

RESEARCH ARTICLE

Survival Analysis for White Non-Hispanic Female Breast Cancer Patients

Hafiz Mohammad Rafiqullah Khan^{1*}, Anshul Saxena², Kemesha Gabbidon², Tiffanie Shauna-Jeanne Stewart³, Chintan Bhatt⁴

Abstract

Background: Race and ethnicity are significant factors in predicting survival time of breast cancer patients. In this study, we applied advanced statistical methods to predict the survival of White non-Hispanic female breast cancer patients, who were diagnosed between the years 1973 and 2009 in the United States (U.S.). **Materials and Methods:** Demographic data from the Surveillance Epidemiology and End Results (SEER) database were used for the purpose of this study. Nine states were randomly selected from 12 U.S. cancer registries. A stratified random sampling method was used to select 2,000 female breast cancer patients from these nine states. We compared four types of advanced statistical probability models to identify the best-fit model for the White non-Hispanic female breast cancer survival data. Three model building criterion were used to measure and compare goodness of fit of the models. These include Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC). In addition, we used a novel Bayesian method and the Markov Chain Monte Carlo technique to determine the posterior density function of the parameters. After evaluating the model parameters, we selected the model having the lowest DIC value. Using this Bayesian method, we derived the predictive survival density for future survival time and its related inferences. **Results:** The analytical sample of White non-Hispanic women included 2,000 breast cancer cases from the SEER database (1973-2009). The majority of cases were married (55.2%), the mean age of diagnosis was 63.61 years (SD = 14.24) and the mean survival time was 84 months (SD = 35.01). After comparing the four statistical models, results suggested that the exponentiated Weibull model (DIC= 19818.220) was a better fit for White non-Hispanic females' breast cancer survival data. This model predicted the survival times (in months) for White non-Hispanic women after implementation of precise estimates of the model parameters. **Conclusions:** By using modern model building criteria, we determined that the data best fit the exponentiated Weibull model. We incorporated precise estimates of the parameter into the predictive model and evaluated the survival inference for the White non-Hispanic female population. This method of analysis will assist researchers in making scientific and clinical conclusions when assessing survival time of breast cancer patients.

Keywords: Breast cancer survival data - statistical probability models - Bayesian inference - predictive inference

Asian Pac J Cancer Prev, 15 (9), 4049-4054

Introduction

Cancer accounts for approximately 20% of all cause specific mortality in the world (WHO, 2013a). Cancer has been identified as the predominant cause of death in developed nations and is also the second most common cause of death in developing nations due to limited resources (Jemal et al., 2011).

Nearly 32.6 million people were diagnosed with cancer in the past 5 years prior to 2012. In the same year, 14.1 million were diagnosed as new cancer cases, and an estimated 8.2 million people died due to cancer. Among these, female breast cancer accounted for nearly 522,000 deaths and 1.67 million new cases. Responsible for nearly

14.3% of all cancer related deaths worldwide, breast cancer is also the most fatal cancer among females aged between 20-59 (Ferlay et al., 2013; WHO, 2013b).

Globally, the projections for new breast cancer cases in the year 2020 indicate an increase of 18%; and in the United States, a 14.4% increase is projected (Ferlay et al., 2013). It has been suggested that during the year 2000 to 2009, the overall annual percentage change (APC) in breast cancer incidence has increased in the U.S. For women aged 40-49, most racial and ethnic groups showed an overall APC increase of 1.1% ($p < 0.001$) (Hou and Huo, 2013).

Breast cancer incidence is increasing among women in the U.S. In the year 2012, an estimated 226,870 new

¹Department of Biostatistics, ²Department of Health Promotion and Disease Prevention, ³Department of Dietetics and Nutrition, ⁴Department of Epidemiology, Robert Stempel College of Public Health and Social Work Florida International University, Florida, USA *For correspondence: hmkhan@fiu.edu

breast cancer cases and 39,510 deaths occurred due to the breast cancer (ACS, 2013; Siegel et al., 2014). In the year 2010, breast cancer patients in the U.S. spent around \$16.5 billion in care and medical services; which is approximately 13% of the total estimated national expenditure on cancer care. It is projected that this figure will reach \$23.34 billion in the year 2020, and is going to be the second largest expenditure in medical care for cancer (Mariotto et al., 2011; Montero et al., 2012). About 14% of women in the U.S. are expected to develop breast cancer in their lifetime and the associated costs of care will affect their survival (ACS, 2013). Among these women, breast cancer affects ethnicities at different rates (DeSantis et al., 2011). A higher incidence of 417 cases per 100,000 is reported among White females compared to other racial groups (White et al., 2013).

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).

The Centers for Disease Control and Prevention (CDC) has identified that between the years 1999-2010, White women accounted for the highest incidence of breast cancer and the third highest mortality rates from the disease (CDC, 2012). However, White women have a higher five-year survival rate when compared to other racial groups. Death rates among White women have declined 2% from 1997 to 2007; this decline is not found in other racial or ethnic groups (DeSantis et al., 2011). Research shows that White women, older than age 40 had higher rates of breast cancer compared to black women in the same age group (Clarke et al., 2012).

According to the American Cancer Society, the incidence of age-adjusted breast cancer vary greatly from state to state for example, they may range from 47.79 cases per 100,000 in Connecticut to 20.25 cases in rural Georgia, the death rates range from 27.6 per 100,000 in Alaska to 17.5 per 100,000 in New Hampshire (ACS, 2013). The contrast in incidence and mortality rates among race, ethnicity and other determinants demonstrate a need for statistical modeling to predict the survival times. Patients diagnosed with breast cancer, visit clinics, healthcare units, and hospitals to receive modern treatments to improve their prognosis. Advancements in modern technology are able to help patients determine survival

days. In addition, there is a huge demand for new uses of statistical analyses in order to facilitate new discoveries, diagnosis, and treatment planning.

The objectives of this paper are: *i*) to analyze demographic variables of the selected sample; *ii*) to demonstrate that the breast cancer survival data follows a specific probability model by using model selection criteria for goodness of fit tests; *iii*) to perform the Bayesian analysis of the posterior distribution for the parameters; and *iv*) to derive Bayesian survival inference for future times by using the best fit model.

Materials and Methods

We used breast cancer data (N=657,712) from Surveillance, Epidemiology and End Results website (SEER, 2010). The data contains information for breast cancer patients from 1973 to 2009, and covers 12 cancer registries among the 50 U.S. states. A stratified random sampling method was used to select nine of the 12 available states to provide a representative sample of White non-Hispanic categories (Figure 1). The total SEER data included 4,269 males and 653,443 females. Among the 608,032 total females, 22,639 were White Hispanic, and 531,562 were White non-Hispanic. Men were not included in this study due to the small chance (0.70% of total patients) of breast cancer occurring within this group.

Selected patients and their demographic characteristics

A random sample of 2,000 White non-Hispanic cancer cases was included in the data analysis representing nine states in the U.S. (Figure 1). Simple random sampling (SRS) methodology was used to select a representative sample of patients and to minimize selection bias.

Health care professionals use various statistical probability modeling techniques to determine the prognosis of cancer patients. Often times the data stored in various cancer registries or databases are utilized. Survival data taken from these databases follow several statistical probability models, for example exponential, Weibull, Exponentiated Exponential (EE), Exponentiated Weibull (EW), Beta Generalized Exponential (BGE), Beta inverse Weibull (BIW) model, among others. For accurate predictions, it is imperative that the data fit the appropriate model. Different racial and ethnic groups may follow different distribution patterns and so it is important to use

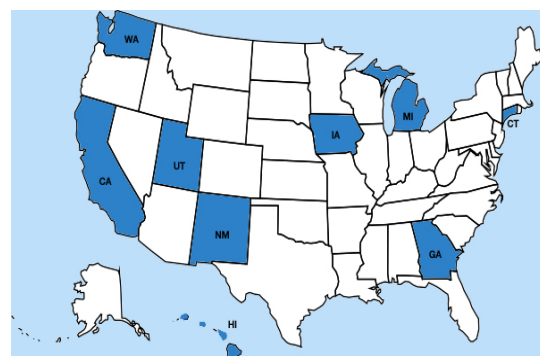


Figure 1. White Non-Hispanic Breast Cancer Cases (n=2,000) Randomly Selected from the Nine States (darker blue color represents the selected states)

statistical methodologies to draw clinical inferences.

Khan et al. (2014a), discussed in details four types of statistical probability models. These models include EEM, EWM, BGEM, and BIW, which were briefly used in this study. There are two parameters, shape ($\alpha > 0$) and scale ($\lambda > 0$) for the EEM (Khan et al. 2014b). The Weibull model has three parameters, $\alpha > 0$ and $\beta > 0$ are the shape parameters, and $\lambda > 0$ is the scale parameter (Khan et al. 2014c). The beta generalized exponential model has four parameters, where the shape parameter, $\alpha > 0$ and the scale parameter, $\lambda > 0$, and additional two parameters, $a > 0$ and $b > 0$ are essential for varying tail weight and to present skewness (Barreto-Souza et al., 2010). The beta inverse-Weibull (BIW) model is another type of statistical probability model, where β is the shape parameter, and two extra parameters, $a > 0$ and $b > 0$, are used to introduce skewness and tail weight (Khan et al. 2014c, 2014d).

In exploration of the posterior probability for the parameters from the EEM, BGEM, EWM and BIWM, an innovative Bayesian method may be used to achieve posterior inference. In the healthcare research field, Bayesian statistics have become more popular because of its use of parametric and model-based inference, and its applicability to clinical diagnostics, potentially improving the field of translational research. Data and model parameters are random variables in the Bayesian estimation technique; data is termed as "observed" and parameters are termed as "unobserved" variables. The joint distribution of the posterior parameters is determined by multiplication of the likelihood and prior. The likelihood relies on the model of underlying process given any values for parameters, it is measured as a conditional distribution that specifies the probability of the observed data. Prior and likelihood combine all the available information about the parameters, and manipulates the joint distribution in many ways and makes inference about the parameters given the data. Given a set of observed data, the Bayesian inference develops the posterior distribution for the parameters which allows population predictions when applied to datasets. For more information regarding Bayesian method and its inference, the readers are referred to other works (Khan et al., 2012a; 2012b; 2013a; 2013b).

SPSS software (IBM SPSS, 2011) was used to gather descriptive statistics. Nine out of 12 states were used to extract data and a geographic map was drawn for White non-Hispanic women cases using the Google fusion table (Gonzalez et al., 2010). Mathematica version 8.0 (Wolfram Research, 2012), an advanced computational software produced a graphical representation of the predictive density for a single future survival time for the selected sample. Furthermore, it was used to obtain additional predictive inferences for the survival times. To assess goodness of fit, summary results of the posterior parameters, and to execute related calculations WinBugs software (MRC Biostatistics Unit, 2013) was used.

Results

Frequency of the sample of breast cancer cases ranged from 366 (Connecticut) to 25 (Hawaii), and thus the percentage of White non-Hispanic females with breast

cancer cases in the analysis varied from 1.3% to 18.3% for various states (Table 1). The second highest percentage of selected patients was observed from Washington and the next highest from Michigan.

The quartiles for the mean age at diagnosis were 52, 63, and 75 years, respectively (Table 2). The mean and median age at diagnosis were close and suggested an approximate normal distribution. The median survival time was 87 months and ranged between 38-160 months for White non-Hispanic females. The majority of the females in the analytical sample were married (55%).

Best fit survival model and its parameters

Khan et al. (2014b) discussed in details about three model selection criteria; Akaike Information Criterion (AIC), Deviance Information Criterion (DIC), and Bayesian Information Criterion (BIC). The DIC, a measure of fit is widely used for the comparison of different models. The Markov Chain Monte Carlo (MCMC) method is used to attain the posterior distribution of parameters for the sample. The DIC values can be either positive or negative, however, the model with smaller values is considered better than those with larger values. As with DIC and AIC, the model containing the lower Bayesian Information Criterion (BIC) values is considered better between any two estimated models. Bayesian Information

Table 1. Frequency Distribution of Nine Selected States for White Non-Hispanic Breast Cancer Cases

States	White non-Hispanic female	
	Frequency	%
Georgia	148	7.4
Hawaii	25	1.3
Iowa	278	13.9
Michigan	327	16.3
New Mexico	96	4.8
Utah	139	7
Washington	348	17.4
California	273	13.6
Connecticut	366	18.3
Total	2,000	100

Table 2. Descriptive Statistics for White Non-Hispanic Breast Cancer Cases

Characteristics	Categories	White non-Hispanic
Age at diagnosis (years)	Mean	63.31
	SD	14.24
	Median	63
	Range	15-104
	Quartile (Q ₁ , Q ₂ and Q ₃)	52, 63 and 75
	Variance	202.74
Survival time (months)	Mean	84.17
	SD	35.01
	Median	87
	Range	38-160
	Quartile (Q ₁ , Q ₂ and Q ₃)	60, 87 and 118
	Variance	1225.5
Marital status at diagnosis	Single	182
	Married	1103
	Separated	29
	Divorced	153
	Widowed	464
	Unknown	69

Criterion (BIC) is an asymptotic result assumed that the distribution of data is an exponential family.

We used log-likelihood function of the models in WinBugs (MRC Biostatistics Unit, 2013) and applied them for White non-Hispanic survival data. The AIC, BIC, and DIC values are calculated and the summary results of the measures of goodness of fits are reported in Table 3.

Table 3 consists of AIC, BIC, and DIC values for the EE, EW, BGE, and BIW models. Comparing the estimated values of all AIC, BIC, and DIC for the models, the EWM fits better for the survival times because it produces the smallest values of AIC and DIC.

Table 4 presents the summary results (Mean, SD, MC Error, Median, and Confidence Intervals) of the parameters in the case of best-fit exponentiated Weibull model for White non-Hispanic female breast cancer cases.

Figure 2 displays the graphical representation of the parameters. It is noted that all the parameters produced skewed distributions. Parameter values when plotted showed that alpha and beta are negatively skewed, and lambda is positively skewed. The range (95% CI) of these parameter values are described in Table 4. After 50,000 iterations, the kernel density appears smoothened.

Table 3. Selection of the Models for White Non-Hispanic Females on the Basis of AIC, BIC, and DIC Criteria

Model criteria	AIC	BIC	DIC
Exponentiated exponential	19826.4	19837.001	19826.378
Exponentiated Weibull	19821.7	19838.102	19818.22
Beta generalized exponential	19828.5	19850.903	19824.524
Beta inverse Weibull	19866.3	19883.102	19860.298

Table 4. Summary Results of the Posterior Parameters in the Case of Exponentiated Weibull for White Non-Hispanic Females Breast Cancer Cases (n=2,000; sample=50,000)

Node	Mean	SD	MC error	Median	95% CI
Alpha	7.25	0.1189	0.001082	7.28	(6.949, 7.385)
Beta	1.092	0.00643	1.74E-04	1.093	(1.076, 1.101)
Lambda	0.0189	5.34E-04	1.49E-05	0.01875	(0.01833, 0.02031)

Table 5. Future Inference Based on the EWM for White Non-Hispanic Breast Cancer Cases Survival Data

	Summary	White non-Hispanic
Raw moments	Mean	90.836
	Standard Error	0.860919
	m_1	90.836
	m_2	9733.54
	m_3	1.21411×10^6
Corrected moments	m_4	1.73738×10^8
	μ_1	90.836
	μ_2	1482.36
	μ_3	60649.6
Skewness & Kurtosis	μ_4	1.02316×10^7
	β_1	1.12925
	β_2	4.65624
	γ_1	1.06266
	γ_2	1.65624
Survival intervals	90%	(33.6239, 156.1253)
	95%	(28.6290, 170.4252)
	98%	(19.9019, 200.7740)
	99%	(15.6533, 219.8770)

Survival inference

We developed a predictive survival model using the results of the best-fit model to demonstrate the survival time of White non-Hispanic women with breast cancer patients. By using the values of model selection criterion found in Table 3, it is identified that the data follows the exponentiated Weibull model. Applying the Bayesian survival model, we assume the data set $x=(x_1, \dots, x_n)$ represent n White non-Hispanic female breast cancer cases survival days that follow the exponentiated Weibull model, and let y be a future survival time, then following Khan et al. (2011), the predictive density of y given the observed White non-Hispanic survival data x , is given by $p(y | x) = \int p(y | \alpha, \beta, \lambda) p(\alpha, \beta, \lambda | x) d\lambda d\beta d\alpha$, where, $p(\alpha, \beta, \lambda | x)$ is the posterior density function, and $p(y | \alpha, \beta, \lambda)$ represents the probability density function for a future survival time (y) that is defined from the best fit exponentiated Weibull model.

Graphical representation of the predictive density is shown in Figure 3 based on the survival times of White non-Hispanic cases.

Figure 3 presents the predictive survival density function. The predictive density (Figure 3) for the survival

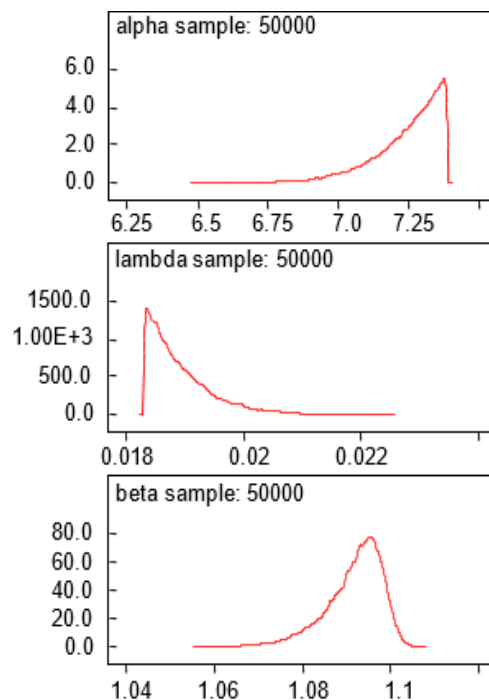


Figure 2. Kernel Density of the Posterior Parameters in the Case of Exponentiated Weibull for White Non-Hispanic Females Breast Cancer Cases (n=2,000)

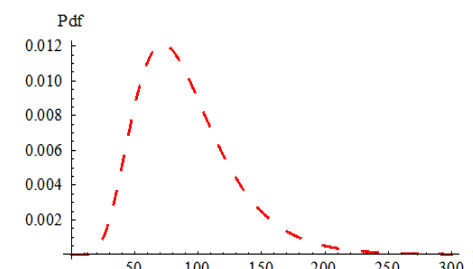


Figure 3. Predictive Density for White Non-Hispanic Female Breast Cancer Patient's Survival Data

times appears to be unimodal and it is positively skewed, ranging from 28.629 to 170.4252 (95% CI).

Table 5 contains summary statistics for future survival times of the patients. We identified that the survival times are higher for future patients compared with existing diagnosed patients. We obtained the raw and corrected moments for the survival inference for future White non-Hispanic female breast cancer patients. Since Kurtosis is <3 , we can assume that the future survival values follow a platykurtic distribution. The data points represent a flat distribution as compared to a normal distribution, which has a wider peak. Since Skewness is >0 , majority of the data falls to the left of the mean, with extreme values to the right.

Discussion

For White non-Hispanic women diagnosed with breast cancer between the years 1973 to 2009 in the U.S., several statistical models were used to show the best-fit for the breast cancer survival data. The sample consisted of 2,000 White non-Hispanic women; stratified random sampling was used at the state level and simple random sampling used within the nine states.

The Mean (SD), age (in years) at diagnosis for breast cancer cases was 63.31 (14.24), with age 15 being the minimum age at diagnosis for White non-Hispanic women. The highest mortality rates among women diagnosed with breast cancer are those that are 50 years or older (SEER, 2010). Survival time ranged from 38 to 160 months, with a Mean (SD), 84.17 (35.01) months. The majority of these cases were married.

To speed up performance of the Bayesian posterior parameters and to draw their corresponding dynamic kernel densities, a reparameterization method was used for the exponentiated Weibull model. After running 50,000 Monte Carlo repetitions reported with negligible MC errors we obtained posterior inference for the parameters.

Given the breast cancer survival model, we were able to determine the inference for posterior parameters using the Bayesian method. By using the Markov Chain Monte Carlo method, the inferences for the posterior parameters for the best-fit model are reported for White non-Hispanic females.

Based on the goodness of fit analysis, the breast cancer survival sample for White non-Hispanic women followed the exponentiated Weibull (EW) distribution. The lowest DIC value was 19818.220, indicating the best goodness of fit. In this case under the selection of EW distribution, Mean (SD) values for α , β , and λ are 7.25 (0.1189), 1.092 (0.00643), and 0.0189 (5.34×10^{-4}), respectively. The dynamic kernel density for each of the parameters is reported for White non-Hispanic females so that one can observe the shape of the kernel density. It was noticed that all parameters displayed skewed distributions.

In the case of the survival inference, the best fit statistical survival model and the Bayesian method were used to derive a predictive model for a single future survival time. A summary table for the predictive mean, standard error, and 95% future survival intervals are provided on the basis of the predictive density. According

to the results, the shape of the future survival model for White non-Hispanic women is positively skewed.

Figure 3 shows the graphical representations of White non-Hispanic female future survival times using the exponentiated Weibull distribution in the Bayesian method. Higher survival times are identified for White non-Hispanic women compared to the existing survival times. For ethnicity, we report the predictive raw and corrected moments, predictive skewness, and kurtosis for future survival time in Table 5. The model is able to predict survival times accurately within 90%-99% confidence intervals while taking into account multiple parameters.

We identified a data-based statistical probability model from the 1973-2009 SEER database to demonstrate the effectiveness of predicting breast cancer survival data for White non-Hispanic women. Statistical probability models are important for posterior model parameters in order to predict survival times among ethnicity and for describing inferences for observations. To best determine a fitted model, methods for measuring the goodness of fit tests are imperative in the selection of the best statistical probability models for survival samples of ethnicity. AIC, BIC, and DIC model selection criteria were used to develop statistical probability model for ethnicity.

These findings will be beneficial to healthcare researchers and practitioners to aid in the prediction of a patient's possible survival time given the patient's current state and medical history. Therefore, the findings may work to improve knowledge, demonstrate scientific discovery, and innovation. This may improve the diagnosis and treatment of breast cancer cases within the United States and the world.

References

- ACS (2013). Cancer Facts and Figures. Atlanta: American Cancer Society. Retrieved from: <http://www.cancer.org/research/cancerfactsstatistics/breast-cancer-facts-figures>. [Accessed September 15, 2013]
- Barreto-Souza W, Santos AHS, Cordeiro GM (2010). The beta generalized exponential distribution. *J Stat Comput Sim*, **80**, 159-72.
- CDC (2012). Breast Cancer Rates by Race and Ethnicity. Available at: <http://www.cdc.gov/cancer/breast/statistics/race.htm>. Accessed June 12, 2012.
- Clarke CA, Keegan THM, Yang J, et al (2012). Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *J Natl Cancer Inst*, **104**, 1094-101.
- DeSantis C, Siegel R, Bandi P, Jemal A (2011). Breast cancer statistics. *CA Cancer J Clin*, **61**, 409-18.
- Ferlay J, Soerjomataram I, Ervik M, et al (2013). Cancer Incidence and Mortality Worldwide: IARC Cancer Base. GLOBOCAN 2012, 1 (11). Lyon, France: International Agency for Research on Cancer. Available from: <http://globocan.iarc.fr>. [Accessed December 20, 2013].
- Gonzalez H, Halevy AY, Jensen CS, et al (2010). Google fusion tables: web-centered data management and collaboration. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1061-66). New York, NY, USA: ACM.
- Graeser MK, Engel C, Rhiem K, et al (2009). Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *J Clin Oncol*, **27**, 5887-92.
- 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.
- IBM SPSS (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.
- Jemal A, Bray F, Center MM, et al (2011). Global cancer statistics. *CA Cancer J Clin*, **61**, 69-90.
- Khan HMR, Saxena A, Shrestha A (2014a). Posterior inference for White Hispanic breast cancer survival data. *J Biomet Biostat*, **5**, 183.
- Khan HMR, Saxena A, Rana S, Ahmed NU (2014b). Bayesian modeling for male breast cancer data. *Asian Pac J Cancer Prev*, **15**, 663-9.
- Khan HMR, Saxena A, Kemesha G, Rana S, Ahmed NU (2014c). Mode-based survival estimates of female breast cancer data. *Asian Pac J Cancer Prev*, **15**, 2893-900.
- Khan HMR, Saxena A, Kemesha G, Elizabeth R, Alice S (2014d). Statistical applications for White Hispanic breast cancer survival data. *Asian Pac J Cancer Prev*, [Epub ahead of print].
- 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, 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.
- 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.
- Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML (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 U.S. managed care perspective. *Breast Cancer Res Treat*, **134**, 815-22.
- MRC Biostatistics Unit (2013). WinBUGS software version 1.4.3: The BUGS Project.
- 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. *Breast Cancer Res Treat*, **123**, 627-35.
- Robertson A, 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.
- SEER. (2010). Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2009), National Cancer Institute, DCCPS, Surveillance Research Program. [Accessed July 10, 2013].
- Siegel R, Ma J, Zou Z, Jemal A (2014). Cancer statistics, 2014. *CA Cancer J Clin*, **64**, 9-29.
- 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 Cancer J Clin*, **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.
- 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. *Fam Cancer*, **13**, 163-72.
- White A, Richardson LC, Krontiras H, Pisu M (2013). Socioeconomic disparities in breast cancer treatment among older women. *J Womens Health (Larchmt)*, **23**, 335-41.
- WHO (2013a). Cause-specific mortality and morbidity: age-standardized mortality rate by cause by World Bank income group. Geneva: WHO.
- WHO (2013b). Charts for the 10 leading causes of death for women worldwide and by income group. Geneva: WHO.
- Wolfram Research (2012). The Mathematica Archive: Mathematica 8.0.