Analysis of Risk Factors of Death among Breast Cancer Patients in Ethiopia: Parametric shared frailty model

Dessie Yismaw (yismawd@gmail.com)
Ethiopian Public Health Institute
Belay Desyebelew
Dire Dawa University

Research Article

Keywords: Breast Cancer, Accelerate Failure Time Model, Parametric Shared Frailty Model, Random Effect, Survival Analysis, Ethiopia

Posted Date: September 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1103204/v1

Analysis of Risk Factors of Death among Breast Cancer Patients in Ethiopia: Parametric shared frailty model

Dessie Yismaw[1*], Belay Desyebelew[2]

[1] Ethiopian Public Health Institute, Assosa, Ethiopia

[2]Department of Statistics, Dire Dawa University, Dire Dawa, Ethiopia

*Correspondence: Dessie Yismaw
yismawd@gmail.com

Abstract

Breast cancer is the most frequently occurring cancer next to cervical cancer in both developed a nd developing countries with high risk in developed countries and low risk in developing countri es. The main objective of the study is to investigate the risk factors, median survival time, compa re the survival curves, compare the performance of AFT and parametric frailty models in modelin g survival time of breast cancer patients. A retrospective study was employed on 392 breast canc er patients registered from 2013-2018 in University of Gondar and Felege Hiwot Referral Hospit als in northwest of Ethiopia. The median survival time of breast cancer patients who live in the ur ban and rural areas are 34 and 35 months, respectively. The family history of breast cancer has 0 .643 times shortened survival time than no family history of breast cancer. *The frailty of Universi ty of Gondar and Felege Hiwot Referral Hospitals were 0.536 ($\mu<1$) and 1.465 ($\mu>1$), respective ly.* The study employed multivariable lognormal-gamma shared frailty model and the results of th e model revealed that age, weight, place of residence, tumor size, stage, number of estrogens, stat us of estrogen receptors, number of progesterone and family history of breast cancer are statistica lly significant factors which affect the survival time of breast cancer patients.

Keywords: Breast Cancer, Accelerate Failure Time Model, Parametric Shared Frailty Model, Random Effect, Survival Analysis, Ethiopia.

1 Introduction

Breast Cancer is a type of cancer which attacks the tissues of breast and the most common type of breast cancer is ductal carcinoma, which begins in the lining of the milk ducts (thin tubes that carry milk from the lobules of the breast to the nipple). Another type of breast cancer is lobular carcinoma, which begins in the lobules (milk glands) of the breast. Invasive breast cancer is breast cancer that has spread from where it began in the breast ducts or lobules to surrounding normal tissue. Breast cancer occurs in both men and women, although male breast cancer is rare (1).

The stage of breast cancer, according to WHO is classified as stage I: The tumor is no larger than two centimeters, stage II: The tumor is around five centimeters in size and may have spread to the lymph nodes under the arm; stage III: The tumor(s) may have spread to lymph nodes, be clumped together and stage IV: Tumor(s) that have spread to other organs in the body.

Globally, in 2008, according to world health organization (WHO) the estimated diagnosis of breast cancer was 1383500 and estimated death with breast cancer was 458400 (33%). The disease is common in both developed and developing countries and the highest rate is in European (EU)-28 countries. The incidence rate was around 80.3 and the mortality rate 14.4 (2). Several studies were conducted in European (EU)-28 countries to support the above facts including the UK (3), France (4-5) and Belgium (6).

The study in African showed that social determinants may influence differences in breast cancer care among the African population (7). The most common cancer in Africa women is breast cancer in the 35-55 year age group (8). The study conducted in sub-Saharan Africa and Eastern African countries showed breast cancer (34 to 46 cases per 100,000) is the second most often occurring cancer following cervical cancer including Ethiopia (9).

Secginli and Nahcivan (2006) identified that alcohol use and smoking are avoidable and family history of breast cancer is non-avoidable risk factor associated with breast cancer. These authors also identified age, dense breast tissue, Estrogen exposure and breast-feeding as risk factors of breast cancer.

Survival time analysis is often conducted with Kaplan-Meier, life-table and the cumulative hazard estimator and semi-parametric method (11-12). These methods require large sample size and it is very hard to get estimates of the hazard function. The above mentioned methods also assume a

homogeneous population (have same risk) and the parameter estimations are based on partial likelihood approach so that the inference is almost exclusively based on asymptotic results (13).

However, some researchers used parametric models which are powerful to detect risk factors and require lower sample size (14).

In USA, a retrospective cohort study was conducted and using a likelihood based criteria for model selection, Weibull model was found to be the best parametric survival model to examine the effect of predictor variables on the survival time of breast cancer patients (15).

In the real world, populations are not homogeneous since the effect of drug might be individual specific or group specific or each subject might have its own biological response, living condition, hereditary factors, even unobserved factors etc. To fill the gap of excluding unobserved factors, we use advanced model which combines measured/observed factors with unmeasured/unobserved factors, and it provides multiplicative effect on the baseline hazard function; namely frailty model. Frailty model provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data (16). The main focus of this study is, therefore, to explore factors that affect the survival time of breast cancer patients by using parametric shared frailty model.

The median time of breast cancer deaths as proposed by Surveillance, Epidemiology, and End Results (SEER) statistics, the average age of diagnosis of breast cancer was 62 years between the years of 2010 and 2014. Furthermore, breast cancer in women is most often diagnosed between the ages of 55 and 64 years (17).

Time to death of breast cancer patients is affected by wide range of factors related with social, cultural and socio-demographic factors such as age, gender, marital status, employment status, level of education, and income (18). A retrospective study which is conducted on 686 breast cancer patients in German to compare the performance of the common parametric models (Exponential, Weibull, Lognormal and Log logistic) revealed that log-normal model is better for breast cancer dataset (19). The authors also found that tumor size, lymph nodes, tumor grade and progesterone receptors have statistically significant effect on the survival time of breast cancer patients. Fortunately, the socio demographic variables were found to be statistically insignificant which

might be due to unmeasured factors or random effects that were not considered in the model. As a result, we have included random effects in our model (shared frailty models).

Faradmal J *et al.* (2012) conducted a retrospective cohort study on 769 women in Iran to determine factors associated with breast cancer patients. Based on CPH and Gamma-frailty models, tumor size, tumor grade, status of involvement of lymph nodes and number of involved lymph nodes were statistically significant (20).

According to Pereira and Stokes (2018) 5 years and 10 years study, the estimated survival probabilities of breast cancer were 80% and 70%, respectively. Mean tumor size at diagnosis was 22 mm with a standard deviation of 18.4 mm and mean age at diagnosis was 58.2 years with a standard deviation of 13.5 years and median survival time was 11 years.

A retrospective study conducted on 114 patients at black lion Specialized Hospital (BLSH) and St. Paul's Hospital Millennium Medical College (SPHMMC) showed that the molecular subtypes were statistically significant in different age ranges ($P< 0.05$). But there was no statistically significant difference in tumor grade, histology stages between the molecular subtypes of breast cancer. The author also concluded that in Ethiopia and other East African countries, hormone receptor negative tumors are not the most common molecular subtypes of breast cancer (22).

2 Materials and Methods

The study was conducted based on data taken from University of Gondar and Felege Hiwot Referral Hospitals found in the west Amhara regional state and are the only referral hospitals which have cancer treatment center in the region. A retrospective cohort study design was employed on 392 breast cancer patients who started treatment between January, 2013 to December, 2018 and have at least one followed up to retrieve relevant information from the medical records to address the objective of the study. The patients' identification numbers were used to select the relevant information from different breast cancer patients' record charts.

2.1 Variables Included in the Study

The response (dependent) and predictor variables used in the model for the estimation of parameters are defined as in Table 1.

The dependent (response) variable was the survival time of breast cancer among breast cancer patients in northwest of Ethiopia. It was measured in months and is the length of time from breast cancer started date to death or censor. Breast cancer patients who stayed alive during the study time, lost to follow–up or died by other than the breast cancer causes were considered as censored. The event status is (0=censored, 1=died due to breast cancer).

Table 1: Operational Definition and Categorization of the Independent Variables

Variables	Definition and categories with codes
Age	Patient age at baseline (continuous)
Residence	Patient's Residence (Urban=0, rural=1)
WHO stage	WHO clinical stage of BC (1=stage I,2=stage II,3=stage III,4=stage IV)
Weight	Weight of patient (in Kg)
Tumor size	Tumor size of patient (mm)
Estrogen	Number of estrogen receptor (1-1144)
Status of estrogen	Status of estrogen (positive=0, negative=1)
Progesterone	Number of progesterone receptor (1-2380)
Status of progesterone	Status of progesterone (positive=0, negative=1)

Hospital was considered as clustering effect in frailty model.

2.2 Cox PH Regression Model

The Cox proportional hazards (CPH) regression model is a broadly applicable and the most widely used method of survival analysis. Survival models are used to quantify the effect of one or more explanatory variables on failure time. The Cox PH model is a semi-parametric model where the baseline hazard h(t) is allowed to vary with time.

$$\mathrm{hi}(\mathrm{t}|\mathrm{x}) = \mathrm{ho}(\mathrm{t})\exp(\beta 1\mathrm{X1i} + \beta 2\mathrm{X2i} + \cdots + \beta \mathrm{pXip}) = \mathrm{ho}(\mathrm{t})\exp(\beta'\mathrm{X}) \qquad (1)$$

Where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero. With p covariates collected in a vector;

$Xi = (xi1, xi2, .. xp)t$ is a vector of explanatory variables for a particular individual and $\beta' = (\beta1, \beta2, \dots, \beta p)$ is vector of regression coefficients.

The corresponding survival function for Cox-PH model is given by:

$$S(t, X, \beta) = \{So(t)\}^{\exp\left(\sum_j^p \beta' X\right)} \quad (2)$$

Where, $So(t)$ is the baseline survival function. In Cox-PH model, there is no distributional assumption made for the survival time; the only assumption is that the hazards ratio does not change over time. In the Cox proportional hazards model, the outcome is described in terms of the hazard ratio and the hazard ratios of two individuals with different covariates X and X^* can be given by:

$$HR = \frac{ho(t)\exp(\beta' x)}{ho(t)\exp(\beta' X^*)} = \exp\{\sum \beta'(X - X^*)\} \quad (3)$$

2.3 Shared Frailty Model

The frailty approach is a statistical modeling concept which aims to account for heterogeneity caused by unmeasured covariates and is a random effect model for time-to-event data. The shared frailty model is used with multivariable survival data where unobserved frailty is shared within groups of individuals, and thus a shared frailty model may be thought of as a random effects model for survival data (25). The shared frailty approach assumes that all failure times in a cluster (GURH ֆ FHRH) are conditionally independent given the frailties. The value of the frailty term is constant over time and common to individuals in the cluster (each hospital). Conditional on the random effect called the frailty and denoted by ui, the survival times in cluster i ($1 \leq i \leq n$) are assumed to be independent and the proportional hazard frailty model assumes:

$$hij(t|xij, \omega i) = h0(t)\exp(\beta' Xij + \omega i) \quad (14)$$

Where $hij(t)$ is the conditional hazard function for the j^{th} subject from the i^{th} cluster (conditional on ωi), ho(t) is the common parametric baseline hazard function, **β** is a vector of unknown regression coefficients, X_{ij} is the vector of covariates, and ωi is the random effect for i^{th} cluster. This model can be rewritten as;

$$hij(t|xij, ui) = h0(t)ui\ \exp(\beta' Xij) \quad (15)$$

Where $ui = \exp(\omega i)$ is called the frailty for the i[th] cluster and the model is called the shared frailty model because subjects in the same cluster share the same frailty factors.

The useful frailty distributions are gamma frailty and inverse Gaussian frailty among others. Consequently, we use gamma and inverse Gaussian shared frailty models for the current study.

2.3.1 Shared Gamma Frailty Distribution

The density function and Laplace transformation of gamma distribution for frailty u can be given as follows:

$$f\mathrm{u}(\mathrm{u}i) = \begin{cases} \dfrac{\mathrm{ui}^{1/\theta-1}\exp\left(-\frac{\mathrm{ui}}{\theta}\right)}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} & ;\ \mathrm{u} > 0, \theta > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{16}$$

Where Γ (.) is the gamma function with the corresponding Gamma distribution gm (μ, θ) with $\mathrm{E}(\mathrm{u}) = 1$ and $\mathrm{Var}(\mathrm{u}) = \theta$. The parameter θ provides information on the variability (heterogeneity) in the population of clusters. The associated Laplace transformation is:

$$\mathrm{L}(\mathrm{s}) = \left(1 + \frac{\mathrm{s}}{\theta}\right)^{-\theta} \ , \theta > 0 \tag{17}$$

If $\theta > 0$, then there is heterogeneity. So, the large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups. The conditional survival function of the gamma frailty distribution is given by (26).

$$\mathrm{S}\theta(\mathrm{t}) = [(1 - \theta \ln\{\mathrm{S}(\mathrm{t})\}]^{\frac{-1}{\theta}} \ , \ \theta > 0 \tag{18}$$

The conditional hazard function of the gamma frailty distribution is given by (26).

$$\mathrm{h}(\mathrm{t}|\theta) = \mathrm{h}(\mathrm{t})[1 - \theta \ln\{\mathrm{S}(\mathrm{t})\}]^{-1} = \frac{\mathrm{h}(\mathrm{t})}{[1-\theta \ln\{\mathrm{S}(\mathrm{t})\}]} \tag{19}$$

Where,$S\ (t), h(t)$ are the survival and the hazard functions of the baseline distribution.

2.3.2 Inverse Gaussian Shared Frailty Distribution

The probability density function of an inverse Gaussian shared distributed random variable with parameter $\theta > 0$ is giv*en by:*

$$\mathrm{fu}(\mathrm{ui}) = \begin{cases} \frac{1}{\sqrt{2\pi\theta}}\ \mathrm{Ui}^{-\frac{3}{2}} \exp\left(-\left(\frac{\mathrm{Ui}-1}{2\theta\mathrm{Ui}}\right)^2\right) & ;\ \theta > 0\ , U > 0 \\ 0, \quad \text{other wise} \end{cases} \tag{20}$$

For identifiability, we assume U has expected value equal to one and its variance is θ. The Laplace transformation of the inverse Gaussian distribution is given by (27).

$L(s) = \exp[\frac{1-\sqrt{1+2\theta s}}{\theta}]$, $\theta > 0$, $s > 0$ And for q≥ 1

$$L^{(q)}(s) = (-1)^q(2\theta s + 1)^{-\frac{q}{2}} \frac{Kq-(1/2)(\sqrt{2\theta^{-1}(s+1/2\theta)})}{K1/2(\sqrt{2\theta^{-1}(s+1/2\theta)})} L(s) \quad (21)$$

Where *K is* the modified Bessel function of the second kind (25). $K\gamma(\omega) = \frac{1}{2}\int_0^\infty t^{\gamma-1}\exp\{-\frac{\omega}{2}(t + \frac{1}{t})\}\, dt$, γ ε R, ω> 0.

For the inverse Gaussian frailty distribution the conditional survival function(27) is given by:

$$S\theta(t) = \exp\{\frac{1}{\theta}\left(1 - \sqrt{1 - 2\theta\ln\{S(t)\}}\right)\} \quad , \theta > 0 \quad (22)$$

For the inverse Gaussian frailty distribution the conditional hazard function is given by (27).

$$h\theta(t) = \frac{h(t)}{\sqrt{1-2\theta\ln\{S(t)\}}} \quad , \ \theta > 0 \quad (23)$$

Where, S (t) and h (t) are the survival and the hazard functions of baseline distributions. With multivariate data, an Inverse Gaussian distributed frailty yields a Kendall's Tau given by:

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{\exp(2/\theta)}{\theta^2}\int_{2/\theta}^{\infty} \frac{\exp(-u)}{u}\, du \quad , \text{where } \tau\epsilon(0,1/2) \quad (24)$$

3 Results

This section presents the result of the statistical analysis and discussions carried out to answer the basic research questions and the objective of the study. The data management and analysis were done using SPSS 23, STAT 14.1, NCSS 12 and R statistical software.

3.1 Descriptive Analysis

Table 2: Baseline characteristics of categorical covariates with their time-to-event status

Variable	Category	Frequency	Status	
			Censored	died
Place of residence	urban	260(66.3%)	211(63.2)	49(84.5%)
	Rural	132(33.7%)	23(36.8%)	9(15.5%)
Gender	female	383(97.7%)	326(97.6%)	57(98.3%)
	male	9(2.3%)	8(2.4%)	1(1.7%)
Study center	FHRH	196(50%)	160(47.9%)	36(62.1%)
	GURH	196(50%)	174(52.9%)	22(37.9%)
Total		392	334(85.2%)	58(14.8%)

In this study, we used equal allocation sampling technique for each referral hospital. From Table 2, it can be seen that 58 (14.8%) patients were died during the last 60 months. Out of the 392 breast cancer patients, 383 (97.7%) were females and 9 (2.3%) were males. The majority (66.3%) of the patients lived at urban area, 211 (63.2%) while the remaining 132 (33.7%) were living in rural areas. As displayed in Table 3, the average age of the 392 breast cancer patients at baseline is 55.81 years with standard deviation of 12.360 years. The average weight of patients is 59.37 kg with standard deviation of 9.769 kg and the average numbers of estrogen and progesterone involved at the baseline are 144.84 and 306.8, respectively.

Table 3: The descriptive statistics of the continuous covariates

Covariates	Mean 95%CI	Mode	Median	Std.dev	Min	Max
Age at baseline	55.81[54.55,56.87]	57	64	12.36	25	80
Weight at baseline	59.37[58.36,60.20]	59	58	9.769	29	80
Number of estrogens involved	144.84[144.80,145]	145.8	123	75.38	12	1472.1
Number of progesterone involved	306.8[306.4,305.32]	237	161	194.33	25.1	897

3.2 Non-Parametric Survival Analysis

3.2.1 The Kaplan- Meier Estimate of Time-to-Death of Breast Cancer

Survival time distributions for time-to-death due to breast cancer is estimated for each group using the K-M method. In order to compare the survival curves of two groups, log-rank test has been employed. The estimated median time and 95% confidence interval for time-to-death with different covariates are summarized in Table 4.

Table 4: Median survival time and confidence intervals for categorical covariates

Covariates	Categories	Median (months)	95% CI
Place of residence	Urban	34	[32.00,36.50]
	Rural	35	[32.39,39.58]
Study center	GURH	37.50	[35.00,40.58]
	FHRH	32	[28.00,34.00]
Histological stage	I	34	[28.00,37.00]
	II	37	[32.50,41.08]
	III	32.5	[29.42,35.58]
	IV	35.5	[29.29,42.00]
Family history	Yes	35	[32.00,37.00]
	No	34	[33.00,36.00]
Tumor size	≤2	34	[20.50,45.00]
	2.01-6.34	35	[34.00,36.00]
	≥6.35+	35	[32.00,37.00]

The median survival time of breast cancer patients from rural is 35 months. Similarly, the median survival time of breast cancer patients from patients from the urban area is 34 months. The median survival time of breast cancer patients who attend at Gondar University referral hospital is 37.5 months and that of the patients in Felege Hiwot referral hospital is 32 months. The median survival time of breast cancer patients for the other covariates can be seen from Table 4. The overall median survival time of breast cancer patients is 35 months.

3.3 Exploratory Data Analysis

The survival time plot for breast cancer patients by place of residence is shown in Figure 1. The figure shows the difference between the survival probabilities of patients from the rural and urban areas. There is a significant difference in survival times between the patients live in Rural and Urban areas (log rank test p=0.002).

The Kaplan-Meier survival probability estimates from 8 months to 60 months were about 0.992 down to 0.725 for patients who live in the rural areas and 0.992 to 0.442 for patients live in urban areas. The plot indicates that the risk of death in the rural was better than patients live in urban areas. The plot also shows that as time increases the hazard ratio for patients came from urban areas has higher value (2.89). The hazard ratio of patients came from urban areas is 1.89 times more than who came from rural areas. The survival time plot for breast cancer patients by different categorical covariates has also been explored.

Figure 1. KM Survival Plot by Place of Residence The figure shows the difference between the survival probabilities of patients from the rural and urban areas.

3.4 Statistical Analysis

In survival analysis, it is recommended to look at the Kaplan Meier curve for categorical independent variables to provide insight into the shape of the survival function for each group and give an idea of whether or not the groups are proportional. We also considered the test of equality across the groups of covariates to explore whether or not include the covariates in the final model. The covariates which are significant in the univariable analysis are included in the multivariable model. The log rank test is used for the purpose of variable selection for the initial multivariable model.

The study employed Lognormal and Log logistic Frailty Models to analyze the breast cancer data obtained from the two Hospitals and used different model selection criteria (Akaike Information Criterion and log-likelihood) to compare and select a model which fits the data well (Table 5). A lower value of AIC and highest value of log likelihood suggests a best model which fits a given data. It can be seen from the table that the log-normal gamma shared frailty model (with log likelihood of -259.18 and AIC of 544.860) is found to be an appropriate model compared to other models. This indicates it is the final model fits our breast cancer data.

Table 5: Comparison of parametric frailty models based on log likelihood and AIC.

Model baseline distribution	Frailty distribution	Log likelihood	No. covariates	AIC	rank
Lognormal	**Gamma**	**-259.18**	**16**	**544.860**	**1**
	Inverse Gaussian	-259.732	16	551.873	2
Log logistic	Gamma	-259.932	16	555.910	3
	Inverse Gaussian	-259.952	16	560.863	4

****AIC: Akaike Information Criterion, Hospital was considered as clustering effect in frailty model.*

3.4.1 Multivariable Analysis with Log-Normal Gamma Frailty Model

Log-normal gamma frailty model is a model which includes a frailty component. The results obtained from the log normal-gamma frailty model are presented in Table 6. The estimated value of theta ($\widehat{\theta}$) is 0.263 and statistically significant. The associated Kendall's tau (τ), which measures dependence within Gondar university and Felege Hiwot referral hospitals (clusters) is estimated to be 0.116 and se (τ) = 0.121. This reveals that it was significant and is reasonable to include the random effect in the model and the likelihood ratio test confirmed the presence of heterogeneity. These results implied that the frailty component had significant contribution to the model and the estimated shape parameter in the log-normal gamma frailty model is 2.615 (σ=2.615) and greater than unity. Hence, the shape of the hazard function is unimodal which implies it increases up to its maximum point and then decreases.

The acceleration factor of age, 0.98 indicates that the survival of time of breast cancer patients decreases by 2 percent for a unit increase in patients' age (Table 6). The acceleration factor of living in rural areas is 1.806 which indicates that the survival time of breast cancer patients who live in rural areas is estimated to be 80.6% longer than patients live in urban areas. The acceleration factors of the other covariates can be interpreted in the same way. The frailties of Gondar University and Felege Hiwot Referral Hospitals are 0.536(μ<1) and 1.465(μ>1), respectively. The values of the frailties indicate that the survival time of breast cancer patients from Gondar University is longer than patients from Felege Hiwot referral hospital.

Table 6: Summary table for log normal-gamma frailty model

Variable	Categories	Estimate($\widehat{B}$)	SE($\widehat{B}$)	$\widehat{\Phi}$	95% CI	p-value

Age		-0.02	0.07	0.98	[0.968,0.994]	0.031**
Sex	male	--		--	--	--
	female	-0.123	0.631	0.884	[0.257, 3.046]	0.8455
Weight		-0.0295	0.009	0.971	[0.970,0.988]	0.0014**
Place of	urban (ref)	----		----		
Residence	Rural	0.591	0.179	1.806	[1.270,2.564]	0.001**
Tumor	(<=2) (ref)	-------		--	--------	-----
	2.01-6.34	-0.567	0.00462	0.567	[0.562,0.572]	0.00*
	>=6.35+	-0.519	0.0489	0.595	[0.541,0.655]	0.035*
Histolo	I (ref)	---		---	---	
gical stage	II	-0.223	0.324	0.800	[0.724,1.234]	0.491
	III	-0.941	0.248	0.390	[0.24,0.634]	0.0015**
	IV	-1.11	0.2594	0.330	[0.198,0.548]	0.00018*
Number of estrogen		-0.015	0.007	0.985	[0.997,0.999]	0.031**
Status estrogen	Negative (ref)	----		----	--------	
	positive	-0.4690	0.182	0.627	[0.438,0.894]	0.00989*
Number progesterone		0.0021	0.0005	1.002	[1.001,1.003]	0.00025*
Status Progesterone	Negative (ref)	------		----	------	
	positive	0.0502	0.165	1.051	[0.761,1.453]	0.761
Family history	No	-----		---	------	
	Yes	-0.442	0.0534	0.643	[0.579, 0.714]	0.032**
Θ=0.263, τ=0.116 λ=2.314 log likelihood= **-259.18**, **σ**=2.615						
Likelihood-ratio test of Ø = 0: chi-square = 209.42, P-value = 0.000*						

*$\hat{\Phi}$ Indicates Acceleration factor with * significant at 5% level; 95%CI: 95%confidence interval for acceleration factor; SE ($\hat{\beta}$): st andard error for $\hat{\beta}$; Ref. Reference.*

Discussion and Conclusion

The main purpose of this study is to determine the risk factors of death among breast cancer patients. Data were extracted from Felege Hiwot and University of Gondar Referral Hospitals. The comparison the shared frailty models were done using the log-likelihood and AIC values. Accordingly, the lognormal gamma frailty model was found to be the final model which fits our data. The findings of the model showed that age, place of residence, weight, histological stages I, III and IV, number of estrogens, estrogen status, number of progesterone, family history of breast cancer and tumor size categories (2.01-6.34 and 6.35+) were significantly associated with survival time of breast cancer patients.

The result of our study revealed that an increase in age and weight shorten the survival time of breast cancer patients which is consistent with the studies of Secginli and Nahcivan (2006); and Fatma(2000). The results also showed that tumor characteristics have direct effect on survival time of breast cancer patients. Place of residence was significant predictive factor for survival time of the patients. This study is consistent with other findings of Ahmad et al (2014). Moreover, the patients who live in the rural areas have more survival time than patients who live in urban areas. The most striking result is that family history of breast cancer is a factor that significantly predicts the survival time of the breast cancer patients ($p=0.032$). This finding was is consistent with a study conducted by Ahmed et al (2014) which showed that patients who have a mother or sister diagnosed with breast cancer have a greater risk of developing this cancer than patients without any breast cancer family history. The number of estrogens, estrogen status and number of progesterone also have significant association with survival time of breast cancer patients. The findings also provide evidence that higher number of estrogen and the status of estrogen are statistically significant.

This paper focuses on Accelerated failure time and shared parametric frailty models which relies on parametric specification of the baseline hazard and the distribution of the frailty. Here, we have considered two clusters based on the study center (hospitals) for potential dependence in the random quantities corresponding to each failure time which is induced by frailty. Out of the total 392 patients who started cancer medicine (treatments), about 14.8% died at the end of the study and the estimated median survival time of breast cancer patients was 35 months.

Then parametric frailty models were fitted to the breast cancer data set and among these parametric frailty models, the Log-normal gamma frailty model is found to be the best fit survival time of

breast cancer patients' dataset. The result of the log-normal gamma shared frailty model showed that age, weight, place of residence, tumor size, histological stages (I, III, IV), and number of estrogens involved, status of estrogen, number of progesterone and family history of breast cancer were found to be significant predictors for survival time of patients in Gondar University and Felege Hiwot Referral Hospitals. Living in the rural areas, histological stage I and the number of progesterone involved prolong the survival time of breast cancer patients. Similarly, age, weight, place of residence, tumor size, histological stages III and IV number of estrogens involved, status of estrogen and family history of breast cancer shorten the time of death of patients with breast cancer.

Limitations and Recommendation

The main limitation of this study was the absence of some basic clinical and demographical variables, such as, living style, level of education, marital status, smoking, menopausal, etc. Further studies should include all important risk factors for the survival time of breast cancer patients as much as possible. The study recommends future researchers to conduct their study with a larger sample size along with large clusters and a more complete data set determine factors affecting the survival time of breast cancer patients. Regular public and professional education are required to increase the awareness of hereditary breast cancer and the importance of family screening, as well as to promote early diagnosis and treatment.

DECLARATINS

Ethics Approval

Before data collection, a letter of ethical clearance written by the College of Natural and Computational Science of University of Gondar was submitted to Both Felege Hiwot and Gondar Referral Hospitals and permission to collect anonymized data was obtained. The data was extracted by trained data clerks in the Hospitals and none of the researchers had access to original cards of the patients. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Consent to publication
Not applicable**.**

Data Availability

The datasets generated for this study are available online at: https://osf.io/fgbwq

Competing Interests

No potential conflict of interest was reported.

Funding

No funding.

Author Contributions

Dessie Yismaw conceived the idea, performed the data cleaning and analysis and interpreted the ensuing results. Belay Desyebelew contributed to the conception and prepared the manuscript. Both the authors read and approved the final draft.

Acknowledgement

We acknowledge University of Gondar and Felege Hiwot Referral Hospitals, Ethiopia for the data they supplied. The manuscript was prepared on the basis of Dessie Yismaw's Masters' thesis.

References

1. WHO 2013. Latest world cancer statistics. *The International Agency for Research on Cancer (IARC).*
2. ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer, 136.
3. Duffy, S. W., Tabar, L., Olsen, A. H., Vitak, B., Allgood, P. C., Chen, T. H., Yen, A. M. & Smith, R. A. 2010. Absolute Numbers of Lives Saved and Overdiagnosis in Breast Cancer Screeninq, from a Randomized Trial and from the Breast Screening Programme in England. *Journal of medical screening,* 17**,** 25-30.
4. Antona, D. 2006. Le tétanos en France en 2002–2004. *BEH,* 7**,** 53-55.
5. Traumatismes, I. D. V. S. D. M. C. E. 2011. *Programme de dépistage du cancer du sein en France: résultats 2007-2008, évolutions depuis 2004*, Institut de veille sanitaire.
6. Altobelli, E., Rapacchietta, L., Angeletti, P. M., Barbante, L., Profeta, F. V. & Fagnano, R. 2017. Breast cancer screening programmes across the WHO European Region: differences

among countries based on national income level. *International journal of environmental research and public health,* 14**,** 452.

7. Mody, G. N., Nduaguba, A., Ntirenganya, F. & Riviello, R. 2013. Characteristics and presentation of patients with breast cancer in Rwanda. The American Journal of Surgery, 205, 409-413.
8. Polat, K., Sahan, S., Kodaz, H. & Günes, S. A new classification method for breast cancer diagnosis: feature selection artificial immune recognition system (FS-AIRS). International Conference on Natural Computation, 2005. Springer
9. Lemlem, S. B., Sinishaw, W., Hailu, M., Abebe, M. & Aregay, A. 2013. Assessment of knowledge of breast cancer and screening methods among nurses in university hospitals in Addis Ababa, Ethiopia, 2011. ISRN oncology,
10. Secginli, S. & Nahcivan, N. O. 2006. Factors associated with breast cancer screening behaviours in a sample of Turkish women: a questionnaire survey. *International journal of nursing studies,* 43**,** 161-171.
11. Kaplan, E. L. & Meier, P. 1958. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53, 457-481.
12. Klein, J. P. & Goel, P. K. 2013. *Survival analysis: state of the art*, Springer Science & Business Media.
13. Andersen, P. K. & Gill, R. D. 1982. Cox's regression model for counting processes: a large sample study. *The annals of statistics***,** 1100-1120.
14. Cox, D. R. & Oakes, D. 1984. *Analysis of Survival Data.,* London Chapman and Hall.
15. Pham, M. H. 2014. Survival Analysis-Breast Cancer. *Undergraduate Journal of Mathematical Modeling: One+ T*
16. Lee, E. T. & Wang , J. W. 2003. Statistical Methods for survival Data Analysis. New Jersey: John Willey.
17. Vaupel, J. W., Manton, K. G. & Stallard, E. 1997. The Impact of Heterogeneity in individual frailty on the dynamics of Mortality. Demography 6**,** 439-454.
18. Siegel, R. L., Miller, K. D. & Jemal, A. 2015. Cancer statistics, 2015. *CA: a cancer journal for clinicians.*

19. Fatma, S. 2000. Five Year Survival Analysis of Patients with Clinical Stages I and IIA Breast Cancer who Received Initial Treatment at North Carolina Hospitals. *NORTH CAROLINA DEPARTMENT OF HEALTH AND HUMAN SERVICES,* 123*,* 1-10.
20. V. Vallinayagam, S. Prathap & P. Venkatesan 2014. Parametric Regression Models in the Analysis of Breast Cancer Survival Data. *International Journal of Science and Technology* 3**,** 1-5.
21. Faradmal J, Talebi A, Rezaianzadeh A & H, M. 2012. Survival Analysis of Breast Cancer Patients using Cox and Frailty Models. *Journal of Research in Health Sciences,* 12**,** 127-130.
22. Perera, M. & Stokes, C. 2018. A Statistical Model with Non-Linear Effects and Non-Proportional Hazards for Breast Cancer Survival Analysis. *Advances in Breast Cancer Research,* 7**,** 65.
23. Endale, H., Daniel, S., Wondemagegnhu, T., Yonas, B., Abebe, B., Markos, A., Thomas, S., Sofia, D. M., Christina, K. & Mats, G. K. 2018. Breast cancer in Ethiopia: evidence for geographic difference in the distribution of molecular subtypes in Africa. *BMC Women's health***,** 1-8.
24. Gupta, R. C., Akman, O. & Lvin, S. 1999. A study of log-logistic model in survival analysis. *Biometrical Journal: Journal of Mathematical Methods in Biosciences,* 41**,** 431-443.
25. Collett, D. 2003. *Modelling survival data in medical research*, Chapman and Hall/CRC.
26. Hougaard, P. 2000. Statistical inference for shared frailty models. *Analysis of Multivariate Survival Data.* Sp
27. Hanagal, D. 2011. Modeling survival data using frailty models. Sage Publications Ltd.
28. Gutierrez, R. G. 2002. Parametric frailty and shared frailty survival models. *Stata Journal,* 2**.**
29. Ahmad , R. B., Sahar , S. M., Majd, H. A., Mohammad , E. A., Nahid , N. & Kimiya, G. 2014. Survival Analysis of Patients with Breast Cancer using Weibull Parametric Model.

Figures

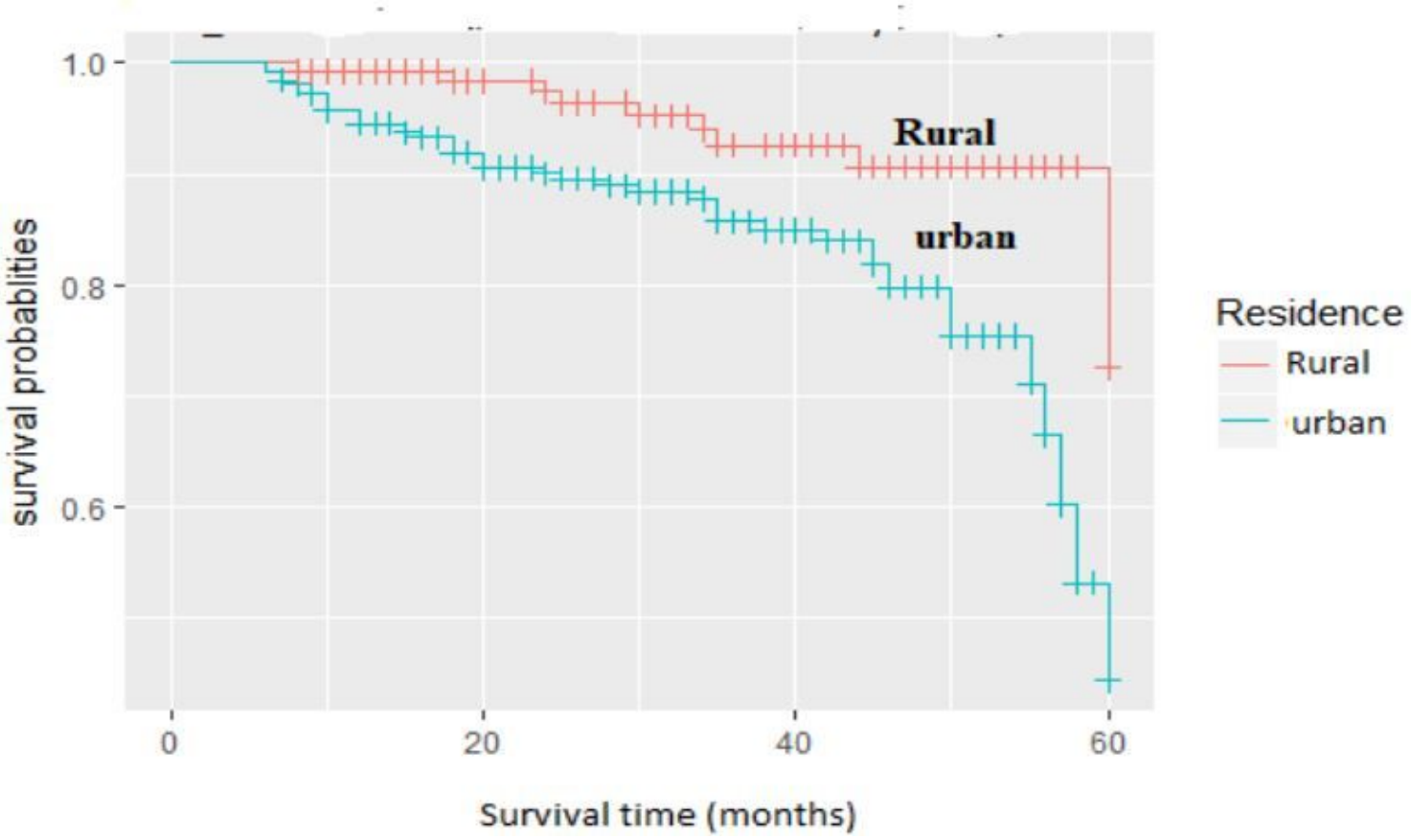

Figure 1

KM Survival Plot by Place of Residence The figure shows the difference between the survival probabilities of patients from the rural and urban areas.