



Survival Analysis of Cancer Patients at Teaching Hospitals in Ethiopia

Gemechu Bekana Fufa^{1*}

¹*Department of Statistics, College of Natural and Computational Sciences, Wollega University, Nekemte, Ethiopia.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JPRI/2020/v32i4531093

Editor(s):

(1) Dr. R. Deveswaran, M. S. Ramaiah University of Applied Sciences, India.

Reviewers:

(1) Mohit Sharma, Sri Guru Ram Das University of Medical Sciences & Research, India.

(2) Bhavin B. Vasavada, Shalby Hospitals, India.

(3) Ramnath Takiar, National Cancer Registry Program (ICMR), India.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/64688>

Received 10 November 2020

Accepted 14 January 2021

Published 01 February 2021

Original Research Article

ABSTRACT

Survival analysis today is widely implemented in the fields of medical and biological sciences. The basic principle behind the survival analysis implies to a statistical approach designed to take into account the amount of time utilized for a study period, or the study of time between entry into observation and a subsequent event. The event of interest pertains to death and the analysis consists of following the subject until death. This study aims to analysis the survival status of cancer patients, to identify the main cancer in women at study area and to determine the survival time of women with cancer after undergoing certain treatments in the selected teaching hospital(s). A random sample of 460 cancer patients was selected using purposive sampling from the study area, in 2016 to 2018. Parametric and non-parametric models, Logistic regression models, Statistical distributions and tests methods of data analysis were used in this study. The demographic information: age, educational level, treatments, economic level, marital status and stages of cancer were included and Comparisons were made among major cancers in patients. The findings of the study suggest that among major types of cancer in patients, breast cancer was the highly affecting the females in the study area. There are 62.15 % survival and 37.35 % of not survived patients in the data and the majority of patients (about 88 %) are diagnosed when they are in stages 1 and 2 and very few (about 8 %) of them are diagnosed in advanced stage of cancer. It also reveals that the average tumor sizes are significantly different for all ages and stages. By

*Corresponding author: E-mail: gemechu.bekana@yahoo.com;

grouping ages into groups of 5, I also stratified the number of patients diagnosed with cancer in different stages and the diagnoses at different stages were different. Finally, Results of survival time shows that the surviving of cancer patient were different at different ages and treatments level.

Keywords: Cancer; parametric and non-parametric; survival analysis; hospital; Ethiopia.

1. INTRODUCTION

1.1 Back Ground of the Study

Cancer is a condition of abnormal and rapid cell destruction inside the tissues making a mass of extra tissues which is known as tumor [1]. The cancer disease is majorly classified into two types based on the tissue or tumor growth: Benign and malignant. Unlike benign tumors which are assumed not harmful, malignant tumors are formed by jumping of cancer cells to other parts of the body. Scientists have stated the reason behind formation of such condition is due to adhesion property of the cancer-causing cells which is stated as the metastasis. The major types of cancers are breast cancer (in women), leukemia (in children), prostate cancer (in men), Lung & Bronchus, Pancreas, Ovary, Non-Hodgkin Lymphoma, Leukemia, Uterine corpus, Liver & intrahepatic bile duct, Brain/Other nervous systems and colon cancer [2]. Our present paper deals with the subject of cancer in women with condition of malignancy. Cancer in women was common cause of mothers' mortality in the world, particularly in Ethiopia. Result from a survey by the Ethiopian health Institute, reported that about 60% women were affected by cancer [3, 4]. Out of major types of cancers, the following are the cancer in the women: breast cancer, Lung & Bronchus, Pancreas, Ovary, Non-Hodgkin Lymphoma, Leukemia, Uterine corpus, Liver & intrahepatic bile duct, Brain/Other nervous systems and colon cancer [5, 6].

In 2012, there were an estimated of 6.2 million worldwide who had survived cancer after being diagnosed within the preceding five years. The highest proportion of cancer survivors who were observed within the past five years were in the developed country, this is due to the availability of services for early stage at diagnosis in those countries. Whereas there were low proportion of survivor in low- and middle-income country's (LMICs), because of resources are limited and breast cancer survivorship issues are only recently being addressed [7] & [8].

Survival analysis today is implemented in almost all fields of sciences. An analysis which is

performed to determine the probability of occurrence of the events associated with death or failure after treatment to the subjects is termed as survival analysis. This classification is applicable with help of machine-learning methods that evolve categorical results with predetermined time intervals. Survival analysis of cancer has acquired good importance for cancer detection in early stages taking into consideration risk factors. Different kinds of survival studies in present day include clinical trials, prospective cohort studies, retrospective cohort studies and retrospective correlative studies [9, 10]

Survival analysis deals with time to event modeling data with censoring. Censoring is mechanism of identification of the data values which do not follow up until end of the experiment. In many cases data considered for survival analysis are right censored which implies that the concerned subjects leave the study before the event has occurred or study ends before the event has occurred. The primary interest is to investigate the time to event or the survival probability. The statistical methods employed in study of survival and hazard probability can be performed parametrically, and non-parametrically based on the nature of the data [11,12].

1.2 Statement of Problems

Cancer incidence in Ethiopia, is relatively high compared to other countries, whereas lower than in most developed regions of the world. It is the most common cause of cancer deaths among women in developing countries particularly in Ethiopia and survival tends to be poor in this country because of a combination of a late stage at diagnosis and limited access to timely and standard treatment [13]. Despite, the government concern on the issue of non-communicable diseases with special emphasis on cancer, in order to reduce the incidence and mortality, the survival status of cancer is still were not known in Ethiopia [14].

Cancer is a major cause of morbidity in Ethiopia. Out of total about 1.65 thousand were cancer in the women [3]. According to the statistical

sources, today in the Ethiopia, approximately one in five women over their lifetime have a risk of developing cancer [11]. The report from health institute shows that out of total population, more than half cancer were in the women [14]. A recent report from the Ministry of Health in Ethiopia shows that more than half of mother's deaths are caused by cancers, so studying the burden of cancer as one of the important causes of death in the country is essential [15]. Hence, this study aimed to analysis the survival status and its predictor of cancer patients at teaching hospital in Ethiopia.

1.3 Objectives of the Study

The Main objective of the study is to analysis the survival status of cancer patient women in the selected teaching hospitals. Specifically, this study aims:

- To identify the main cancer in the women at teaching Hospital.
- To analysis the survival time of woman with cancer after undergoing certain treatments.
- To evaluate the survival probability cancer patients.
- To compare the survival time of cancer patients in the selected teaching hospital (s).

2. DATA AND METHODOLOGIES

2.1 Data Collection Methods

This study was conducted at teaching hospitals in Jimma and Black lion (Addis Ababa), Ethiopia. The study area was selected purposively and random sample of the cancer patients were taken. The data were collected from secondary data documented at the selected hospitals. About 460 sample of cancer patients' women were included in to the study.

2.2 Methodologies

2.2.1 Survival and hazard functions

Survival time can be estimated as a variable which calculates the time between the starting point and ending point of event of interest or time of interest. In medical field [16] it is termed as the period elapsing between the completion or institution of any procedure and death. The survival time and event data are collected on

practical grounds which is either censored or truncated.

The survival function can be expressed with help of another distribution used commonly in statistical techniques; namely cumulative probability function CDF denoted as $F(t)$. The survivor function is defined as the complement of the CDF which is formulated in the relationship below

$$S(t) = (T > t) = 1 - (t) \quad (1)$$

Similarly Hazard function is an alternative representation of the distribution of T or the instantaneous occurrence of the event and is defined as

$$h(t) = \lim_{dt \rightarrow 0} \frac{Pr(t < T \leq t + dt | T > t)}{dt} \quad (2)$$

The above expression is termed as the instantaneous rate of occurrence for the conditional probability that the event will occur in the time interval between t and $(t+dt)$ as it has not occurred before. By the prior computation of the conditional probability in the numerator and application of limits gives the hazard function as

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

In other words, the hazard function can be stated as the rate of the occurrence of the event at time t equals to the probability density at time t divided over the probability of the surviving to that duration without experiencing the event. The above formula can be expressed using the relation between density and survival function as follows

$$h(t) = -\frac{d}{dt} \log S(t) \quad (4)$$

The above expression of hazard function is integrated using limits 0 to t and applying the boundary condition $S(0) = 1$ (which implies event not occurred at time 0) to obtain relation between hazard and survival function. The survival function gives the probability that a subject will survive past time t . As t ranges from 0 to infinity, the survival function has the following properties:

- At time $t = 0$, $S(t) = 1$. In other words, the probability of surviving past time 0 is 1 .
- At time $t = \text{infinity}$, $S(\text{infinity}) = 0$. As time goes to infinity, the survival curve goes to 0 .

Survival models for the analysis of data have three main characteristics: (i) the dependent variable or response is the waiting time until the occurrence of a well-defined event, (ii) observations are censored, in the sense that for some units the event of interest has not occurred at the time the data are analyzed, and (iii) there are predictors or explanatory variables whose effect on the waiting time we wish to assess or control [17].

2.2.2 Non-parametric and parametric analysis

2.2.2.1 Non-parametric (NP) analysis

Estimating the distribution of the dependent variable without making assumptions about its shape is an important first step in analyzing a dataset. Given the importance of the distribution of the dependent variable it is valuable to “let the data speak for itself” first [1]. Estimating the probabilities without making any assumptions on its shape is called non-parametric analysis [18].

The function used to represent the distribution is the Survivor function. Nonparametric methods do not require the knowledge of the underlying distribution of the failure time. Hence it provides an edible way to deal with the data in many practical situations. The seminar paper by Kaplan and Meier [8] is the benchmark in survival analysis especially from nonparametric point of view. It compelled the application of descriptive statistics and improved the development of all existing NP approaches with censored data. The survivor function is calculated by dividing the number of survivors by the total number of subjects for every time.

2.2.2.2 Kaplan-meier estimator

The Kaplan-Meier estimator originally was derived as an NP maximum likelihood estimator of $F(t)$. Because of the latter method of derivation, it is also called as the product-limit (PL) estimator. If the data was not censored then the empirical survival function is given by:

$$S(t) = \frac{1}{n} \sum_{i=1}^n I\{t_i > t\} \quad (5)$$

Where I is termed as the indicator function which takes a value of one if the condition $t_i > t$ is true or zero otherwise [19]. The estimator is simply the proportion alive at t . For the censored data, assume the ordered times of death as $t_1 < t_2 < t_3 < \dots < t_m$ and d_k be the death occurred at t_k . Let n_k be the number of persons alive just before t_k .

This is the number exposed to risk at that time. The Kaplan-Meier (KM) or product limit estimate of the survivor function is

$$\hat{S}(t) = \prod_{i:t(i) < t} (1 - \frac{d_i}{n_i}) \quad (6)$$

The justification of the estimate is explained as follows. In order to survive until the time t one must first survive until the time t_1 . And the conditional probability of surviving from t_2 to t_1 given already survived t_1 is to be satisfied. The Kaplan-Meier (KM) is a step function with jumps at the observed times.

2.2.2.3 The nelson-aalen estimator

For estimating a cumulative hazard $H(t)$, one simple approach is to find an estimator of $S(t)$ and take minus the log. An alternative approach is to estimate the cumulative hazard directly using the Nelson-Aalen estimator. The Nelson Aalen estimator is a non-parametric estimator of the cumulative hazard rate function from censored survival data [20]. Consider a sample of n individuals from a right censored survival population. Our observation of the survival times for these individuals will typically be subject to right censoring meaning that for some individuals we only know that their true survival times exceed certain censoring times. The censoring is assumed to be independent in the sense that the additional knowledge of censorings before any time t does not alter the risk of failure at t . The Nelson-Aalen estimator is a step function with the location of the steps placed at each observed death time and the vertical size of the steps is the inverse of number at risk, where number at risk is the number of patients just before the death that are still observed to be alive. With larger samples the Nelson-Aalen estimator will get closer to the true cumulative hazard. The Nelson-Aalen estimator is given by

$$\hat{H}(t) = \sum_{t_j < t} \frac{d_j}{r_j} \quad (7)$$

Where d_j is the subjects who die at time t_j and r_j is the number of subjects at risk just prior to time t_j . The variance of the estimator is given by

$$Var(\hat{H}(t)) \approx \sum_{t_j < t} \frac{d_j}{r_j^2} \quad (8)$$

The advantage of non-parametric analysis is that the results do not rest on the assumptions. The disadvantage is that we can only compare limited number of groups which implies it is very difficult to see the impact of multiple explanatory

variables on the subjects [21]. Another disadvantage of non-parametric techniques is that it can only deal with the quantitative explanatory variables like GDP, rich and poor countries etc.

2.3 Parametric Analysis

Parametric survival analysis assumes functional form of probability distribution for the variables that provides the influence of explanatory variables on survival time. The strength of this analysis is the estimation is relatively easy and survival curves are smoother as they draw information from whole data. This parametric analysis is carried out using two different approaches which are regression parametric models (Accelerated Failure Time models) and Proportional Hazard (PH models). These models are provided as the common scales for the distributions in parametric survival models. Both PH and AFT models were analyzed on basis of t -scale over the distributions with interval $(0, \infty)$, whereas the AFT models were also interpreted on the basis of (t) - scale over the distributions termed as pure AFT models.

When comparing parametric models, the Akaike Information Criterion (AIC) and loglikelihood values can be used to select the best parametric model. The best fit model is the one with smaller AIC and largest Log-likelihood. Once the model is identified we will perform a residual analysis check that lets us assess the absolute goodness of fit of the identified parametric model.

The common variables selected for analysis of cancer patient's survival are survival status (dead or survive), treatments, marital status, stages of cancer, age, economic level of the patients, grade, of tumor, place of residence (rural and urban), educational level and other. Similarly, the variables were coded as follows (Table 1).

2.3.1 Logistic regression

Logistic regression is mostly used to predict a categorical (usually dichotomous) variable from a given set of independent variables. If all the independent variables are continuous, we usually employ discriminant analysis for modeling the data. In case if all or few independent variables are categorical, logistic regression analysis is the best choice.

In this study, for identifying the the survival status of the cancer patients. 1 = if survive and 0 = if dead. In logistic regression analysis, it is

assumed that the explanatory variables affect the response through a suitable transformation of the probability of the success. This transformation is a suitable link function of P , and is called the logit-link, which is defined as:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta X \quad (9)$$

Where $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the model parameters and X will the predictor/independent chosen variables. The transformed variable, denoted by $\text{logit}(P)$ is the log-odds and is related to the explanatory variables.

2.3.2 Nonlinear differential, reproduction number

In this work I have considered nonlinear differential equations in four dimensional dynamics systems. I divide the whole cancer patients of the selected hospitals in to four treatment groups: The No treatment (NT), treatment with Radiation (TR), treatment with Radiation and Surgery (TRS) and treatment with surgery (TS) groups to study the treatment status of cancer patients in selected hospitals.

I found that the basic reproduction number, which depends on five parameters, is as follows.

$$R_0 = \frac{\beta_1 + \varepsilon_1}{\theta + \mu + \delta + 1 - \varepsilon_1} \quad (10)$$

2.3.3 Exponential distribution

In regression models it is common practice that the dependent variable depends on the explanatory variables only through a linear function. Because of its historical significance, mathematical simplicity and important properties, the exponential distribution is one of the most popular parametric models. This is the simplest possible distribution with one parameter which is derived treating the hazard function as a constant and of monotonic value over baseline hazard function denoted as $h(t) = \lambda$.

$$h(t) = \exp(\beta_0) \exp(X_i \beta) \quad (11)$$

So for the exponential distribution the instantaneous failure rate is independent of t so that the conditional chance of failure does not depend on how long the individual has been on trial. This is referred to as the memory less property of the exponential distribution.

$$h(t) = \exp(\beta_0) \exp (X_i\beta) \tag{12}$$

So for the exponential distribution the instantaneous failure rate is independent of t so that the conditional chance of failure does not depend on how long the individual has been on trial. This is referred to as the memory less property of the exponential distribution.

As we can see from table 2 above, the breast cancer is the highest (18.65%) followed by ovary (16.96%) and the smallest in percentages was Brain/Other nervous systems cancer (2.43) in the study area. This study focused on the analysis of breast and ovary cancer, because the women in the study area were highly affected by these diseases.

3. RESULTS AND DISCUSSIONS

3.2 Socio-Demographic Characteristics of the Study Participants

3.1 Descriptive Analysis

This part of the paper discusses the major findings of the study. It begins with analysis of the descriptive result on the main types of cancer in the women of study area followed by discussion of results obtained from estimation methods used under each objective of the study. The result shows that, breast cancer is the highest followed by ovary in the study area. The percentages of types of cancers in women were as follows:

Among enrolled Between January 1st 2016 to December 31st 2018, 460 cancer patients were selected for this study. Cards of four hundred sixty (385 censored and 75 death) cancer women were included in the present study. The mean age of participants at the time of diagnosis was 42.61years with SD \pm 12.28 years. Slightly nearly half, 451 (44.5%) of the age group was less than 40 years old. The socio-demographic characteristics of the study participants are shown below (Table 3).

Table 1. Variables used in survival modeling

No	Variables	Code classification
1	Age	X_1
2	Grade	X_{2i} , $i = 1, 2, 3, 4, 5$ $i = 1$: well differentiated $i = 2$: moderately differentiated $i = 3$: poorly differentiated $i = 4$: under differentiated $i = 5$: cell type not determined
3	Numprims	X_3
4	Treatments	X_{4i} , $i = 1, 2, 3, 4$ $i = 1$: No treatment $i = 2$: Radiation $i = 3$: Radiation and Surgery $i = 4$: Surgery (reference)
5	Stages	X_{5i} , $i = 1, 2, 3, 4$ (reference)
6	Tumor Size	X_6

Table 2. Summary of major cancer in women of study area

Types of cancer in women	Percentages (%)	Minimum age	Maximum age
Lung & Bronchus	15.87	18	61
Breast	18.65	21	63
Colon & Rectum	8.96	23	61
Pancreas	13.52	19	45
Ovary	16.96	18	48
Non-Hodgkin Lymphoma	7.78	21	51
Leukemia	6.70	22	56
Uterine corpus	5.17	21	65
Liver & intrahepatic bile duct	2.96	21	45
Brain/Other nervous systems	2.43	22	48
All sites	100%		

3.2.1 Logistic regression analysis

The Initial Log Likelihood Function, (-2Log Likelihood or -2LL) is a statistical measure like total sums of squares in regression. The initial - 2LL value is 21332.017 at step 0, before any variables have been added to the model.

According to Table 4 Age of the cancer patients, economic level, place of residence, stages of the cancer, treatments, tumor size and education level are significant at 5%, P-value < 0.05, whereas other variables were insignificant variables, Sig. > 0.05.

When no event times are censored, a non-parametric estimator of S(T) is $1 - F_n(t)$, where $F_n(t)$ is the empirical cumulative distribution function. When some observations are censored, we can estimate S(t) using the Kaplan-Meier product-limit estimator. The Kaplan-Meier method is based on individual survival times and assumes that censoring is independent of survival time (that is, the reason an observation is censored is unrelated to the cause of failure).

The Kaplan-Meier estimator of survival at time t is shown in Equation 6. Here $t_j, j = 1, 2, \dots, n$ is the total set of failure times recorded (with t^+ the maximum failure time), d_j is the number of failures at time t_j , and n_j is the number of individuals at risk at time t_j . The Kaplan-Meier survival estimates as a function of time was as follows.

From the survival function and hazard function analysis, the survival probability of surviving past time t, is 0.85 (survival value) and the relative likelihood of the event occurring at time t, $f(t)$, conditional on the subject's survival up to that time t, $S(t)$. The hazard rate thus describes the instantaneous rate of failure at time and ignores the accumulation of hazard up to time t is 0.92.

Considering the cancer in women (breast and ovary) survival data, we are interested in knowing how long women with cancer will survive after undergoing certain treatments. Treatments include *radiation or surgery or both radiation and surgery or no treatment*. Also, we would like to know the effectiveness of

Table 3. Socio-demographic characteristics of cancer patients at study area (n = 460)

Covariate	category	Vital status at last contact		Total No. (%)
		Censored No. (%)	Death No. (%)	
Age in years	<40	160 (43.36)	30(40.54)	190 (42.51)
	40-49	125 (58.14)	16 (21.62)	141 (59.00)
	50-59	55 (33.74)	14 (18.92)	69 (31.22)
	60-69	18 (4.76)	6 (8.11)	28 (6.09)
	>70+	11 (16.92)	8 (10.81)	19 (23.46)
Place of residence	Urban	215 (56.88)	24 (29.27)	239 (51.96)
	Rural	163 (43.12)	58 (70.73)	221(48.04)
Marital status	Single	65 (17.71)	16 (19.05)	81 (17.96)
	Married	236 (64.31)	41 (48.81)	277 (61.42)
	Divorced	24 (6.54)	13 (15.48)	37 (8.20)
	Widowed	42 (11.44)	14 (16.67)	56 (12.42)

Table 4. Test of Significance of Independent Variables Using Wald Test

	Analysis of maximum likelihood estimates							
	DF	Estimate (β)	S. E	Wald Chi-square	Pr>Chisq	OR	95% CI	
							lower	upper
Intercept	1	25.05	-	32.010	.000	-	-	-
Age (X ₁)	1	0.360	0.383	3.120	0.001	1.958	1.650	2.350
Economic level (X ₂)	1	0.077	0.021	4.720	0.004	4.982	4.620	5.630
place of residence (X ₃)	1	0.089	0.028	6.118	0.004	1.782	1.480	2.630
Stages (X ₄)	1	- 2.045	0.051	24.873	0.002	3.871	3.320	4.302
Treatments (X ₅)	1	- 3.086	0.315	23.019	0.000	6.721	6.250	7.201
Tumor Size (X ₆)	1	1.321	0.284	22.060	0.000	5.472	5.120	5.892
Educational level (X ₇)	1	1.055	0.356	2.037	0.002	1.842	0.652	1.250

treatments when implemented in different stages of cancer. Firstly, we considered the effectiveness of treatments on survival for the overall data. From the Table 6 women who are treated with radiation have a median survival of 30 months with 95 % CI (28, 32) months. Interestingly, women those who are treated with both treatments has the same median survival as of women who received surgery.

As we can see from the results of Table 7, women in stage-4 cancer can be advised to stay away from any treatment for a better survival. Financially, this could really save so much for women. Further we investigated the effect of treatment stage wise. The median survival for those who are in stage I and stage II who did not receive any treatment for cancer is not reported because the KM estimator for these data never reached a failure probability greater than 35.28 % or a survival probability lower than 64.72 % for the former case and data never reached a failure probability greater than 45.29 % or a survival probability lower than 54.71 % for the latter case.

We also found that the numerical value of the reproduction number based on the real data collected from selected hospitals is $R_0 = 0.7657 < 1$. This in principle shows that the cancer treatment is low in the selected hospitals. We investigate two equilibrium point namely

stable disease free and equilibrium point and unstable endemic equilibrium point. To control the cancer using the treatment steps, we then identify control parameter which gives insight to decreases or stop of the cancer in women (breast and ovary cancer). The basic control parameters that can decreases the cancer by treatment are $\beta_1 = 0.06577$ which is the rate of transmission from no treatment to treatment with radiation and $\beta_2 = 0.04537$ which is the rate of transmission from no treatment in to treatment with radiation and surgery.

3.2.2 Selection of best fit parametric model

We use the model selection criteria to select the best parametric model. From the previous sections, by performing the residual analysis for the fitted parametric models log-logistic parametric model performed better than other models. We identify that the log-logistic model has the lowest AIC and highest likelihood values performs better than other models. Thus, the following result shows the goodness of fit.

The survival and fitted survival functions for exponential parametric model are given by equations below. Table 9 has the analysis of the maximum likelihood estimation of parameters of the Exponential model for cancer (breast and ovary) patients.

Table 5. Kaplan-Meier survival estimates as a function of time

Time	Subject at risk (n_j)	fail (dj)	Censored	Surv. Prob $P_j = (1-d_j/n_j)$	Cumulative survival $S_j = P_j \times P_{j-1}$
0	296	3	6	0.97	0.97
1	287	5	7	0.96	0.96
2	275	3	4	0.97	0.97
3	268	4	9	0.95	0.95
4	255	3	6	0.96	0.96
5	246	2	5	0.97	0.97
etc	-	-	-	-	-

Table 6. Treatment wise KM estimates for median survival

Treatment	No treatment	Radiation	Radiation & Surgery	Surgery
Median Survival	-	30	25	25
95% CI	-	[28, 30)	[21, 30)	[21, 29)

Table 7. Stage vs. Treatment Product-Limit Estimates for median survival

Stages	No treatment	Radiation	Radiation & Surgery	Surgery
Stage I	-	178(172, -)	104(53, -)	62(42, 100)
Stage II	-	145(140, 149)	128(45, -)	43(32, 66)
Stage III	93(81, 103)	52(47, 57)	32(24, 110)	27(17, 31)
Stage IV	34(28, 39)	23(21, 27)	17.5(12, 22)	14.5(11, 18)

Table 8. Goodness of fit for parametric models

Selected Distribution	Log-Likelihood	AIC
Gamma	-43730.415	87506.83
Log-Normal	-43957.836	87959.67
Weibull	-43961.92163	87967.84
Log-Logistic	-44259.26034	88560.52
Exponential	-43714.80892	87473.62

Table 9. Analysis of MLEs for Exponential Model

Analysis of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimates	S. E	95% Confidence Limits		Pr >ChiSq
Intercept	1	5.36	0.083	5.1960	5.5219	<.0001
Age	1	-0.032	0.006	-0.0334	-0.0309	<.0001
Grade	1	0.317	0.035	0.2485	0.3857	<.0001
Grade	2	0.137	0.029	0.0794	0.1943	<.0001
Grade	3	-0.172	0.056	-0.2293	-0.1153	<.0001
Grade	4	-0.211	0.023	-0.3206	-0.1011	<.0001
Grade	5	0.000	-	-	-	-
Numprims	1	-0.085	0.013	-0.1116	-0.0586	<.0001
Treatment	1	0.940	0.849	0.8499	1.0494	<.0001
Treatment	2	0.589	0.0424	0.506	0.676	<.0001
Treatment	3	0.236	0.0738	0.0914	0.380	<.0014
Treatment	4	0.000	-	-	-	<.0001

4. DISCUSSION

Log-rank test was performed to test equality of survival curves for the presence of any significant differences in survival time among various levels of the categorical variables considered in the study. In this study, the test statistics showed that there is a significant difference in survival function for different categorical variables. Accordingly, the Kaplan-Meier analysis indicated significant evidence of differences in survival times. It is found that the median survival time for those who had clinical stage I, II or III at baseline had a longer survival time than those in advanced clinical stage (IV), this difference was statistically significant with p-value < 0.000. The median survival time for those who have negative lymph node status had a longer was than those who had positive lymph node at baseline. This difference was statistically significant with (p-value = 0.000) [22,23].

According to the results of this study, those women whose surgical margin involved at baseline were found to be a strong predictor of mortality among breast cancer women. Women who had surgical margin involved at diagnosis were 3.13 times higher risk of death as compared to women who had surgical margin free at time of diagnosis. Most of previous studies also found twice or more risk of mortality

in patients with surgical margin involved compared to those with surgical margin free which reflect surgical margin was an important determinant of survival. Such findings could be explained by the fact that residual disease at the surgical margins could increase the risk of local recurrence and possibly death through the years [24-26].

5. CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The results of the analysis showed that the breast cancer was higher than those others of cancer in the women in the study area. The mean percentage of breast cancer (18.65 %), ovary (16.96 %) and ovary (15.87 %) were the highest percentages comparing to the other variables. This study found evidence that some of the variables considered have significant influence on survival status of patients. The Age, economic level, place of residence, Stages, treatment level and educational level were significant variables among the stated variables.

The coefficients in the logistic regression are based on what is called the odds-ratio. It is the factor by which the odds change when the independent variable increases by one unit. If β

is positive, this factor will be greater than 1, which means that the odds are increased; if β is negative, the factor will be less than 1, which means that the odds are decreased. When β is 0, the factor equals 1, which leaves the odds unchanged. From the results for positive β , the Odds-ratio (ORs) are greater than 1 implies that the odds are increased and for negative β , the Odds-Ratio (ORs) are less than 1 implies that the odds are decreased.

Women who are treated with radiation alone have a median survival of 30 months and women treated with surgery alone and both radiation & surgery reported a median survival of 25 months. Non-parametric method for survival, based on the treatment indicated that the combination of radiation and surgery has the same effect on survival as treated with surgery alone. Also, the women in stage-4 cancer can be advised to stay away from any treatment for a better survival. Financially, this could really save so much for women. Further we investigated the effect of treatment stage wise. It is an interesting observation that women who are identified with malignant breast cancer tumor, but have not received any treatment has more survival rate when compared to women who are treated with either radiation or surgery or combination of both.

5.2 Recommendation

Based on this study finding, the following recommendations should have been forwarded with each respective body:

- The federal minister of health should expand breast cancer early screening programs (breast self-examination, clinical breast exam and mammography) to improve treatment results in cancer in women.
- Could be strengthened awareness in collaboration with public medias about breast cancer prevention, screening and treatment is crucial.

CONSENT

As per international standard or university standard, patients' written consent has been collected and preserved by the authors.

ETHICAL APPROVAL

As per international standard or university standard written ethical approval has been collected and preserved by the authors.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Blossfeld HP, Rohwer G. Techniques of event history modeling: New approaches to causal analysis. Lawrence Erlbaum Associates Publishers; 2002.
2. Hamashima, Chisato, et al. Survival analysis of patients with interval cancer undergoing gastric cancer screening by endoscopy. *PLoS One*. 2015;10(5): e0126796.
3. Ethiopia Health Organization. Cancer Facts & Figures. 2016-2017;2017. Available:<http://www.who.int/mediacentre/factsheets/fs297/en/> Accessed on: 12/31/17.
4. González-Manteiga W, Cadarso-Suarez C. Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Communications in Statistics-Theory and Methods*. 1994;4(1):65-78.
5. American Cancer Society. Breast Cancer Facts & Figures 2005-2006. Atlanta: American Cancer Society, Available:<http://www.cancer.org/>
6. Andersen PK, Borgan Ø, Hjort NL, Arjas E, Stene J, Aalen O. Counting process models for life history data: A review [with discussion and reply]. *Scandinavian Journal of Statistics*. 1985;97-158.
7. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh J, Comber H, Forman D, Bray F: Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European Journal of Cancer*. 2013;49(6):1374-1403.
8. Kunz, Pamela L, et al. Long-term survivors of gastric cancer: a California population-based study. *Journal of Clinical Oncology*. 2012;30(28):3507-3515.
9. Eil YK, Nishimoto R, Mediansky L, Mantell J, Hamovitch M, Social relations, social support and survival among patients with cancer. *Journal of Psychosomatic Research*. 1992;36(6):531-541.
10. Imbeah ZEP, Afrane BA, Kretchy I. A et al, Prevalence and self-management of female sexual dysfunction among women in six regions of Ghana: A cross-sectional study. *Journal of Women's Health, Issues and Care*. 2015;4(6).

11. Yang, Dongyun, et al. Survival of metastatic gastric cancer: Significance of age, sex and race/ethnicity. *Journal of Gastrointestinal Oncology*. 2011;2(2):77.
12. Ng Z, Ong M, Jegadeesan T, Deng S, and Yap C. Breast cancer: exploring the facts and holistic needs during and beyond treatment. *Healthcare*. 2017;5(2):26.
13. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Botstein D. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752.
14. Ethiopia Health institute. *Cancer and Transmit Disease in Ethiopia*; 2018.
15. Siegel R, Naishadham D, Jemal A. *Cancer statistics, 2012*. CA: A Cancer Journal for clinicians. 2012;62(1):10-29.
16. Cantor A. *SAS survival analysis techniques for medical research*. SAS Institute; 2003.
17. Rodríguez, G. Statistical issues in the analysis of reproductive histories using hazard models. *Annals of the New York Academy of Sciences*. 2007;709(1):266-279.
18. Aheng ZS. A. Stigmatization is a Hindrance to the Fight Against Breast Cancer in Ghana, ModernGhana, Accra, Ghana; 2015. Available: <https://www.modernghana.com/news/659540/1/stigmatization-isa-hindrance-to-the-fight-against-breast-cancer-in-ghana-d.html>
19. Rodríguez G. *Lecture Notes on Generalized Linear Models*; 2007. Available: <http://data.princeton.edu/wws509/notes/>
20. Buis ML. *An introduction to survival analysis*. Department of Social Research Methodology Vrije Universiteit Amsterdam [Online]; 2006.
21. Allison PD. *Fixed effects regression methods for longitudinal data: Using SAS*. SAS Institute; 2005.
22. Atlanta GA. American Cancer Society, 66 p., pdf. Available from. *Journal of Consumer Health on the Internet*. 2012; 16(3):366-367.
23. Chen Q, Oppenheim A, Wang D. *Survival Analysis and Data Mining*; 2008.
24. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. *Journal of the National Cancer Institute*. 2004; 96(3):218-228.
25. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Kropp S. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genetics*, 2008;4(4):e1000054.
26. Tarver T. *Cancer Facts & Figures 2012*. American Cancer Society (ACS); 2012.

© 2020 Fufa; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/64688>