)$ denotes the vector of regression coefficients. However, we consider $\omega = (\gamma, \eta)$ as the vector of model parameters. By substituting eqns. (6) and (7) into eqn. (14), and use of eqn. (15) the log-likelihood function is

$$l(\omega) = \sum_{i=1}^n \delta_i \left\{ \exp(X_i^T \eta) f_{GB-S}(t) \exp \left\{ -\exp(X_i^T \eta) \Phi \left(\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right) \right\} \right\} + \sum_{i=1}^n (1 - \delta_i) \left\{ \exp \left\{ -\exp(X_i^T \eta) \Phi \left(\frac{1}{\alpha} \left[\left(\frac{t}{\beta} \right)^\lambda - \left(\frac{\beta}{t} \right)^\lambda \right] \right) \right\} \right\} \quad (16)$$

We can derive the model parameters by numerical maximization of the log-likelihood function obtained from eqn. (16) using the mathematical and statistical software's. The computational program is available from the authors upon request. Under suitable regularity conditions, the maximum likelihood (ML) estimator $\hat{\omega}$ can be approximated by the multivariate normal distribution with mean vector ω and covariance matrix $\Sigma(\hat{\omega})$ estimated at $\hat{\omega}$, namely the information matrix which is as follows:

$$\Sigma(\omega) = \left[-\frac{\partial^2 l(\omega)}{\partial \omega \partial \omega^T} \right]^{-1} \tag{17}$$

The required second derivatives can be computed numerically. Different models may be compared by penalizing over-fitting using the Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC). The model with the smallest value of any of these criteria is taken as the preferred model for describing the given data set. The quasi-Newton and Nelder-Mead methods (we can use option TECH in PROC NLP of SAS) were utilized to obtain the maximum likelihood estimates and asymptotic tests for the parameters were performed [29].

Simulation study

We considered the proposed model with the PGB-S distribution for the event times (T) with the parameters α 2, $\beta=2$ and λ (0.4, 0.5, 0.6). For each individual $i, i=1, \dots, n$, the number of causes (N_i) of the event of interest for the i th individual is generated from the Poisson distribution with parameter $p_{0i} = \exp(-\varphi) = \exp(-\exp(\eta_0 + \eta_1 x_i))$. Single binary covariate x was considered with values drawn from a Bernoulli distribution with a parameter of 0.5. We took η_0 .5 and η_1 -1 so that the cured fraction for the two levels of x_i are $p_0^{(0)} = 0.192$ and $p_0^{(1)} = 0.545$ respectively.

The censoring times are selected from the uniform distribution on the interval (0, τ), where τ was adjusted in order to control the proportion of censored observations. The proportion of censored observations was equal to 40% on average.

We took the sample sizes to be $n=50, 100, 500$ and 1000. For each sample, we calculated the average of maximum likelihood estimates of the cured fraction ($p_0^{(0)}$ and $p_0^{(1)}$), standard deviations (SD) and square root of mean square errors (SRMSE) of the maximum likelihood estimates for 1000 conducted simulations. The simulated data of PGB-S cure models are shown in Table 1. As the sample size increases,

N		$p_0^{(0)}$			$p_0^{(1)}$		
		Average	SD	SRMSE	Average	SD	SRMSE
$\lambda=0.4$	50	0.221	0.024	0.037	0.5756	0.0228	0.038
	100	0.220	0.019	0.034	0.5730	0.0243	0.037
	500	0.220	0.017	0.033	0.5743	0.0211	0.036
	1000	0.219	0.018	0.032	0.5739	0.0224	0.036
$\lambda=0.5$	50	0.212	0.018	0.027	0.5775	0.0197	0.038
	100	0.211	0.016	0.025	0.5699	0.0250	0.035
	500	0.213	0.017	0.027	0.5689	0.0225	0.033
	1000	0.213	0.018	0.027	0.5679	0.0224	0.032
$\lambda=0.6$	50	0.211	0.020	0.027	0.5672	0.0201	0.030
	100	0.210	0.019	0.026	0.5648	0.0194	0.028
	500	0.210	0.018	0.025	0.5553	0.0222	0.024
	1000	0.208	0.018	0.024	0.5545	0.0225	0.024

Table 1: Averages of maximum likelihood estimates, standard deviation and square root of mean square error (RMSE) of cured fractions $p_0^{(0)}$ and $p_0^{(1)}$ for the simulated data of PGB-S cure rate model.

the average of maximum likelihood estimates becomes closer to the true values of cured fraction, with the SDs and SRMSEs decreasing.

Results

We discuss an example employing the modeling approach presented in Section 2. The data set includes 557 women with breast cancer who were visited and treated at Cancer Research Center in Shahid Beheshti University of Medical Sciences, Tehran, Iran during 1992 to 2012; the patients were followed up until October 2014. The patients or patients' family members were contacted via phone calls to confirm their health status (i.e. whether they are still alive or not) and to fill any gaps in their medical records. We had to exclude some patients because of some reasons, first their medical records had incomplete information, second they were related to male and third, their cause of death was not breast cancer. The mean of observed time (t) is 4.07 ± 0.94 years with ranges from 3.23 to 211.97 months, (from 0.2692 to 17.66 years) and refers to the time until the patient's death or the censoring time. The Patients who died from other causes, as well as patients who were still alive at the end of the study were censored observations (83%). Stage of cancer (I, n=126; II, n=252; III, n=163; IIII, n=16), grade of cancer (I, n=61; II, n=312; III, n=184), Estrogen Receptor (absent, n=169; present, n=388) and Progesterone Receptor (absent, n=369; present, n=188) are taken as covariates. The Kaplan-Meier estimate of the surviving function is given in Figure 4. The presence of a flat line above 0.4 indicates that the cure models are suitable choices for analyzing this data set.

In the next step the maximum likelihood estimates of the coefficients as well as AIC and SBC selection criteria the PGB-S model are given in Table 2. We also reported the results of Hashimoto. According to both criteria, the PGB-S model stands out as the best one. In order to compare the nested models, which is the case when comparing the PGB-S and PB-S models, one can compute the maximum values of log-likelihoods to obtain the LR statistics for testing $H_0: \lambda=0.5$ versus $H_0: \lambda \neq 0.5$. The result of this test shows that the null hypothesis can be rejected ($p < 0.05$).

The quantile-quantile plot (QQ plot) of the normalized randomized quantile logarithm residuals from Dunn and Smyth [30], Rigby and Stasinopoulos [31] in Figure 5 suggests that the PGB-S model has yielded a good fit. From Table 2 for two models, the covariate stage of cancer had a significant effect (III and IV) on the reduction of the cured fraction. This estimate can easily be computed from Table 2 due to the parametrization of the cured fraction.

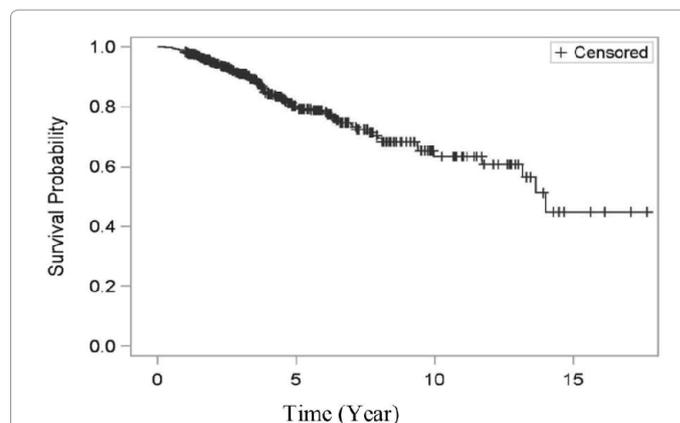


Figure 4: Kaplan-Meier estimate of the surviving function to the breast cancer data set.

Variable	PB-S (Hashimoto et al. [26])		PGB-S (This Article)	
	Estimate	SD	Estimate	SD
$\eta_{\text{intercept}}$	-1.390283	0.552006	-1.219967	0.554269
$\eta_{\text{Stage 4 to 1}}$	2.664943	0.486006	2.734685	0.511989
$\eta_{\text{Stage 3 to 1}}$	1.612516	0.415637	1.629502	0.420455
$\eta_{\text{Stage 2 to 1}}$	0.304332	0.434209	0.303773	0.438291
$\eta_{\text{Grade 3 to 1}}$	0.727723	0.484344	0.711619	0.493874
$\eta_{\text{Grade 2 to 1}}$	0.261274	0.463585	0.244934	0.470529
$\eta_{\text{ER+ to ER-}}$	-0.377558	0.373236	-0.362323	0.391168
$\eta_{\text{PR+ to PR-}}$	0.105269	0.368441	0.113465	0.384354
β	146.358428	22.145824	146.970379	34.521112
α	1.437391	0.046891	0.001651	0.136312
λ	----	---	0.000748	0.061687
AIC	1179.5		1164.56	
SBC	1239.8		1212.10	

Table 2: Maximum likelihood estimates for the parameters (SD) of the PB-S and PGB-S models with the cure rate fitted to the breast cancer data set.

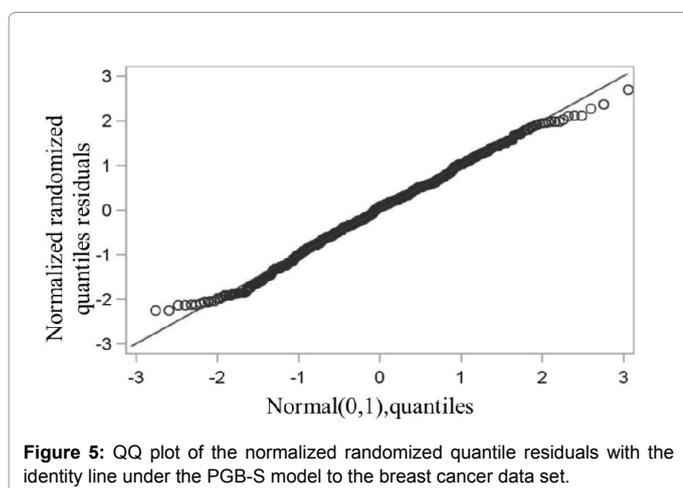


Figure 5: QQ plot of the normalized randomized quantile residuals with the identity line under the PGB-S model to the breast cancer data set.

Discussion and Conclusion

In the analysis of lifetime data we could have the presence of cure fraction and covariates, especially in medical applications. The Birnbaum-Saunders distribution has been extensively used for modeling in several fields such as medical sciences, biological studies, engineering and insurance. Because the generalized Birnbaum-Saunders (GB-S) distribution is a highly flexible lifetime model, so it would be a better fit to the data model which was introduced by Owen [7].

In this paper, a PGB-S model for analyzing survival data with cure fraction was proposed which is conceived inside a latent competing cause's scenario with cure fraction. We introduced a four-parameter continuous model called the PGB-S distribution which extends the GB-S one. We proposed cure rate model has the structure of the mixture cure rate model, and the PGB-S distribution represents the distribution associated with the individuals at risk.

The interpretation of the covariates turns out to be straightforward due to the parameterization of the cure fraction. Looking at the strong points of the PGB-S model, it can be employed as an interesting option to explain/predict the survival time for long-term individuals.

References

- Kleinbaum, David G, Klein, Mitchel (2005) *Survival Analysis: A self-Learning Text*, Springer, New York.

- Birnbaum ZW, Saunders SC (1969) A new family of life distributions. *Journal of Applied Probability* 6: 319-327.
- Johnson NL, Kotz S, Balakrishnan N (1995) *Continuous univariate distributions*. Wiley series in probability and mathematical statistics: applied probability and statistics. New York.
- Díaz-García JA, Sánchez VL, Narro A (2002) A new family of life distributions based on Birnbaum-Saunders distribution, Mexico.
- Díaz-García JA, Leiva-Sánchez VC (2005) A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference* 128: 445-457.
- Owen WJ (2004) Another look at the Birnbaum-Saunders distribution.
- Owen WJ (2006) A new three-parameter extension to the Birnbaum-Saunders distribution. *IEEE Transactions on Reliability* 55: 475-479.
- Leiva V (2016) Genesis of the Birnbaum-Saunders Distribution. In *The Birnbaum-Saunders Distribution*: 1-15.
- Elsayed E, Cheng Y (2013) Accelerated Life Testing Model for a Generalized Birnbaum-Saunders Distribution.
- Barros M, Paula GA, Leiva V (2008) A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis* 14: 316-332.
- Barros M, Paula GA, Leiva V (2009) An R implementation for generalized Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis* 53: 1511-1528.
- Boag JW (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)* 11: 15-53.
- Berkson J, Gage RP (1952) Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* 47: 501-515.
- Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*: 1041-1046.
- Kuk AY, Chen CH (1992) A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79: 531-541.
- Peng Y, Dear KBG, Denham JW (1998) A generalized F mixture model for cure rate estimation. *Statistics in Medicine* 17: 813-830.
- Yakovlev AY, Tsidikov AD (1996) *Stochastic models of tumor latency and their biostatistical applications* 1.
- Chen MH, Ibrahim JG, Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* 94: 909-919.
- Yin G, Ibrahim JG (2005) A general class of Bayesian survival models with zero and nonzero cure fractions. *Biometrics* 61: 403-412.
- Cooner F, Banerjee S, Carlin BP, Sinha D (2007) Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association* 102: 560-572.
- Rodrigues J, de Castro M, Cancho VG, Balakrishnan N (2009) COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference* 139: 3605-3611.
- Castro Md, Cancho VG, Rodrigues J (2009) A Bayesian Long-term Survival Model Para-metrized in the Cured Fraction. *Biometrical Journal* 51: 443-455.
- Borges P, Rodrigues J, Balakrishnan N (2012) Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data. *Computational Statistics and Data Analysis* 56: 1703-1713.
- Cancho VG, Louzada F, Barriga GD (2012) The Geometric Birnbaum-Saunders regression model with cure rate. *Journal of Statistical Planning and Inference* 142: 993-1000.
- Cordeiro GM, Cancho VG, Ortega EM, Barriga GD (2016) A model with long-term survivors: negative binomial Birnbaum-Saunders. *Communications in Statistics-Theory and Methods* 45: 1370-1387.
- Hashimoto EM, Ortega EM, Cordeiro GM, Cancho VG (2014) The Poisson Birnbaum-Saunders model with long-term survivors. *Statistics* 48: 1394-1413.

27. Casella G, Berger RL (2002) *Statistical inference 2*, (Duxbury Pacific Grove, CA).
28. Klein JP, Moeschberger ML (2005) *Survival analysis: techniques for censored and truncated data*, (Springer Science & Business Media).
29. Betts JT (2010) *Practical methods for optimal control and estimation using nonlinear programming*, (SIAM).
30. Dunn PK, Smyth GK (1996) Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5: 236-244.
31. Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54: 507-554.