RESEARCH ARTICLE

Model-Based Survival Estimates of Female Breast Cancer Data

Hafiz Mohammad Rafiqullah Khan[1]*, Anshul Saxena[2], Kemesha Gabbidon[2], Sagar Rana[3], Nasar Uddin Ahmed[4]

Abstract

Background: Statistical methods are very important to precisely measure breast cancer patient survival times for healthcare management. Previous studies considered basic statistics to measure survival times without incorporating statistical modeling strategies. The objective of this study was to develop a data-based statistical probability model from the female breast cancer patients' survival times by using the Bayesian approach to predict future inferences of survival times. Materials and Methods: A random sample of 500 female patients was selected from the Surveillance Epidemiology and End Results cancer registry database. For goodness of fit, the standard model building criteria were used. The Bayesian approach is used to obtain the predictive survival times from the data-based Exponentiated Exponential Model. Markov Chain Monte Carlo method was used to obtain the summary results for predictive inference. Results: The highest number of female breast cancer patients was found in California and the lowest in New Mexico. The majority of them were married. The mean (SD) age at diagnosis (in years) was 60.92 (14.92). The mean (SD) survival time (in months) for female patients was 90.33 (83.10). The Exponentiated Exponential Model found better fits for the female survival times compared to the Exponentiated Weibull Model. The Bayesian method is used to obtain predictive inference for future survival times. Conclusions: The findings with the proposed modeling strategy will assist healthcare researchers and providers to precisely predict future survival estimates as the recent growing challenges of analyzing healthcare data have created new demand for model-based survival estimates. The application of Bayesian will produce precise estimates of future survival times.

Keywords: Bayesian method - model development - predictive inference

Introduction

Globally, cancer has become a leading cause of death in both developed and developing nations, accounting for nearly 20% of all cause-specific mortality (WHO, 2013a). It is the most common cause of death among economically developed countries and the second most common cause of death among developing nations (Jemal et al., 2011). The survival rate for cancer in developing countries lags behind that of developed nations due to numerous factors that diminish life expectancy; these include poverty, poor health care delivery, and late stage diagnosis (Jemal et al., 2011). In the year 2012, cancer accounted for an estimated 8.2 million deaths worldwide and approximately 14.1 million new cases. About 32.6 million people were living with cancer, diagnosed within 5 years of 2012. Of these, breast cancer alone was responsible for 522,000 deaths and 1.67 million new cases, making it the most prevalent type of cancer. Furthermore, breast cancer is the deadliest cancer among females aged 20-59 years worldwide (Ferlay et al., 2013; WHO, 2013b).

Female breast cancer accounts for approximately 25% of the total cancer cases and 14.3% of all cancer deaths. Projections for new breast cancer cases (all ages) suggest an 18% worldwide increase in 2020 as compared to 2012. In the United States (U.S.), the projected number of new cases for the year 2020 is 14.4% more than the estimated new cases in 2012 (Ferlay et al., 2013). In addition, a recent study reported a significant increasing trend of overall annual percentage change (APC) in breast cancer incidence rate from 2000 to 2009 in the U.S. Among women aged 40-49 years, the overall APC was around 1.1% (p <0.001) for most race and ethnic categories (Hou and Huo, 2013). With regards to race and ethnicity, white women face a higher incidence of breast cancer than any other racial or ethnic group at 417 cases per 100,000, while black women reported the highest mortality rates associated with breast cancer, demonstrating 171 cases per 100,000 (White et al., 2013).

In addition to being the most common type of cancer worldwide, breast cancer is also the most common cancer among women in the U.S.; accounting for an estimated

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

226,870 new cases and 39,510 deaths in the year 2012 (Siegel et al., 2012; ACS, 2013; USCS, 2013). In 2010, the total cost associated with breast cancer was around $16.5 billion as compared to the estimated $124.57 billion total national expenditure on cancer care. Projected cost scenarios for the year 2020 suggest medical expenditures approximating $23.24 billion, making it the second largest overall increase (32%) in medical expenditures related to cancer care. The quality of life and healthcare costs are related to the length of survival of breast cancer patients. (Mariotto et al., 2011; Montero et al., 2012).

Approximately 2.6 million U.S. women with a history of breast cancer were alive (survived) in January 2008, more than half of whom were diagnosed less than 10 years earlier (ACS, 2012). Moreover, 1 in 8 women in the United States can expect to develop breast cancer over the course of their entire life (ACS, 2012). The survival depends on a multitude of factors and attributes which include early detection, age of the person, obesity, socio-economic status, stage of the cancer, type of cancer, overall health of the person, access to the effective treatment modality, and support during different stages of life after diagnosis (Kwan et al., 2010; Protani et al., 2010; Peairs et al., 2011; Sprague et al., 2011; Yang et al., 2011).

The guidelines for the screening of breast cancer are updated periodically. The current guidelines recommend that females aged between 20 and 39 years old consider clinical breast examination (CBE) every 3 years. Women 40 and older, who do not currently show any symptoms, should continue to receive a CBE annually along with their periodic general medical examination (Smith et al., 2013). Women with a familial history of breast cancer should start screening regularly prior to age 40. Most clinically recognizable signs and symptoms of breast cancer appear in the advanced stages of the disease, making it important to detect and diagnose breast cancer early, in order to improve prognosis and better manage the disease (Robertson et al., 2011; Smith et al., 2013).

There are several treatment options available for breast cancer that include surgery, chemotherapy, radiation therapy, hormonal therapy, bisphosphonates, and targeted therapy (Goldhirsch et al., 2013; Khatcheressian et al., 2013; NCI, 2013 Graham et al., 2014). *i*) Surgery (breast-conserving therapy [BCT] or modified radical mastectomy [MRM]), including radiotherapy is usually the standard choice of treatment (Litiere et al., 2012; Goldhirsch et al., 2013). The type of surgery used is predicated on the type of breast cancer (NCI, 2013); (a) BCT is an operation to remove the tumor without removing the affected breast. The National Institute of Health recommends breast conservation surgery for most women who are in the early stages of breast cancer (White et al., 2013). Lumpectomy and partial mastectomy with or without lymph node dissection are types of breast conservation surgery (Barton, 2013; Rao et al., 2013), (b) MRM is the procedure to remove the entire affected breast along with lymph nodes under the arm (Zurrida et al., 2011). Usually, a sentinel lymph node biopsy is performed before surgery and if cancer cells are found then surgery is conducted (Giuliano et al., 2011 Feigelson et al., 2013; Graham et al., 2014); *ii*) Chemotherapy also has a major role in this treatment. When chemotherapy is given before the surgery, it reduces the amount of tissue to be removed (Gampenrieder et al., 2013; Graham et al., 2014); *iii*) radiation therapy uses high energy radiation which kills or inhibits cancer cells from growing (NCI, 2013); *iv*) hormonal therapy is the treatment which prevents hormonal action and stops the growth of the cancer cells (Burstein et al., 2010; NCI, 2013), and *v*) the targeted therapy includes use of substances that identify and kill cancer cells without harming normal cells. Monoclonal antibodies and the tyrosine kinase inhibitors are the main sources for targeted therapy (Weiner et al., 2010; Gampenrieder et al., 2013; Goel et al., 2013; NCI, 2013).

After the diagnosis, breast cancer patients survive for a specific time-period. The length of this period (years, months, weeks, or days, etc.) is from the beginning of follow-up until death due to cancer and is treated as survival time. The survival time for each patient is recorded. Recorded survival times for diagnosed cancer patients are known as survival data. Analysis of survival data is very important in assessing and monitoring the progress of a patient's cancer survival. For detailed information on survival analysis, the reader can refer to Kleinbaum and Klein (2012).

Hospitals and various cancer registries store and record cancer survival data for future analytical purposes. As survival data is collected through various different sources, the outcome analysis differs based on the application, analytic tools, and methods used. This type of data however, requires new and innovative statistical methods and analysis to understand their scientific contribution. These statistical analyses can be used to draw inferences about the existing survival data and its probability model. Results from such analysis can help investigators in identifying the factors, which contribute to the poor prognosis of cancer. This method can also be extended to study the survival of a patient at various future time intervals, for example 1, 5 or 10 years cancer survival rate. These models can also have various extended applications such as, studying the recurrence rates of cancers, comparing two treatments, finding an effective or better treatment, and identifying factors that are affecting the progression of disease. Thus, cancer survival analysis using these techniques is an important application of statistics in medicine. Medical professionals, public health experts, and health policy majors can benefit from this increased knowledge, and can properly allocate resources necessary to improve the quality of life of cancer patients.

The high rates of disease burden worldwide demonstrate the public health significance of predicting the survival of breast cancer. However, there is currently not enough research done to build predictive inferences for precisely estimating survival days of cancer patients, especially using information driven novel approach of the Bayesian model. This paper addresses the need for a new approach and its application in using real life data.

The healthcare data extracted from several lab experiments and these data follow various right skewed statistical probability distributions, for example: exponential, gamma, Weibull, Exponentiated Exponential (EE), and Exponentiated Weibull (EW). Statistical and

computational techniques are urgently needed to analyze such data and to develop new scientific conclusions.

Among the various types of right skewed models, the Exponentiated Exponential Model (EEM) and the Exponentiated Weibull Model (EWM) are frequently used in modeling the data from biomedical sciences. The EEM is a generalization of the exponential distribution and received tremendous and widespread attention (Gupta and Kundu, 1999). The EEM has two parameters (scale and shape, where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter). Moreover, it was observed that many other properties of EEM are quite similar to that of the Weibull family, suggesting the possible use of EEM as an alternative of the EWM. It is also concluded that certain findings from the exponentiated exponential (EE) were a better fit to the data than the Weibull model (Gupta and Kundu, 2001). The EWM has been extensively used for analyzing survival data (Nassar and Eissa, 2003). It is noted that when $\beta=1$, the EWM (where $\alpha > 0$ and $\beta > 0$ are the shape parameters, and $\lambda > 0$ is the scale parameter, respectively) reduces to the EEM.

Predictive inference is a statistical inference where existing data is available from numerous lab experiments and future unavailable data is extracted by using statistical novel methods. Among all those novel methods, the Bayesian method is widely used to explore the posterior probability for the parameters, as well as future observations, for example for cancer incidence (Jafari-Koshki et al., 2014), mortality (Liu et al., 2012) or survival (Khan et al., 2014). In the Bayesian estimation technique, model parameters and data are considered random variables with joint probability distribution, which is stated by a probabilistic model. In the Bayesian method, data are considered as 'observed variables' and parameter as 'unobserved variables', and multiplying likelihood and prior gives the joint distribution of the parameters and is called the posterior distribution. The 'prior' contains the parameter value(s) information and later data is investigated as a probability distribution, whereas, the likelihood depends on the model of underlying process, and is measured as a conditional distribution that specifies the probability of the observed data, provided any certain values for the parameters. Prior and likelihood combine all the information that is necessary to make inferences about the parameters. The purpose of the Bayesian inference is to develop the posterior distribution of the parameters given a set of observed data and to obtain future survival estimates. For further information regarding the Bayesian method, the readers are referred to Baghestani et al. (2009) Khan, (2012a; 2012b; 2012c; 2012d; 2013a; 2013b), and Khan et al., 2013), among others.

The main objectives of this paper are to i) study demographic and socio-economic variables of the selected breast cancer patients; ii) review the widely used right skewed models EEM and EWM; iii) prove that the breast cancer sample survival data follow a specific statistical probability model by using model selection criteria for the goodness of fit tests; iv) utilize a novel Bayesian analysis to obtain the posterior distribution of the parameters; and v) obtain the predictive inference for future survival times and the likelihood of females getting breast cancer.

Materials and Methods

We used the breast cancer patients' data (N=657,712) from the Surveillance, Epidemiology and End Results (SEER; 1973-2009) website (SEER, 2013). SEER data contain breast cancer patients' information mostly from 12 states in the U.S. A stratified random sampling scheme was employed to draw a sample from the randomly selected nine states to represent race categories. The data was stratified according to gender (males=4,269 and females=653,443) and then a simple random sampling (SRS) method was applied to select a sample of 500 females. Tables 1, Table 2 and Table 3 present the summary results of the descriptive statistics based on the selected patients.

Table 1. Frequency Distribution of Selected Female Breast Cancer Patients

State	Female	
	Count	Percentage
California	85	17
Connecticut	71	14.2
Georgia	34	6.8
Hawaii	26	5.2
Iowa	72	14.4
Michigan	77	15.4
New Mexico	17	3.4
Utah	34	6.8
Washington	84	16.8
Total	500	100

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

Characteristics	Categories		Female
Age at diagnosis (years)	Mean		60.9
	SD		14.9
	Median		62
	Range		26-97
	Quartiles	Q_1	50
		Q_2	62
		Q_3	72.8
Survival time (months)	Mean		90.3
	SD		83.1
	Median		82.5
	Range		1-430
	Quartiles	Q_1	31
		Q_2	82.5
		Q_3	142.8

Table 3. Race, Ethnicity, and Marital Status of Female Breast Cancer Patients, SEER (1973-2009)

Characteristics	Categories	Female
Race	White	412
	Black	49
	Other	39
Hispanic Origin	Non-Hispanic	489
	Hispanic	11
Marital status at diagnosis	Single	63
	Married	259
	Divorced/Separated	57
	Widowed	102
	Unknown	19

Statistical Model Fitting

There are many methods available to measure the goodness of fit of each model. The most popular methods currently used by several researchers to compare various models 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 (Congdon, 2007). As a criterion, it uses model fit and complexity. It shows how good the model predictions fit the given data, while it represents the complexity of fitness given each model of the data. Although, DIC is used as the global fitness of model, it can also be partitioned to understand more details of model inadequacy (Spiegelhalter et al., 2002). The values of DIC can be positive as well as negative and models with lower values are considered better. Besides DIC, effective numbers of parameters is considered the secondary criterion of goodness of fit. (Akaike, 1973), generalized his work over factor analysis and time series analysis by introducing information criterion, which later became popular as Akaike's information criterion or AIC. Sakamoto, Kitagawa and Ishiguro (1986), students of Akaike, gave many interesting examples using AIC in a book named Akaike Information Criterion Statistics. Compared to cross-validation, AIC was superior in terms of originality. During that time, the maximum likelihood method was more popular among statisticians and AIC is quite closer to the maximum likelihood method. AIC could be applied to the results without any additional calculation. Akaike and his colleagues effectively combined the AIC to Bayesian framework in 1977 and 1978. DIC is similar to AIC and provide the same results as AIC when models with only fixed effects are fitted. The Bayesian Information Criterion (BIC) is an asymptotic result assuming 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 BIC penalizes free parameters more than AIC. As in AIC, when given any two estimated models, 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 female survival times.

A new reparameterization method of the parameters was used for the Birnbaum-Saunders Lifetime Model (Ahmed et al., 2008). One may utilize a reparameterization method by considering the log-likelihood function from the EEM. Assuming the data $X=(X_1, X_2, ..., X_n)$ represents n breast cancer patients' survival times, then a reparameterization method may be applied considering the log-likelihood function from the EEM.

Similarly, one may obtain the log-likelihood function from the EWM. By using the reparameterization method, one would obtain better performance of the posterior distribution for the parameters. The following Table 4 presents the selection of EEM compared to EWM on the basis of AIC, BIC, and DIC criterions.

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. Comparing the estimated values of all AIC, BIC, and DIC based on the EEM and EWM in the case of females, the EEM fits better for the female survival times because it produces smaller values of AIC, BIC, and DIC. Thus, for female survival data fits better for EE distribution as compared to EW distribution because estimated values of AIC, BIC, and DIC generated smaller values by using the EEM.

In the Bayesian approach, the knowledge of the distribution of the parameters is updated using 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. It is observed that the EE fits the breast cancer survival data; we therefore, attempt to obtain the posterior summary results for the parameters and their probability distributions.

The results of the posterior distribution parameters α and λ are estimated using the Markov Chain Monte Carlo (MCMC) method. MCMC 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 EEM and then its parameter values are assigned to the appropriate theoretical probability distributions. The WinBugs software (MRC Biostatistics Unit, 2013) is used to obtain the summary results (Mean, SD, 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 times 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 1 displays the graphical representation of the parameters' behavior in the case of the EEM based on the female survival data. After 80,000 Monte Carlo repetitions, it is noted that the shape parameter plays approximately symmetrical distribution.

Table 4. Selection of The Exponentiated Exponential Model on the Basis of AIC, BIC, and DIC Criterions

Criterions	Statistical Models	
	EEM	EWM
AIC	5601.71	5722.18
BIC	5608.14	5734.82
DIC	5599.69	5718.3

Table 5. Summary Results in the Posterior Parameters in the Case of Best Fit Exponentiated Exponential Model for Female Survival Data

Node		Median	95% Confidence interval	Iterative sample
alpha	0.0	1.1	0.980 ...	80,000
lambda	6.0	0.01	0.0095 ...19)	80,000

Results

Due to the current economic crisis, health care costs are constantly increasing at an alarming rate. It is important for health care researchers and providers to identify populations at risk of acquiring diseases. The challenge is to identify and provide intervention without significantly increasing the cost of diagnosis or treatment, while the population is healthy or asymptomatic. Recently, predictive inference has been the popular technique to conduct high-risk assessments at low costs. Health care providers and researchers also use the predictive inference to improve current health care services.

Predictive inference applies to available healthcare data, for instance, it can be used to identify people who have high medical need and are 'at risk' for above-average future medical service utilization. To date, there is no standardized process to address this problem; however, there is a novel Bayesian method, which can predict the breast cancer survival times based on the past data collected from patients.

The Bayesian predictive method is growing more popular, finding new practical applications in the fields of health sciences, engineering, environmental sciences, business, economics, 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. For more about Bayesian predictive approach see (Khan et al. (2010; 2011a; 2011b).

In this section, a predictive survival inference for the breast cancer patients is developed by using a novel Bayesian method. It is found that the female cancer patients' survival data follow the EEM by using the criterions AIC, BIC, and DIC. In this section by considering the female survival times, which constitute the EEM, the predictive inference for future survival

Table 6. Summary Results of Predictive Inference for the Female Breast Cancer Survival Data

Summary statistics	Predictive intervals	Predictive skewness and kurtosis
Mean=97.7451	90%: (4.510940, 301.187675)	β_1=2.52878
SE=4.04025	95%: (3.285315, 371.157430)	β_1=5.86616
	98%: (2.024840, 415.134570)	Y_1=1.59021
	99%: (1.727886, 446.996028)	Y_2=2.86616

*SE=Standard Error

model is discussed.

Assuming the data $X=(X_1, ..., X_n)$ represents n, female breast cancer patient's survival times, which follow the EEM. The posterior probability can be defined by multiplying the likelihood function and the prior density for the parameters. Bayes and empirical Bayes estimates of survival and hazard functions of a class of distribution is discussed in details by Ahsanullah and Ahmed (2001). Ahmed and Tomkins (1995) estimated lognormal mean by making use of uncertain prior information. (Khan et al., 2004; 2011b) derived the Bayesian predictive inference from the Weibull Life Model by means of a conjugate prior distribution for the scale parameter and a uniform prior distribution for the shape parameter. Considering the assumption from Khan et al. about the prior knowledge of the parameters, we obtained summary of the predictive results for future survival times.

A numerical integration command 'NIntegrate' in conjunction with the symbolic computational software Mathematica version 8.0, Wolfram Research (2012), is applied to obtain the predictive results. The Mathematica package is also utilized to carry out all related calculations.

The summary results of female predictive mean, standard error, and predictive intervals for future survival time are given in Table 6. The predictive shape characteristics i.e., the estimated values of skewness, and kurtosis are also presented in the same table. Based on the results one would conclude that the predictive probability model forms a right skewed model.

These findings are important for health care researchers and providers in order to characterize future disease patterns, and to make effective future plans in our healthcare industry.

Discussion

There is a large amount of underutilized clinical data available in the healthcare industry. This includes clinical, imaging, biochemical, cellular, and genomic data, which require newer and more advanced statistical analysis. The implementation of newer analytical approaches is necessary to identify statistical probability distributions for drawing scientific conclusions about future patterns and mortality rates of diseases.

We used statistical models for the breast cancer patients' survival data of patients diagnosed during 1973 to 2009 in the U.S., and determined the best fit probability model by using the Bayesian method. We used a representative sample consisting of 500 female breast cancer patients diagnosed during 1973-2009. It is found that the EEM best fits the female cancer survival data.
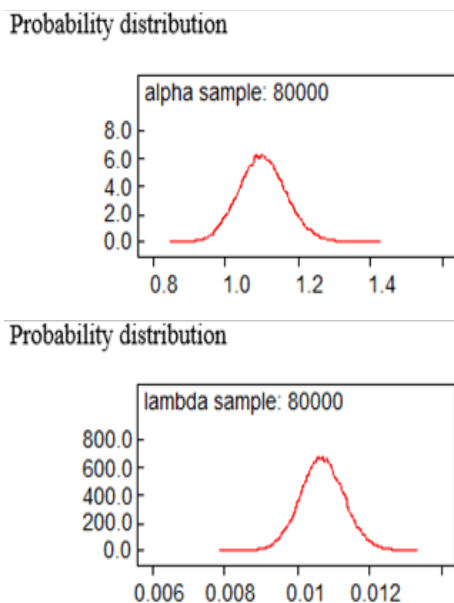
Figure 1. Posterior Probability Distribution for the Parameters in the Case of Best Fit Exponentiated Exponential Model

The mean (SD) age at diagnosis for breast cancer patients was 60.92 (14.92) years for females. The minimum age at diagnosis for females was 26 years. In the sample, the oldest female at the time of diagnosis was 97 years. The mean (SD) survival time for female cases was 90.33 (83.10) months. In this sample, among 500 females, race was distributed as 82% White, nearly 10% Black, and 8% listed as 'other'. There were nearly 3% people of Hispanic origin. The majority of these patients were married.

Clinical features of the breast cancer showed that most of the patients had tumor Grade I or II of either laterality. Approximately 97% of the breast cancer diagnosis among all the female samples was confirmed by histology and more than 85% of the cases had malignancy in the breast cancer tumor. In the selected sample, around 20% of the patients died of breast cancer while others passed away from unrelated causes.

It is found that the breast cancer data from the female sample followed exponentiated exponential distribution with the DIC value of 5599.69. The mean (SD) for and values are 1.10 (0.06) and 0.01 (6.01×10⁻⁴), respectively. The confidence intervals for posterior parameters are given in Table 5. The dynamic probability distributions for each of the parameters are reported in Figure 1 so that one can observe the shape of the distributions. The predictive mean survival time, standard error, predictive intervals, and measures of skewness, and kurtosis are reported in Table 6. Results show that the shape of the future survival times follows a positive skewed distribution.

In the recent progress of biomedical science, huge amounts of data have been collected from thousands of subjects. Characteristics and disposition of scientific data can be described by using various statistical probability models. These models are crucial in making statistical inferences about the parameters and future disease patterns that affect health. Thus, these novel methods of making statistical inferences can be very helpful in early diagnosis and intervention planning.

We selected stratified random sample of breast cancer patients from the SEER (1973-2009) database registry. The methods for measuring the goodness of fit tests are used to select the best statistical probability model for the female based on breast cancer survival data. To develop statistical probability model for survival days of females, we used model selection criterions, AIC, BIC, and DIC to measure the best fit to the breast cancer survival data. We found that the EEM best fits the female survival data. A detail analysis of the posterior models for the parameters is described with their summary results.

The log-likelihood functions of both models EE and EW are used to reparameterize the original parameters to accelerate better performance of the Bayesian posterior parameters and to draw its corresponding dynamic probability distributions. The summary results of the posterior parameters are reported by using the MCMC method. The results are obtained after running 80,000 Monte Carlo repetitions.

A computational software package, 'Mathematica version 8.0', is used to attain the future survival time and also to obtain the related predictive inferences. WinBugs software is used to check the goodness of fit tests, to obtain the summary results of the posterior parameters, to determine the dynamic probability distributions of the parameters, and also to carry out all related calculations.

The predictive mean, standard error, predictive intervals, and measures of skewness and kurtosis are reported for the future survival time. Based on the results of skewness and kurtosis one would comment that the shape of the future survival model for the female is positively skewed. These findings will be extremely helpful for the healthcare researchers and providers to predict a patient's possible future medical outcome given the patient's current state and past history. Thus, it will effectively combine knowledge, discovery, and innovation from the breast cancer patients on the basis of nine states to provide an enhanced and improved rationale for the diagnosis and treatment of breast cancer patients all over the U.S.

References

American Cancer Society (ACS) (2012). Breast Cancer Facts & Figures 2011-2012. Atlanta: American Cancer Society.

ACS (2013). Cancer Facts and Figures. Atlanta: American Cancer Society.

Ahmed SE, Tomkins RJ (1995). Estimating lognormal mean using uncertain prior information, *Pakistan J Statistics*, **11**, 67-92.

Ahmed SE, Budsaba K, Lisawadi S, Volodin AI (2008). Parametric estimation for the Birnbaum-Saunders lifetime distribution based on a new parameterization, *Thailand Statistician*, **6**, 213-40.

Ahsanullah M, Ahmed SE (2001). Bayes and empirical Bayes estimates of survival and hazard functions of a class of distributions. *Lecture Notes in Statistics*, **148**, 81-7.

Akaike H (1973). Information theory and an extension of the maximum likelihood principle. N.Petrov and E. Caski (eds), Budhapest: Addademiai Kiado, Proceedings of the 2nd International Symposium on Information Theory.

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.

Barton MK (2013). Potential survival benefit with lumpectomy for patients with early stage breast cancer. *CA Cancer J Clin*, **63**, 217-8.

Burstein HJ, Prestrud AA, Seidenfeld J, et al (2010). American Society of Clinical Oncology clinical practice guidelines: Update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J Clin Oncol*, **28**, 3784-96

Congdon P (2007). Mixtures of spatial and unstructured effects for spatially discontinuous health outcomes. *Computational Statistics and Data Analysis*, **51**, 3197-212.

Feigelson HS, James TA, Single RM, et al (2013). Factors associated with the frequency of initial total mastectomy: results of a multi-institutional study. *J Am Coll Surg*, **216**, 966-75.

Ferlay J, Soerjomataram I, Ervik M, et al (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer.

Gampenrieder SP, Rinnerthaler G, Greil R. (2013). Neoadjuvant chemotherapy and targeted therapy in breast cancer: past, present, and future. *J Oncol*, **2013**, 732047.

Gilks WR, Richardson S, Spiegelhalter DJ (1996). Markov Chain Monte Carlo in Practice. Chapman & Hall, Suffolk.

Giuliano AE, Hunt KK, Ballman KV, et al (2011). Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA*, **305**, 569-75.

Goel S, Gupta N, Walcott BP, et al (2013). Effects of vascular-endothelial protein tyrosine phosphatase inhibition on breast cancer vasculature and metastatic progression. *J Natl Cancer Inst*, **105**, 1188-201.

Goldhirsch A, Winer EP, Coates AS, et al (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*, **24**, 2206-23.

Graham LJ, Shupe MP, Schneble EJ, et al (2014). Current approaches and challenges in monitoring treatment responses in breast cancer. *J Cancer*, **5**, 58-68.

Gupta RD, Kundu D (1999). Generalized exponential distributions. *Aust N Z J Stat*, **41**, 173-88.

Gupta RD, Kundu D (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical J*, **43**, 117-30.

Hou N, Huo D (2013). A trend analysis of breast cancer incidence rates in the United States from 2000 to 2009 shows a recent increase. *Breast Cancer Res Treat*, **138**, 633-41.

Jafari-Koshki T, Schmid VJ, Mahaki B (2014). Trends of breast cancer incidence in Iran During 2004-2008: A Bayesian space-time model. *Asian Pac J Cancer Prev*, **15**, 1557-61.

Jemal A, Bray F, Center M, et al (2011). Global Cancer Statistics. *CA Cancer J Clin*, **61**, 69-90.

Khatcheressian JL, Hurley P, Bantug E, et al (2013). Breast cancer follow-up and management after primary treatment: American Society of Clinical Oncology clinical practice guideline update. *JCO*, **31**, 961-65.

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 (2012c). Predictive inference for a future response using symmetrically trimmed sample from the half-normal model, *Comput Stat Data Anal*, **56**, 1350-61.

Khan HMR (2012d). Improved predictive inference for responses based on location parameter median from the Rayleigh model given a doubly censored sample, Communications in Statistics - *Theory and Methods*, **41**, 4427-38.

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 Stat*, **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, Albatineh AN, Tesfa S, et al (2011a). Predictive inference from two-parameter Weibull model given a doubly censored sample of hepatocellular carcinoma of liver patients, *J App Stat Sci*, **19**, 115-28

Khan HMR, Albatineh A, Alshahrani S, et al (2011b). Sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients, *Comput Stat Data Anal*, **55**, 3093-103.

Khan HMR, Haq MS, Provost SB (2004). Predictive distributions for responses from the Weibull life testing model. *J Stat Theory App*, **3**, 53-73.

Khan HMR, Provost SB, Singh A (2010). Predictive inference from a two parameter Rayleigh life model given a doubly

censored sample, *Commun Stat Theory Methods*, **39**, 1237-46.

Khan HM, Saxena A, Rana S, Ahmed NU (2014). Bayesian method for modeling male breast cancer survival data.. *Asian Pac J Cancer Prev*, **15**, 663-9.

Kleinbaum DG, Klein M. (2012). Survival Analysis: A Self-Learning Text (3rd ed.). New York: Springer.

Kwan ML, Kushi LH, Weltzien E, et al (2010). Alcohol consumption and breast cancer recurrence and survival among women with early-stage breast cancer: the life after cancer epidemiology study. *JCO*, **28**, 4410-16.

Litière S, Werutsky G, Fentiman IS, et al (2012). Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol*, **13**, 412-19.

Liu SZ, Zhang F, Quan PL, et al (2012). Time trends of esophageal cancer mortality in Linzhou city during the period 1988-2010 and a Bayesian approach projection for 2020. *Asian Pac J Cancer Prev*, **13**, 4501-4.

Mariotto AB, Yabroff KR, Shao Y, et al (2011). Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst*, **103**, 117-28.

Montero AJ, Eapen S, Gorin B, Adler P (2012). The economic burden of metastatic breast cancer: a US managed care perspective. *Breast Cancer Res Treat*, **134**, 815-22.

MRC Biostatistics Unit (2013). WinBUGS software version 1.4.3: The BUGS Project.

Nassar MM, Eissa FH (2003). On the exponentiated Weibull distribution. *Commun Stat Theory Methods*, **32**, 1317-36.

National Cancer Institute (NCI) (2013). Breast Cancer Treatment (PDQ®). Health Professional version, updated: November 19, 2013.

Peairs KS, Barone BB, Snyder CF, et al (2011). Diabetes mellitus and breast cancer outcomes: a systematic review and meta-analysis. *JCO*, **29**, 40-6.

Protani M, Coory M, Martin JH (2010). Effect of obesity on survival of women with breast cancer: systematic review and meta-analysis. *Breast Cancer Res Treat*, **123**, 627-35.

Rao R, Euhus D, Mayo HG, Balch C (2013). Axillary node interventions in breast cancer: a systematic review. *JAMA*, **310**, 1385-94.

Robertson C, Arcot Ragupathy SK, Boachie C, et al (2011). The clinical effectiveness and cost-effectiveness of different surveillance mammography regimens after the treatment for primary breast cancer: systematic reviews registry database analyses and economic evaluation. *Health Technol Assess*, **15**, 1-322.

Sakamoto Y, Ishiguro M, Kitagawa G (1986). Akaike Information Criterion Statistics. Tokyo: D. Reidel.

SEER (2013). Surveillance, Epidemiology and End Results. Cancer of the Breast - *SEER Stat Fact Sheets*.

Siegel R, Naishadham D, Jemal A (2012). Cancer statistics, 2012. *CA Cancer J Clin*, **62**, 10-29.

Smith RA, Brooks D, Cokkinides V, et al (2013). Cancer screening in the United States, 2013. *CA Cancer J Clin*, **63**, 87-105.

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B Stat Methodol*, **64**, 191-232.

Sprague BL, Trentham Dietz A, Gangnon RE, et al (2011). Socioeconomic status and survival after an invasive breast cancer diagnosis. *Cancer*, **117**, 1542-51.

U.S. Cancer Statistics Working Group (USCS). (2013). United States Cancer Statistics: 1999-2010 Incidence and Mortality Web-based Report. Atlanta: U.S. DHHS, CDC and NCI.

Weiner LM, Surana R, Wang S (2010). Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nat Rev*

Hafiz Mohammad Rafiqullah Khan et al
Immunol, **10** 317-27.

White A, Richardson LC, Krontiras H, Pisu M (2013). Socioeconomic disparities in breast cancer treatment among older women. *J Women's Health*.

Wolfram Research (2012). The Mathematica Archive: Mathematica 8.0, Wolfram Research Inc, Illinois.

World Health Organization (WHO) (2013a). Cause-specific mortality and morbidity: Age-standardized mortality rate by cause by World Bank income group, 2008. Geneva: WHO.

WHO (2013b). Charts for the 10 leading causes of death for women worldwide and by income group, 2011. Geneva: WHO.

Yang XR, Chang-Claude J, Goode EL, et al (2011). Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the breast cancer association consortium studies. *J Natl Cancer Inst*, **103**, 250-63.

Zurrida S, Bassi F, Arnone P, et al (2011). The changing face of mastectomy (from mutilation to aid to breast reconstruction). *Int J Surg Oncol*, **2011**, 980158.