Indian Journal of Science and Technology, Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/97350, July 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

# Cure Models based on Weibull Distribution with and without Covariates using Right Censored Data

**Madaki Umar Yusuf[1]* and Mohd Rizam Abu Bakar[2]**

[1]Department of Mathematics and Statistics, Yobe State University, Nigeria; bomayu84@gmail.com
[1, 2]Department of Mathematics, Universiti Putra Malaysia Darul-Ehsan, Serdang Selangor, Selangor, Malaysia; mrizam@upm.my

## Abstract

In this paper we use a methodology based on the Weibull distributions covariates in the presence of cure fraction models, censored data and covariates. **Objective:** The objective of the study is to check the performance of mixture and non-mixture cure models based on LPML. **Methods/Analysis:** Two models were explored here in which are the mixture and non-mixture cure fraction models. Inferences for the models are obtained under the Bayesian approach via Markov Chain Monte Carlo (MCMC) where the posterior estimates were obtained by using Metropolis-Hastings sampling methods in the presence of covariates and without covariates considering a real life time dataset and comparing the two cure models using the Log Pseudo Maximum Likelihood estimates (LPML) and some related special cases of the distribution. **Findings/Conclusion:** We observed that the Weibull distribution has the least LPML value while its special cases where the two models are quite similar having the highest values on the other hand, the Mixture fits better than the non-mixture having the highest (LPML) based on the results obtain from all the models suggesting that the standard parametric cure (mixture) model fits the AML data which shows a great indication of similarity with the covariates and flexibility of the models.

**Keywords:** Bayesian Analysis, Cure Models, MCMC Algorithm, Right Censored Data, Survival Analysis, Weibull Distribution

## 1. Introduction

The cure fraction model[1,2] is also used to be called as an extension to the usual parametric survival models, to account for the fraction of individuals who will not experience the event of interest. It can also be called a long-term survival models regarding the kind of event been specified. The two most common cure models or long-term survivors are the mixture and non-mixture models[2,7]. Although, this model seems attractive and it is used widely in reliability and survival analysis areas.

## 2. Some Reviews

The long-term survival models play a very ideal role in the area of survival and reliability analysis. In recent years it is becoming more increasingly famous for analyzing data in clinical trials. The model had extensively suggested by authors like[23,25,27]. More recently statisticians and biometricians conducted some research based on the novelty of these models, [4] estimates some parameters in by Bounded Cumulative Hazard (BCH) model using Left censored data[3], proposed a model which they incorporates some regression covariates effects by non-mixture BCH model approach. [6]Proposed a method using a standard model for analyzing a clustered and interval censored survival time data by incorporating some fluctuation stochastic effects in both the PH and logistic components. [9]Demonstrated that the Beta-Weibull density can be expressed as a mixture of Weibull densities and provide some expressions for their MGF, while[28], proposed a new approach on mixture model via latent cure rate markers in the cure model context. [28]Suggested a mixture model

*Author for correspondence

with random effects to cause specific survival data of female breast cancer patients. They proposed two sets of random effects to capture the regional variation in the cure fraction and survival of uncured patients, respectively. [23]Proposed some ideas based on their notions which include the following:

- They suggested that when covariates are included in the analysis, on the other hand mixture model does not require a proportional hazard context.
- Mixture model provides improper posterior distributions for some types of non-informative improper priors when covariates are included through the parameter π via a standard regression model.
- Mixture model does not seem to outline the biological process for obtaining the failure time in a situation where cancer relapse is involved.

Parametric models, such as Weibull, Gompertz, exponential and many more can be employed. In general the main limitation of the parametric cure fraction models are sometimes hard to find a distribution flexible enough to fit the data. On this context, the non-parametric techniques are considered to be more attractive under the violation of parametric assumptions. It has been demonstrated by[10,16] that any parametric family of distribution can be incorporated into larger families through an application of the probability integral transform. In addition, the beta modified Weibull distribution by[6] is also a generalization of the Weibull distribution new family due to its flexibility in accommodating different forms of the risk function seems to be an important family that can be used in a variety of problems in modeling survival data. A Bayesian formulation of the cure fraction model is given by several authors as such[2,3]. A cure rate model based on the Beta-Weibull distribution was proposed by[6] techniques for estimation of cure rates when there are partially observed or missing covariates have been discussed by[1,3,4,7,11,12,14,33].

## 3. Methods

Let T be a random variable representing the time until the occurrence event of interest and let t > 0 be an observation from T. In the mixture cure fraction model which was first proposed by[7], the probability that time-to-event is larger than that some specified time t given by the survival function:

$$S(t) = p + (1-p)S_O(t), \qquad t > 0 \qquad (1)$$

Where p is the proportion of "long-term survivors" or cured patients", regarding the event of interest (0<p<1) and So(t) is the standard parametric survival curve for the susceptible ones who will undergo the event. Thus, the mixture cure fraction model assumes that at the beginning of the follow-up period there is a group of individuals who will never undergo the event. The cumulative distribution function for T is:

$$\begin{aligned} F(t) &= P(T \le t) = 1 - S(t) \\ &= (1-p)[(1-p)S_O(t)] \\ &= (1-p)F_O(t), \end{aligned} \qquad (2)$$

As such, the $\lim_{\top}(t \to \infty)\ [\![F_{\downarrow}O\ (t))]\!]$ indicates that $\lim_{t\to\infty} F(t) = 1 - p$. The probability density function (p.d.f) for T is:

$$f(t) = \frac{dF(t)}{dt} = (1-p)f_O(t), \qquad (3)$$

Where $f_o(t)$ is the baseline probability function for those who will undergo the event being specified. Let us assume right censored data and non-informative missing type. If we consider a simple random sample given as $(t_i, \delta_i)$ of size n, having i=1,....,n, for the contribution of the i-th subject for the likelihood function of the mixture model is given by:

$$\begin{aligned} L_i &= [f(t_i)]^{\delta_i}\,[S(t_i)]^{1-\delta_i} \\ &= [(1-p)[f_O(t_i)]]^{\delta_i}\,[p + (1-p)S_O(t_i)]^{1-\delta_i}, \end{aligned} \qquad (4)$$

Where $\delta_i$ is a censoring indicator variable, that is, $\delta_i = 1$ for an observed lifetime and $\delta_i = 0$ for a censored lifetime. Moreover, the log-likelihood function is:

$$\begin{aligned} L_1(\theta) &= x\ln(1-p) + \sum_{i=1}^{n} \delta i \ln S_O(t_i) \\ &+ \sum_{i=1}^{n} (1-\delta i)\ln[p + (1-p)So(t_i)], \end{aligned} \qquad (5)$$

Where $x = \sum_{i=1}^{n} \delta i$ is the number of uncensored observations. On the other hand, the non-mixture formulation

has been suggested by some author[2]. This model is define as an asymptote for the cumulative hazard and hence the cure fraction. In this case, the survival function for the non-mixture otherwise called Bounded Cummulative Hazard (BCH) model[3] is given by:

$$S(t) = pF_O(t_i) = \exp\big(\ln(p)\big)S_O(t_i), \quad (6)$$

Where $0 < p < 1$ is the probability of cured individuals and where $F_O(t) = 1 - S_O(t)$ with the hazard function:

$$h(t) = -\big(\ln(p)\big)f_O(t_i),$$

Note that the $\lim_{t\to\infty} F_O(t) = 1$ implies that $\lim_{t\to\infty} S(t) = p$.

$$L_i = [h(t_i)]^{\delta i}\,[S(t_i)],$$

$$= [-ln(p)[f_O(t_i)]]^{\delta i} \exp\big(\ln(p)\big)F_O(t_i), \quad (7)$$

Where the hazard function $\frac{h(t)\big(f(t)\big)}{S(t)}$ can be interpreted as the risk of an event immediate after time t conditional on surviving up until time t.

$$L_{II}(\theta) = x ln(-lnp) + \sum_{i=1}^{n} \delta i\ ln f_O(t)(t_i)$$

$$+(lnp)\sum_{i=1}^{n}[1 - S_O(t_i)], \quad (8)$$

Where $x = \sum_{i=1}^{n} \delta_i$ is the number of uncensored observations.

## 3.1 The Weibull Distribution

The Weibull distribution is a very well-known and famous distribution named after a great Swedish Physicist, mathematician, engineer and scientist Ernst Hjalmar Waloddi Weibull in 1951[25]. He used it in 1939 to analyze the breaking strengths of materials in Physics"[26]. The distribution has a great advantage over any parametric models over the usual Cox proportional hazards (semi-parametric) model is that these models are more informative and flexible. Because of its relative flexible property of hazard function and efficiency in estimating the parameters in the survival functions, ever since, it has been a widely general used for analyzing lifetime data analysis, modeling and simulations. However, only monotonically increasing and decreasing hazard functions can be generated from the classic two-parameter Weibull distribution[5]. It can be used to model devices with decreasing, constant, or increasing failure rate. This versatility is one of the main objectives for the wideness use of the Weibull distribution in survival, engineering and reliability analysis[3]. The procedure employed to analyze survival data include the Kaplan-Meier method for estimating the survival function, the log-rank test and the cox proportional hazard for more refined tools to be able to illustrate and analyze more complex related model[20]. In health studies, the use of parametric models for survival data analysis has been increasing in the recent years or decades in response to a need data structures, as such bivariate events times[25]. Using parametric models, it is straightforward process to derive the hazard functions and to obtain predicted survival times. Moreover, Cox model assumes that the ratio of the hazard functions of two different levels of a covariate is constant over time (proportional hazards) and this is not always achieved when we deal with real data. Parametric models can be more flexible to deal with non-proportional hazards. The Weibull distribution represents a generalization of some special cases as follows:

- Exponential distribution: A one-parameter exponential distribution[28] is a special case of Weibull distribution with $\gamma = 1$.
- The Beta distribution: A two-parameter distribution[16] is a special case of Weibull distribution with $\alpha = 1, \beta = 1$.
- The Rayleigh distribution: A one-parameter distribution[17] is a special case of Weibull distribution with $\alpha = 1$.
- The Beta-Weibull (BW) distribution: A four-parameter distribution[12] is a special case of Weibull distribution with $\alpha = 1, \beta = 1, \gamma = 1$ and $\lambda=1$.
- The Exponentiated-Weibull (EW) distribution: A three-parameter distribution[30] is a special case of Weibull distribution with $\beta = 1, \gamma = 1$ and $\lambda = 1$.
- The Beta-Exponential (BE) distribution: A three-parameter distribution[32] is a special case of Weibull distribution with $\alpha = 1, \beta = 1$and $\gamma = 1$.
- The Modified Weibull (MW) distribution: A three-parameter distribution[31] is a special case of Weibull distribution with $\alpha = 1, \gamma = 1$ and $\lambda = 1$.
- The Generalized Modified Weibull (GMW) distribution: A four-parameter distribution[29] is a

special case of Weibull distribution with α = 1, β = 1, γ = 1 and λ = 1.

As a special case of the distribution for those that will experience the event of interest by the baseline probability density function with two parameters given by:

$$f_O(t) = \gamma\lambda t^{\gamma-1}\exp[-\lambda t\gamma], \qquad (9)$$

The baseline survival function is given by:

$$S_O(t) = \exp[-\lambda t\gamma], \qquad (10)$$

Where the γ, λ > 0, that is, they are all positive. The corresponding hazard function is given by:

$$h_O(t) = \frac{f_O(t)}{S_O(t)}$$

$$= \frac{\gamma\lambda t^{\gamma-1}\exp[-\lambda t\gamma]}{\exp[-\lambda t\gamma]} = \gamma\lambda t^{\gamma-1}, \qquad (11)$$

## 3.2 The Log-Likelihood functions for the Susceptible Individuals

If we assume the mixture model (3), the log-likelihood function for θ = γ, λ, p, will be given as:

$$l_I(\theta) = x\ln(1-p) + x\ln\gamma + x\ln\lambda + (\gamma-1)w - M_1(\gamma) - M_2(p,\gamma,\lambda), \qquad (12)$$

where,

$$x = \sum_{i=1}^{n}\delta i$$

$$w = \sum_{i=1}^{n}\delta i\ ln(t_i)$$

$$M_1(\gamma) = \sum_{i=1}^{n}\delta i\ ti\gamma$$

$$M_2(p,\gamma,\lambda) = \sum_{i=1}^{n}(1-\delta i)\ln[p + (1-p)\exp[-\lambda t\gamma]]$$

Alternatively, if we assume the non-mixture model (6), on the other hand the log- likelihood function for p, γ, λ will be given as:

$$l_{II}(\theta) = x\ln(-\ln p) + x\ln\gamma + x\ln\lambda + (\gamma-1)w - \lambda M_1(\gamma) + (\ln p)M_2(\gamma,\lambda), \qquad (13)$$

where,

$$x = \sum_{i=1}^{n}\delta i$$

$$w = \sum_{i=1}^{n}\delta i\ ln(t_i)$$

$$M_1(\gamma) = \sum_{i=1}^{n}\delta i\ ti\gamma$$

$$M_2(p,\gamma,\lambda) = \sum_{i=1}^{n}[1-\exp[-\lambda t\gamma]]$$

Due to the intricate or complexity case of the likelihood functions $l_I(\theta)$ and $l_{II}(\theta)$, the estimation of the parameter by maximization or direct method will be extremely a difficult task. So to overcome in dealing with this type of problem, we consider the use of Bayesian Inference based on Markov Chain Monte Carlo (MCMC) methods. We can probably also take in the incorporate a vector of covariates Xi that may be closely related and associated with the proportion p of cure rate fraction models by replacing p in the log-likelihood function $l_I(\theta)$ and $l_{II}(\theta)$, by:

$$p_i = (\exp(x'\eta))/(1+\exp(x'\eta)), \qquad (14)$$

Where $x'_i = (1, x_{1i}, x_{2i}, \ldots, x_{Ji})$ is the vector of the observations of J covariates for the i-th individual and $\eta' = (1, \eta_1, \eta_2, \ldots, \eta_{1K})$ is the vector of unknown parameters. Furthermore, we can study the effect of the vector of covariates Wi on the parameter λ by replacing λ in the mixture and non-mixture log-likelihood functions that is, $l_I(\theta)$ and $l_{II}(\theta)$, given by:

$$\lambda_i = \exp(w_i'\zeta), \qquad (15)$$

Where $[w'] i = (1, w_1 1i, w_1 2i\ [\ldots, w] Ki$ is the vector of the observations of K covariates for the i-th individual and $\zeta' = (1, \zeta_1, \zeta_2, \ldots, \zeta_{1K})$ is the vector of unknown parameters.

# 4. Bayesian Analysis

We assume an earlier uniform U (0,1) prior distribution for the extent probability p of long-term survivors and the Gamma function having (0.001,0.001) for the scale parameter λ and the shape parameter γ, where the Gamma (a,b) denotes a gamma distribution with mean and variance where a and b are known hyper parameters. We likewise expect gamma to assume a prior distribution for the parameters p, γ, and λ. If we consider the covariates, using the non-informative prior distribution for the unknown parameters in the models.

### 4.1 Log Pseudo Marginal Likelihood Measure (LPML):

Comparison of mixture and non-mixture assuming different distributions was accessed by log Pseudo marginal likelihood measure. The LPML is derived from the Conditional Predictive Ordinate (CPO) statistics[13]. For the i-th observation, the $CPO_i$ is given by: An MCMC approximation of $CPO_i$ is given by:

$$\widehat{CPO} = \left[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{f\left(\frac{D_j}{\Theta_b}\right)}\right]^{1}, \qquad i = 1,\ldots..,n, \qquad (16)$$

$$f\left(\frac{D_i}{y[i]}\right) = \int f\left(\frac{D_i}{\Theta f\left(\frac{\Theta}{D_i}\right)d\Theta}\right)$$

Where Θ is the incomplete vector of parameters, Di is each instance of the full data D, $D_{[i]}$ is D without the current observation i and $f\left(\frac{\Theta}{D_i}\right)$ is the posterior density of Θ given $D_i$, $i = 1,\ldots,n$.

Where B is the number of iterations during the implementation of the MCMC procedure after the burn -in period and Θb is vector of the samples that will be obtained at the 4th to 5th iterations[15]. For a given model, the LPML value is given by:

$$LPML = \sum_{i=1}^{m}\log\widehat{CPO}, \qquad (17)$$

The larger the value of LPML, the better is fit of the model[15]. Alternatively, the Pseudo Bayes factor (PMF)[15] comparing models m and m' is:

$$PMF'_{mm} = \exp\left(\widehat{LPML}_m - \widehat{LPML}_{m'}\right). \qquad (18)$$

If we assume cure fraction models or the long-term survivors above, therefore we would consider the Gamma prior distribution function having Gamma (0.001, 0.001) for the regression parameters αo and βo, and a Gaussian prior distribution N (0, 100) for the regression parameters α1 and β1, l = 1,.....k. We likewise assume a prior independence distribution for the parameters. We consider Log Pseudo Maximum Likelihood (LPML), proposed by[15]. The LPML is an approximation of the Bayes factor and its objective is to incorporate the complexity of a model. LMPL is easy to calculate and applicable to a wide range of statistical models. It is based on the posterior distribution of the log-likelihood following the original suggestion of[15], for model choice in the Bayesian framework into the selection criterion. Metropolis-Hastings sampling algorithm which is very vital, flexible and versatile method for simulating intricate or non-intricate cases to sample an arbitrarily parametric class distributions. It is also use involving the Acceptance-Rejection sampling method, including the Gibbs algorithm serve as special cases of the Metropolis-Hastings sampling algorithm[8]. We also obtained the Highest Probability Density (HPD) intervals for parameters of interest[13]. A $100(1-\omega)\%$ HPD interval for a generic parameter θ is a subset of the parameter space C given by $C = \left\{\theta : \pi\left(\frac{\theta}{D}\right) \geq k\right\}$ where $\pi\left(\frac{\theta}{D}\right)$ is the posterior distribution for $\theta$ given the data D and k is the largest number such that $\int_{\pi(\theta p)\geq k}^{k}, \pi\left(\frac{\theta}{D}\right) = 1-\omega.$

## 5. Acute Myelogenous Leukemia (AML) Data

A right censored survival data of 33 patients from Acute Myelogenous Leukemia data was used[35]. Also measured was the patient's white blood cell count at the time of diagnosis. The patients were also factored into two (2) groups according to the presence or absence of a morphologic characteristic of white blood cells.

In the analysis we consider the cure fraction models introduced above in the presence or not of covariates in the models. So in the first analysis we assume the cure fraction models not considering any covariates in the models[14]. All cases in this work, we generated 30,000 samples for each parameter of interest given automatically by R software[24].

We assume a burn-in-sample of size 10,000 to minimize the effect of the initial values used in the simulation process. We simulated another 200,000 Gibbs a sample, taking every 100th sample to have approximately uncorrelated values, which gives the posterior summaries of interest was based on the final Gibbs sample of size 2,000 samples. The convergence of the approach algorithm was monitored using standard procedures, as well as the simulation process.

**Table 1.** The posterior summaries for the parameters of the models not including with covariate and considering the data set of 43 Acute Myelogenous Leukemia patients

| | Estimated Parameter | Posterior median | 95% H DP[a] | LP M L[b] | HW[c] p value | Geweke's p value |
|---|---|---|---|---|---|---|
| | γ | 2.1034 | (0.6784,1.2345) | -742.403 | 0.463 | 0.429 |
| Mixture | λ | -0.3865 | (-1.1256,0.6364) | | 0.4569 | 0.343 |
| | p | -0.2306 | (-0.4566,0.2381) | | 0.152 | 0.651 |
| | | | | | | |
| | γ | 1.9980 | (3.5975,5.2487) | -730.774 | 0.232 | 0.076 |
| Non-mixture | λ | -0.5640 | (-1.1376,0.3364) | | 0.099 | 0.223 |
| | p | -0.6876 | (0.0166,0.2681) | | 0.772 | 0.435 |

**Table 2.** The posterior summaries with covariate and considering the data set of 43 Acute Myelogenous Leukemia patients

| Cure Models | Estimated Parameter | Posterior median | 95% H DP[a] | LP M L[b] | HW[c] p value | Geweke p value |
|---|---|---|---|---|---|---|
| | γ | 1.0129 | (0.3684 , 2.1799) | -866.403 | 0.453 | 0.729 |
| | ζ0 | -0.3526 | (-1.1376, 0.3364) | | 0.069 | 0.533 |
| | ζ1 | -0.6706 | (-0.0166,-0.2681) | | 0.152 | 0.651 |
| Mixture | ζ2 | -0.9889 | (-0.5611, 0.2922) | | 0.148 | 0.307 |
| | η0 | -0.1337 | (-1.1376,0.3364) | | 0.799 | 0.343 |
| | η1 | -0.4843 | (-0.5416,0.2922) | | 0.618 | 0.707 |
| | η2 | -0.9124 | (-1.1376,-0.3364) | | 0.779 | 0.243 |
| | | | | | | |
| | γ | 0.9930 | (0.5975,11.2487) | -864.774 | 0.232 | 0.076 |
| | ζ0 | -0.5640 | (-1.2356,0.4534) | | 0.099 | 0.223 |
| | ζ1 | -0.6876 | (0.0756,0.2681) | | 0.372 | 0.353 |
| Nonmixture | ζ2 | -1.0025 | (0.4561,1.2722) | | 0.618 | 0.307 |
| | η0 | -0.1466 | (-2.1576,1.3767) | | 0.099 | 0.256 |
| | η1 | -0.4688 | (-0.0550,1.2368) | | 0.834 | 0.435 |
| | η2 | -0.9530 | (1.0250,0.1167) | | 0.568 | 0.872 |

## 5. Results

In Table 1 above, it shows the inference results considering Bayesian approaches using MCMC methods, we have used the Log Pseudo Maximum Likelihood estimates (LMPL) and given automatically by R software[24]. From the fitted survival models, we conclude that the Mixture and non-mixture models are very well fitted by the survival times with similarities in the intervals. From the results obtained above, we observed that the Bayesian inferences give similar results. Overall, Mixture model is the better fitted by the data (with larger Monte Carlo estimates for Log Pseudo Maximum Likelihood estimates (LMPL). While the Mixture model on the other hand (the first order approximation of non-mixture model) gives the larger value of LMPL and also the inference results considering a standard 2-parameter Weibull distribution not considering any covariates, where the Bayesian estimates where obtained by Markov Chain Monte Carlo (MCMC) using Metropolis-Hastings random walk sampling to obtain the posterior summaries.

In Table 2 above, the estimated parameters with covariates for the posterior summary of distribution was shown where Metropolis-Hastings samples drawn from

**Table 3.** The posterior summaries assuming mixture model and cure fraction without covariate considering the dataset of 43 Acute Myelogenous Leukemia patients

| Cure Models | Estimated Parameter | Posterior median | 95% H DPa | LP M Lb | HWc p value | Geweke's p value |
|---|---|---|---|---|---|---|
| Weibull | γ | 1.0129 | (0.3684,2.1799) | -756.403 | 0.453 | 0.729 |
| | λ | -1.0381 | (0.5337,1.6854) | | 0.338 | 0.607 |
| | p | -0.3526 | (-1.1376,0.3364) | | 0.069 | 0.533 |
| | | | | | | |
| Exponential | γ | 0. 9930 | (1.5975,0.8487) | -728.684 | 0.232 | 0.076 |
| | p | 1.0456 | (0.5611,1.2922) | | 0.618 | 0.307 |
| | | | | | | |
| Beta | α | | | | | |
| | β | 1.2006 | (0.0166,0.2681) | | 0.392 | 0.251 |
| | p | -0.4126 | (-1.1376,0.3364) | | 0.099 | 0.343 |
| | | | | | | |
| Rayleigh | γ | 1.5563 | (1.4012,1.8571) | -7654.557 | 0.244 | 0.265 |
| | p | -0.3126 | (-1.1376,4.3364) | | 0.023 | 0.223 |
| | | | | | | |
| BW | α | 0.9930 | (2.5975,3.2487) | -858.664 | 0.232 | 0.076 |
| | β | 1.0456 | (0.4611,1.3922) | | 0.632 | 0.267 |
| | γ | -0.5640 | (1.1376,0.3364) | | 0.199 | 0.323 |
| | λ | -0.6876 | (0.0166,0.2681) | | 0.372 | 0.351 |
| | p | -1.0025 | (0.5611,1.2922) | | 0.618 | 0.307 |
| | | | | | | |
| EW | β | 0.9456 | (0.2311,1.3922) | -875.521 | 0.932 | 0.247 |
| | γ | 1.0649 | (1,9664,5.2017) | | 0.483 | 0.347 |
| | λ | -0.3126 | (-0.1376,0.3364) | | 0.329 | 0.243 |
| | p | 1.2006 | (0.0166,0.2681) | | 0.392 | 0.251 |
| | | | | | | |
| BE | α | 1.6540 | (1.4782,1.5581) | -845.557 | 0.302 | 0.265 |
| | β | -0.3126 | (-1.1376,4.3364) | | 0.023 | 0.323 |
| | γ | -0.0806 | (-1.0166,-0.2881) | | 0.362 | 0.281 |
| | p | -1.9084 | (-1.5611,-0.4922) | | 0.618 | 0.237 |
| | | | | | | |
| MW | α | 0.9129 | (0.3684,2.1799) | -856.403 | 0.453 | 0.729 |
| | γ | 1.0381 | (0.5337,1.6754) | | 0.338 | 0.607 |
| | λ | 1.2932 | (0.1207,3.2681) | | 0.224 | 0.251 |
| | p | -0.3526 | (-1.1376,0.3364) | | 0.069 | 0.533 |
| | | | | | | |
| GMW | α | 0.9433 | (2.5645,2.2423) | -878.723 | 0.356 | 0.081 |
| | β | 1.0726 | (0.533,1.2354) | | 0.532 | 0.275 |
| | γ | -0.3640 | (1.2476,0.5404) | | 0.327 | 0.428 |
| | λ | -0.2536 | (0.0264,0.3720) | | 0.153 | 0.629 |
| | p | -1.0130 | (0.4631,1.4512) | | 0.428 | 0.327 |

**Table 4.** The posterior summaries, without cure fraction or covariates in the special cases

| Cure Models | Estimated Parameter | Posterior median | 95% H DPa | LP M Lb | HWc p value | Geweke's p value |
|---|---|---|---|---|---|---|
| Weibull | γ | 1.0129 | (0.3684,2.1799) | -726.403 | 0.453 | 0.639 |
| | λ | -1.0381 | (0.5337,1.6854) | | 0.338 | 0.607 |
| | | | | | | |
| Exponential | γ | 0. 9930 | (1.5975,0.8487) | -708.684 | 0.232 | 0.076 |
| | | | | | | |
| Beta | α | 1.0649 | (1.9664,1.2017) | -786.561 | 0.483 | 0.023 |
| | β | 1.2006 | (0.0166,0.2681) | | 0.392 | 0.251 |
| | | | | | | |
| Rayleigh | γ | 1.5563 | (1.4012,1.8571) | -745.877 | 0.244 | 0.265 |
| | | | | | | |
| BW | α | 0.9930 | (2.5975,3.2487) | -835.674 | 0.232 | 0.076 |
| | β | 1.0456 | (0.4611,1.3922) | | 0.632 | 0.267 |
| | γ | -0.5640 | (1.1376,0.3364) | | 0.199 | 0.323 |
| | λ | -0.6876 | (0.0166,0.2681) | | 0.372 | 0.351 |
| | | | | | | |
| EW | β | 1.0456 | (0.4611,1.3922) | -823.521 | 0.632 | 0.267 |
| | γ | 1.0649 | (1,9664,5.2017) | | 0.483 | 0.347 |
| | λ | -0.3126 | (-0.1376,0.3364) | | 0.329 | 0.243 |
| | | | | | | |
| BE | α | 1.6540 | (1.4782,1.5581) | -845.557 | 0.302 | 0.265 |
| | β | -0.3126 | (-1.1376,4.3364) | | 0.023 | 0.323 |
| | γ | -0.0806 | (-1.0166,-0.2881) | | 0.362 | 0.281 |
| | | | | | | |
| MW | α | 1.0129 | (0.3684,2.1799) | -856.403 | 0.453 | 0.729 |
| | γ | 1.0381 | (0.5337,1.6854) | | 0.338 | 0.607 |
| | λ | 1.2932 | (0.1107,3.2681) | | 0.224 | 0.251 |
| | | | | | | |
| GMW | α | 0.9433 | (2.5645,2.2423) | -878.723 | 0.356 | 0.081 |
| | β | 1.0726 | (0.533,1.2354) | | 0.618 | 0.307 |
| | γ | 0.5640 | (1.1376,0.3364) | | 0.099 | 0.223 |
| | λ | -0.6876 | ((0.0166,0.2681) | | 0.372 | 0.351 |

**Table 5.**

| Cure Models | Estimated Parameter | Posterior median | 95% H DPa | LP M Lb | HWc p value | Geweke's p value |
|---|---|---|---|---|---|---|
| Weibull | γ | 1.0129 | (0.3684,2.1799) | -845.403 | 0.453 | 0.729 |
| | ζ0 | -0.3826 | (- 0.9376,0.3764) | | 0.569 | 0.933 |
| | ζ1 | -0.3716 | (-0.3466,0.2681) | | 0.152 | 0.651 |
| | ζ2 | -0.9889 | (-0.4645,-0.7324) | | 0.648 | 0.757 |
| | η0 | -0.1337 | (-1.3476,0.5240) | | 0.799 | 0.343 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | η1 | -0.4843 | (-0.5416,0.2922) | | 0.618 | 0.707 |
| | η2 | -0.9124 | (-1.1376,-0.4564) | | 0.779 | 0.245 |
| | | | | | | |
| Exponential | γ | 0.9930 | (4.5975,11.2487) | -828.774 | 0.232 | 0.076 |
| | η0 | -0.1466 | (-2.1576,1.3767) | | 0.099 | 0.256 |
| | η1 | -0.4688 | (-0.0550,1.2368) | | 0.834 | 0.435 |
| | η2 | -0.9830 | (2.1180,0.1287) | | 0.768 | 0.645 |
| | | | | | | |
| Beta | α | 1.0649 | (1.9364,4.2317) | -785.521 | 0.483 | 0.007 |
| | β | 1.2006 | (0.0166,0.2681) | | 0.392 | 0.251 |
| | η0 | -0.1466 | (-2.1576,1.3767) | | 0.099 | 0.256 |
| | η1 | -0.4688 | (-0.0550,1.2368) | | 0.834 | 0.435 |
| | η2 | -0.9530 | (1.0250,0.1167) | | 0.568 | 0.872 |
| | | | | | | |
| Rayleigh | γ | 1.654 | (1.4712,1.8571) | -785.557 | 0.101 | 0.265 |
| | η0 | -0.1466 | (-2.1576,1.3767) | | 0.099 | 0.256 |
| | η1 | -0.4688 | (-0.0550,1.2368) | | 0.834 | 0.435 |
| | η2 | -0.9530 | (1.0250,0.1167) | | 0.568 | 0.872 |
| | | | | | | |
| BW | α | 0.9840 | (4.5975,11.2487) | -858.774 | 0.232 | 0.076 |
| | β | 1.0456 | (0.5611,1.2922) | | 0.618 | 0.307 |
| | γ | -0.5640 | (1.2360,0.4564) | | 0.099 | 0.223 |
| | ζ0 | -0.3526 | (-1.4617,0.2340) | | 0.069 | 0.533 |
| | ζ1 | -0.6706 | (-0.0166,0.2681) | | 0.152 | 0.651 |
| | ζ2 | -0.9383 | (-0.5611,-0.2352) | | 0.305 | 0.148 |
| | η0 | -0.1337 | (0.9876,0.3634) | | 0.799 | 0.343 |
| | η1 | -0.4843 | (-0.325,0.6783) | | 0.418 | 0.607 |
| | η2 | -0.9530 | (1.0250,0.7667) | | 0.568 | 0.872 |
| | | | | | | |
| EW | β | 1.0649 | (1,9664,5.2017) | -835.521 | 0.483 | 0.007 |
| | γ | -0.3126 | (-0.1376,0.3364) | | 0.099 | 0.243 |
| | ζ0 | -0.3526 | (-1.1376,0.3364) | | 0.069 | 0.533 |
| | ζ1 | -0.6706 | (-0.0166,0.2681) | | 0.152 | 0.651 |
| | ζ2 | -0.9889 | (-0.5611,-0.2922) | | 0.148 | 0.307 |
| | η0 | -0.1337 | (-1.1376,0.3364) | | 0.799 | 0.343 |
| | η1 | -0.4843 | (-0.5416,0.2922) | | 0.618 | 0.707 |
| | η2 | -0.9124 | (-1.1976,-0.3364) | | 0.479 | 0.487 |
| | | | | | | |
| BE | α | 1.654 | (1.4712,1.8571) | -825.557 | 0.113 | 0.265 |
| | β | -0.3126 | (-1.1376,4.3364) | | 0.023 | 0.223 |
| | γ | -0.0806 | (-1.0166,-0.2881) | | 0.362 | 0.281 |
| | η0 | -0.1466 | (-2.1576,1.3767) | | 0.099 | 0.256 |
| | η1 | -0.4688 | (-0.0550,1.2368) | | 0.834 | 0.435 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | η2 | -0.9530 | (1.0250,0.1167) | | 0.568 | 0.872 |
| | | | | | | |
| MW | α | 1.0129 | (0.3684,2.1799) | -866.403 | 0.453 | 0.729 |
| | γ | 1.0381 | (0.5337,1.6854) | | 0.338 | 0.607 |
| | ζ0 | -0.3526 | (-1.1376,0.3364) | | 0.069 | 0.533 |
| | ζ1 | -0.6706 | (-0.0166,0.2681) | | 0.152 | 0.651 |
| | ζ2 | -0.9889 | (-0.5611,-0.2922) | | 0.148 | 0.307 |
| | η0 | -0.1337 | (-1.1456,0.3364) | | 0.799 | 0.343 |
| | η1 | -0.4843 | (-0.5416,0.2922) | | 0.638 | 0.817 |
| | η2 | -0.9124 | (-1.1376,-0.3364) | | 0.779 | 0.243 |
| | | | | | | |
| GMW | α | 0.9240 | (4.5975,11.2687) | -856.764 | 0.232 | 0.176 |
| | β | 1.0256 | (0.5611,1.4922) | | 0.618 | 0.507 |
| | γ | -0.5440 | (1.2760,0.4574) | | 0.129 | 0.343 |
| | ζ0 | -0.3526 | (-1.4617,0.2640) | | 0.069 | 0.553 |
| | ζ1 | -0.6706 | (-0.0166,0.2581) | | 0.152 | 0.851 |
| | ζ2 | -0.9383 | (-0.2611,-0.3652) | | 0.205 | 0.458 |
| | η0 | -0.1337 | (0.8876,0.3334) | | 0.699 | 0.543 |
| | η1 | -0.4843 | (-0.725,0.6483) | | 0.518 | 0.607 |
| | η2 | -0.9530 | (1.0450,0.7567) | | 0.668 | 0.853 |

the joint posterior distribution and we used the results of the median over the mean, for this reason where by most of the special cases of the distribution were skewed in nature. Also we noted that all p values from (HW) Heidelberger and Welch convergence diagnostic criteria do not reject the null hypothesis of stationary of the chains, since they are all larger or equal than 0.10. While the Geweke's p values also suggest convergence, on the other hand these results shows that, among all the models considered Weibull distribution has the least (LPML) Log Pseudo Marginal Likelihood value, while GMW, BW, MW, BE and EW distributions have similar LPML values, where a strong evidence shows that these models are better fitted by the data than the model based on the standard Beta, Exponential, Weibull and Rayleigh distributions fitting the distributions in the presence of cure fraction for the[23].

Figure 1 shows the graph plots of the Kaplan-Meier estimates for the survival function versus the respective predict values obtained from the parametric mixture models for each probability distribution of interest: 1. Weibull, 2. Expo-nential, 3. Beta, 4. Rayleigh, 5. BW, 6. EW, 7. BE, 8. MW and GMW distributions for (Acute Myelogenous Leukaemia (AML) dataset). It clearly shows that by having a diagonal straight lines in the plots this represents a well perfect relationship and agreement between Kaplan-Meier estimates and predicted values.

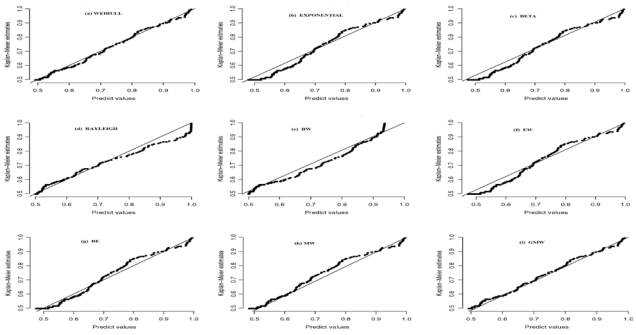

**Figure 1.** Shows the graph plots of the Kaplan-Meier estimates.

The Table 3 below shows the inferences for the posterior summaries assuming the mixture model and cure fraction based on the Weibull distribution and its special cases, among all the models considered Weibull distribution has the least (LPML) Log Pseudo Marginal Likelihood value, while GMW, BW, MW, BE and EW distributions have similar LPML values, where a strong evidence shows

that these models are better fitted by the data than the model based on the standard Beta, Exponential, Weibull and Rayleigh distributions fitting the distributions in the presence of cure fraction for the[23], without including the cure fraction p shows great indication of similarity with Table 4in terms of the LMPL results which proves a better fit in the results[20]. We ignored the non-mixture tables as both output were similar.

The Table 4 below shows the results in summary of the posterior for the models based on the Weibull distribution and its special cases, without including the cure fraction p. The GMW, BW, MW, BE, EW and Beta distributions fitted of models not including p gives larger LMPL values compare with the fit of models including the cure fraction this shows that the cure fraction models are more are suitable, flexible and straight to model Acute Myelogenous Leukemia study data.

From Table5 below they clearly show that the posterior summary for the mixture and non-mixture-models based on the Weibull distribution and its special cases where the two models are quite similar but the Mixture models fits better having the highest (LPML) based on the results obtain. Also, the 95% credible intervals for η2 do not include the zero value suggesting that the AML data shows a great indication of similarity with the covariates. We also obtained Bayesian estimates for the cure fractions of each risk group considering the simulated samples for η0, η1 and η2 and the relations. The groups have different cure fractions. We can obtain Bayesian estimates for the cure fractions of each risk group considering the simulated samples for η0, η1 and η2 and the relations.

## 6. Conclusion:

In the life time data analysis, where we have the presence of cure fraction and covariates for a data with this structure can be appropriately analyzed using different parametrical formulations, as mixture and non-mixture models. In this paper, we showed that parametric models based on the BW distribution and its special cases can be useful to analyze medical data sets. We adopt the limitation of the BW distribution is that the survival function cannot be expressed in a closed form10. This problem can become more critical when we consider covariates, since the likelihood function then becomes more complex. Data with this structure can be appropriately analyzed using different parametrical formulations, as mixture and non-mixture models. In this context of Bayesian analysis using the Markov Chain Monte Carlo (MCMC) methods, we obtained the LMPL above and the corresponding credible intervals and the posterior means, which were automatically generated by R software on the fitted survival models, we interpret and conclude the Mixture and non-mixture Models that they were very much clear fitted and suited the survival statistical models considering the results obtain from Tables and figure. As a layman statistician we observe that the values were similar and closer to each other likewise the LMPL results. We also showed that the use Bayesian methodology is a way to get the inferences for the parameters of the model, where the model estimation is facilitated by the use of the R package MCMC pack. An advantage of Bayesian approach over other methods is that it explicitly incorporates the prior opinion for the parameters when more covariates were considered. In clinical applications, the knowledge of a specialist on the proportion of patients who are to the event can be incorporated into a prior distribution for the cure fraction p, results to a more relatively inferences.

## 7. References


1. Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: An alternative to two-component mixture model. J Amer Statist Assoc. 2003; 98(464):1063–78.
2. Achcar JA, Coelho-Barros EA, Mazucheli J. Cure fraction models using mixture and non-mixture models. Tatra Mountains Mathematical Publications. 2012; 51:1–9.
3. Ibrahim NA, Taweab F, Arasan J. A parametric non-mixture cure survival model with censored data. Computational Problems in Engineering. Springer International Publishing. 2014; 307:231–8.
4. Aljawadi B, Bakar MRA, Ibrahim NA, Midi H. Parametric estimation of the cure fraction based on BCH Model using left-censored data with covariates. Modern Applied Science. 2011; 5(3):103–10.
5. Bakar MRA, et al. Bayesian approach for joint longitudinal and time-to-event data with survival fraction. Bulletin of the Malaysian Mathematical Sciences Society. 2009; 32(1):75–100.
6. Chen MH, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association. 1999; 94:909–19.
7. Berkson J, Gage R. Survival curve for cancer patients following treatment. Journal of American Statistical Association. 1952; 47:501–15.

8. Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. The American Statistician. 1995; 49(4):327–35.
9. Cordeiro GM,Carrasco JMF, Ortega EM. A generalized modified Weibull distribution for lifetime modelling. Statistical Computation Anals. 2008; 53(2):450–62.
10. Cordeiro GM, Nadarajah S, Ortega EM. General results for the beta Weibull distribution. Journal of Statistical Computation and Simulation. 2013; 83(6):1082–114.
11. Copelan EA, Biggs JC, Thompson JM, Crilley P, Szer J, Klein JP, Atkinson K. Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with BuCy2. NCBI. 1991; 78(3):838–43.
12. Famoye F, Lee C, Olumolade O. The beta Weibull distribution. J Stat Theory Appl. 2005; 4:121–36.
13. Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. 1992. No. TR-462.
14. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. CRC Press; 2003.
15. Geisser S, Eddy WF. A predictive approach to model selection. Journal of the American Statistical Association. 1979; 74(365):153–60.
16. Gupta AK, Nadarajah S. Editors. Handbook of beta distribution and its applications. CRC Press. 2004.
17. Irvine T. The Rayleigh distribution. 2012.
18. Kim S, Xi Y, Chen MH. A new latent cure rate marker model for survival data. The Annals of Applied Statistics. 2009; 3(3):1124–46.
19. Lee C, Famoye F, Olumolade O. BetaWeibull distribution: Some properties and applications to censored data. J Mod Appl Stat Methods. 2007; 6(1):173–86.
20. Martinez EZ, Achcar JA, Jcome AA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. Computer Methods and Programs in Biomedicine. 2013; 112(3):343–55.
21. Mudholkar GS, Srivastava DK, Kollia GD. A generalization of the Weibull distribution with application to the analysis of survival data. Journal of American Statistical Association. 1996; 91:1575–83.
22. Eugene N, Lee C, Famoye F. Beta-normal distribution and its applications. Commun Stat Theory Methods. 2002; 31(4):497–512.
23. Seppa K, Hakulinen T, Kim HJ, Laara E. Cure fraction model with random effects for regional variation in cancer survival. Statistics in Medicine. 2010; 29(27):2781–93.
24. R software. Available from: http://CRAN.R-project.org/package
25. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. Springer; 2002.
26. Weibull WA. A statistical distribution of wide applicability. Journal of Applied Mechanics. 1951; 18:293–7.
27. Wahed AS, Luong TM, Jeong JH. A new generalization of Weibull distribution with application to a breast cancer data set. Statistics in Medicine. 2009; 28(16):2077–94.
28. Xiang L, Ma X, Yau KK. Mixture cure model with random effects for clustered and interval censored survival data. NCBI. 2011; 30(9):995–1006.
29. Carrasco JMF, Ortega EMM, Cordeiro GM. A generalized modified Weibull distribution for lifetime modeling. Computational Statistics and Data Analysis. 2008; 53:450–62.
30. Pal M, Ali MM, Woo J. Exponentiated Weibull distribution Statistica. 2006; 66 (2):139–47.
31. Sarhan AM, Zaindin M. Modified Weibull distribution. Applied Sciences. 2009; 11:123–36.
32. Nadarajah S, Kotz S. The beta exponential distribution. Reliability Engineering and System Safety. 2006; 91(6):689–97.
33. Yusuf MU, Bakar MRBA. A Bayesian estimation on right censored survival data with mixture and non-mixture cured fraction model based on Beta-Weibull distribution. In Innovations through Mathematical and Statistical Research: Proceedings of the 2nd International Conference on Mathematical Sciences and Statistics (ICMSS2016) (Vol. 1739, No. 1, p. 020079).,AIP Publishing. June 2016.