International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Survival Models on Breast Cancer Patients - A Cross Sectional Data from the Ocean Road Cancer Institute (ORCI)

Dr. Ramkumar .T Balan[1], Hassan Swedy Lunku[2], Ismail Juma[3]

Associate Professor, Department of Statistics, CNMS, UDOM

Department of Local Government Accounting and Finance, LGTI

Department of Local Government Accounting and Finance, LGTI

Abstract: ***The aim of this study was to evaluate a semi-parametric survival model on time of survival of breast cancer patients and a case study is considered for describing the model. The study is on survival times of breast cancer patients treated at ORCI which may help to reduce breast cancer outcomes. Data of socio-demographic characteristics of patients, their reproductivity, stages of disease, treatments undergone and follow up of the treatment were considered to develop the model. Cox proportional hazard regression model is adopted considering the covariates incurring in the system. The Wald statistics evaluates whether the coefficients of given variable is statistically significant and found that stage of disease was significant while other covariates -history of multiple sexual partners, smoking cigarette, alcohol use, cancer grade fails to be significant .Also Kaplan Meier Estimate is derived and median survival time is estimated. The results of this paper showed that non parametric and semi parametric survival models were better to predict survival time of breast cancer patients.***

1. Introduction

Cancer is highly curable if diagnosed at an early stage of the disease and many researchers have been studying the relationship between variables and survival times of cancer patients (Maranga, et al, 2013). According to GLOBALCAN 2018, the global cancer burden estimated to have risen to 18.1 million new cases and 9.6 million deaths in 2018, increasing cancer burden sue to factors such as population growth and ageing. Interested facts based on large number of patients with early stage cancer contain small development of secondary malignant of cancer at the diagnosis time. Established that medically cancer patients treated with chemotherapy and radiotherapy improves survival of patients with early stage of cancer disease and thus treatment options for early cancer stage such as stage I and II include chemotherapy while addition of radiotherapy and surgery for patients with late stages (ACS, 2016).

Breast cancer is a major cause of death and cancer related mortality among women. The prevalence of breast cancer reported as increasing in most of the Asian and African countries (Park et al., 2011), Tanzania share same experience. The cause origin of breast cancer is largely unknown therefore, there is no definite primary prevention strategy established. The main preventive strategy is focusing on the early detection and early intervention to improve the survival rates. Despite the increasing incidence, the survival rates of breast cancer patients in many developed countries substantially improved (Coleman et al., 2008)(Blamey et al., 2007).

Breast cancer lead cause of women cancer deaths in Tanzania where affects approximately more than 4,000 women and causes higher deaths of 7% every year. Prevention intervention had to be taken up by 2025 Tanzania Development Vision (TDV) and projected that the number of new cases will rise to almost 6000 and deaths 4000 annually (ORCI, 2015). High breast cancer prevalence shifted from developed countries to developing and poorer countries with less medically equipped and spreading fast among the poor citizens who have no or low financial situations to manage medical treatment provided. According to GLOBAL CAN 2018 registry cancer report revealed breast cancer is the leading type of cancer worldwide in terms of new cases where approximately 2.1 million diagnoses estimated in 2018 contributing about 11.6% of the total cancer incidence burden and leading cause of women cancer deaths for 15.0%.

Many analyses of cancer survival data prefer to use Cox Proportional Hazards (CPH) model which had influence on development of cancer covariates in the field of survival analysis. However the CPH relies on the constant ratio between hazards over time interval assumption of covariates. Simplest ways to extend CPH model with inclusion of the interactions between the covariates and survival time of the patients with either linear or polynomial function. Survival of cancer patients is important indicator of the treatment response and consequences to attempts managing breast cancer on the patient's survival duration in Tanzania are not known while breast cancer cases increases annually (ORCI, 2010, 2015). Currently, study on breast cancer survival rate is scarce in Tanzania context. Therefore this study comes up with evaluation of survival times into semi parametric survival model considering covariates such as residence location, education level, patient age, number of pregnancies and children alive, contraceptive drugs and smoking behavior, alcohol intake, grade and stage of breast cancer, treatment mode given, marital status, duration of symptoms.

The main aim of this study was to evaluate a semi-parametric survival models times of breast cancer patients treated at ORCI for the available data in order to show its

applicability and workability. Study provides information on survival times of breast cancer patients treated at ORCI which will help in reducing breast cancer outcomes through improvement of disease management and inform policy and decision makers on patients' survival periods with different covariates hence will increase the patient benefit in medical and ethical ways.

Data

A retrospective review of female patients with diagnosed proven breast cancer treated at ORCI and used the available information from medical files from January 2014 to December 2015 and followed up to December 2017. Study was done at ORCI providing specialized cancer treatment, in Tanzania, of radiotherapy and chemotherapy as well as the palliative care. It also runs outpatient clinics, nuclear medicine services, diagnostic imaging as well as screening for various types of cancers. ORCI consists of four separate clinics which are new-patient and out-patient clinic, follow-up clinic, and treatment clinic consist of radiation and chemotherapy treatment clinics. Ideal institute for the study since ORCI specialized and has capacity to handle referrals of patients for cancer treatment and this makes ORCI the choice for a large number of patients with different cancer diseases.

Sampling procedure and sample size

Included women diagnosed with proven breast cancer and treated at ORCI between January 2014 and December 2015 and status of followed up to December 2017 indicated in medical files, for death certificate indicated with dates of death for expired patients. The non-probabilistic sampling technique preferred that was purposive sampling method and using WHO manual on sample size determination in health studies by Lwanga and Lemeshow, 1991. The sample size estimated was 196 patients by using the following criteria such as test survival rate of 70%, anticipated population proportion of 80%, 5% significance level, 95% confidence level and 90% power of test, 7% absolute precision required on either side of the proportion.

Data abstraction, collection and analysis:

Study instruments consisted of files which divided into the following sections: a) Socio-demographic characteristics b) Reproductive data c) Stage, treatment and follow up. Since the secondary data is the only source of the required information of the study then data abstracted from medical files by using open source kobo toolbox and statistical analysis was done using Statistical Package R.

2. Methods

Cox's Proportional Hazards Regression Model

Semi parametric model know as Cox Proportional Hazard (CPH) regression model determine effect of survival times of patients and covariates on the Hazard Ratio (HR) and the baseline hazard rate unspecified hence assumption included. If we have several number explanatory covariates of study, we can express the hazard (or risk) of dying at time t or checking for association between various covariates and lower or higher survival rates. The final model from a Cox regression analysis yield an equation for the hazard as a function of several explanatory covariates and main point about CPH model is to compare the HRs of individuals who have different covariates, hence known as Proportional hazards.

CPH assumptions:

1) It does not assume knowledge of absolute risk.
2) Estimates relative rather than absolute risk.

Important survival model was published by D.R. Cox in 1972 and his paper presented is one of most frequent articles used in medical statistics research which usually associated with death. The suggested model depend on hazard rate in patients survival time (t) as shown as follows

$$\lambda\ (t;\ z) = \lambda_o\ (t)\ \exp(\underline{\beta Z}) = \lambda_o\ (t)\ \exp(\sum_{i=1}^{p} \beta_i Z_i)$$

Where λ_o (t): initial hazard function when all values of Z = 0 also known as baseline hazard

$\underline{\beta}$: are unknown's regression coefficients

$\underline{Z:}$ is the p-dimensional vector of covariates

Survival function can be written as follows; S (t; Z) = $\{S_0(t)\}exp(\sum_{i=1}^{p} \beta_i Z_i)$. Where $exp(\sum_{i=1}^{p} \beta_i Z_i)$is the proportional hazard function.

The estimation for CPH model parameter β is the same for any transformation and only rank statistics can carry information about the parameter β when baseline hazard λ_0 is completely unknown. It follow that the rank statistic is sufficient to estimate parameters β and to apply the rank statistic to get inferences about β study used marginal distribution of the ranks and marginal likelihood estimation and not maximum likelihood estimation. Standard maximum partial likelihood method to obtain the estimates of parameter cannot be applied directly due to a very high dimension space of covariates and use of penalized partial likelihood to deal with non linearity problem. Generally, survival analysis involves the modeling of time to death of patients since in this study death considered as an event of interest. Patients' survival times function as a property of random variable which map a set of events associated with patient death used in wider range of application including patients' death in study.

3. Results

Total of 196 female patients with initial diagnosis of breast cancer treated at ORCI between January 2014 and December 2015 were followed up to December 2017, all patients were alive at the diagnosis time. 79 (40.3%) of them were confirmed dead within that period, 37 (18.9%) were still alive and 80 (40.8%) were lost to follow up, 79 (40.3%) patients death was caused by diagnosed breast cancer where 13(6.6%) patient had family history of having cancer incidence in their family, 77 (39.3%) reside in rural area while the remaining 119 (60.7%) reside in urban areas. All patients were referred from other hospital across the country where 137 (69.9%) referred from Muhimbili National Hospital (MNH) while the remaining 59(30.1%) from other health institute or hospitals located in Tanzania. 121 (61.7%) patients were married while patients with divorced, separated and widow marital status were 10(5.1%), 7(3.6%) and 39 (19.9%) respectively and 13(6.6%) and 6 (3.1%) were not married but live with a spouse and single. 172 (87.8%) patients were not engaged with multiple sexual partners while 31 (15.8%) patients used contraceptive drug

for family planning and 64 (32.7%) and 8 (4.1%) patients had history of alcohol intake and smoking behavior.

85 (43.3%) patients were categorized as post menopause and the remaining 111 (56.6%) as pre menopause due to their age limit at the diagnosis period where patients with less than or equal to 50 years categorized as pre menopause and above 50 as post menopause. 105 (53.6%) patients were diagnosed with filtrating ductal carcinoma and 27(13.8) with invasive ductal carcinoma and remaining 64 (32.6%) with others grade of breast cancer. Patients diagnosed with different stage of cancer disease where 15 (7.7%), 32 (16.3), 51 (26%) and 98 (50%) with respective stage I, II, III and IV. There is evidence that most of the patients did not engaged in regular screening for breast cancer since 54.2% The treatment given to patients indicate that 71.9% of patients received Chemotherapy, 14.3 received radiotherapy and 13.8% other mode of treatment such as chemo-radiotherapy, parental I.M./I.V. drugs, intravenous fluids. The median and mean duration of the breast cancer from the diagnosis period were 9 and 12.81 months respectively, mean age of patients were 50.09 years, mean number of pregnancies and children born by patients were respectively 4.15 and 4.02.

Table 1: ORCI Patients diagnostic details from January 2014-December 2015

Covariates	Number of patients	Percentage
Referrals: Yes	196	100
MNH	137	69.9
Others	59	30.1
Location of residence		
Rural	77	39.3
Urban	119	60.7
Marital status		
Divorced	10	5.1
Not married	13	6.6
Married	121	61.7
Separated	7	3.6
Single	6	3.1
Widow	39	19.9
Menopause status category		
Pre menopause	111	56.6
Post menopause	85	43.3
Multiple sexual partners		
Yes	24	12.2
No	172	87.8
Contraceptive drug user		
Yes	31	15.8
No	165	84.2
Smoking behavior		
Yes	8	4.1
No	188	95.9
Alcohol intake		
Yes	64	32.7
No	132	67.3
Grade of cancer		
Infiltrating ductal carcinoma	105	53.6
Invasive ductal carcinoma	27	13.8
Others	64	32.6
Cancer stage		
I	15	7.7
II	32	16.3
III	51	26.0
IV	98	50.0
MODE OF TREATMENT		
Chemotherapy	141	71.9
Radiotherapy	28	14.3
Others	27	13.8
Breast		
Both	4	2.0
Left	99	50.5
Right	93	47.4
Family history		
Yes	13	6.6
No	183	93.4
Survival status		
Alive	37	18.9
Dead	79	40.3
Lost	80	40.8
Death caused by cancer		
Yes	79	40.3
No	117	59.7

Cox Proportional Hazards Regression models

The basic Cox PH model attempted to fit survival data with covariates z to hazard function of the form; $h(t/z) = h_0(t)exp\{\beta \cdot z\}$ where β is unknown vector estimated parameters of covariates and $h_0(t)$ is the baseline hazard which is non-parametric. Primary interest lies in finding the parameters β which is found by partial likelihood. The estimate of the survival function for an individual with covariates z obtained via $\widehat{S(t/z)} = \widehat{S_0(t)}^{\exp(\hat{\beta}' z)}$. Cox PH regression models show effect of covariates on the hazard rate. Cox regression performed using the Breslow method since we don't have extensive tied death times and Breslow method is efficient computationally even though same results obtained by using exact and Efron methods, the Hazard ratio for CPH assumed to be constant over time. Results in table below show that cancer stage of patients significantly affects the survival times of patients.

Table 2: CPH Estimation using Breslow method

Covariates	Coef	exp(coef)	se(coef)	z	P
Location of residence	-0.06339	0.93858	0.26906	-0.24	0.814
Multiple sexual partners	0.10302	1.10851	0.39428	0.26	0.794
Contraceptive drug user	-0.07309	0.92952	0.35208	-0.21	0.836
Stage cancer II	0.34834	1.41672	1.16243	0.30	0.764
Stage cancer III	1.63024	5.10510	1.03840	1.57	0.116
Stage cancer IV	2.48001	11.94144	1.01981	2.43	0.015
Smoking behavior	-0.35023	0.70453	0.75967	-0.46	0.645
Breast	-0.20245	0.81673	0.19391	-1.04	0.296
Alcohol intake	-0.13569	0.87311	0.25910	-0.52	0.600
Menopause status category	-0.35518	0.70104	0.41823	-0.85	0.396
Age	0.01303	1.01312	0.01601	0.81	0.416
Number of pregnancies	0.06875	1.07117	0.23347	0.29	0.768
Number of children	-0.14709	0.86321	0.23734	-0.62	0.535
Duration symptom in month	-0.00769	0.99234	0.00963	-0.80	0.425

The -2loglikelihood of 701.6146, concordance of 0.921 with standard error of 0.064 and R-square of 0.423 with maximum possible R-square of 0.904 indicated the significance of the mixture model of categorical and continuous covariates. The p-value for all three tests (likelihood, Wald and Score) were significant (less than the specified level of significance, 5%), indicating that the model is significant. These tests evaluate the omnibus null hypothesis that all coefficients were 0, and the omnibus null hypothesis was soundly rejected.

The column marked Z gives the Wald statistic value on table 9 shown above. It corresponds to the ratio of each regression coefficients to its standard error. The Wald statistics evaluates whether the coefficients of given variable is statistically significantly different from zero, from the table below shows the stage were significantly at 5% other covariates such as location of patients residence, history of multiple sexual partners, smoking cigarette, alcohol use, cancer grade, contraceptive drug use, menopause status category fails to be significant since p-value were greater than level of significance.

Similarly, the p-value of stage IV was 0.015 was significant at 5% and the hazard ratio of 2.48001 indicating a strong relationship between patient cancer stage IV and increased risk of death. Holding other covariates constant, a higher value of stage IV associated with a poor survival. By contrast, the p-value for location of residence, multiple sexual partner, contraceptive drug use, smoking history, alcohol intake history, menopause category cancer grade were higher than level of significance, respective hazard ratio of 0.06339, 0.10302, 0.07309, 0.35023, 0.13569 and 0.35518 with 95% confidence interval include 1 indicates that the covariates makes small contribution to the difference in HR after adjusting for stage, and only trend toward significance.

Stepwise regression

To check whether all covariate deserve to be included in the model, stepwise regression was used. Remove covariates with lowest AIC and excluded in subsequent analysis with minimum model AIC and minimum log likelihood was 350.7551. After removing the covariates that did not significantly deserve to be in the model at 5% level of significance, table below shows that breast cancer stages significantly affects the survival times of patients, cancer grade options significantly affected the survival of patients(<0.05). There was evidence of steady increase in the risk of death with advancements in stage II, III and IV of 0.764, 0.116 and 0.015 respectively than those diagnosed at stage I.

Table 3: Stepwise regression

	Df	AIC
None		392.31
Stage	3	318.62

Testing for proportional hazards assumption

Proportional hazards assumption assumes that the hazard for any individual is a fixed proportion of the hazard for any other individual. In Cox PH regression, survival curves must have constant relative hazard functions that are proportional over time. Residuals used to investigate the lack of fit of a model to a given subject. For Cox regression, there is no easy analog to the usual observed minus predicted residual of linear regression. Schoenfeld (1982) proposed the first set of residuals for use with Cox regression and exact results were not really different than approximate methods looking for non-straight line and indicate violation of Cox proportional hazards assumptions for all covariate fitted. Table below shows the global test gives a p-value greater than 5% that is not significant suggesting that the assumption has not been violated for Cox model.

Table 4: Assessment of the Cox PH assumption

Covariates	Rho	Chisq	p
Location of residence	-0.04735	2.11e-01	0.6458
Multiple sexual partners	-0.00210	4.81e-04	0.9825
Contraceptive drug user	-0.09702	7.64e-01	0.3819
Stage cancer II	-0.10746	9.48e-01	0.3303
Stage cancer III	0.00804	5.30e-03	0.9420
Stage cancer IV	-0.02068	3.48e-02	0.8520
Smoking behavior	0.13029	1.52	0.2181
Breast	0.01624	1.66e-02	0.8975
Alcohol intake	0.02558	5.70e-02	0.8113
Menopause status category	0.24035	4.40	0.0359
Age	-0.22333	3.97	0.0462
Number of pregnancies	-0.01274	1.68e-02	0.8969
Number of children	0.03899	1.54e-01	0.6950
Duration symptom in month	0.04956	2.31e-01	0.6308
GLOBAL	NA	1.19e01	0.6164

The median survival times of the CPH model estimated to be 738 days with 79 patients died due to breast cancer observed out of 196 patients with 95% lower confidence limit of 414 days. As shown on the figure below is the survival function graph of the CPH model with upper and lower confidence bound.

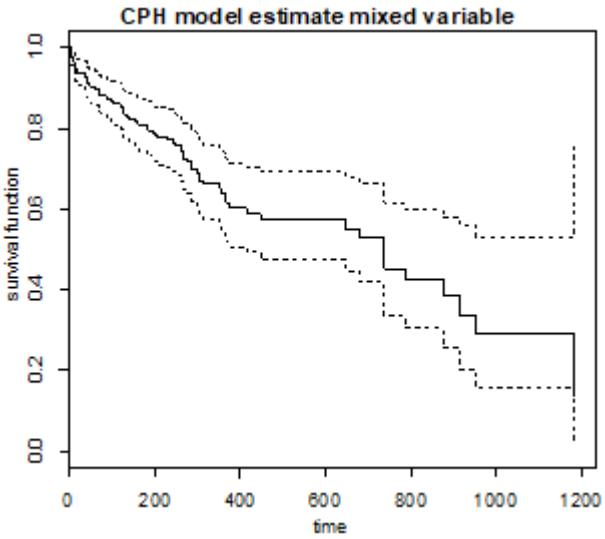
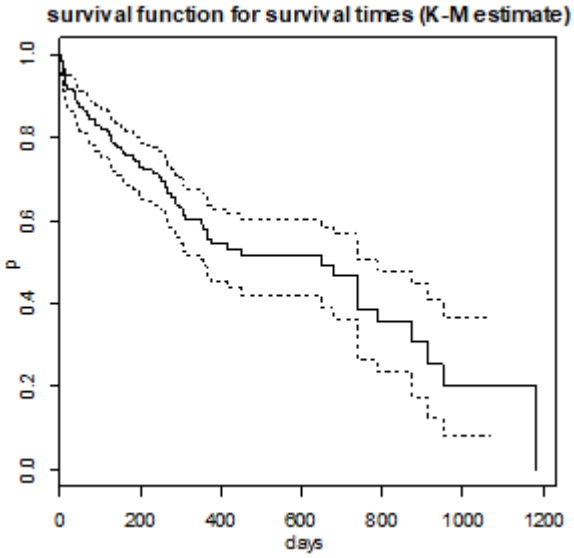


Figure 1: Survival function of the CPH model

Kaplan-Meier estimate and pointwise bounds

The Kaplan-Meier estimate of the survival function corresponds to the non-parametric maximum likelihood estimate (MLE). Right censoring was done on observations that were either alive or lost to follow up. The Kaplan-Meier estimator used since it incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by observed survival and censored times. The figure shows a sharp drop in the survival of patients within the first few days, indicating that most of the patients experience the event early before 800 days. As shown on the figure below there were no difference on estimated median survival times for both Kaplan-Meier and Fleming-Harrington estimates rather the difference on mean survival times where 584.229 days observed for Kaplan-Meier estimates and 590.3 days for Fleming-Harrington with 95% confidence interval.



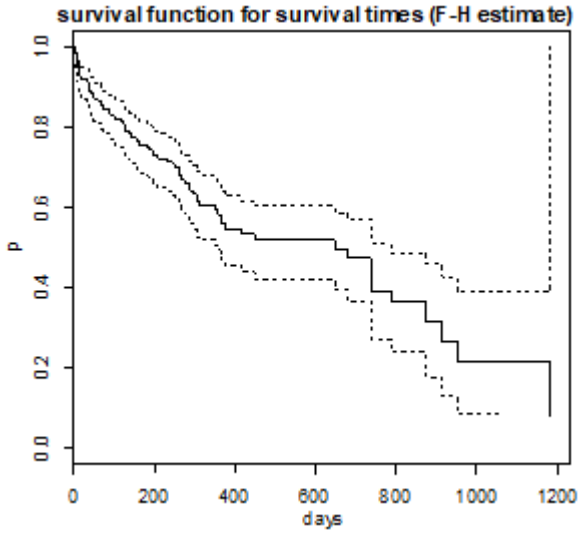


Figure 2: Kaplan-Meier and Fleming-Harrington Survival functions plots

Survival curve gives a visual representation of the survival trend, drops in survival curve occur whenever the terminal event occurs to patient where 79 death due to breast cancer out of 196 observed with true mean survival times of 584.229 days and median of 647.00 days (standard error of true mean= 43.774) while that of Fleming-Harrington median and true mean of survival times were 647.00 and 590.30 days respectively (with standard error of true mean= 43.7).

Table 5: Patients Means and Medians for Survival Time

Covariates	Mean				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Smoking behaviour								
YES	577.500	155.220	273.270	881.730	.	.	.	.
NO	577.186	44.106	490.738	663.634	647.000	158.691	335.965	958.035
Alcohol intake								
YE	613.367	68.272	479.554	747.181	738.000	305.125	139.954	1336.046
NO	569.247	50.818	469.645	668.850	647.000	186.343	281.768	1012.232
Contraceptive drug user								
YES	463.307	65.201	335.513	591.101	367.000	80.810	208.613	525.387
NO	593.850	47.225	501.289	686.412	647.000	173.391	307.154	986.846
Multiple sexual partner								
YES	514.255	76.435	364.442	664.069	647.000	254.145	148.875	1145.125
NO	586.729	46.815	494.971	678.487	447.000	163.254	127.022	766.978

There is significant evidence from observation that patients with multiple sexual partners to have lower median and mean survival times and hence patients who diagnosed with breast cancer engaged with multiple sexual partners dies early. Patients who use contraceptive methods of family planning had higher mean and median survival times than the one who didn't use. The result further indicated that patients who had history of smoking behavior and alcohol intake had higher mean survival times of 577.500 and 613.367 days. The confidence bands, which are a bit more generalized, constructed. These bands bounds on an entire range of time, that is, for a 95% confidence band, the probability that any part of the true curve is out of the confidence bands is 0.05. The figure below is the K-M estimate cumulative hazard with 95% confidence band with Nelson-Aalen estimator the curve shows no changes

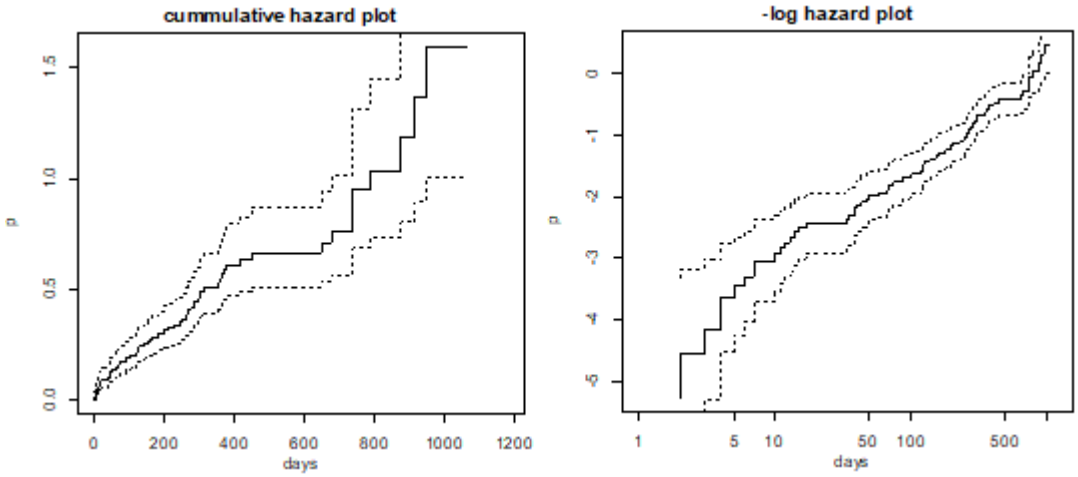


Figure 3: K-M estimates cumulative and negative log hazards with 95% CI

The mean survival times at stage I was 512.769 days, 962.017 days at stage II, 627.131 days at stage III and 417.171 days at stage IV and this shows that patients with late stage had lower survival times than other stage as shown on table below. Survival probability for each stage differs and 58 patients with late stage IV died and had lower rate of 0.050 than other stages where 1 patient with Stage I and survival probability of 0.923 was higher, 3 patients with stage II and survival probability of 0.897 and 17 patients with stage III had survival probability of 0.424.

Table 2: Patients breast cancer stages characteristics

Cancer stages	Stage I	Stage II	Stage III	Stage IV
Probability of survival	0.923	0.897	0.424	0.050
Mean survival times (days)	512.769	962.017	627.131	417.171
Number of patients died	1	3	17	58

Using log rank test, Peto and Peto modification of the Gehan-Wilcoxon test, the survival at different cancer stages of diagnosis were significantly different since p-values for both were less than 5% indicating that the survival at different stages of diagnosis is different and depict that the risk of death was higher at advanced stage as shown on the figure below since patients diagnosed at later stage had lower survival times compared to other stages of survival function at different cancer stages. Other covariates were not significant since p-values for both test were greater than 5% and depict that other covariate considered were not different on survival times.

Table 7: Mantel-Haenszel and Peto and Peto tests

Covariates	Mantel-Haenszel test			Peto and Peto test		
	Degrees of freedom	Chi-square	p-value	Degrees of freedom	Chi-square	p-value
Stage of cancer	3	27.2	5e-06	3	24.9	2e-05
Location of residence	1	0	1	1	0.1	0.8
Multiple sexual partners	1	0	0.9	1	0	0.9
Contraceptive drug user	1	0.1	0.7	1	0	0.8
Smoking behavior	1	0.2	0.6	1	0	0.9
Breast location	2	5.2	0.08	2	4.9	0.09
Alcohol intake	1	0.1	0.8	1	0	1
Menopause status category	1	0	0.9	1	0	0.9

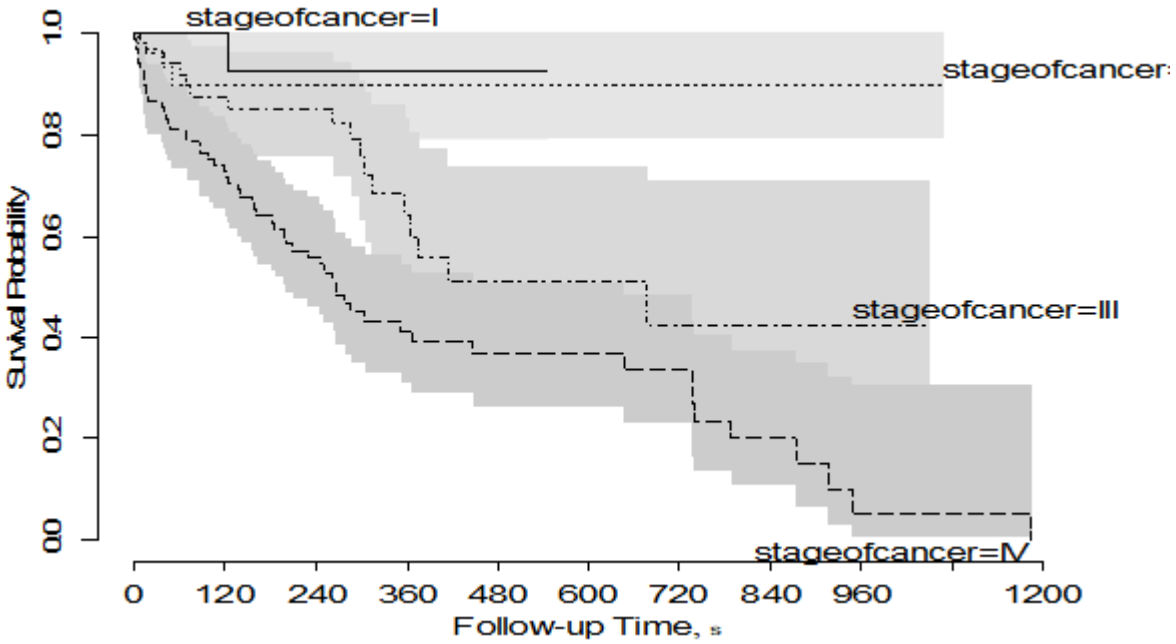


Figure 4: Survival function at different cancer stages

4. Discussion

Detection of breast cancer at early stages through regular screening programs of women and comprehensive treatment should be taken up to improve the overall survival of the patients than waiting for patient to reach the latter stage of the disease. Findings shows that the survival of patients was poor and patients with latter cancer stage i.e. stage IV had an increased risk of death compared to those with earlier stages such as stage I, II and stage III. The probability of survival of patients with breast cancer was 0.204 and for stage I was 0.923, stage II was 0.897, for stage III was 0.424 and stage IV was 0.050. Chance of survival for stage IV was lower compared to others due to patients' late arrival at treatment institute due to delay of being referred to other hospitals. Improve awareness is necessary in controlling breast cancer since patient resides in rural areas had higher median survival times of 647.00 days than those resides in urban areas with median survival times of 447.00 days and can be done by having health education on breast cancer introduced in teaching. Carry out regular screening programs to create awareness and this will encourage women to attend regular screening therefore increase chances of diagnosis at the precancerous or earlier stages which is curable with lower cost. The results of this paper showed non parametric median survival times for Kaplan-Meier estimates and Flemming-Harrington estimates approach each other of 647 days and were lower compared to that of semi parametric method of 738 days.

Table 8: Semi and non parametric survival time estimates

True mean (rmean)	SE (rmean)	Median	0.95LCL	0.95UCL	
584.2	43.5	647.0	357.0	787.0	Kaplan-Meir estimate
590.3	44.7	647.0	357.0	787.0	Flemming-Harrington estimates
		738.0	414	NA	CPH model

Tanzania government should intervene on reducing burden of cancer treatment cost on the patients since it is very necessary because cancer treatment is very expensive for an average Tanzanian and the poor cannot afford it and in most cases patients are undertreated because they cannot meet the high costs of longer treatments or tend to leave the treatment facilities. There is also need for improved and more efficient equipment required in cancer treatment and more institute to be introduced than being dependant on one specialized institute. These interventions will greatly improve the survival of patients diagnosed with cancer. For CPH model fitted the stage of breast cancer of patients affected the survival days of the patients and therefore, among the covariates determined, the cancer stage significantly affected the survival of patients for both semi parametric and non parametric models than the other covariates.

5. Conclusion

Estimating survival functions for different diseases interested statisticians for number of years and since survival function gives information on the probability of a time-to-event of interest which was death caused by breast cancer for this study. This paper prefer Cox regression as semi parametric and non parametric models instead of parametric models for survival analysis because of few assumptions but under certain circumstances the parametric models give more precise estimates due to its condition but it is not valid when the PH assumption does not hold at all.

References

[1] Woolson, R.F., (1981); *Rank test and a one sample log rank test for comparing observed survival data to standard population*. Biostatistics, 37:687-696.
[2] Leung, K.M., R.M. Elashoff and A.A. Afifi, (1997); *Censoring issues in survival analysis*, Annual Review of Public Health, 18:83-104

[3] Kaplan, E. L. and Meier, P. (1958): *Nonparametric estimation from incompletes observations*. Journal of American Statistical Association, 53:457-451.
[4] Agresti A. (2002); *Categorical Data Analysis, 2nd edition*, John Wiley and Sons, New York
[5] Izenman A.J. and Tran L.T. (1990); *Estimation of the survival function and hazard rate*, Journal of Stat planning and inference, V24, p233-247
[6] Nihal A and Tekin M.S. (2007); *Cox regression model with Non proportional hazard applied to lung cancer survival*, Hacettepe Journal of Mathematics and Statistics, vol. 36, No. 2, p157-167
[7] Monem A.M. (2014); *Survival analysis by using Cox regression model with application*, International Journal of Scientific&Technology research, vol. 3, issue 11.
[8] Andersen P.K., Borgan O., Gill R.D., Keiding N. (1993); *Statistical Models Based On Counting Processes*, Springer, New York.
[9] Schafer J. L. (2002); *Analysis of Incomplete Multivariate Data*, London, Chapman and Hall
[10] Rubin D. B. (1987); *Multiple imputation for Non response in Surveys*, New York, John Wiley and Sons
[11] Lee E. T. and Wang J. W. (2003); *Statistical Methods for Survival Data Analysis*, 3rd edition, Wiley & Sons Inc, New Jersey.
[12] Swaminathan R, Lucas E and Sankaranarayanan R (2011); *Cancer survival in Africa, Asia, the Caribbean and Central America: Database and Attributes*, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon
[13] Swaminathan R, et al (2002); *Effect of loss to follow up on population based cancer survival rates in developing countries.* Int J cancer suppl. 13:172
[14] Collet D. (2003); *Modelling survival data in medical research*, Chapman and Hall, London
[15] WHO (2007), *The World Health Organization's fight against cancer; Strategies that prevent, cure and care*, WHO cancer brochure
[16] WHO/ICO, *HPV Information Centre on HPV and Cervical Cancer (HPV Information Centre)*, 2010, WHO
[17] World Health Organization (WHO), 2008, *The Global Burden of Disease: 2004 Update*. Geneva: World Health Organization, 2008
[18] WHO, *Human Papillomavirus (HPV) and Cervical Cancer, s.l.: World Health Organization (WHO).* 2013.
[19] Schoenfeld D. (1982), *Residuals for the proportional hazards regression model*, Biometrika, 69(1): 239-241
[20] Agrest A., (2007); *Introduction to categorical data analysis*, Wiley & Sons Inc. Hoboken, New Jersey
[21] Hosmer D.W and Lemeshow S (1999); *Applied survival analysis: Regression modeling to time to event data*, John Wiley & Sons Inc, New York
[22] Sharma S. (1996); *Applied multivariate techniques*, John Wiley & Sons Inc. Canada
[23] Lwanga S.K. and Lemeshow S. (1991); *Sample size determination in health studies: A practical guide*, WHO manual
[24] Theaune, T. and Gambach (2001); *Modelling survival data- Extending the Cox model*, Springer
[25] Zhou Mai (2000); *Understanding the Cox model with time-change covariates*, the American Statistician
[26] Zhou Mai (2000); *Use software R to do survival analysis and simulation- a tutorial*, the American statistician
[27] Kalbfleish & Prentice (2002); *The statistical analysis of failure time data*, 2nd edition, Wiley and Sons