

## RESEARCH ARTICLE

# Bayesian Method for Modeling Male Breast Cancer Survival Data

Hafiz Mohammad Rafiqullah Khan<sup>1\*</sup>, Anshul Saxena<sup>2</sup>, Sagar Rana<sup>3</sup>, Nasar Uddin Ahmed<sup>4</sup>

### Abstract

**Background:** With recent progress in health science administration, a huge amount of data has been collected from thousands of subjects. Statistical and computational techniques are very necessary to understand such data and to make valid scientific conclusions. The purpose of this paper was to develop a statistical probability model and to predict future survival times for male breast cancer patients who were diagnosed in the USA during 1973-2009. **Materials and Methods:** A random sample of 500 male patients was selected from the Surveillance Epidemiology and End Results (SEER) database. The survival times for the male patients were used to derive the statistical probability model. To measure the goodness of fit tests, the model building criterions: Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC) were employed. A novel Bayesian method was used to derive the posterior density function for the parameters and the predictive inference for future survival times from the exponentiated Weibull model, assuming that the observed breast cancer survival data follow such type of model. The Markov chain Monte Carlo method was used to determine the inference for the parameters. **Results:** The summary results of certain demographic and socio-economic variables are reported. It was found that the exponentiated Weibull model fits the male survival data. Statistical inferences of the posterior parameters are presented. Mean predictive survival times, 95% predictive intervals, predictive skewness and kurtosis were obtained. **Conclusions:** The findings will hopefully be useful in treatment planning, healthcare resource allocation, and may motivate future research on breast cancer related survival issues.

**Keywords:** Breast cancer survival data - statistical inference - Bayesian inference - exponentiated exponential

*Asian Pac J Cancer Prev*, 15 (2), 663-669

### Introduction

According to the World Health Organization (WHO), breast cancer accounts for 519,000 deaths worldwide. Surprisingly, 69% of those deaths were in developing countries, refuting the belief that breast cancer is a disease of developed countries (WHO, 2004). The symptoms of breast cancer include presence of a lump, which is commonly pea sized, in or near the breast or in the underarm that changes the appearance of the overlying skin and causes pain (Merck Manual of Diagnosis and Therapy, 2012). The complications of the breast cancer include chest pain, mastitis, metastasis, nipple discharge, and side effects of radiation therapy and chemotherapy (Hurria, 2012). The metastasis of the breast cancer occurs via hematogenous spread or lymphatic spread, generally to lung, pleura, liver, bone, adrenal glands, gastro intestinal track, pancreas and peritoneum (Mülle et al., 2001).

Early diagnosis of breast cancer is very important as its signs and symptoms appear in the later stages of the

disease. Early diagnosis ensures early management of the breast cancer and decreases the mortality associated with the disease. Male breast cancers occur in patients experiencing gynecomastia due to hormonal problems and overabundance of estrogen precursors. These are usually due to side effects of therapeutic drugs, hormonal imbalance during puberty or side effects of androgenic and anabolic steroid abuse. According to the National Cancer Institute estimates for 2013, about 2,240 males are diagnosed annually and 410 males die of breast cancer in US (ACS, 2013). We are focusing on the survival estimations for male breast cancer patients because of paucity of statistical studies for male survival times.

In recent years, the healthcare industry collected an unprecedented amount of phenomic data (clinical, imaging, biochemical, cellular), and genomic data (DNA sequencing and microarray analysis), from thousands of subjects which are not fully utilized by the researchers and are not even available for public usage. These large databases require a new direction of statistical analysis

<sup>1</sup>Department of Biostatistics, <sup>2</sup>Department of Health Promotion and Disease Prevention, <sup>4</sup>Department of Epidemiology, Robert Stempel College of Public Health and Social Work, <sup>3</sup>Division of Research Information and Data Coordinating Center, Florida International University, Florida, USA \*For correspondence: [hmkhan@fiu.edu](mailto:hmkhan@fiu.edu)

and the implementation of newer analytical approaches to identify data based statistical probability distributions and to draw a scientific conclusion about future situations of disease patterns and death rates.

The data extracted from health related lab/experiments may follow several statistical probability distributions including exponentiated exponential and exponentiated Weibull. Statistical and computational techniques are immensely necessary to understand such informative data and to make scientific conclusions that are reliable.

The Exponentiated Exponential Model (EEM) is frequently used in modeling the data for engineering and biomedical sciences. The EEM is a generalization of the exponential distribution, which was introduced by Gupta and Kundu (1999) and received tremendous and widespread attention. The EEM has two parameters (scale and shape). A random variable  $x$  is said to have the exponentiated exponential distribution if its probability the density function (pdf) is given by

$$p(x|\alpha, \lambda) = \begin{cases} \alpha \lambda \exp\{-\lambda x\} (1 - \exp\{-\lambda x\})^{\alpha-1}, & x \geq 0; \alpha, \lambda > 0, \\ 0 & \text{elsewhere,} \end{cases} \quad (1)$$

where  $x > 0, \lambda > 0$  and  $\alpha > 0$ . We can write  $x \sim EE(\alpha, \lambda)$ . The distribution of Eqn. (1) was introduced by Gupta and Kundu (1999). In the particular, if  $\alpha = 1$ , is the exponential distribution.

In 1939, Waladdi Weibull, a Swedish physicist, introduced Weibull distribution. In 1993, Mudholkar and Srivastava presented the first exponentiated Weibull model (EWM) that contained distributions with bathtub shaped and unimodal failure rates. Since then, it has been extensively used for analyzing the lifetime data by Nassar and Eissa (2003), and Choudhury (2005).

The probability density function (pdf) for the exponentiated Weibull model is given by

$$p(x|\alpha, \lambda, \beta) = \begin{cases} \alpha \beta \lambda x^{\beta-1} \exp\{-\lambda x^\beta\} (1 - \exp\{-\lambda x^\beta\})^{\alpha-1}, & x \geq 0; \alpha, \beta, \lambda > 0, \\ 0 & \text{elsewhere,} \end{cases} \quad (2)$$

where  $\alpha$  and  $\beta$  are the shape parameters, and  $\lambda$  is the scale parameter respectively.

Statistical inference to predict future outcome using past observations is known as predictive inference. Most of the studies on EEM and EWM have addressed the estimation of parameters and reliability, and hazard functions but, there are not enough studies which are related to the posterior distribution of the parameters and predictive inference for future outcome. The derivation of predictive inferences is necessary for future advances in biomedical science.

Setting up a full probability model	Calculating posterior distribution	Evaluating the fitness of the model
<ul style="list-style-type: none"> <li>• Comprises all observable as well as nonobservable quantities in a problem</li> <li>• Model should be consistent with underlying scientific problem</li> </ul>	<ul style="list-style-type: none"> <li>• The conditional probability distribution of unobserved quantities given the observed data</li> </ul>	<ul style="list-style-type: none"> <li>• Check the fitness of model with the data</li> <li>• Are conclusion are reasonable?</li> <li>• If necessary, expand the model and repeat all the steps</li> </ul>

Figure 1. The Process of Bayesian Data Analysis, Gelman (2004)

The purpose of Bayesian inference is to develop the posterior distribution of the parameters given a set of observed data. According to Gelman (2004), the Bayesian method can be divided into three components.

For further information regarding Bayesian method, the readers are referred to Berger (1985), Geisser (1993), Bernardo and Smith (1994), Ahsanullah and Ahmed (2001), Gelman et al. (2004), and Baklizi (2004; 2008). Additional applications of Bayesian method for predictive inference have been discussed by Raqab (1995; 2009), Thabane (1998), Thabane and Haq (2000), Khan et al. (2004; 2013), Ali-Mousa and Al-Sagheer (2006), Baghestani et al. (2009), and Khan (2012a; 2012b; 2013a; 2013b).

The main goals of this paper are to: *i*) study some demographic and socio-economic variables; *ii*) review right skewed models EEM and EWM; *iii*) display different scenarios of the EEM and EWM by changing the parameter; *iv*) a justification to prove that the given sample data follows the EEM and EWM by using model selection criteria for goodness of fit tests; *v*) draw a Bayesian analysis of the posterior distribution of the parameters; *vi*) derive Bayesian predictive model for future survival time from the EEM and EWM; *vii*) utilize the predictive models in the breast cancer survival data sets to obtain the predictive inference for future response and the likelihood of males getting breast cancer and; *viii*) to obtain the predictive intervals for the future survival times.

This paper is organized as follows: A real breast cancer survival data example is discussed in details in section 2. Measures of goodness of fit tests and log-likelihood functions for both models are presented in Section 3. Section 4 addresses the Bayesian predictive model including the likelihood function, posterior density function, predictive density for response given a sample of observations from the EWM. In section 5 the results and discussion are included. The last section, which is 6, gives the concluding remarks.

## Materials and Methods

### A real data example

We formally requested and acquired the breast cancer patients' data (N=657,712) from the Surveillance, Epidemiology and End Results (SEER; 1973-2009) website. SEER data contain breast cancer patients' information mostly on 12 states in the USA. A representative probability random sampling scheme was employed to draw a sample from the randomly selected nine states to represent race categories.

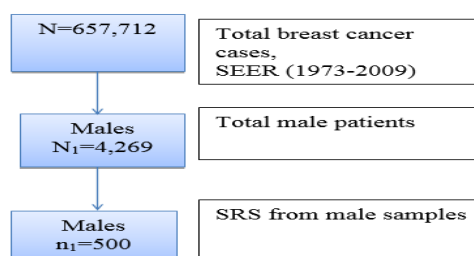
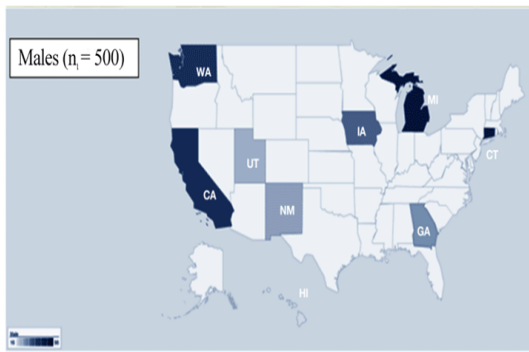


Figure 2. Selection of Male Patients from SEER (1973-2009)



**Figure 3. Number of Male ( $n_1=500$ ) Patients Randomly Selected from Nine States (darker color represents a high density and lighter color is the low density) of Breast Cancer Patients**

**Table 1. Frequency Distribution of Selected Male Breast Cancer Patients**

State	N	%
California	70	14
Connecticut	94	18.8
Georgia	42	8.4
Hawaii	16	3.2
Iowa	52	10.4
Michigan	98	19.6
New Mexico	33	6.6
Utah	27	5.4
Washington	68	13.6
Total	500	100

**Table 2. Age at Diagnosis and Survival Time of Male Breast Cancer Patients, Classified by Mean, Standard Deviation (SD), Median, Range, and Quartiles, SEER (1973-2009)**

Characteristics	Categories	Male
Age at diagnosis (years)	Mean	66.19
	SD	12.45
	Median	67
	Range	30-99
	Quartiles 1/2/3	57/67/75
Survival time (months)	Mean	84.14
	SD	75.25
	Median	65
	Range	1-407
	Quartiles	28/65/119

\*N1=4,269,  $n_1=500$

**Table 3. Race, Ethnicity, and Marital Status of Male Breast Cancer Patients**

Characteristics	Categories	Male
Race	White	436
	Black	45
	Other	19
Hispanic Origin	Non-Hispanic	483
	Hispanic	17
Marital status at diagnosis	Single	54
	Married	334
	Divorced/Separated	33
	Widowed	57
	Unknown	22

This data was split according to gender (where males=4,269) and then a simple random sampling (SRS) method was used to select a sample of size 500 male breast cancer patients.

*Measures of goodness of fit*

The main methods currently used by several researchers to compare various models goodness of fit are: Akaike Information Criterion (AIC), Deviance Information Criterion (DIC), and Bayesian Information Criterion (BIC). The widely used method, DIC, is a Bayesian measure of fit which is used for overall comparison of different models, for example, public data by Congdon (2005; 2007). It shows how good the model predictions fit the given data, while it represents the complexity of fitness given each model of the data. Akaike (1973) generalized his work over factor analysis by introducing information criterion which later became popular as Akaike's information criterion or AIC. Bayesian information criterion (BIC) is an asymptotic result assumed that the data distribution is an exponential family and can only be used to compare estimated models when numerical values of the dependent variable are identical for all estimates being compared. The model with lower value of BIC is preferred over others. AIC, BIC, and DIC values are reported in Table 4 for the EEM and EWM based on the male survival times.

Assuming the data  $x=(x_1, x_2, \dots, x_n)$  represents  $n$  male breast cancer patients survival times, then one may obtain the log-likelihood function from the EWM specified by (2) which is given by

$$\text{Log } L((\alpha, \beta, \lambda)|x) = n \log(\alpha) + n \log(\beta) + n \log(\lambda) + (\alpha - 1) \sum_{i=1}^n \log(1 - \exp\{-\lambda x_i^\alpha\}) - \lambda \sum_{i=1}^n x_i^\alpha + (\beta - 1) \sum_{i=1}^n \log(x_i)$$

Ahmed et al. (2008) used a new reparameterization method for the Birnbaum-Saunders lifetime model. Later, Achcar et al. (2013) considered a reparameterization some skewed models. One may utilize a reparameterization method by considering the log-likelihood functions from the EWM that is described as follows: Assume  $\rho_1 = \log(\alpha)$ ;  $\rho_2 = \log(\beta)$ ; and  $\rho_3 = \log(\lambda)$ . Furthermore, we assume that  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  are independently distributed. To obtain non-informative prior for  $\rho_1$ ,  $\rho_2$  and  $\rho_3$ , let a uniform prior distribution for  $\rho_i$  be  $U(-a_i, a_i)$ ,  $\forall_{i=1,2,3}$ . Then the joint posterior density is given by

$$p(\rho_1, \rho_2, \rho_3|x) = p(\rho_1, \rho_2, \rho_3) \times \exp\{n(\rho_1 + \rho_2 + \rho_3) - e^{\rho_3} \sum_{i=1}^n x_i^{e^{\rho_2}} + (e^{\rho_1} - 1) \sum_{i=1}^n \log(1 - \exp\{-e^{\rho_3} x_i^{e^{\rho_2}}\}) + (e^{\rho_2} - 1) \sum_{i=1}^n \log(x_i)\}$$

Similarly, one may obtain the log-likelihood function from the EEM (1) or from the log-likelihood function of the EWM (when  $\beta=1$ ) and may use the reparameterization method to derive the joint posterior density for  $\rho_1$  and  $\rho_2$ .

By using the reparameterization method, one would obtain better performance of the posterior distributions for the parameters and their results are reported in Table 5. The posterior kernel densities for the parameters are given in Figure 5.

Table 4 consists of AIC, BIC, and DIC values for the EE and EW models. This is a common way to test the goodness of fit models. Lower values of AIC, BIC, and

DIC infer better model fit of the data. Here, goodness of fit of survival times of male patients is being tested. In case of males, the data fits the EW distribution better than the EE distribution. The estimated value of BIC is the highest (12235.73) while the DIC value is the lowest (12225.30) for males in the case of EEM. The estimated value of BIC is the largest (5540.67) while the DIC value is the least (5530.30) for males in the case of EWM. Comparing the estimated values of all AIC, BIC, and DIC based on the EEM and EWM in the case of males, the EWM fits better for the male survival times because it produces smaller values of AIC, BIC, and DIC.

Figure 4A portrays the different scenarios of the EE model where each graph is composed of three superimposed models with certain parameters. Figure 4A shows scenarios of the EEM with some values of the parameters. The left graph in Figure 4A contains three inverted J-shaped graphs that have a fixed  $\alpha$  and increasing  $\lambda$  parameters. The red line graph shows more variability compared to the green and purple line graphs. The right graph in Figure 4A, contains three right skewed graphs when  $\alpha$  is increasing and  $\lambda$  is fixed. In the right graph,  $\lambda$  is kept constant while values of  $\alpha$  is increasing. It is observed that the green line is showing more variability as compared to the purple and red lines.

Figure 4B portrays the different scenarios of the EW model where each graph is composed of four superimposed models with certain parameters. The graphical scenario of EW models where in the left graph,  $\alpha$  is kept fixed with increased values of  $\beta$  and  $\lambda$ . The left graph displays three approximately normally distributed graphs, and one positively skewed graph. In the right graph,  $\lambda$  is kept fixed with decreased values of  $\alpha$  and  $\beta$ . The right graph displays three approximately normally distributed graphs where one is very wide, and one positively skewed graph. In Figure 4B left graph, it is noted that the red line is showing more variability compared to the other three normal graphs and in the right graph, the orange line is more variable as

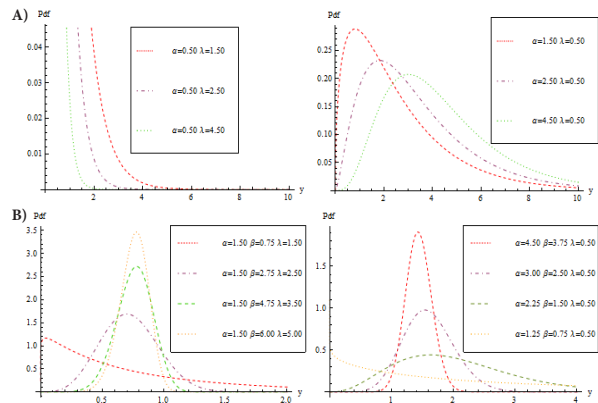


Figure 4. Scenarios of the A) EEM and B) EWM with Some Values of the Parameters

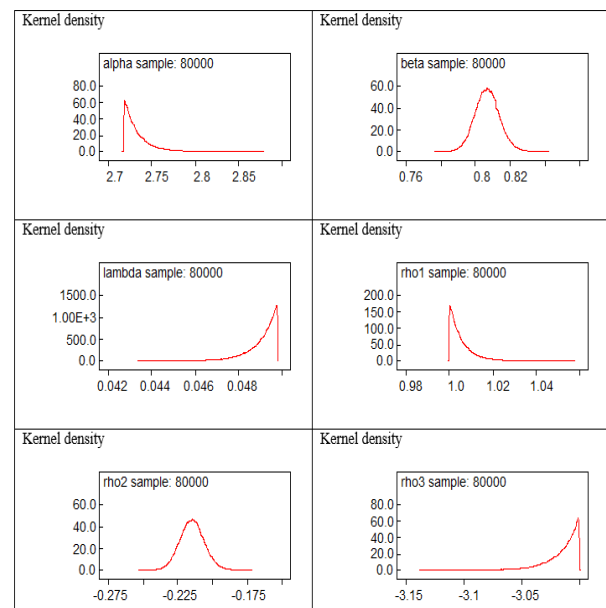


Figure 5. Kernel Density of the Posterior Parameters in the Case of Exponentiated Weibull for Male Breast Cancer Patients ( $n_1=500$ )

Table 4. Selection of the Exponentiated Exponential and Exponentiated Weibull Models for Male on the Basis of AIC, BIC, and DIC Criteria

Model Criteria:	EEM	EWM
AIC	12227.3	5532.25
BIC	12235.73	5540.67
DIC	12225.3	5530.3

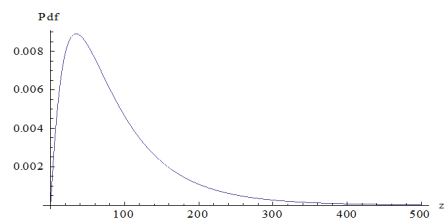


Figure 6. Predictive Density for a Single Future Survival Time

Table 5. Summary Results of the Posterior Parameters in the Case of Exponentiated Weibull for Male Breast Cancer Patients

Node	Mean	SD	MC error	Median	95%CI	Start	Sample
alpha	2.733	0.01425	8.62E-05	2.728	(2.719, 2.771)	1001	80000
beta	0.8069	0.007026	3.33E-05	0.8068	(0.7934, 0.821)	1001	80000
lambda	0.04908	6.87E-04	4.10E-06	0.04929	(0.04724, 0.04977)	1001	80000
rho1	1.005	0.00519	3.14E-05	1.004	(1, 1.019)	1001	80000
rho2	-0.2146	0.008704	4.12E-05	-0.2147	(-0.2314, -0.1972)	1001	80000
rho3	-3.014	0.01419	8.46E-05	-3.01	(-3.052, -3)	1001	80000

\*Dbar=post.mean of  $-2\log L$ ; Dhat= $-2\log L$  at post.mean of stochastic nodes; \*\* $\rho_1=(1, 2)$ ,  $\rho_2=(-1, 1)$ ,  $\rho_3=(-4, -3)$

	Dbar	Dhat	pD	DIC
zeros	5529.27	5528.25	1.029	5530.3
total	5529.27	5528.25	1.029	5530.3

compared to the other three graphs.

In Bayesian approach, the knowledge of the distribution of the parameters is updated through the use of the observed data, resulting in what is known as the posterior distribution of the parameters. In the case of breast cancer data, we are interested in estimating the posterior distribution of the parameters assuming that observed random variables form an appropriate theoretical probability distribution.

Table 4 indicates the summary results of the posterior distribution of the parameters from the exponentiated Weibull by using the breast cancer males' patient data. By setting the values of the  $\rho_1, \rho_2$  and  $\rho_3$  the results of the posterior distribution parameters  $\alpha, \beta$ , and  $\lambda$  are estimated using the MCMC method. Markov chain Monte Carlo is a class of algorithms used in statistics for generating samples from a probability distribution (Gilks et al., 1996). The log-likelihood function is derived from the EWM and then by its parameter values which are assigned to appropriate probability distribution. The WinBugs software is used to obtain the summary results (Mean, SD, MC Error, Median, and Confidence Intervals) of the parameters. The early iterations are ignored in order to remove any biases of estimated values of the parameters resulting from the survival values of  $x$  utilized to initialize the chain, a process that is called burn-in. After removing the burn-in samples, the remaining samples are treated as if the samples are from the original distribution. The procedure was conducted by 80,000 Monte Carlo repetitions to produce the inference for the posterior parameters in Table 5.

Figure 5 displays the graphical representation of the parameters' behavior in the case of exponentiated Weibull given the male survival data. It is noted that the shape parameter  $\beta$  plays approximately symmetric distribution and other model parameters are following skewed distributions.

#### The Bayesian predictive survival model

The Bayesian predictive method is growing more popular, finding new practical applications in the fields of health sciences, environmental sciences, and social sciences, among others. The Bayesian predictive approach which is used for the design and analysis of survival research studies in the health sciences is now widely used to reduce healthcare cost and to successfully allocate health care resources.

In this section, a predictive survival model for the

**Table 6. Predictive Inference Based on 500 Male Survival Data Points**

Summary statistics	Predictive intervals	
Mean=88.237	90%: (3.960938, 201.724312)	
SE=3.17231	95%: (2.145168, 249.461554)	
	98%: (1.605262, 290.309872)	
	99%: (1.427024, 320.156072)	
*Raw moments	Corrected moments	Skewness and kurtosis
$m_1=88.2397$	$\mu_1=m_1=88.2397$	$\beta_1=2.90855$
$m_2=12818.10$	$\mu_2=5031.80$	$\beta_2=6.83539$
$m_3=2.6278 \times 10^6$	$\mu_3=608728$	$\gamma_1=1.70545$
$m_4=6.8362 \times 10^8$	$\mu_4=1.73065 \times 10^8$	$\gamma_2=3.83539$

breast cancer patients is developed by using a novel Bayesian method. It is found that the male cancer patients' data follow the EWM by using AIC, BIC, and DIC.

Assuming the data  $x=(x_1, \dots, x_n)$  represents  $n$  male breast cancer patients survival times that follow the EWM, let  $z$  be a future response (or future survival days), then the predictive density of  $z$  given the observed data  $x$  is

$$p(z|x) = \int p(z|\alpha, \beta, \lambda) p(\alpha, \beta, \lambda | x) d\alpha, d\beta, d\lambda,$$

where  $p(\alpha, \beta, \lambda | x)$  is the posterior density function, and  $p(z|\alpha, \beta, \lambda)$  represents the probability density function of a future response ( $z$ ) that may be defined from model (2). The posterior density is given by

$$p(\alpha, \beta, \lambda | x) \propto L(\alpha, \beta, \lambda | x) p(\alpha, \beta, \lambda),$$

where  $L(\alpha, \beta, \lambda | x)$  is the likelihood function,  $p(\alpha, \beta, \lambda)$  is the prior density for the parameters.

To derive the likelihood function, let  $x_1, \dots, x_n$  be a random sample of size  $n$  from model (2). Thus,  $x=(x_1, \dots, x_n)$ ' forms an observed sample. Then given a set of data  $x=(x_1, \dots, x_n)$  from (2), the likelihood function is given by

$$L(\alpha, \beta, \lambda | x) \propto (\alpha\beta\lambda)^n \exp\left\{-\sum_{i=1}^n (-\lambda x_i^\beta)\right\} \left[\prod_{i=1}^n (x_i^{\beta-1})\right] \left[\prod_{i=1}^n (1-\exp(-\lambda x_i^\beta))\right]^{\alpha-1}$$

Ahmed (1992) discussed an estimation theory under uncertain prior information. Ahsanullah and Ahmed (2001) discussed in details on Bayes and empirical Bayes estimates of survival and hazard functions of a class of distribution. Ahmed and Tomkins (1995) estimated lognormal mean by making use of uncertain prior information. Khan et al. (2004) derived the Bayesian predictive model from the Weibull life model by means of a conjugate prior for the scale parameter and a uniform prior for the shape parameter. It is assumed that the prior density for the scale parameter ( $\lambda$ ) is given by

$$p(\lambda) \propto \lambda \exp\{-\lambda\}, \lambda > 0, \quad (3)$$

Considering Khan et al. (2004), the shape parameters,  $\alpha$  and  $\beta$ , have a uniform prior over the interval  $(0, \alpha)$  and  $(0, \beta)$ , respectively, which is given below:

$$p(\alpha, \beta) \propto 1/\alpha\beta, \alpha, \beta > 0, \quad (4)$$

Thus, the joint prior density is

$$p(\alpha, \beta, \lambda) \propto \lambda \exp\{-\lambda\}/\alpha\beta, \alpha, \beta, \lambda > 0, \quad (5)$$

Considering the prior density in (5), the posterior density of  $\alpha, \beta$ , and  $\lambda$  is given by

$$p(\alpha, \beta, \lambda) \propto (\alpha\beta)^{n-1} \lambda^{n+1} \exp\left\{-\sum_{i=1}^n (\lambda x_i^\beta)^2 - \lambda\right\} \left[\prod_{i=1}^n (x_i^{\beta-1})\right] \left[\prod_{i=1}^n (1-\exp(-\lambda x_i^\beta))\right]^{\alpha-1}$$

A numerical integration command 'NIntegrate' in conjunction with the symbolic computational software Mathematica version 8.0, Wolfram Research (2012), is applied to plot the predictive density graph. The predictive means, standard deviations, predictive intervals, and the measures of skewness and kurtosis are obtained. The Mathematica package is also utilized to carry out all related calculations.

Figure 6 shows the graphical representation of the

predictive density based on the male breast cancer patients' survival days. It is noted that the predictive density formed right skewed model.

The summary results of male predictive means, standard errors, and predictive intervals for a future survival day are given in Table 6. The predictive shape characteristics, raw moments, corrected moments, and measures of skewness and kurtosis are also presented in the same table.

These findings are very important towards health care researchers and providers to characterize future disease patterns and to make an effective future plan in our health industry.

## Results

We used a random sample consisting of 500 breast cancer male patients diagnosed during 1973-2009 in the USA. The mean $\pm$ SD, of age at diagnosis for breast cancer patients was 66.19 $\pm$ 12.45 for males. The minimum age at diagnosis for males 30 years. In the sample, oldest males at the time of the diagnosis was 99 years. The mean $\pm$ SD, of survival time for male was 84.14 $\pm$ 75.25 months. For males, the survival time ranged from 1 to 407 months. In this sample, among 500 males, race was distributed as 87% White, 9% Black, and 4% other among male cases. There were nearly 3% people of Hispanic origin. The majority of these patients were married. The basic statistical calculations helped us to conclude that the breast cancer in male is diagnosed at a higher age. The late diagnosis makes survival time look shorter after diagnosis.

In the case of the goodness of fit analysis, the breast cancer data from male sample followed exponentiated Weibull distribution with lower DIC value as 5530.30. For EW distribution, mean $\pm$ SD values for  $\alpha$ ,  $\beta$  and  $\lambda$  are 2.73 $\pm$ 0.01, 0.81 $\pm$ 0.01 and 0.05 $\pm$ 6.87 $\times 10^{-4}$ , respectively. Rho values for males are  $\rho_1=(1, 2)$ ,  $\rho_2=(-1, 1)$ ,  $\rho=(-4, -3)$ . The Markov Chain errors and confidence intervals for posterior parameters are given in Table 5. The dynamic kernel densities for each of the parameters are reported in Figure 5 so that one can observe the shape of the kernel density.

Figure 6 shows the graphical representation of the single future survival time. The shape of the future distribution of survival times is positively skewed. The raw and corrected moments are described in Table 6 based on males' survival times.

## Discussion

A Statistical probability modeling has become essential for the modern practice of medicine, pharmaceutical, and clinical studies, the effective future direction of health care, and the health profession of education. Statistical probability models can describe the nature of data from any scientific field. From such statistical models, one can make statistical inferences about the parameters as well as the future observations because they play a very important role in decision making for future patterns of diseases in human health.

In this study, we selected a random sample of male

breast cancer patients from the SEER (1973-2009) database. The methods for measuring the goodness of fit tests are used to select the best statistical probability models for the male based on breast cancer survival data. The sample is composed of male data from nine randomly selected states out of 12 states in the USA. To develop statistical probability models for survival days of males, we used model selection criteria, AIC, BIC, and DIC to measure the best fit to the breast cancer survival data. We found that the EWM best fits the male survival data. A detail analysis of the posterior models for the parameters is described with their summary results.

A novel reparameterization method is performed in terms of the log-likelihood functions for the EW models to accelerate better performance of the Bayesian posterior parameters and to draw its corresponding dynamic kernel densities. The summary results of the posterior parameters are reported with very less MC errors which are negligible by using the MCMC method. The results are obtained after running 80,000 Monte Carlo repetitions. The results of the posterior distribution of parameters using the breast cancer patients' data will contribute a new addition to the knowledge of the model parameters.

A summary table for the predictive means, standard deviations, and the predictive intervals are obtained for the predictive density of a single future survival time. The measures of skewness and kurtosis of a future response from the predictive model are also given. Based on the results of skewness and kurtosis one would comment that the shape of the future survival model for the males is positively skewed. These findings will be extremely helpful for the healthcare researchers and providers to predict a male patient's possible future medical outcome given the patient's current and past history of reported conditions.

Descriptive statistics are obtained by using the SPSS software. An advanced computational software package, 'Mathematica version 8.0', is used to show the graphical representations of the EE and the EW models, to display the predictive density and also to achieve additional predictive inferences for male survival times. WinBugs software is used to check the goodness of fit tests, to obtain the summary results of the posterior parameters, to determine the kernel densities of the parameters, and also to carry out all related calculations.

## References

- Achcar JA, Coelho-Barros EA, Cordeiro GM (2013). Beta distributions and related exponential models: a Bayesian approach. *Braz J Probab Stat*, **27**, 1-19.
- American Cancer Society (2013). Cancer Facts and Figures 2013. Atlanta: Am Can Soc. Available online Exit Disclaimer.
- Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In *Breakthroughs in Statistics*, pp. 610-624. Springer New York.
- Ahmed SE, Budsaba K, Lisawadi S, Volodin AI (2008). Parametric estimation for the Birnbaum-Saunders lifetime distribution based on a new parameterization. *J Thai Statistician*, **6**, 213-40.
- Ahmed SE (1992). Asymptotic estimation of reliability in a life-testing model. *J Ind Math Soc*, **41**, 7-18.

- Ahmed SE, Tomkins RJ (1995). Estimating lognormal mean using uncertain prior information. *Pak J Stat*, **11**, 67-92.
- Ali-Mousa MAM, Al-Sagheer SA (2006). Statistical inference for the Rayleigh model based on progressively type-II censored data. *Statistics*, **40**, 149-57.
- Ahsanullah M, Ahmed SE (2001). Bayes and empirical Bayes estimates of survival and hazard functions of a class of distributions. *Lect Notes Stat*, **148**, 81-7.
- Baghestani AR, Hajizadeh E, Fatemi SR (2009). Bayesian analysis for survival of patients with gastric cancer in Iran. *Asian Pac J Cancer Prev*, **10**, 823-6.
- Baklizi A (2004). Preliminary test estimation in the Rayleigh distribution using minimax regret significance levels. *Pertanika J Sci & Techno*, **12**, 45-50.
- Baklizi A (2008). Estimation of  $P(X \leq Y)$  using record values in the one and two parameter exponential distributions. *Commun Stat-Theor M*, **37**, 692-8.
- Berger JO (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- Bernardo JM, Smith AFM (1994). *Bayesian Theory*. Wiley, New York.
- Choudhury A (2005). A simple derivation of moments of the exponentiated Weibull distribution. *Metrika*, **62**, 17-22.
- Congdon P (2005). Bayesian predictive model comparison via parallel sampling. *Comput Stat Data Anal*, **48**, 735-53.
- Congdon P (2007). Mixtures of spatial and unstructured effects for spatially discontinuous health outcomes. *Comput Stat Data Anal*, **51**, 3197-212.
- Geisser S (1993). *Predictive Inference: An Introduction*. Chapman and Hall, New York.
- Gelman (2004). Parameterization and Bayesian modeling. *J Am Stat Assoc*, **99**, 537-45.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, New York.
- Gilks WR, Richardson S, Spiegelhalter DJ (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Suffolk.
- Gupta RD, Kundu D (1999). Generalized exponential distributions. *Aust NZ J Stat*, **41**, 173-88.
- Huria A (2012). Patterns of relapse and long-term complications of therapy in breast cancer survivors. UpToDate.
- Khan HMR (2012a). Estimating predictive inference for responses from the generalized rayleigh model based on complete sample. *J Thai Statistician*, **10**, 53-68.
- Khan HMR (2012b). Several priors based inference from the exponential model under censoring. *JP J Fund Appl Stat*, **2**, 1-13.
- Khan HMR (2013a). Comparing relative efficiency from a right skewed model. *JP J Biostat*, **9**, 1-26.
- Khan HMR (2013b). Inferential estimates from the one-parameter half-normal model. *J Thai Statistician*, **11**, 77-95.
- Khan HMR, Albatineh AN, Kheradia TN (2013). Local predictive estimates towards MLEs from the generalized half-normal model. *J Stat Adv Theo Appl*, **9**, 73-98.
- Khan HMR, Haq MS, Provost SB (2004). Predictive distributions for responses from the Weibull life testing model. *J Stat Theo Appl*, **3**, 53-73.
- Merck Manual of Diagnosis and Therapy (2012). Merckmanuals.
- Mudholkar GS, Srivastava DK (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans Rel*, **42**, 299-302.
- Mülle A, Homey B, Soto H, et al (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature*, **410**, 50-6.
- Nassar MM, Eissa FH (2003). On the exponentiated Weibull distribution. *Commun Stat-Theor M*, **32**, 1317-36.
- National Cancer Institute (2012). What You Need To Know About™ Cancer: Understanding Cancer.
- Raqab MZ (1995). On the maximum likelihood prediction of the exponential distribution based on double type II censored samples. *Pak J Stat*, **11**, 1-10.
- Raqab MZ (2009). Distribution-free prediction interval for the future current record statistics. *Stat Papers*, **50**, 429-39.
- Thabane L (1998). *Contributions to Bayesian Statistical Inference*. Ph.D. Thesis, the University of Western Ontario, Canada.
- Thabane L, Haq MS (2000). Bayesian prediction from a compound statistical model: An actual application. *Pak J Stat*, **16**, 110-25.
- Weibull W (1939). A statistical theory of the strength of materials. Available online Exit Disclaimer.
- World Health Organization (2004). *World health report 2004: Global Burden of Disease*. Geneva, World Health Organization.
- Wolfram Research (2012). *The Mathematica Archive: Mathematica 8.0*, Wolfram Research Inc, Illinois.