FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 6 No. 3, June, 2022, pp 95 - 99
DOI: https://doi.org/10.33003/fjs-2022-0603-975

RISK FACTOR ANALYSIS OF BREAST CANCER PATIENTS IN A NIGERIAN TERTIARY HOSPITAL

***[1]Aako, O. L., [2]Adewara, J. A. and [1]Are, S. O.**

[1]Mathematics and Statistics Department, Federal Polytechnic, Ilaro, Ogun State, Nigeria
[2]Distance Learning Institute, University of Lagos, Akoka, Nigeria

*Corresponding authors' email: olubisi.aako@federalpolyilaro.edu.ng

ABSTRACT

This research attempts to determine the survival rate and identify the risk factors affecting the survival of breast cancer patients. The study used 334 breast cancer patients from University College Hospital (UCH), Ibadan. The Kaplan-Meier estimator result shows that 70 patients survived after treatment, the median survival age is 58 years old and the median survival time is 1423 days which is equivalent to 4 years with survival probability of 0.455. Three survival models namely; Cox proportional, Weibull and Log-logistic regression models were fitted to the data. Cox proportional model has the lowest value of Akaike Information Criterion (AIC) (724.6865) and Bayesian Information Criterion (BIC) (745.1765) is considered the best fitted model. The result of Cox proportional model shows that the prognostic factors affecting the risk of dying from breast cancer are Histology (MC), Tumor stage II, Surgical type (MRM) and Surgical type (SM).

Keywords: Breast cancer, Cox model, Parametric model, Prognostic factor, Survival

INTRODUCTION

There are overwhelming different types of cancers known that affect humans, especially, female. The results of findings confirm Breast Cancer the commonest type of cancer (McPherson et al., 2000). Breast cancer is a menace among women in advanced as well as evolving countries and the next leading source of cancer mortality after lung cancer (Montesarrat et al., 2009). Breast cancer is a complex disease and it is difficult to predict what course it will take (Anders et al., 2009). Early diagnosis of breast cancer is important in handling the disease and in determining the prognosis (El Saghir et al., 2011). Hence, timely screening is the unsurpassed approach to regulate the effect of breast cancer and increase the survival of women suffering from breast cancer. Assessing the survival probability and the factors that amplify the survival of women with breast cancer might aid early discovery actions, improve treatment and care in the area of health.

Several studies have been done in survival data analysis using parametric and Semi-parametric survival models. Moghimi-Dehkordi et al. (2008) used Cox, Weibull, Exponential and Lognormal survival models to assess factors affecting the survival of stomach cancer patients. Popoola et al. (2012) studied survival of breast cancer patients in Nigeria in order to determine the trend movement of the disease, test for the significance in the distribution of survival time and estimate the probability of survival. Zare et al. (2012) studied parametric log-logistic model of breast cancer in Southern Iran. Vallinayagam et al. (2014) evaluated common parametric models in modeling breast cancer survival data. Lognormal model was considered the best among the models. Nikpour et al. (2016) used parametric models to evaluate factors influencing the survival of gastric cancer patients using cox, Weibull, Exponential, Log-normal, Logistics and gamma models. Weibull model was identified as the best for the survival data.

Teshnizi et al. (2017) assessed the risk factors affecting pediatric cases of acute leukemia using cox, exponential, Weibull, log-logistics, log-normal, Gompertz and generalized gamma regression models. Syahila et al. (2017) studied the survival of breast cancer patients using different parametric models. Log-logistics model was considered as the best fitted model for the data. Ezekiel and Aako (2018) compared Cox and Weibull models in assessing the factors affecting survival of asthmatic patients. The result of AIC and Log likelihood showed that Weibull regression model fit the asthma patients' data more than Cox Proportional hazard model. Adeboye et al. (2020) assessed Cox, Weibull, Log-logistic, Log-normal and Gompertz models in determining the prognostic factors affecting the length of stay of asthmatic patients in hospital. Log-normal regression model was considered the best model for the data.

Musa et al. (2020) examined the factors which affect under-five mortality in North Central Geo-Political Zone of Nigeria, using Cox proportional hazards regression model. The results showed that contraceptive used of mother, birth weight and type of toilet facility being used by the family and place of residence were significantly associated with under-five child mortality. Falgore et al. (2022) accessed some risk factors of Rheumatoid Arthritis using logistic regression. The results showed that alcoholism and age have a significant association with Rheumatoid Arthritis. In this study, prognostic factors that could lead to early death of an individual with breast cancer would be investigated using semi-parametric and parametric survival regression models. The breast cancer patient's survival probability would be determined with the application of Kaplan-Meier estimator.

MATERIALS AND METHODS

Data Source

Secondary data consisting of records of 334 patients who have suffered from breast cancer at the cancer units of University College Hospital (UCH), Ibadan from January 2016 - December 2020 will be considered in this study. UCH serves as the major referral center for most Nigeria hospitals.

Kaplan-Meier (KM) Non-Parametric Model

The Kaplan-Meier estimator (Kaplan and Meier, 1958) is used to estimate the survival function from lifetime data. In medical research, it is used to measure the fraction of patients living for a certain amount of time after treatment.

The survival function $S(t)$ of the estimator is:

$$\hat{S}(t) = \prod_{i:ti<t} < t\frac{ni-di}{ni} \quad (1)$$

where t_i is a time when at least one event happened, d_i is the number of event (deaths) that happened at time t_i and n_i is the individual known to survive at time t_i.

Cox Proportional Semi-Parametric Regression Model

The Cox regression model is suitable for modeling the time to a specified event, based upon the values of given covariates. Cox regression uses censoring which is vital in analyzing real life situation (Kleinbaum and Klein, 2012). It is called a semi-parametric model because though the; distribution of the outcomes may not be known, the regression parameters are known.

The Cox proportional hazard is given by:

$$h(t;x) = h_o(t)e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p} \quad (2)$$

where $h(t;x)$ = expected hazard at time t, $h_o(t)$ = baseline hazard and $X_1, \ldots, X_p$ are independent variables.

Weibull Regression Model

Weibull regression model is a parametric approach which has its advantage in the flexibility and simplicity of hazard functions and survival functions. The shape and scale parameter of the distribution are β and λ respectively.

The hazard of Weibull distribution is:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1} \quad (3)$$

where $\lambda > 0$ and $\beta > 0$.

The probability density function $f(t)$ and survival function $S(t)$ of the distribution are respectively:

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}e^{-(\lambda\beta)\beta} \quad (4)$$

and

$$S(t) = e^{-(\lambda\beta)\beta} \quad (5)$$

Log-logistics Parametric Model

The log-logistic model is a statistical regression model for a non-negative outcome variable. It is applicable in cases where the logarithmic outcome variable follows a logistics distribution and especially in a situation where the hazard rates first increase and later decrease.

The survival function of log-logistic model is

$$S(t) = 1 - F(t) = \left[1 + \left(\frac{t}{\alpha}\right)^{\beta}\right]^{-1} \quad (6)$$

the probability density function is

$$f(t) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{2}} \quad (7)$$

and so, the hazard function is

$$h(t) = \frac{f(t)}{S(t)} = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{1+\left(\frac{t}{\alpha}\right)^{\beta}} \quad (8)$$

Model selection Criterion

Akaike Information Criteria (AIC)

Model selection criteria used in this study are Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC).

$$AIC = -2 * \log(likelihood) + 2(k) \quad (9)$$

$$BIC = -2 * \log(likelihood) + k\log(n) \quad (10)$$

where n = sample size and k = number of parameters.

RESULTS AND DISCUSSION

Description of Variables

The time until death of a breast cancer patient (days) is considered as the dependent variable and Age, Gender, Surgery type, Survival status, Histology, Tumor stage of breast cancer patients are considered as independent variables. The Survival status of Cancer Breast Patients was coded as "1" Death of breast cancer patient and "0" survived breast cancer patient. The Tumor stage was coded as "0" patients with tumor stage I, "1" patients with tumor stage II and "2" patients with tumor stage III. Gender was coded as "0" for male and "1" for female.6 Surgery type was coded as "0" for modified radical lumpectomy (LP), "1" for mastectomy (MRM) and "2" for simple mastectomy (SM) and "3" for others. Histology was coded as "0" for infiltrating ductal carcinoma (IDC), "1" for infiltrating lobular carcinoma (ILC) and "2" for mucinous carcinoma (MC). The descriptive analysis is as presented in Table 1.

Table 1: Descriptive analysis of breast cancer patients

Factor	Frequency	Percentage (%)
Gender		
Female	330	98.8
Male	4	1.2
Total	334	100.0
Tumor stage		
I	64	19.2
II	189	58.6
III	81	24.3
Total	334	100.0
Histology		
MC	12	3.4
ILC	89	26.6
IDC	233	69.8
Total	334	100.0
Surgery type		
MRM	96	28.7
LP	66	19.8
SM	67	20.1
Other	105	31.4
Total	334	100.0

From Table 1, 330 (98.8%) were females while the remaining 4 (1.2%) were males. 64 (19.2%) patients at stage I, 189 (58.6%) patients at stage II and 81 (24.3%) at stage III. 12 (3.4%) of the patients are affected with MC type of breast cancer, 89 (26.6%) of the patients are affected with ILC type of breast cancer and 233 (69.8%) are affected with IDC type of breast cancer. 96 (28.7%) of the patients undergone MRM, 66 (19.8%) have undergone LP surgery type, 67 (20.1%) have undergone SM surgery type and 105 (31.4%) have undergone other types of surgery.

Table 2: Descriptive Statistics of Age

	Min	Q_1	Median	Mean	Q_3	Max
Age	29	49	58	58.9	68	90

Table 2 shows the basic summary of age of the patient, it can be deduced that the mean age of the patients with breast cancer is approximately 60 years, the minimum age is 29 years while the maximum age 90 years and the median age is 58 years.

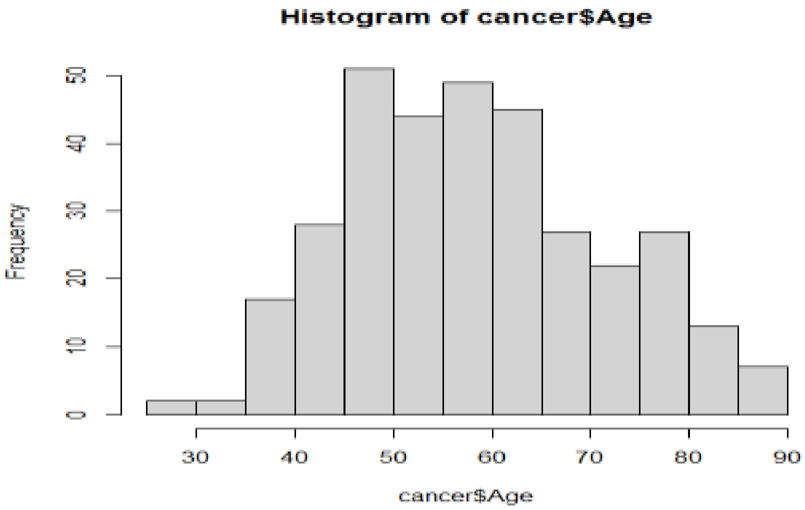

Figure 1: Histogram plot of Age Breast cancer's patient

Figure 1 is the histogram plot of the breast cancer patients age which shows that the patients that commonly have breast cancer is between 45 to 65.

Table 3: Median for Survival Time of the Data

n= observation	events	median	lower limit	upper limit
334	72	1423	1223	NA

It can be deduced from Table 3 that out of (334) patients admitted between years 2015-2020, 72 survived after treatment and the median survival time is 1423 days which is equivalent to 4 years.

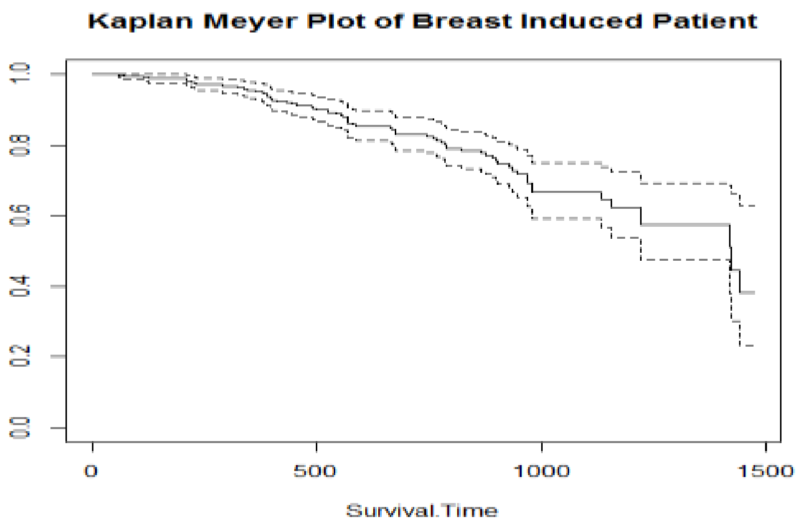

Figure 2: Kaplan-Meier Estimate Plot

Figure 2 shows that the probabilities of surviving is decreasing as the time is progressing, the dashed lines on the survival plot indicate the upper and lower confidence intervals. The chance of surviving breast cancer after median surviving time is 0.455.

Table 4: Comparison of cox and parametric regression models

	Cox	Log-Logistics	Weibull
Scale	NA	0.833	1.03
Age	-0.0033	0.0023	0.0047
Gender (Male)	-0.0568	0.2112	0.1781
Histology (ILC)	-0.0230	-0.0483	-0.1532
Histology (MC)	0.3363	-0.4614	-0.4386
Tumour_Stage (II)	0.1519	-0.2203	-0.2228
Tumour_Stage (III)	-0.0277	-0.0202	0.0503
Surgery_type (MRM)	0.5914	-0.6917	-0.6726
Surgery_type (Other)	0.7184	-0.6591	-0.6615
Surgery_type (SM)	0.2217	-0.1724	-0.2629
AIC	**724.6865**	1388.605	1398.711
BIC	**745.1765**	1430.527	1440.633

Table 4 presents the coefficients of Cox, Log-logistic and Weibull models. Cox model have negative coefficients for age, gender, ILC and Tumor stage III and positive coefficients for MC, Tumor stage II, MRM, SM and Other. Log-Logistics model have positive coefficients for age and gender and negative coefficients for ILC, MC, Tumor stage II, Tumor stage III, MRM, SM and Other. Weibull model have negative coefficients for ILC, MC, Tumor stage II, MRM, SM and Other and positive coefficients for age, gender and Tumor stage III. An increase in variable with negative coefficient decreases the survival rate of the patients while an increase in variable with positive coefficient increases the survival rate of the patients. The AIC and BIC criterion of Weibull model were the highest (1398.711, 1440.633) follow by Log-logistics model (1388.605, 1430.527) while Cox model has the lowest (724.6865, 745.1765). Hence, Cox proportional model is the best model to assess factors affecting the survival of breast cancer patients.

Table 5: Coefficient and Hazard Ratio of Cox proportional model

	Coefficient	Hazard Ratio (HR)	SE (coef)	p-value	L 95% C.I	U 95% C.I
Age	-0.0033	0.9967	0.0095	0.791	0.9784	1.015
Gender (MALE)	-0.0568	0.9447	1.0201	0.986	0.1279	6.976
Histology (ILC)	-0.0230	0.9769	0.2787	0.641	0.5658	1.687
Histology (MC)	0.3363	1.3998	0.6126	0.933	0.4214	4.650
Tumour_Stage (II)	0.1519	1.1640	0.3662	0.889	0.5679	2.386
Tumour_Stage (III)	-0.0277	0.9727	0.4245	0.999	0.4233	2.235
Surgery_type (MRM)	0.5914	1.8065	0.4200	0.319	0.7930	4.115
Surgery_type (Other)	0.7184	2.0511	0.3995	0.614	0.9374	4.488
Surgery_type (SM)	0.2217	1.2482	0.4487	0.441	0.5180	3.008

Considering the result of Cox proportional model in Table 5, a unit decrease in the age of a breast cancer patient is likely to decrease the risk of dying by 1% (HR = 0.99, 95% CI: 0.98 – 1.02), male patients have a hazard rate of 6% (HR = 0.94, 95% CI: 0.13 – 6.98) less than female in the reference. The hazard rate for Histology (ILC) is decrease by 2% (HR = 0.98, 95% CI: 0.57 – 6.98) in reference to IDC category while that of Histology (MC) increased by 40% (HR = 1.40, 95% CI: 0.42 – 4.65). The hazard rate for Tumour stage (II) is 16% (HR = 1.16, 95% CI: 0.57 - 2.39) more than Tumor stage (I) in the reference while that of Tumor stage (III) is 3% less (HR = 0.97, 95% CI: 0.42 - 2.24). The hazard rate of Surgery type (MRM) is 81% (HR = 1.81, 95% CI: 0.79 - 4.11) more than Surgery type (LP) in the reference while that of Surgery type (SM) is 25% more (HR = 1.25, 95% CI: 0.52 – 3.01).

CONCLUSION

The study of the factors that affect the rate at which breast cancer survived were considered. The result of Kaplan Meier estimator shows that 72 patients survived the breast cancer while the median survival time of the patient is 1423, the survival plot shows that as the probabilities of surviving decreases, the survival time increases. The performance of semi-parametric Cox proportional model, parametric Weibull and Log-Logistic models were critically examined to determine the best fitted model. Cox proportional hazard model which has the lowest AIC (724.6865) and BIC

(745.1765) was declared the best fitted model. The result of Cox proportional analysis shows that as the age increases, there is a decrease in the chance of surviving. A movement in the Tumor Stage of the breast cancer increases the likelihood of death of the patients and vice versa. Also, a change in the type of surgery performed for the patients will increase their rate of death. Hence, the prognostic factors affecting the risk of dying from breast cancer are Histology (MC), Tumor stage II, Surgical type (MRM) and Surgical type (SM).

REFERENCES

Adeboye, N. O., Ajibode, I. A. and Aako, O. L. (2020). On the Survival Assessment of Asthmatic Patients Using Parametric and Semi-Parametric Survival Models. *Occupational Diseases and Environmental Medicine*, **8**, 50-63. DOI: 10.4236/odem.2020.82004.

Anders, C. K., Johnson, R., Litton, J., Phillips, M. and Bleyer, A. (2009). Breast cancer before age 40 years. *Semin Oncol.; **36**(3): 237-249*. doi: 10.1053/j.semioncol.2009.03.001.

El Saghir, S., Adebamowo, A., Anderson, O., Carlson, W., Bird. A., Corbex, M., Badwe, A., Bushnaq, A., Eniu, A., Gralow, R., Harness, K., Masetti, R., Perry, F., Samiei, M., Thomas, B., Wiafe-Addai, B. and Cazap, E. (2011). Breast cancer management in low resource countries (LRCs): consensus statement from the Breast Health Global Initiative. *Supplement*, ***20***(2), S3-S11.

Ezekiel, I. D. and Aako, O. L. (2018). Comparison of Cox's and Weibull Regression Models in Assessing the Prognostic Factors for Survival of Asthmatic Patients. *International Journal of Current Innovation Research. 4, **11**(A), 1390-1394*. DOI: 10.24327/IJCIR.

Falgore, J.Y., Kajuru, J. Y., Usman, M. and Ramadan, K. M. (2022). Accessing of Some Risk Factors of Rheumatoid Arthritis Using Logistic Regression. *FUDMA Journal of Sciences (FJS), **6**(1), 2022,102-106*. DOI: https://doi.org/10.33003/fjs.

Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. American Statistical Association*, **53**, *457-481*.

Kleinbaum, D. G. and Klien, M. (2012). *Survival Analysis: A Self – Learning Text.* 3rd edition, Statistics for Biology and Health, USA. Springer.

McPherson, K., Steel, M., and Dixon, M. (2000). ABC of Breast Diseases: breast cancer-epidemiology, risk factors, and genetics. *British Medical Journal, **21***(4), 624.

Moghimi-Dehkordi, B., Safaee, A., Pourhoseingholi, M. A., Fatemi, R., Tabeie, Z. and Zali, M. R. (2008). Statistical Comparison of Survival Models for Analysis of Cancer Data. *Asian Pacific Journal of Cancer Prevention*, ***9**, 417-420.*

Montsarrat, R., Sandra, L., Ester, V., Montserrat, M., Misericordia, C., Rafael, M., Roger, P. and Josep-Alfons, E. (2009). Effectiveness of Early Detection of Breast Cancer Mortality Reduction in Catalonia (Spain). *BMC Cancer*, ***9**(*3), doi:10.1186/1471-2407-9-326.

Musa, M. C., Asiribo, O. E., Dikko, H. G., Usman, M. and Sani, S. S.(2020). Modelling the Determinants of Under-five Child Mortality Rates Using Cox Proportional Hazards Regression Model. *FUDMA Journal of Sciences (FJS)*, **4**(4), 401–408. DOI: https://doi.org/10.33003/fjs.

Nikpour, A., Charati, J. Y., Maleki, I., Ranjbaran, H. and Khalilian, A. (2016). Parametric Model Evaluation in Examining the Survival of Gastric Cancer Patients Andits Influencing Factors. *Global Journal of Health Science; **9**(3), 260-265.* doi:10.5539/gjhs.v9n3p260.

Popoola, A., Ogunleye, O., and Ibrahim, N., (2012). Five Year Survival of Patients with Breast Cancer at the Lagos State University Teaching Hospital, Nigeria. *Online Journal Medical and Medical Science*, ***1***, 24-31.

Syahila, E. A., Mohd Asru,l A. A., Sie, L. K. and Siti Afiqah, M. J. (2017). Analysis of Survival in Breast Cancer Patients by Using Different Parametric Models. *IOP Conf. Series: Journal of Physics, Conf. Series **890***(1), 1-6. DOI: 10.1088/1742-6596/890/1/012169

Teshnizi, S. H. and Ayatollahi, S. M. T. (2017). Comparison of Cox Regression and Parametric Models: Application for Assessment of Survival of Pediatric Cases of Acute Leukemia in Southern Iran. Asian. *Pac J Cancer Prev, **18** (4), 981-985.*

Vallinayagam, V., Prathap, S, and Venkatesan, P, (2014). Parametric Regression Models in the Analysis of Breast Cancer Survival Data. *International Journal of Science and Technology*, ***3***(3).

Zare, N., Doostfatemeh, M. and Rezaianzadeh, A. (2012): Modeling of Breast Cancer Prognostic Factors using a Parametric Log-Logistic Model in Fars province, Southern Iran. *Asian Pacific Journal of Cancer Prevention*, ***13**, 1537.*