

## RESEARCH ARTICLE

Editorial Process: Submission:07/21/2020 Acceptance:04/07/2021

# Utilization of a Mixture Cure Rate Model based on the Generalized Modified Weibull Distribution for the Analysis of Leukemia Patients

Mohamed Elamin Omer<sup>1,2\*</sup>, Mohd Abu Bakar<sup>2</sup>, Mohd Adam<sup>2,3</sup>, Mohd Mustafa<sup>2</sup>

### Abstract

**Objective:** Cure rate models are survival models, commonly applied to model survival data with a cured fraction. In the existence of a cure rate, if the distribution of survival times for susceptible patients is specified, researchers usually prefer cure models to parametric models. Different distributions can be assumed for the survival times, for instance, generalized modified Weibull (GMW), exponentiated Weibull (EW), and log-beta Weibull. The purpose of this study is to select the best distribution for uncured patients' survival times by comparing the mixture cure models based on the GMW distribution and its particular cases. **Materials and Methods:** A data set of 91 patients with high-risk acute lymphoblastic leukemia (ALL) followed for five years from 1982 to 1987 was chosen for fitting the mixture cure model. We used the maximum likelihood estimation technique via R software 3.6.2 to obtain the estimates for parameters of the proposed model in the existence of cure rate, censored data, and covariates. For the best model choice, the Akaike information criterion (AIC) was implemented. **Results:** After comparing different parametric models fitted to the data, including or excluding cure fraction, without covariates, the smallest AIC values were obtained by the EW and the GMW distributions, (953.31/969.35) and (955.84/975.99), respectively. Besides, assuming a mixture cure model based on GMW with covariates, an estimated ratio between cure fractions for allogeneic and autologous bone marrow transplant groups (and its 95% confidence intervals) were 1.42972 (95% CI: 1.18614 - 1.72955). **Conclusion:** The results of this study reveal that the EW and the GMW distributions are the best choices for the survival times of Leukemia patients.

**Keywords:** Acute lymphoblastic leukemia- cure fraction model- maximum likelihood estimation- right

*Asian Pac J Cancer Prev*, 22 (4), 1045-1053

### Introduction

Acute lymphocytic leukemia (ALL) is a form of blood and bone cancer where lymphoid progenitor cells proliferate in the bone marrow, blood, and other sites. While ALL represents 80% of children leukemia, it is uncommon in grown-ups (20% of cases). In the United States, the incidence of ALL is estimated at 1.64 per 100,000 persons (National Cancer Institute, 2020). According to the American Cancer Society database (2019), an estimated 5930 new cases were diagnosed, with 1,500 deaths due to ALL in 2019. The formulation of treatment for grown-ups ALL has been adjusted from pediatric conventions. Shockingly, while cure tends to be 90% for standard-hazard pediatric ALL, the long-term survival rate is humbler in grown-ups (Terwilliger and Abdul-Hay, 2017).

The main treatment of ALL is chemotherapy, which comprises of induction, intensification, and long-term maintenance, with the central nervous system (CNS)

prophylaxis provided at different times during the therapy. Induction therapy aims to accomplish total remission and to re-establish ordinary hematopoiesis. Following the induction treatment, patients underwent three cycles of consolidation treatment of methotrexate with leucovorin rescue and L-asparaginase. Registered individuals as high-risk disease and a corresponded donor, then received allogeneic stem cell transplantation (allo-SCT). The rest were randomly assigned to standard intensification/maintenance or autologous bone marrow transplants.

In survival analysis, analysts generally utilize product-limit estimates or log-rank test (Bradburn et al., 2003a), semi-parametrical models (for instance, Cox proportional hazards model), or regular parametrical models considering several well-known distributions in the existence of covariates (Cox, 1972). The Weibull distribution is widely used in cancer research (Bradburn et al., 2003b). Since its risk function is flexible and its parameters are easy to estimate. However, data sets of medical studies often necessitate more advanced

<sup>1</sup>Department of Mathematics, College of Science, Sudan University of Science and Technology, Khartoum, Sudan. <sup>2</sup>Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia. <sup>3</sup>Institute of Mathematical Research, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia. \*For Correspondence: elamin\_omer@yahoo.com

parametric models. As a consequence, to resolve this problem, several authors in the literature have proposed new classes of parametric distributions based on Weibull distribution such as the EW (Mudholkar and Srivastava, 1993), the GMW (Carrasco et al., 2008), the log-beta Weibull (Ortega et al., 2013), and the generalized alpha power inverse Weibull (Basheer, 2019) distributions. Another common scenario in survival data analysis, particularly in cancer research, is when a fraction of a population is not exposed to the event of interest. For this situation, patients were divided into two groups: those who were exposed to the event under study, and those who were not exposed to it and, therefore, were not at risk. These patients are viewed as cured or immunized.

The existence of cured subjects in a sample data is commonly proposed by a Kaplan-Meier curve, which displays a tall and a steady level with dense censoring at the right extreme (Corbière et al., 2009). To model the proportion of cured subjects, many authors have proposed several statistical methods. For further reading, we refer interested readers, for instance, to (Boag, 1949; Berkson and Gage, 1952; Goldman, 1984; Haybittle, 1965; Farewell, 1982; Maller and Zhou, 1992; Abu Bakar et al., 2008; Lu, 2010; López-Cheda et al., 2017). Moreover, the maximum likelihood estimation technique has been suggested by some authors such as (Farewell, 1982; Yamaguchi, 1992; Ghitany and Maller, 1992; Peng et al., 1998; Sy and Taylor, 2000) amongst others.

## Materials and Methods

### Leukemia data

In this paper, we considered a leukemia dataset, presented by Kersey (1987) and available in smcure package in R software (Cai et al., 2012). This dataset consists of 91 patients with high-risk ALL and is divided into two subsets; the first subset (Group 1) contains 46 patients who exposed to allogeneic bone marrow transplants and the second subset (Group 2) includes 45 patients who received autologous bone marrow transplants. The event of interest is time to death. We found that there is 24.17 % of censored observations, in other words, 28.26 % if we assume the patients received allogeneic bone marrow transplant and 20 % if we consider the patients received autologous bone marrow transplant.

### Mixture cure model

In the context of survival analysis, a mixture cure model assumes that the survival function for the entire population can be expressed as a mixture of the cured and the uncured patients and is given by

$$S(t) = \rho + (1 - \rho)S_{uc}(t), \quad 0 < \rho < 1 \quad (1)$$

where  $\rho$  is a proportion of “cured patients” or “long-term survivors” and  $S_{uc}(t)$  represents the survival function for susceptible patients (Boag, 1949). The density function for the random time  $T$  is given by:

$$f(t) = (1 - \rho)f_{uc}(t) \quad ,$$

where  $f_{uc}(t)$  is the probability density function for the uncured subjects.

Assume for each subject belongs to a random sample of size  $m$ , we observe the pair  $(t_j, d_j)$ ,  $j = 1, \dots, m$ , then the  $j^{th}$  subject contribution for the likelihood function can be given by

$$L_j = [f(t_j)]^{d_j} [S(t_j)]^{1-d_j} = [(1-\rho)f_{uc}(t_j)]^{d_j} [\rho + (1-\rho)S_{uc}(t_j)]^{1-d_j},$$

where  $d_j$  is a censoring indicator variable which is defined as

$$d_j = \begin{cases} 1, & \text{event of interest occurs} \\ 0, & \text{censoring.} \end{cases}$$

### The generalized modified Weibull distribution (GMW)

In the present study, we consider GMW distribution for uncured patients. The GMW distribution was proposed by Carrasco (2008), which is a four - parameter distribution with probability density function  $f_{uc}(t)$  and survival function  $S_{uc}(t)$  given respectively by:

$$f_{uc}(t) = \frac{\varepsilon \psi t^{\eta-1} (\eta + \lambda t) \exp[\lambda t - \varepsilon t^\eta \exp(\lambda t)]}{\{1 - \exp[-\varepsilon t^\eta \exp(\lambda t)]\}^{1-\psi}}, \quad t > 0, \quad (2)$$

and

$$S_{uc}(t) = 1 - \{1 - \exp[-\varepsilon t^\eta \exp(\lambda t)]\}^\psi, \quad t > 0, \quad (3)$$

where  $\varepsilon > 0, \eta \geq 0, \lambda \geq 0$ , and  $\psi > 0$ . This distribution has a capability to model monotonic and non-monotonic failure rates, and it is denoted by  $X \sim GMW(\varepsilon, \psi, \eta, \lambda)$ . The  $\varepsilon$  parameter is the scaling parameter, whereas  $\psi$  and  $\eta$  are shape parameters. The parameter  $\lambda$  is a sort of quickening operator in the lack of time. When the time increments,  $\lambda$  is working as an operator of weakness in the individual's survival time (Carrasco et al., 2008). The respective risk function is as follows:

$$h_{uc}(t) = \frac{\varepsilon \psi t^{\eta-1} (\eta + \lambda t) \exp[\lambda t - \varepsilon t^\eta \exp(\lambda t)] \{1 - \exp[-\varepsilon t^\eta \exp(\lambda t)]\}^{\psi-1}}{1 - \{1 - \exp[-\varepsilon t^\eta \exp(\lambda t)]\}^\psi} \quad (4)$$

The GMW distribution (Carrasco et al., 2008) represents a generalization of some special cases which are given as follows:

- Weibull distribution (W): in the case where  $\lambda=0$  and  $\psi=1$ , the equation (2) becomes

$$f_{uc}(t) = \varepsilon \eta t^{\eta-1} \exp(-\varepsilon t^\eta), \quad t > 0,$$

which is the probability density function of a Weibull distribution with two parameters, and, furthermore, if  $\eta=1$  and  $\eta=2$ , the cases correspond to exponential (E) and Rayleigh (R) distributions, respectively.

- Extreme value distribution (EV): for  $\psi=1$  and  $\eta=0$ , GMW reduces to

$$f_{uc}(t) = \varepsilon \lambda \exp[\lambda t - \varepsilon \exp(\lambda t)], \quad t > 0,$$

which is type I extreme value distribution (Kotz and Nadarajah, 2000). Nevertheless, when assuming survival

data some care is needed since its support dispersion over the entire real line (Lai et al., 2003). From the equation (3), we can derive the survival function for this distribution which is given by  $S_{uc}(t) = \exp[-\varepsilon \exp(\lambda t)]$ . Consequently, we have  $S_{uc}(t) = \exp(-\varepsilon)$  when  $t=0$ , and, in other words,  $S_{uc}(t=0) \neq 1$  as it is anticipated considering survival outcomes. This distribution is also included in the present study regardless of these issues.

- Exponentiated Weibull distribution (EW): if we let  $\lambda=0$  in the equation (2), we obtain the density

$$f_{uc}(t) = \varepsilon \psi \eta t^{\eta-1} \exp(-\varepsilon t^\eta) [1 - \exp(-\varepsilon t^\eta)]^{\psi-1}, \quad t > 0,$$

which is the probability density function of EW distribution proposed by Mudholkar (1993). If  $\eta=1$ , in addition to  $\lambda=0$ , the particular case coincides with the exponentiated exponential (EE) distribution. Gupta (2001) debated some statistical properties of the EE distribution. If  $\eta=2$  besides  $\lambda=0$ , the case coincides with the generalized Rayleigh (GR) distribution (Kundu and Raqab, 2005).

- Modified Weibull distribution (MW): if we take  $\psi=1$ , the GMW distribution becomes

$$f_{uc}(t) = \varepsilon t^{\eta-1} (\eta + \lambda t) \exp[\lambda t - \varepsilon t^\eta \exp(\lambda t)], \quad t > 0,$$

which is the density of a MW distribution with three-parameter as suggested by Lai (2003). Sarhan (2009) presented the estimation of the parameters and properties of MW distribution using the maximum-likelihood estimation approach.

#### The log - likelihood function

Let  $\Omega = (\rho, \varepsilon, \psi, \eta, \lambda)$ , then the log-likelihood function for  $\Omega$  regarding the mixture model in Equation (1) takes the form:

$$\begin{aligned} l(\Omega) = & \sum_{j=1}^m [\ln(1-\rho)] d_j + (\ln \varepsilon) \sum_{j=1}^m d_j + \sum_{j=1}^m (\ln \psi) d_j + (\eta-1) \sum_{j=1}^m d_j \ln t_j \\ & + \sum_{j=1}^m d_j \ln(\eta + \lambda t_j) + \lambda \sum_{j=1}^m d_j t_j - \varepsilon \sum_{j=1}^m d_j t_j^\eta \exp(\lambda t_j) \\ & - \sum_{j=1}^m (1-\psi) d_j \ln \{1 - \exp[-\varepsilon t_j^\eta \exp(\lambda t_j)]\} \\ & + \sum_{j=1}^m (1-d_j) \ln \{\rho + (1-\rho)(1 - \{1 - \exp[-\varepsilon t_j^\eta \exp(\lambda t_j)]\}^\psi)\}. \end{aligned} \quad (5)$$

#### Covariates Influence

To link the cure fraction  $\rho$  with a vector of covariates  $x_j$  considering the mixture cure model (1), we substitute  $\rho_j = \frac{\exp(x_j \beta)}{1 + \exp(x_j \beta)}$  for  $\rho$  in the Equation (5), where  $\beta$  represents a vector of unknown coefficients.

In order to study the effect of the transplant type as a covariate, we assume the following model:

$$\ln \left( \frac{\rho_j}{1 - \rho_j} \right) = \beta_0 + \beta_1 x_j$$

where  $x_j$  is a binary variable associated with the transplant (1 for the autologous transplant; 0 for the allogeneic transplant). The parameter  $\beta_1$  is linked to the influence of the transplant on the cure fraction. If the zero belongs to the 95% confidence intervals for  $\beta_1$ , we can deduce that there is no proof the transplant has influence.

Moreover, let  $\Phi = (\varphi_0, \varphi_1)$  be a vector of unknown values, we can include the kind of transplant as a covariate in the shape control parameter  $\psi$  by substituting

$\psi = \exp(\varphi_0 + \varphi_1 x_j)$  in the Equation (5). In this case, the parameter  $\varphi_1$  is associated with the transplant effect on the shape of the survival curve.

#### Model choice

Comparison between mixture cure models presuming various distributions, was evaluated using the Akaike information criterion (AIC) introduced by Akaike (1974) in the early 1970s. The AIC is defined by,  $AIC = -2l + 2q$  where  $-2l$  being the negative-two-log maximum likelihood and  $q$  is the number of free parameters in the model. A lower AIC value suggests preferable model fit.

## Results

The product-limit estimates of survival function considering the leukemia data are demonstrated in Figure 1a. There is a plateau in the right tail of the curve with a height closer to 0.228. Figure 1b presents the estimated survival curves for allogeneic and autologous groups, where steady plateaus are noticed after around 734 days and about 1,256 days of follow-up for the autologous and allogeneic groups, respectively.

In Table 1, we have maximum likelihood estimates for the parameters of the mixture cure model assuming the GMW distribution and its particular cases and excluding covariates, 95% confidence intervals for the estimated parameter, and AIC values. This table shows that the models considered the EW and the GMW distributions for susceptible patients are provide the smallest AIC values (953.3 and 955.94, respectively). To obtain a more visible illustration for the model fitting, considering different probability distributions for the leukemia data, Figure 2 displays product-limit estimators for survival function versus the corresponding anticipated values obtained by the mixture cure models for every proposed distribution (results from Table 1).

The survival functions and the respective risk functions provided by the mixture cure model fit based on GMW distribution and its sub-distributions (results in Table 1) are shown in Figure 3 (panels (a) and (b)) and (panels (c) and (d)), respectively. Figure 3, panel (b) shows that the survival curves obtained by the models based on the EW and GMW distributions are the nearest to the Kaplan-Meier estimates' curves.

Table 2 presents the results of the models based on GMW distribution and its sub-distributions, excluding the cure rate  $\rho$ . This table also illustrates that the smallest AIC values are obtained by EW and GMW distributions.

The results in Table 3, obtained by considering the mixture cure model based on the GMW distribution and the covariate is included both in  $\rho$  and  $\psi$ . The values of  $\rho_0$  and  $\rho_1$  included in Table 3, were calculated from the formulas;

$$\rho_0 = \exp(\beta_0) / [1 + \exp(\beta_0)]$$

$$\rho_1 = \exp(\beta_0 + \beta_1) / [1 + \exp(\beta_0 + \beta_1)]$$

Table 1. Maximum Likelihood Estimates, Considering the Mixture Cure Model based on the GMW Distribution and Its Special Cases, and Excluding Covariates.

Model	Parameter	Estimate	95% Confidence interval	AIC*
GMW	$\epsilon$	0.32873	(0.28630, 0.37116)	955.94
	$\psi$	9.84054	(8.74790, 10.93318)	
	$\eta$	0.43651	(0.41644, 0.45658)	
	$\lambda$	6.34E-10	(-5.644e-05, 5.644e-05)	
	$\rho$	0.23391	(0.22445, 0.24336)	
W	$\epsilon$	0.00389	(0.00345, 0.00432)	963.68
	$\eta$	1.05514	(1.03570, 1.07457)	
	$\rho$	0.23838	(0.22911, 0.24765)	
E	$\epsilon$	0.00528	(0.00515, 0.00543)	962.03
	$\rho$	0.23746	(0.22818, 0.24674)	
R	$\epsilon$	0.00001296	(1.261e-05, 1.331e-05)	1041
	$\rho$	0.24148	(0.23225, 0.25070)	
EV	$\epsilon$	0.39382	(0.38045, 0.40717)	1098.58
	$\lambda$	0.00294	(0.00289, 0.00298)	
	$\rho$	0.23582	(0.22654, 0.24511)	
EW	$\epsilon$	0.79738	(0.75529, 0.84406)	953.31
	$\psi$	26.16607	(24.52016, 28.02442)	
	$\eta$	0.31734	(0.30823, 0.32571)	
	$\rho$	0.22932	(0.21974, 0.23871)	
EE	$\epsilon$	0.00643	(0.00622, 0.00664)	961.45
	$\psi$	1.33095	(1.28320, 1.37871)	
	$\rho$	0.23932	(0.23007, 0.24858)	
GR	$\epsilon$	0.000005751	(5.427e-06, 6.076e-06)	981.53
	$\psi$	0.37329	(0.36226, 0.38433)	
	$\rho$	0.23853	(0.22927, 0.24780)	
MW	$\epsilon$	0.00402	(0.00359, 0.00446)	965.78
	$\eta$	1.04395	(1.02112, 1.06678)	
	$\lambda$	7.19E-11	(-8.379e-05, 8.379e-05)	
	$\rho$	0.24639	(0.23689, 0.25589)	

\*Akaike information criterion

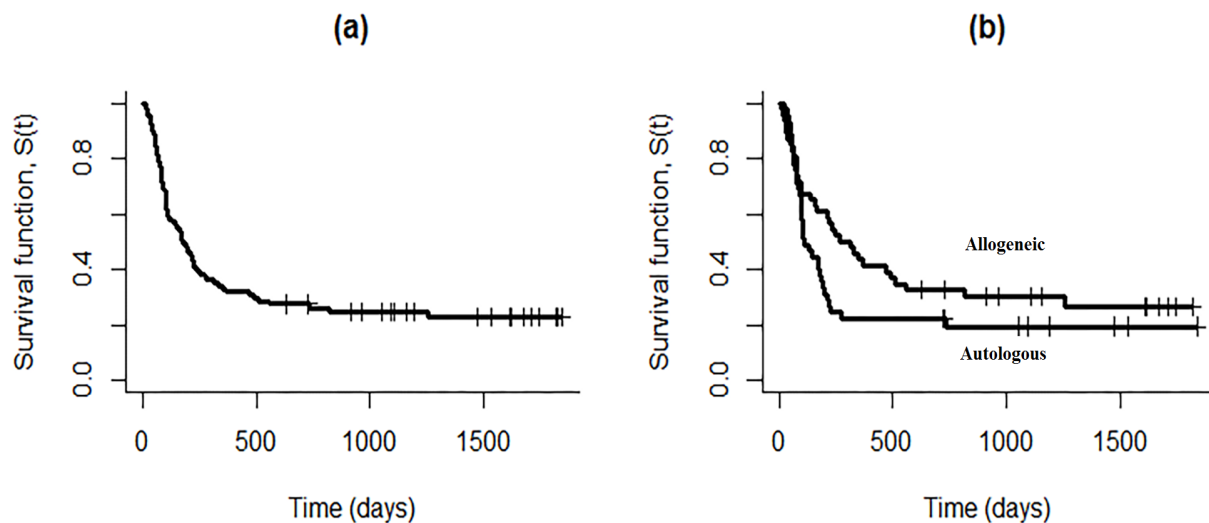


Figure 1. (a) Product-Limit Estimates of the Overall Survival Function for the Leukemia Patients' Data. (b) Kaplan-Meier estimates of survival functions for each type of transplant.

Table 2. Maximum Likelihood Estimates Excluding the Cure Rate  $\rho$  and not Including Covariates

Model	Parameter	Estimate	95% Confidence interval	AIC*
GMW	$\epsilon$	1.28517	(1.24678, 1.32357)	975.99
	$\psi$	26.6112	(25.64237, 27.58015)	
	$\eta$	0.18987	(0.18431, 0.19544)	
	$\lambda$	6.96E-12	(-1.6922e-05, 1.6922e-05)	
W	$\epsilon$	0.01837	(0.01682, 0.01993)	997.5
	$\eta$	0.63401	(0.62130, 0.64672)	
E	$\epsilon$	0.00127	(0.00124, 0.00131)	1015.88
R	$\epsilon$	1.53E-06	(1.4721e-06, 1.5921e-06)	1237.23
EV	$\epsilon$	0.31512	(0.30447, 0.32578)	1190.94
	$\lambda$	0.00125	(0.00123, 0.00128)	
EW	$\epsilon$	2.75704	(2.72275, 2.79133)	969.35
	$\psi$	178.18387	(177.2982, 179.0695)	
	$\eta$	0.12794	(0.12554, 0.13035)	
EE	$\epsilon$	0.00094	(0.00090, 0.00099)	1007.14
	$\psi$	0.56532	(0.54870, 0.58195)	
GR	$\epsilon$	5.18E-07	(4.1004e-07, 6.2647e-07)	1025.91
	$\psi$	0.28599	(0.27616, 0.29551)	
MW	$\epsilon$	0.00749	(0.00722, 0.00778)	1013.84
	$\eta$	0.70024	(0.68874, 0.71176)	
	$\lambda$	5.09E-10	(-5.1854e-05, 5.1855e-05)	

\*Akaike information criterion

and they represent the cure fractions assuming the patients exposed to allogeneic transplants and autologous transplants, respectively. This table shows that the estimated values of cure rates for the allogeneic and

autologous patients groups are 0.27415 and 0.19175, respectively.

The death risk functions considering the GMW mixture cure model with a covariate associated with the

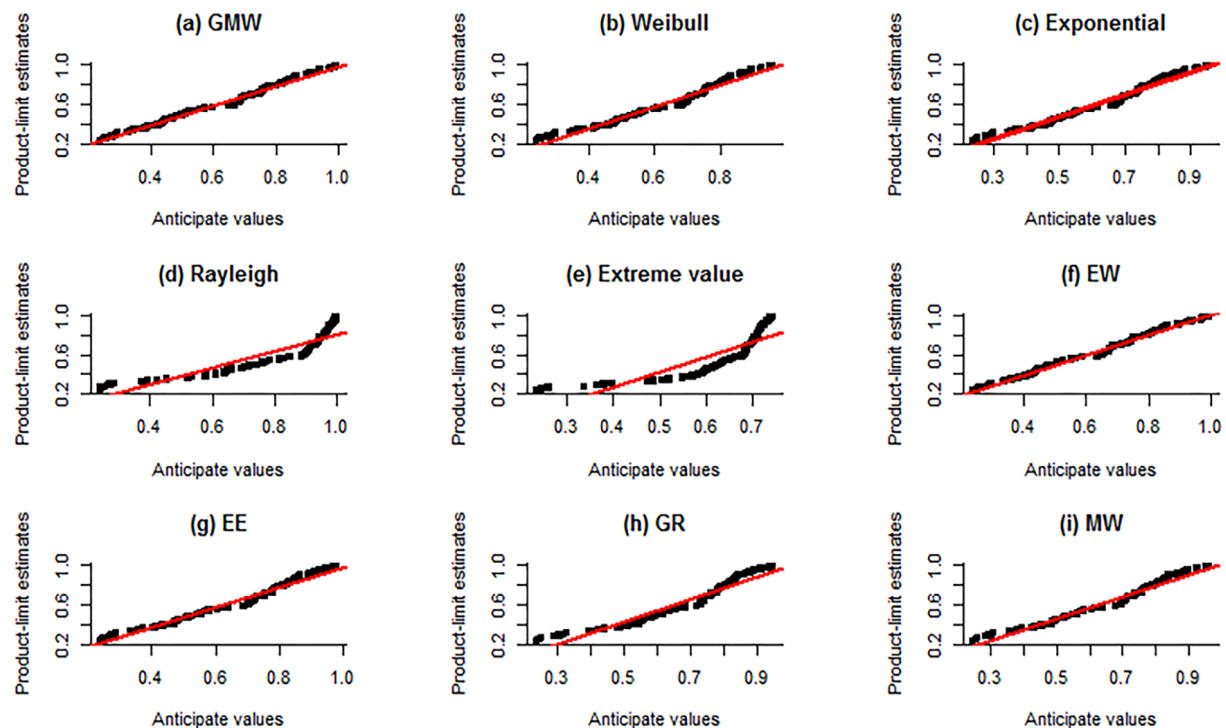


Figure 2. Product-Limit Estimates of the Survival Function versus the Anticipated Values Calculated from the Mixture Models for each Probability Distribution (Table 1, results). Perfect consent between product-limit estimates and anticipated values is shown by the diagonal lines.



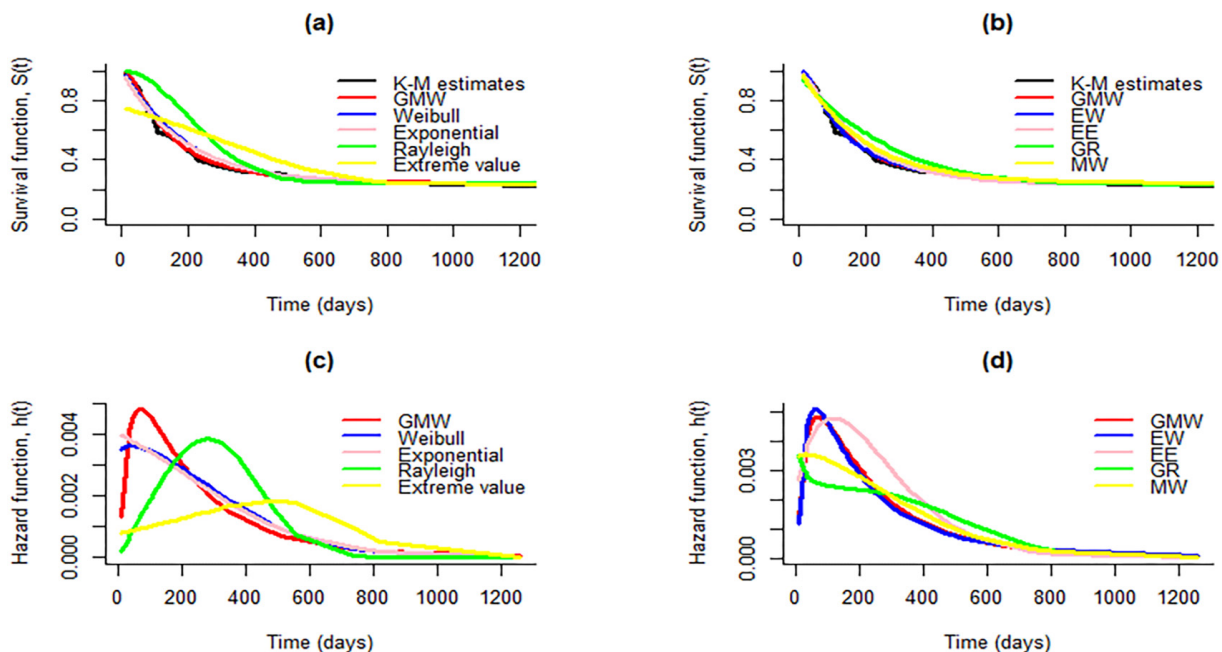


Figure 3. Fitted Survival Curves Obtained from the Mixture Model Based on the GMW Distribution and Its Sub-Distributions for the Leukemia Data (Panels (a) and (b)). Respective hazard functions are shown in panels (c) and (d). For comparisons, curves considering the GMW distribution are displayed in all plots.

Table 3. Maximum Likelihood Estimates, Considering the Mixture Model with the GMW Distribution where Acovariate Including in Cure Fraction  $\rho$  and in the Parameter of Shape  $\Psi$ .

Parameter	Estimate	95% Confidence interval	AIC*
$\epsilon$	0.004364	(0.00320, 0.00553)	970.9
$\phi$	-3.00E-05	(- 0.07801, 0.07794)	
$\phi_1$	-3.95E-05	(- 0.05267, 0.05259)	
$\eta$	1.03346	(0.99103, 1.07589)	
$\lambda$	3.35E-09	(- 9.121e-05, 9.122e-05)	
$\beta$	-0.97365	(-1.04205, - 0.90526)	
$\beta_1$	-0.46498	(- 0.56860, - 0.36137)	
$\rho$	0.27415	(0.26075, 0.28797)	
$\rho_1$	0.19175	(0.16650, 0.21983)	
$\rho_0/\rho_1$	1.42972	(1.18614, 1.72955)	

kind of transplant is incorporated both in the cure fraction  $\rho$  and the parameter of shape  $\psi$  (results from Table 3), are shown in Figure 4. The graph in this figure displays that the hazard functions are overlapping at the time of around 382 days after the transplants.

## Discussion

The present study aims to select a suitable distribution for survival times of susceptible leukemia patients. To achieve this goal, we proposed a mixture cure model based on a GMW and its sub-distributions. This model expands many distributions broadly utilized in the field of survival data analysis. The GMW distribution is adaptable to accommodate various forms of hazard rate functions, such as bathtub-shaped failure rates data.

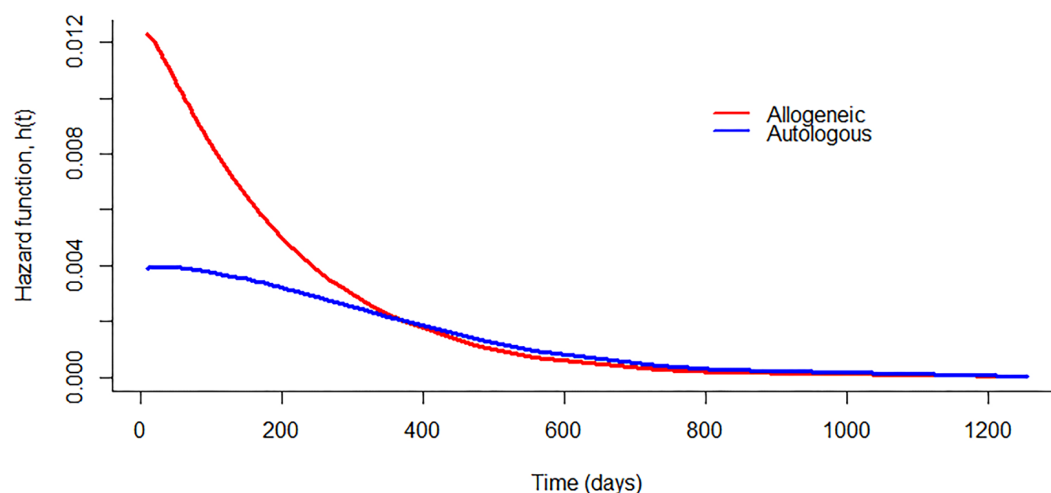


Figure 4. Risk Functions Obtained by the Mixture Cure Model Based on the GMW distribution where a Covariate (Kind of Transplant ) is Linked Both to the Cure Fraction  $\rho$  and in the Parameter of Shape  $\Psi$

Cure fraction models are developed to estimate the probability of being cured. In the absence of cured individuals, cure models can be reduced to the classical survival models. There are two major types of cure models. The first one is the mixture cure model, which is a common approach for modelling data with long-term survivors. In this model, the population is considered as a mixture of cured patients and uncured patients. The advantage of the mixture cure model is that it enables covariates to have different effects on cured patients and on survival times of susceptible patients. It is then possible to consider various covariates in the two parts of the model (incidence and latency) and assess the influence of the same covariate(s) on the two components. This property distinguishes the mixture cure model from the other cure models.

On the other hand, the mixture cure model cannot verify the property of proportional hazard functions. Besides, it does not appear to have a biological interpretation meaning, especially in the cancer recurrence. The other type of cure model is often known as the promotion time cure model or as the non-mixture cure model. It was first proposed by Yakovlev et al., (1993). This model assumes that after the cancer initial treatment, there are some cancer cells left inside a patient's body which may develop slowly over time and produce a discernible relapse of cancer. In some cases, the mixture cure model and the promotion time cure model are mathematically related to each other.

The graph of Kaplan-Meier survival curves in this study indicates that survival models that disregard the rate of cured subjects  $\rho$  would not be appropriate for the analysis of this data. In addition, this graph illustrates that the probability of being cured for the allogeneic group is better than the survival probability of the autologous group. Moreover, it displays stable plateaus at the right tail of each curve, which indicates that there may exist unsusceptible patients in the two treatment groups.

The present study found that the mixture cure models based on the EW and the GMW distributions display a better fit to the leukemia data since these models provide the closest anticipated values to the empirical values. However, in a different setting to the present study, Peng et al., (2001) proposed for the autologous group, the log-normal mixture cure model provides the best fit compared to exponential, Weibull, and gamma mixture cure models.

The results of the current study show that the estimated values for the cure proportion  $\rho$  obtained by the cure models based on the Weibull, exponential, Rayleigh, extreme value, exponentiated exponential, and generalized Rayleigh distributions are greater than the value of  $\rho$  shown in the panel (a) of Figure 1. Applying the models based on the EW and GMW distributions, one can obtain more accurate estimated values for the parameter  $\rho$ , and this means that we have extra proof of a good fit when assuming these models.

The curves of risk functions obtained by the EW and the GMW distributions are extremely near one another. These curves indicate that there is a high danger of death during the period instantly after the transplant. After this pinnacle, the hazard continues to decline until the end of the observation period.

In the present study, comparing between Table 1 and Table 2, we note that the AIC values provided by the fit of models excluding  $\rho$  are greater than the values of AIC obtained by the fit of models not excluding the cure rate. As anticipated, this also reveals that the mixture cure model is very convenient for analysing the data at hand.

Kutal et al., (2018) found that the estimated cure rates for the allogeneic group obtained by a mixture cure model based on Weibull and exponentiated exponential distributions are 0.239 and 0.242, respectively. Moreover, Lázaro et al., (2020) reported that the approximated cure fractions for allogeneic and autologous groups are 0.270 and 0.198, respectively. These results seemed relatively consistent with the present findings in terms of the values of the estimated cure fractions for the allogeneic and autologous patients groups.

It is interesting to note that the 95% confidence intervals for  $\rho_0/\rho_1$  (1.18614, 1.72955) do not contain the value 1, suggesting that there is proof of differences between the population cure rates presuming patients treated by allogeneic and autologous bone marrow transplants. Furthermore, the 95% confidence intervals for  $\beta_1$  (-0.56860, -0.36137) not include zero.

The results of this study reveal that the hazard functions considering the mixture cure model with the GMW distribution and the covariate is linked both to  $\rho$  and  $\psi$ , are not proportional. Thus, the Cox proportional hazard model approach is not appropriate for analysing this data. This result is in agreement with the findings of Kersey et al., (1987). The existence of cure individuals commonly is not presumed by the standard Cox model (Cox, 1972). Nevertheless, the literature presents several expansions of this model that consider cured subjects. However, these approaches are unsuitable for the evaluation of non-proportional risk functions.

The curves of the hazard functions have different shapes, indicating that applying the parametric models based on generalized probability distributions for the study of lifetime data like the data in hand can be very beneficial (Martinez et al., 2013). Because these models can incorporate numerous types of hazard rate functions.

It is worth mentioning that the zero belongs to the 95% confidence intervals for  $\phi_1$  (-0.05267, 0.05259) shown in Table 3, indicating no evidence of a significant difference between the shapes of population risk functions.

The findings of the present study show that the risk of doom is higher in the time around 11 days after the transplant assuming the patients that were exposed to allogeneic bone marrow transplants. While the hazard of dying is higher around 31 days after the transplant, presuming that the patients were treated by the autologous bone marrow transplants. Around 400 days after the transplantation, the hazard of death is almost the same.

In conclusion, parametric models incorporating a cure fraction with a specified distribution for survival times of susceptible subjects are appropriate instruments to analyse survival data with long-term survivors because these models do not presume hazards proportionality, and they can estimate measures that are readily construed by practitioners and health experts, as the fractions of immune individuals and the mean survival time. Mixture

cure model with the EW or the GMW distribution for survival times of uncured leukemia patients provides the best results compared with the Weibull, exponential, Rayleigh, extreme value, exponentiated exponential, generalized Rayleigh, and modified Weibull distributions.

## Author Contribution Statement

Mohamed Elamin Omer was involved in the conception and design of the study, data analyses, interpretation of results, and wrote the first draft of the manuscript. Mohd Rizam Abu Bakar was involved in the conception and design of the study, carried out the data analyses, interpretation of results, and drafting the manuscript. Mohd Bakri Adam and Mohd Shafie Mustafa were involved in the conception and design of the study, data acquisition, and interpretation of results. All authors read and approved the final manuscript and take responsibility for the work.

## Acknowledgments

### Ethical conduct of research

The study's dataset (bmt) is openly available in R software 3.6.2-Package 'smcure'.

### Conflict of Interest

The authors have no known conflict of interest.

## References

- Abu Bakar MR, Salah KA, Ibrahim NA, Haron K (2008). Cure fraction, modelling and estimating in a population-based cancer survival analysis. *Malaysian J Math Sci*, **2**, 113-34.
- American Cancer Society (2019). Cancer Facts and Figures. Retrieved from <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html>.
- Akaike H (1974). A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr*, **6**, 716-23.
- Basheer AM (2019). Alpha power inverse Weibull distribution with reliability application. *J Taibah Univ Sci*, **13**, 423-32.
- Berkson J, Gage RP (1952). Survival curves for cancer patients following treatment. *J Am Stat Assoc*, **47**, 501-15.
- Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc Series B Stat Methodol*, **11**, 15-44.
- Bradburn MJ, Clark TG, Love SB, Altman DG (2003a). Survival Analysis Part II: Multivariate data analysis - An introduction to concepts and methods. *Br J Cancer*, **89**, 431-6.
- Bradburn MJ, Clark TG, Love SB, Altman DG (2003b). Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. *Br J Cancer*, **89**, 605-11.
- Cai C, Zou Y, Peng Y, Zhang J (2012). smcure: An R-package for estimating semiparametric mixture cure models. *Comput Methods Programs Biomed*, **108**, 1255-60.
- Carrasco JMF, Ortega EMM, Cordeiro GM (2008). A generalized modified Weibull distribution for lifetime modeling. *Comput Stat Data Anal*, **53**, 450-62.
- Corbière F, Commenges D, Taylor JMG, Joly P (2009). A penalized likelihood approach for mixture cure models. *Stat Med*, **28**, 510-24.
- Cox DR (1972). Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*, **34**, 187-202.
- Farewell VT (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041-6.
- Ghitany ME, Maller RA (1992). Asymptotic results for exponential mixture models with long-term survivors. *Stat J Theor Appl Stat*, **23**, 321-36.
- Goldman AI (1984). Survivorship analysis when cure is a possibility: A Monte Carlo study. *Stat Med*, **3**, 153-63.
- Gupta RD, Kundu D (2001). Exponentiated exponential family: An alternative to Gamma and Weibull Distributions. *Biom J*, **43**, 117-30.
- Haybittle JL (1965). A two-parameter model for the survival curve of treated cancer patients. *J Am Stat Assoc*, **60**, 16-26.
- Kotz S, Nadarajah S (2000). Extreme Value Distributions: Theory and Applications. In 'Univariate Extreme Value Distributions', Imperial College Press, 57 Shelton Street, Convent Garden, London WC2H 9HE, pp 1-60.
- Kersey JH, Weisdorf D, Nesbit ME, et al (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *N Engl J Med*, **317**, 461-7.
- Kundu D, Raqab MZ (2005). Generalized Rayleigh distribution: different methods of estimation. *Comput Stat Data Anal*, **49**, 187-200.
- Kutal DH, Qinan L (2018). A non-mixture cure model for right-censored data with Fréchet Distribution. *Stats*, **1**, 176-88.
- Lai CD, Xie M, Murthy DNP (2003). A modified Weibull distribution. *IEEE Trans Reliab*, **52**, 33-7.
- Lázaro E, Armero C, Gómez-Rubio V (2020). Approximate Bayesian inference for mixture cure models. *TEST*, **29**, 750-67.
- López-Cheda A, Cao R, Jácome MA, Keilegom IV (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comput Stat Data Anal*, **105**, 144-65.
- Lu W (2010). Efficient estimation for an accelerated failure time model with cure fraction. *Stat Sin*, **20**, 661-74.
- Maller RA, Zhou S (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, 731-9.
- Martinez EZ, Achcar JA, Jácome AAA, Santos JS (2013). 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*, **112**, 343-55.
- Mudholkar GS, Srivastava DK (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans Reliab*, **42**, 299-302.
- National Cancer Institute (2020). SEER cancer statistics review, 1975-2017: Leukemia, annual incidence rates (acute lymphocytic leukemia). Retrieved from [https://seer.cancer.gov/csr/1975\\_2017/browse\\_csr.php?sectionSEL=13&pageSEL=sect\\_13\\_table.06](https://seer.cancer.gov/csr/1975_2017/browse_csr.php?sectionSEL=13&pageSEL=sect_13_table.06).
- Ortega EMM, Cordeiro GM, Kattan MW (2013). The log-beta Weibull regression model with application to predict recurrence of prostate cancer. *Stat Pap*, **54**, 113-32.
- Peng Y, Dear KBG, Denham JW (1998). A generalized F mixture model for cure rate estimation. *Stat Med*, **17**, 813-30.
- Peng Y, Dear KBG, Carriere KC (2001). Testing for the presence of cured patients: a simulation study. *Stat Med*, **20**, 1783-96.
- Sarhan AM, Zaindin M (2009). Modified Weibull distribution. *Appl Sci*, **11**, 123-36.
- Sy JP, Taylor JMG (2000). Estimation in a Cox Proportional Hazards Cure Model. *Biometrics*, **56**, 227-36.
- Terwilliger T, Abdul-Hay M (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J*, **7**, e577.



- Yakovlev AY, Asselain B, Bardou VJ, et al (1993). A simple stochastic model of tumour recurrence and its application to data on premenopausal breast cancer. In *Biometrie et analyse de donne'es spatio-temporelles*, **12**, pp 66-83.
- Yamaguchi K (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in Japan. *J Am Stat Assoc*, **87**, 284-92.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.