RESEARCH ARTICLE

Statistical Estimates from Black Non-Hispanic Female Breast Cancer Data

Hafiz Mohammad Rafiqullah Khan[1]*, Boubakari Ibrahimou[1], Anshul Saxena[2], Kemesha Gabbidon[2], Faheema Abdool-Ghany[1], Venkataraghavan Ramamoorthy[3], Duff Ullah[1], Tiffanie Shauna-Jeanne Stewart[3]

Abstract

Background: The use of statistical methods has become an imperative tool in breast cancer survival data analysis. The purpose of this study was to develop the best statistical probability model using the Bayesian method to predict future survival times for the black non-Hispanic female breast cancer patients diagnosed during 1973-2009 in the U.S. Materials and Methods: We used a stratified random sample of black non-Hispanic female breast cancer patient data from the Surveillance Epidemiology and End Results (SEER) database. Survival analysis was performed using Kaplan-Meier and Cox proportional regression methods. Four advanced types of statistical models, Exponentiated Exponential (EE), Beta Generalized Exponential (BGE), Exponentiated Weibull (EW), and Beta Inverse Weibull (BIW) were utilized for data analysis. The statistical model building criteria, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC) were used to measure the goodness of fit tests. Furthermore, we used the Bayesian approach to obtain the predictive survival inferences from the best-fit data based on the exponentiated Weibull model. Results: We identified the highest number of black non-Hispanic female breast cancer patients in Michigan and the lowest in Hawaii. The mean (SD), of age at diagnosis (years) was 58.3 (14.43). The mean (SD), of survival time (months) for black non-Hispanic females was 66.8 (30.20). Non-Hispanic blacks had a significantly increased risk of death compared to Black Hispanics (Hazard ratio: 1.96, 95%CI: 1.51–2.54). Compared to other statistical probability models, we found that the exponentiated Weibull model better fits for the survival times. By making use of the Bayesian method predictive inferences for future survival times were obtained. Conclusions: These findings will be of great significance in determining appropriate treatment plans and health-care cost allocation. Furthermore, the same approach should contribute to build future predictive models for any health related diseases.

Keywords: Model development - Bayesian method - statistical inference - breast cancer survival - black non-Hispanic

Asian Pac J Cancer Prev, **15 (19)**, 8371-8376

Introduction

Cancer is one of the leading causes of mortality and morbidity worldwide (Sankaranarayanan et al., 2013). As of 2008, there were approximately 12.7 million cancer cases and 7.6 million cancer related deaths globally (Sankaranarayanan et al., 2013). In addition, there were an estimated 6 million cancer cases and 3.4 million deaths in women alone during the same timeframe. Furthermore, in the year 2012, 14.1 million people were diagnosed with cancer and 8.2 million died from cancer (Jemal et al., 2011). Within those cases, female breast cancer accounted for 1.67 million cases and over 500,000 deaths (Ferlay et al., 2012; WHO, 2013). Breast cancer is the 2nd leading cause of death in developing nations and the leading cause of death in developed nations such as the United States (Sankaranarayanan et al., 2013; Jemal et al., 2011). As of 2010, there were 692,600 cases in more developed regions and 690,900 cases in less developed regions (Sankaranarayanan et al., 2013). However age-standardized rates were 2.5 times greater in developed areas compared to the developing areas (96.3 per 100,000 women years in developed regions vs 39.2 per 100,000 women years in less developed regions) (Sankaranarayanan et al., 2013). Globally, breast cancer ranks the second highest in both incidence and prevalence following lung cancer. According to the World Health Organization (WHO) breast cancer was responsible for 612,000 deaths in the year 2010 (WHO, 2010).

In the United States, breast cancer is the second most

*[1]Department of Biostatistics, [2]Department of Health Promotion and Disease Prevention, [3]Department of Dietetics and Nutrition, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, USA *For correspondence: hmkhan@fiu.edu*

commonly diagnosed cancer among women (DeSantis et al., 2011). Approximately 230,480 new cases of invasive breast cancers and 39,520 breast cancer deaths were estimated to occur by the year 2011 (DeSantis et al., 2011). Based on these predictions, one in eight women in the U.S. would develop breast cancer during their average life expectancy (DeSantis et al., 2011). Between the years 2004 and 2008, breast cancer death rates were increasing across various ethnic and racial groups in the U.S. (DeSantis et al., 2011). Another striking feature includes the shift of increasing death rates from the more affluent areas to the poverty stricken areas (DeSantis et al., 2011). As per the American Institute of Cancer Research (AICR), economically disadvantaged women accounted for greater than 39,000 deaths in the U.S. in the year 2011 (Alteri et al., 2011). In 2012, approximately 226,870 diagnosed and 39,510 deaths in the United States (Siegel et al., 2012; ACS, 2013). Approximately 14% of women in the US are expected to develop breast cancer in their lifetime (ACS, 2013).

The highest incident rates of breast cancers have been reported in women between the ages of 50 and 64, with 23,360 in situ cases and 81,970 invasive cases diagnosed in 2011 alone (WHO, 2010). There are several risk factors associated with increased incidence and disease severity including family history, socio-economic status, education levels, access to mammograms, access to health care resources, and race and ethnicity (Liu, 2012). Both genetic and non-genetic factors are associated with breast cancers (Mavaddat, 2010). Some of the most commonly associated non-genetic factors include menstrual problems, reproductive history, BMI, alcohol intake, and physical activity (Bradbury, 2007). There are also significant racial and ethnic variations in breast cancer death rates, severity, progression, and other concrete variables (Liu et al., 2012). White women have greater years of potential life lost (YPLL) due to breast cancer compared to all other races within the U.S., which includes African-Americans, Hispanics, Asia-Pacific Islanders, American Indians, and Alaska Natives (Liu et al., 2012). Though White non-Hispanic women are more likely to be diagnosed with breast cancers than Black-non Hispanic women, the latter report greater mortality rates (Siegel et al., 2012). The Center for Disease Control and Prevention (CDC) 2010 reports showed that breast cancer continues to be the most frequently diagnosed cancer amongst the non-Hispanic Black female population. It also continues to be the leading cause of cancer-related deaths among non-Hispanic Black women aged 45-64 years for the same year (Liu et al., 2012).

The survival of breast cancer patients also depends on factors such as genetics, age at diagnosis, stage of the cancer, access to care, weight, physical activity status, alcohol consumption, disease co-morbidities, social, economic, environmental factors, and ethnicity (Graeser et al., 2009; Kwan et al., 2010; Protani et al., 2010; Peairs et al., 2011; Sprague et al., 2011; ACS, 2013). Screening guidelines have also evolved based on the research findings correlating breast cancer-screening and survival times. Presently, it is recommended that women between ages 20 to 39 complete a clinical breast examination (CBE) every 3 years. Those who are asymptomatic but aged 40 years or older are recommended to receive CBE every year (Robertson et al., 2011; Smith et al., 2013). Women who have a history of breast cancer in their families should start screening on a regular basis before age 40. The most recognizable signs and symptoms of breast cancer often appear in the later stages of the disease, making it imperative to detect, diagnose, and treat breast cancer early (Walker et al., 2013).

Genetic factors are complicated and very difficult to estimate (Mavaddat et al., 2010). Genetic risk is estimated in terms of the familial clustering of the disease measured through familial relative risk (FRR) (Mavaddat et al., 2010). FRR is defined as the risk for the relative of an affected individual to have the disease compared to risk of the general population. FRR is five-folds higher for women over age 60 compared to women under age 40 (Mavaddat et al., 2010). FRR is associated with both the age at diagnosis of the index case and the current age of the relative who is being estimated for the risk (Mavaddat et al., 2010). FRR also shows racial disparities, as white populations have the highest predicted FRR, followed by Hispanics with intermediate values and American Indians with the lowest FRR (Mavaddat et al., 2010).

The survival time is defined as the period between diagnosis of a disease and death because of the disease. Several statistical modeling techniques are used to fit the data (survival days) and calculate the probability of survival. Using newer methods can provide better estimation of future probability. There are several databases including hospital records, cancer registries and morbidity records that contain variables related to population based disease severity and risk, burden, number of potential years of life lost due to the disease, productivity lost due to the disease etc., thereby facilitating the estimation of disease impact, survival rates, and other inferential factors.

The main objective of this paper is to estimate statistical survival inference for breast cancer associated with the female non-Hispanic black population. These estimates are based on large volumes of survival data accrued over the years. The most important initial step included the determination of appropriate statistical model based on repeated sampling trials. We also used a novel Bayesian predictive model to determine the patients' collective future survival outcomes. In addition, we have summarized our inferences in simplified terms for better comprehension. Our findings would go a long way in understanding predictive survival inferences from data based survival model. This would help clinicians to predict future outcomes of the disease given the current survival rates. It would serve as a tool to stratify individuals for appropriate health care decisions as well as understanding the extent and impact of the disease without the use of costly individual based medical procedures.

Furthermore, our results may increase the pressure on allocating resources towards newer, faster, and better screening procedures as well as for rechanneling the research money towards understanding better curative and palliative strategies for the disease in an ethnic group context.

Materials and Methods

Surveillance, Epidemiology and End Results (SEER, 1973-2009) website includes data registries from twelve states stored in protected databases (SEER, 2013). This data included 4,269 males and 608,032 females (N=608,032) breast cancer patients. The breast cancer is comparatively uncommon in males; hence, only female cases were included in this study. There were total 53,531 black non-Hispanic cases.

Numerous skewed statistical probability models have been used for inferential statistics in the case of data analysis. There are some limitations of applying such probability models in large-scale data sets. In the present study, we have described four advanced statistical probability models. These include Exponentiated Exponential (EE), Beta Generalized Exponential (BGE), Exponentiated Weibull (EW), and Beta Inverse Weibull (BIW). These models reduce to other skewed statistical probability models by substituting specific values for the parameters. These models have been discussed in detail by Khan et al. (2014a, 2014b, 2014c, 2014d).

The most commonly used criterion for measuring the goodness of fit are the Akaike Information Criterion (AIC), Deviance Information Criterion (DIC), and Bayesian Information Criterion (BIC). DIC is a Bayesian measure of fit that is used for comparison of different models. The model with lower values are considered a better fit than others (Khan et al., 2014a, 2014b, 2014c, 2014d).

There are 2,000 data points and they are treated as survival times for the corresponding breast cancer patients. Symbolically, survival data set is defined by t, where t contains 2,000 survival data points and n=2,000. The first survival time is defined by t_1, the second survival time is defined by t_2, the third survival time is defined by t_3, and so on. Therefore, the survival data set can be defined as $\mathbf{t}=(t_1, t_2, ..., t_n)$. A reparameterization method from certain model was proposed by Khan et al. (2014a, 2014b, 2014c, 2014d). This method was used to describe the distribution of the parameters in terms of the log-likelihood functions.

The likelihood function is used to estimate the parameters from any statistical probability models. It is beneficial to work with the log-likelihood function (the natural logarithm of the likelihood function) because of the probability model associated with exponent terms. The logarithm is a monotonically increasing function, which contributes to achieving the maximum value for the parameters. The log-likelihood technique is often used in substitution of the maximum likelihood for the estimation of parameters from a statistical model. These estimates of the parameters are approximately unbiased estimates and are used in predicting future survival inferences. Considering survival data $\mathbf{t}=(t_1, t_2, ..., t_n)$ from the four models, Khan et al. discussed the log-likelihood functions (2014a, 2014b, 2014c, 2014d).

Predictive modeling is the process of estimating future risk of diseases from currently available healthcare data. To date, there is no standardized process for predictive modeling. The Bayesian method may be used to predict the breast cancer survival days based on past data. A predictive survival model for breast cancer patients may be developed by using a novel Bayesian method.

Let us assume the data $\mathbf{t}= (t_1, ..., t_n)$ represents n black non-Hispanic female breast cancer patients survival days and let z be a future response (or future survival time). The predictive density of z for the observed data $\mathbf{t}$ is

$$\mathbf{p(z|t)= \int\int\int p(z|}\alpha, \beta, \lambda) \mathbf{p(}\alpha, \beta, \lambda\mathbf{|t)}\, d\lambda d\beta\, d\alpha,$$

where $p(\alpha, \beta, \lambda|\mathbf{t})$ is the posterior density function, and $\mathbf{p(z|}\alpha, \beta, \lambda)$ represents the probability density function of a future survival time (z) that may be defined from the best fit model. The posterior density is given by

$$\mathbf{p(}\alpha, \beta, \lambda\mathbf{|t)} \propto \mathbf{L(}\alpha, \beta, \lambda\mathbf{|t)\ p(}\alpha, \beta, \lambda)$$

where $L(\alpha, \beta, \lambda|\mathbf{t})$ is the likelihood function and $p(\alpha, \beta, \lambda)$ is the prior density for the parameters (Khan et al., 2013a, 2013b, 2012a, 2012b, 2011, 2004).

Then given a set of data $\mathbf{t}= (t_1, ..., t_n)$, the likelihood function is given by

$$L(\alpha, \beta, \lambda \mid \mathbf{t}) \propto (\alpha\beta\lambda)^n \exp\left\{-\sum_{i=1}^{n}(\lambda t_i^{\beta})\right\} \left[\prod_{i=1}^{n}(t_i^{\beta-1})\right] \left[\prod_{i=1}^{n}\left(1 - \exp\left\{-(\lambda t_i^{\beta})\right\}\right)^{\alpha-1}\right].$$

The Bayesian posterior model has been discussed by Khan et al. (2014b). For the Bayesian prediction, we assume a conjugate prior for the scale parameter and uniform priors for the shape parameters.

Considering the prior knowledge, the posterior density is given by

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

IBM SPSS version 20 software was used to obtain summary of the results. WinBugs software was used for statistical analysis of goodness-of-fit of four models and predictive survival inference.

Results and Discussion

Using simple random sampling method, 2,000 black non-Hispanic cases were selected from 53,531 black non-Hispanic cases. Tables 1-3 contain the descriptive statistics for this analytical sample. It is observed from Table 1 that the sample included most of the cases from Georgia and Michigan. Lowest number of cases was selected from the state of Hawaii.

Around 36% of subjects in this sample were married, 22% single and 21% were divorced. The mean

Table 1. Frequency Distribution of the Selected Black Non-Hispanic Patients from Nine States (n=2,000)

States	Black non-Hispanic Count	Percent
Georgia	612	30.6
Hawaii	5	0.25
Iowa	23	1.15
Michigan	730	36.5
New Mexico	12	0.6
Utah	7	0.35
Washington	77	3.85
California	352	17.6
Connecticut	182	9.1
Total	2,000	100

Table 2. Age at Diagnosis, Survival Time, and Marital Status for Female Black non-Hispanic Breast Cancer Patients (n=2,000).

Age at diagnosis (years)	Mean	58.32
	SD	14.43
	Median	57
	Quartiles	47,57,69
	Variance	208.24
Survival time (months)	Mean	66.76
	SD	30.2
	Median	62
	Quartiles	40,62,90
	Variance	911.31
Marital status at diagnosis	Single	437
	Married	727
	Separated	52
	Divorced	273
	Widowed	427
	Unknown	84

Table 3. Selection of the Models for Black non-Hispanic Females (n=2,000) on the Basis of AIC, BIC, and DIC Criterions

Model criterions	AIC	BIC	DIC
Exponentiated exponential	19083	19098.7	19082.948
Exponentiated Weibull	19082.7	19099	19080.75
Beta generalized exponential	19086.6	19109	19082.614
Beta Inverse Weibull	21591.9	21608.7	21585.89

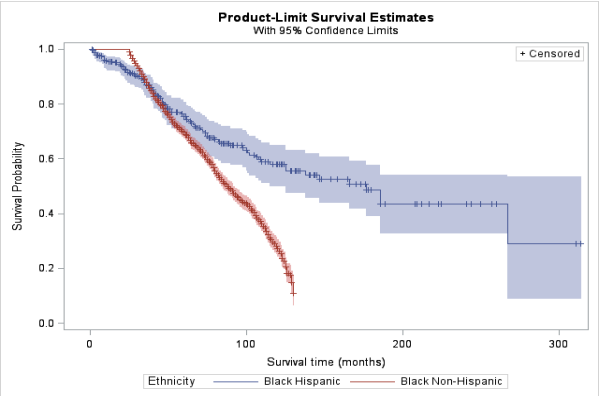

Figure 1. Kaplan-Meier Survival Curves (with 95% Confidence Limits) for Female Black Non-Hispanic where Black Hispanic Breast Cancer Patients are Considered as Reference Variable

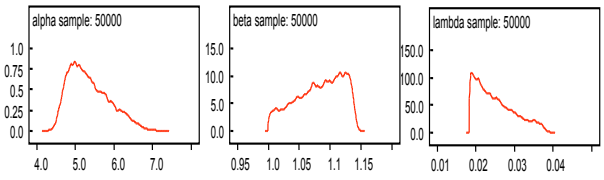

Figure 2. Posterior Parameter Densities in the Case of EW for Black Non-Hispanic Female Breast Cancer Patients (n=2,000).

survival time for non-Hispanic black was 66.76 months (SD=30.20). All black Hispanics breast cancer cases (n=298) were included from the SEER database as a referent group to calculate hazard ratios by ethnicity using Cox Proportional Regression. Hazard ratios compare the probability of an event occurring in one group versus another and take into account the time elapsed until the event occurs. In survival analysis, the event under consideration is death and "alive" status was used as a censoring variable. Statistical significance was established if the 95% confidence interval did not include the integer one. Non-Hispanic black had a significantly increased risk of death compared to Black Hispanics (Hazard ratio: 1.96 [95% Wald (Sandwich) Confidence Limits: 1.51-2.54; 95% Profile Likelihood Confidence Limits: 1.57-2.48]). These results are consistent with the survival curve, confirming the longer survival among Hispanic black and shorter survival among non-Hispanic black women.

Figure 1 shows the Kaplan-Meier estimates of the survival function for Hispanic blacks and non-Hispanic blacks (with 95% confidence limits as the shaded area). The graph shows that the probability of survival decreases over time for both groups. Nevertheless, the slope of the curve is steeper for non-Hispanic blacks, with most participants dying within 130 months after diagnosis.

We calculated the values for AIC, BIC, and DIC. Table 3 shows the summary results of the measures of goodness of fits for the AIC, BIC and DIC. Based on the Black non-Hispanic female survival times, we derived the posterior distributions for the parameters.

In Table 3, the AIC, BIC, and DIC values for the EE, EW, BGE, and BIW models was tabulated. The predictive inference for future survival model is discussed below by considering the female survival times, which constitute the EW model. EW better fits the survival times, because it produces the smallest values for AIC and DIC though the values for BIC are very close to EE and EW models. Table 4 contains the summary results of the parameters for black non-Hispanic female breast cancer patients after repeating the sample of 50,000 times. Figure 2 contains a graphical representation of the parameters behaviors. The posterior distributions of the parameters are approximately skewed distributions.

The predictive inference for future survival model is obtained by considering the female survival times, which constitute the EW model. The graphical representation of the predictive density based on the black non-Hispanic patients' survival times can be seen in Figure 3. It should be noted that predictive density formed a right skewed model.

The summary results of predictive inference are given in Table 5 . It also includes predictive shape of the survival distribution by determining raw moments, corrected moments, and measures of skewness and kurtosis.

Table 4. Summary Results of the Posterior Parameters in the Case of EW for Black non-Hispanic Female Breast Cancer Patients (n=2,000)

Node	Mean	SD	MC error	Median	95% CI	Iterative Sample
alpha	5.364	0.5314	0.03398	5.27	(4.583, 6.542)	50000
beta	1.083	0.03749	0.002551	1.088	(1.008, 1.138)	50000
lambda	0.02502	0.005097	3.48E-04	0.02391	(0.01854, 0.03645)	50000

Table 5 Predictive Inference for Black Non-Hispanic Female Breast Cancer Patients' Survival Data

		Black non-Hispanic
Summary statistics	Mean	69.7324
	Standard error	0.58029
Raw moments	m_1	69.7324
	m_2	5536.09
	m_3	519515
	m_4	5.60874×10^7
Corrected moments	μ_1	69.7324
	μ_2	673.488
	μ_3	39542.1
	μ_4	1.7636×10^6
Skewness & Kurtosis	β_1	5.11835
	β_2	3.88813
	γ_1	2.26238
	γ_2	0.88813
Predictive intervals	90%	(35.101, 150.305)
	95%	(31.220, 155.125)
	98%	(29.023, 159.452)
	99%	(27.170, 166.776)

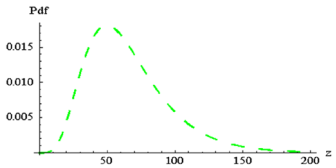

Figure 3. Predictive Density for a Single Future Survival Time with Respect to Black Non-Hispanic Survival Data

The descriptive statistics (frequency distribution) are shown in Table 1. Overall, Table 1a shows the state wise frequency and its corresponding percentages for the selected patients. The table shows the count of 2,000 patients for nine different states; Georgia, Hawaii, Iowa, Michigan, New Mexico, Utah, Washington, California, and Connecticut. Of the 2,000 patients that participated, 730 originated from Michigan, which accounted for 36.5% of the sample. Georgia had the second highest count of participants; 612, totaling 30.6% of patients. The remaining seven states contained 32.9% of the participants. California with 17.6% (352), Connecticut with 9.10% (182), Washington with 3.85% (77), Iowa with 1.15% (23), New Mexico, Utah, and Hawaii with 0.60%, 0.35%, and 0.25% respectively. Majority of patients were from the eastern region of the United States, Michigan and Georgia.

Table 1b, shows descriptive statistics (mean, standard deviation, median, quartiles, and variance) for some demographic characteristics (age at diagnosis, survival times, and marital status at diagnosis) of the selected random sample of Black non-Hispanic breast cancer patients. The mean age of diagnosis for Black non-Hispanics was 58.32 years, with a standard deviation of 14.43; (a median was observed of 57, with a variance of 208.24). The 25th percentile (lower quartile) was reported to be 47, the 50th percentile (second quartile) was reported to be 57, and the 75th percentile (upper quartile) was 69. Survival time, was another characteristic that was observed among the sample, the mean survival time was 66.76 months, which can be converted into 5.56 years with a standard deviation of 30.20; (a median was observed of 62, with a variance of 911.31). The 25th percentile (lower quartile) was reported to be 40, the 50th percentile (second quartile) was reported to be 62, and the 75th percentile (upper quartile) was 90. Marital status was the last demographic characteristic observed, this was classified into six categories and their respective frequencies are reported in Table 1b. The above frequencies show that majority of patients selected were married (727), accounting for about 36.35%. About 21.85% of the selected sample reported that they were single at the time of diagnosis, the second most reported status. Frequency for those separated, divorced, widowed, and unknown were; 52, 273, 427, and 84 respectively.

We used four types of advanced skewed statistical models to determine the best-fit statistical probability model for the survival data of black non-Hispanic female breast cancer patients. Upon completion of the analysis, the EW model demonstrated a better fit compared to the other models for black non-Hispanic survival data. Following the EW model, the breast cancer survival sample for black non-Hispanics had the lowest DIC value of 19080.750.

We used the Bayesian method to determine the inference for posterior parameters for the breast cancer survival model. Table 2 summarizes the use of MCMC to make statistical inferences for the posterior parameters for black non-Hispanic females. Under the selection of EWM, Mean (SD) values for α, β, and λ are 5.364 (0.5314), 1.083 (0.03749), and 0.02502 (0.005097), respectively.

Figure 2, shows the shape of kernel density for each of the parameters for black non-Hispanic females. These findings assist us in observing the posterior shape of the kernel density for the best-fit model parameters.

Black non-Hispanic females' future survival times, is graphically represented in Figure 3. The data for future survival times follow a positively skewed distribution. Table 3 summarizes the predictive raw and corrected moments, predictive skewness, kurtosis, and predictive intervals for future response for black non-Hispanic.

A sample of 2,000 Black non-Hispanic females were selected, where the mean survival age was 69.73 with a standard error of 0.58. Inferential statistics (raw moments, corrected moments, skewness and kurtosis) were calculated and reported. From table 3, the first 4 raw moments were observed, where =69.7324, m_2= 5536.09, m_3= 519515, and m_4= 5.60874×10^7. The first four corrected moments, were $\mu 1$= 69.73, μ_2= 673.488, μ_3= 39542.10, and μ_4= 1.7636×10^6. Skewness and kurtosis were reported to be; β_1= 5.11835, β_2= 3.88813, γ_1= 2.26238, and γ_2= 0.88813, respectively. Statistical significance is established based on 95% predictive survival intervals. Black non-Hispanics had a significantly increased risk of death. The predictive survival intervals were calculated for 90%, 95%, 98%, and 99%, where (35.101, 150.305), (31.220, 155.125), (29.023, 159.452), and (27.170, 166.776), respectively.

Given the patient's current and past medical profile, the best-fit model assists healthcare providers to predict a patient's future survival outcome. Therefore, health

care professionals can use both current knowledge and prediction capacity to adequately inform patients and their families, determine more appropriate treatment trajectory, enhance current healthcare facilities, and improve the allocation of resources, essentially improving the quality of life. This increase in scientific knowledge will provide key insight in the field of health services research. Enhancing health care facilities and improving how resources are spent are key aspects of health services research as the aim is to develop new technologies that advance the science and health care fields. Exclusion of three states will allow other researchers to perform external validation of our findings, since the information-theoretic criteria are essentially internal validations. Furthermore, future studies should be concentrated on identifying the best areas of intervention in order to significantly affect patient prognosis and improve survival times, primarily where health disparities have been identified.

References

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

Alteri R, Bandi P, Brinton L, Casares C, Cokkinides V, Gansler T. Breast cancer facts and figures 2011-2012. American Cancer Society, Atlanta, 2011.

Bradbury AR, Olopade OI (2007). Genetic susceptibility to breast cancer. *Reviews Endocrine Metabolic Disorders*, **8**, 255-67.

DeSantis C, Siegel R, Bandi P, Jemal A (2011) Breast cancer statistics, *CA*, **61**, 408-18.

Ferlay J, Soerjomataram I, Ervik M, et al (2012). Cancer incidence and mortality worldwide: IARC Cancer Base. GLOBOCAN, 1(11). Lyon, France: International agency for research on cancer; 2013.

Graeser MK, Engel C, Rhiem K, et al (2009). Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *J Clinical Oncol*, **27**, 5887-92.

Jemal A, Bray F, Center MM, et al (2011). Global cancer statistics. *CA*, **61**, 69-90.

Khan HMR, Saxena A, Shrestha A (2014). Posterior inference for the white Hispanic breast cancer survival data. *J Biomet Biostat*, **5**, 1-6.

Khan HMR, Saxena A, Rana S, Ahmed NU (2014). Bayesian modeling for male breast cancer data. *Asian Pac J Cancer Prev*, **15**, 663-69.

Khan HMR, Saxena A, Kemesha G, Rana S, Ahmed NU (2014). Mode-based survival estimates of female breast cancer data. *Asian Pac J Cancer Prev*, **15**, 2893-900.

Khan HM, Saxena A, Gabbidon K, Stewart TS, Bhatt C (2014d). Survival analysis for white non-Hispanic female breast cancer patients. *Asian Pac J Cancer Prev*, **15**, 4049-54.

Khan HMR (2012). Estimating predictive inference for responses from the generalized Rayleigh model based on complete sample. *J Thai Statistician*, **10**, 53-68.

Khan HMR .(2012) Several priors based inference from the exponential model under censoring. *JP J Fund Appl Stat*, **2**, 1-13.

Khan HMR (2013). Comparing relative efficiency from a right skewed model. *JP J Biostat*, **9**, 1-26.

Khan HMR (2013). Inferential estimates from the one-parameter half-normal model. *J Thai Statistician*, **11**, 77-95.

Khan HMR, Albatineh AN, Alshahrani S, Jenkins N, Ahmed NU (2011) Sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients. *Computational Statistics Data Analysis*, **55**, 3093-103.

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

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. *J Clin Oncol*, **28**, 4410-6.

Liu L, Zhang J, Wu AH, Pike MC, Deapen D (2012). Invasive breast cancer incidence trends by detailed race/ethnicity and age. *Int J Cancer*, **130**, 395-404.

Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010). Genetic susceptibility to breast cancer. *Molecular Oncology*. **4**, 174-91.

Peairs KS, Barone BB, Snyder CF, et al (2011). Diabetes mellitus and breast cancer outcomes: a systematic review and meta-analysis. *J Clin Oncol*, **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. *Br Ca Res Treat*, **123**, 627-35.

Robertson A, Ragupathy SK, Boachie C, et al (2011). The clinical effectiveness and costeffectiveness of different surveillance mammography regimens after the treatment forprimary breast cancer: systematic reviews registry database analyses and economic evaluation. *Health Technology Assessment*, **15**, 1-322.

Sankaranarayanan R, Ferlay J (2013). Burden of breast and gynecological cancers in low-resource countries. *Breast and Gynecological Cancers*, 1-17.

Siegel R, DeSantis C, Virgo K, et al (2012). Cancer treatment and survivorship statistics. *CA*, **62**, 220-41.

Smith RA, Brooks D, Cokkinides V, Saslow D, Brawley OW (2013). Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening. *CA*, **63**, 88-105.

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

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

Walker MJ, Mirea L, Cooper K, et al (2013). Impact of familial risk and mammography screening on prognostic indicators of breast disease among women from the Ontario site of the Breast Cancer Family Registry. *Familial Cancer*, **13**, 163-72.

WHO (2013). Charts for the 10 leading causes of death for women worldwide and by income group, Geneva. Retrieved from http://www.who.int/nmh/publications/ncd_report_full_en.pdf. Accessed January 2014.

WHO (2010). World health statistics. Retrieved from http://www.who.int/gho/publications/world_health_statistics/EN_WHS10_Full.pdf?ua=1, Accessed January 2014.