

# Survival Analysis of Recurrent Events on Women Breast Cancer: The case of Tikur Anbessa Specialized Hospital, Ethiopia

Tashome Fenta Biru (✉ [tashomefenta@gmail.com](mailto:tashomefenta@gmail.com))

Assosa University <https://orcid.org/0000-0002-2296-0569>

Geremew Muleta Akessa

Jimma University

Yasin Negash Jabir

Jimma University

Selamawit Endale Gurmu

Assosa University

---

## Research Article

**Keywords:** Breast cancer, Counting Approach, Recurrent events, Shared frailty model

**Posted Date:** April 12th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-993296/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Survival Analysis of Recurrent Events on Women Breast Cancer: The case of Tikur Anbessa Specialized Hospital, Ethiopia

**Tashome Fenta Biru<sup>1,\*</sup>, Geremew Muleta Akessa<sup>2</sup>, Yasin Negash Jabir<sup>2</sup>, Selamawit Endale Gurmu<sup>3</sup>**

<sup>1</sup>Department of Statistics, College of Natural Science, Assosa University, Assosa, Ethiopia

<sup>2</sup>Department of Statistics, College of Natural Science, Jimma University, Jimma, Ethiopia

<sup>3</sup>Department of Statistics, College of Natural Science, Assosa University, Assosa, Ethiopia

Email address: [tashomefenta@gmail.com](mailto:tashomefenta@gmail.com) (Tashome Fenta Biru), [Yohager@yahoo.com](mailto:Yohager@yahoo.com) (Geremew Muleta Akessa), [yasinnegash51@yahoo.com](mailto:yasinnegash51@yahoo.com) (Yasin Negash Jabir), [selamawit.endale@gmail.com](mailto:selamawit.endale@gmail.com) (Selamawit Endale Gurmu)

\*Corresponding Author

## Abstract

**Background:** Breast cancer is the most commonly diagnosed cancers worldwide. It is a cancer that develops from breast tissue and most common invasive cancer in women. Recurrent events data have been increasingly important in clinical studies where individuals experience an event more than once and it is a major clinical indicator, which represents the principal cause of breast cancer-related deaths. The aim of the study was to investigate determinants of the recurrence of breast cancer.

**Methods:** To reach the aim, 421 women with breast cancer were included in the study based on data taken from medical record card of patients enrolled starting from 1<sup>st</sup> January 2013 to 30<sup>th</sup> January 2019. A retrospective study has been applied to obtain data on women breast cancer that recorded in oncology department of Tikur Anbessa specialized Hospital. Unmeasured shared similarities due to the impact of multiple events were modeled using a random effect. Cox-PH model and Shared frailty model were used to identify which factor was significantly affecting the recurrence of breast cancer.

**Results:** From the total of 997 recurrent events, about 609 (61.1%) of them experienced recurrence of breast cancer. The shared log-normal frailty model was chosen as the best fit for this breast cancer data set based on the value of Likelihood cross validation criterion. From the result of shared log-normal frailty model age, stage of breast cancer, tumor size, histology grade, breast feeding and oral contraceptives were significantly associated with recurrence of women breast cancer.

**Conclusion:** generally shared log-normal frailty model shows that the stage (II, III, IV), tumor size ((3-5) cm, >5 cm), histology grade (poorly differentiated) and oral contraceptive were significantly increases the risk of recurrence of breast cancer. While, breast feeding was significantly decreases the risk of recurrence of breast cancer. It is recommended that policy maker, ministry of health and Tikur Anbessa Specialized Hospital are expected to make interventions based on these hazardous groups for recurrence of breast cancer.

**Key words:** Breast cancer, Counting Approach, Recurrent events, Shared frailty model

## **1. Background of the Study**

Breast cancer is a cancer that develops from breast tissue and most common invasive cancer in women [1]. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among women worldwide, with an estimated 1.7 million cases and 521,900 deaths [2]. It is a major life threatening and has become the major public health problem of great concern and the most common cause of cancer death among women in less developed countries [3]. This cancer accounts for 25% of all cancer cases and 15% of all cancer deaths among women. Both in developed and developing countries breast cancer is a major health problem account for about one- half of all cancer cases and more than 324, 300 deaths occurred respectively [4].

Breast cancer is the most common leading cause of cancer death problems in Africa [5]. In sub-Saharan Africa, breast cancer is responsible for one in four diagnosed cancers and one in five cancer deaths in women. Recent data estimate that in 2012, 94,000 women developed breast cancer and 48,000 died from it in sub-Saharan Africa. It has been estimated that by 2050, the incidence of Breast cancer in Africa will be double of current 2012 estimate [6]. A study done on global burden of cancer showed 2.4 million women were diagnosed With 523,000 related deaths due to breast cancer in 2015 [7]. Approximately 60% of deaths due to breast cancer occur in developing countries [8].

In Ethiopia breast cancer is the most leading cancer occurring among women. It is estimated that around 9,900 Ethiopian women have breast cancer with thousands of more cases unreported as women living in rural areas often seek treatment from traditional healers before seeking help from the government health system. A retrospective study conducted in TASH from 1997-2012 indicates that of total 16,622 new cancer cases registered 3460 were new cases of breast cancer with prevalence of 20.8% and approximately 216 cases per annual [9].

## 2. Methodology

### Study Population and Setting

The population of this study was all women with breast cancer who had been registered at TASH starting from 1<sup>st</sup> January 2013 to 30<sup>th</sup> January 2019. All the data had been carefully reviewed from the registration log book and patients' registration card; if any inadequate information counters it has been checked from the file and excluded from analysis if proven to be inadequate.

Data structure for modeling recurrent events on breast cancer has been checked carefully to identify first, second, third and additional times of time to recurrence on breast cancer (i.e. recurrent event data).

Table below illustrates data structures required for modeling recurrent events of time to recurrence of breast cancer. The time from the start of the follow up to first time to recurrence, second time to recurrence, third time to recurrence and the time to the last censorship in patients who did not have recurrence time of breast cancer was considered to model recurrent event data. On the other hand, data from breast cancer patients with both one and more time to recurrence were accounted when modeling recurrent event data. The time interval for recurrent events of time to recurrence for each patient was given by the difference between two successive recurrence times of breast cancer patient.

Table 1 : Data Structure for recurrence of Breast Cancer

ID	Start	Stop	Time	Event status	Tumor size	Treatment taken
1	0	8	8	1	1	Combination of $\geq 2$
1	10	16	6	1	2	Radiation
1	23	32	9	0	1	Chemotherapy
2	0	9	9	1	3	Combination of $\geq 2$
2	13	19	6	1	3	Surgery
2	21	31	10	0	2	Hormonal
3	0	7	7	1	1	Surgery
3	14	24	10	0	3	Chemotherapy

The response variable for this study is time to recurrence of women breast cancer in months. The following variables were considered for their influence on recurrence of women breast cancer; age, alcohol, smoking status, residence, histology grade, breastfeeding, treatment taken, stage, oral contraceptives, obesity, tumor size, family history, menopause status and marital status.

## **Method of Data Analysis**

### **Descriptive Statistics**

The description of survival data utilizes nonparametric methods to compare the survival functions of two or more groups and Kaplan-Meier plot(s) would be employed for this purpose [10]. The frequency distribution table was also used to summarize the data based on the study variables.

### **Survival Data Analysis**

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs [11].

### **Cox Proportional Hazard Model**

The most frequently applied analysis method for recurrent data is the model by Andersen and Gill (1982) which is based on the common Cox proportional hazards model [12]. The Andersen-Gill model assumes independence between all observed event times irrespective whether these event times correspond to the same/different patients. The assumption of proportional hazards is that the hazard of time-to-event at any given time for an individual in one group is proportional to the hazard at that time for an individual in the other group. Further assumptions of the Cox PH model are: the ratio of the hazard function for two individuals with different sets of covariates does not depend on time, time is measured on a continuous scale, censoring occurs randomly and uninformative censoring.

Modeling recurrent survival data can be carried out using a Cox PH model with the data layout constructed so that each subject has a line of data corresponding to each recurrent event. The model is typically use to carry out the counting Process approach is the standard Cox PH Model. For recurrent survival data, a subject remains in the risk set for more than one time interval until his or her last interval, after which the subject is removed from the risk set [13]. The hazard function for the  $j^{\text{th}}$  individual in the  $i^{\text{th}}$  group can be obtained by:-

$$\lambda_{ij}(t) = \lambda_0(t)e^{\beta'X} \quad (1)$$

Where  $\lambda_0(t)$  is baseline hazard,  $\beta$  is a  $p \times 1$  vector of regression coefficients and  $X_{ij}$  is the value of  $x$  for the  $j^{\text{th}}$  individual in  $i^{\text{th}}$  group  $i=1 \dots G$ , and  $j=1 \dots n_i$

### Checking the Assumption of Proportional Hazards

Schoenfeld residuals are useful to check the proportionality of the covariates over time, that is, to check the validity of the proportional hazards assumption.

### 3.4.2 Frailty Models

Inference for Cox proportional hazards model was developed under the assumption that the observations are statistically independent, at least conditionally upon covariates. However, this assumption may be violated. Thus in many epidemiological studies, failure times are clustered into groups such as families or geographical units: some unmeasured characteristics shared by the members of that cluster, such as genetic information or common environmental exposures could influence time to the studied event [14]. In a different context, correlated data may come from recurrent events, i.e. events which occur several times within the same subject during the period of observation. In frailty models, dependence is produced by sharing an unobserved variable which is treated as a random effect, or frailty [15, 16].

#### Shared Frailty Model

A natural extension of the univariate frailty model will be multivariate frailty models where individuals are allowed to share the same frailty value. The assumption of a shared frailty model is that both individuals in a pair share the same frailty  $Z$ , and this is why the model is called the shared frailty model. Shared frailty models are appropriate when you wish to model the frailties as being specific to groups of subjects, such as subjects within families. Here a shared frailty model may be used to model the degree of correlation within groups [17].

#### Shared Gamma Frailty Model

The functional form of the one parameter gamma distribution is given by:

$$f_z(z) = \frac{Z^{\frac{1}{\theta}-1} \exp(-Z/\theta)}{\theta^{1/\theta} \Gamma(\frac{1}{\theta})} \quad (2)$$

where  $\Gamma(\cdot)$  is gamma function with Laplace transform that is given by:

$$L(s) = (1 + \theta s)^{-1/\theta}, \theta > 0$$

Thus, the expectation and variance of the Frailty variable will be 1 and  $\theta$  respectively. The shared gamma frailty model (conditional hazard) for individual  $j$  in cluster  $i$  is:

$$\lambda_{ij}(t|X_{ij}, Z_i) = z_i \lambda_0(t) e^{x_{ij}' \beta} = Z_i h(t_{ij})$$

Where  $h(t_{ij}) = \lambda_0(t)e^{x_{ij}'\beta}$  in the Cox regression model for individual  $j$  in cluster  $i$ . The  $Z_i$  are independent identically distributed following a gamma distribution, like in the univariate frailty models.

### Shared Log-Normal Frailty Model

The log-normal frailty model stems mainly from the link with mixed models, where the standard assumption is that the random effects follow a normal distribution. Let  $W \sim N(0, \sigma^2)$  be a normally distributed random effect and let the frailty be given by  $Z = e^W$ . The corresponding frailty has a lognormal distribution. Its function has the form:-

$$f_Z(z) = \frac{1}{z\sqrt{2\pi\sigma^2}} e^{\left(\frac{-(\log z)^2}{2\sigma^2}\right)}, \sigma > 0, \quad Z > 0 \quad (3)$$

The shared log-normal frailty model has the form:-

$$\lambda_{ij}(t) = \lambda_0(t)X_{ij}'\beta + W_i$$

Where  $\lambda_{ij}(t)$  the hazard function is for the  $j^{\text{th}}$  individual from the  $i^{\text{th}}$  group  $\lambda_0(t)$  is the baseline hazard at time  $t$ ,  $X_{ij}$  is the vector of  $p$  covariates recorded for the individual and  $W_i$  is the random effect for the  $i^{\text{th}}$  group. In this model  $\lambda_0(t)$  can be left arbitrary. The  $w_i$ 's,  $i = 1 \dots G$  are a sample (independent and identically distributed) from a density  $f_w(\cdot)$ .

### Penalized Likelihood Approach

Semi-parametric hazard models without frailty terms are fitted by maximization of the partial likelihood [14]. For semi-parametric Frailty models, however, we need to account for the contribution of the unobserved frailty terms. An appropriate estimation method could be used to fit semi parametric frailty models that are the expectation-maximization (EM) algorithm and the penalized likelihood approach [18, 19]

### Choice of Smoothing Parameter

Sometimes it is sufficient to choose the smoothing parameter heuristically, by plotting several curves and choosing the one that seems most realistic. An empirical estimate of the smoothing parameter can be provided or the smoothing parameter can be chosen by maximizing cross-validation [20].

$$\overline{CV}(k) = \frac{1}{n} l_j(\hat{\eta}(k)) - \frac{1}{n} ([\hat{I}(\hat{\eta}) + 2k\Omega]^{-1} \hat{I}(\hat{\eta}))$$

Where:  $l_j$  is the log-likelihood contribution of individual  $j$ .

In this study, the goodness of fit of the Cox and Frailty models is provided by an approximate likelihood cross-validation criterion (LCV) [21]. Likelihood cross-validation criterion (LCV) is approximately equivalent to Akaike's criterion. Lower values of LCV indicate a better fitting model.

### **Computational Procedure (Algorithm)**

The estimated parameters for the models we employed were obtained by the robust Marquardt algorithm which is a combination of the Newton-Raphson algorithm and steepest descent algorithms. It is more stable than the Newton-Raphson algorithm but preserves its fast convergence property near the maximum. The iteration stops when the difference between two consecutive log-likelihoods is small, the coefficients are stable and the gradient is small enough [22].

### **Assessing Model Adequacy**

Regardless of which type of model is fitted and how the variables are selected to be in the model, it is important to evaluate how well the model fits the data. A survival model is adequate if it represents the survival patterns in the data to an acceptable degree. This aspect of a model is known as goodness of fit. Residuals are a useful method for checking the fit of a statistical model. Residuals are central to the evaluation of model adequacy in any setting. Cox-Snell residual is the most widely used residual in the analysis of survival data [23]. The Cox-Snell residual for the  $i^{\text{th}}$  individual is given by:

$$r_{ci} = \exp(\beta' X_{ij}) \widehat{H}_0(t_i) = \widehat{H}_i(t_i) = -\log(\widehat{S}_i(t_i))$$

Where  $\widehat{H}_0(t_i)$  is an estimate of the baseline cumulative hazard function at time  $t_i$ , the observed survival time of that individual,  $\widehat{H}_i(t_i)$  and  $\widehat{S}_i(t_i)$  are the estimated values of the cumulative hazard and survivor functions of the  $i^{\text{th}}$  individual at  $t_i$ . The hazard function follows approximately 45 degree line at which plot depicts for  $-\log(\widehat{S}_i(t_i))$  vs time a straight line and/or for plot  $\log[-\log(\widehat{S}_i(t_i))]$  vs.  $\log(\text{time})$  a straight line through the origin with slope =1, we can say the model fit the data better.

### **3. Analysis and Results**

#### **Descriptive Analyses**

The data for this study has been taken from 421 women that have received treatments for breast cancer more than one time, at oncology department of TASH, Addis Ababa, Ethiopia, between 1<sup>st</sup> January 2013 and 30<sup>th</sup> January 2019 were considered. The outcome was time to recurrent event of breast cancer. From the total of 997 events, about 609 (61.1 %) of the experienced recurrence of Breast cancer and the remaining 388 (38.9 %) were censored (see Table 2 of the appendix).

#### **Cox Proportional Hazard Model**

First fitted Cox proportional hazard model for each risk factor before proceeding to more complicated models. Variable with p-value less than or equal to 25% in the univariable analysis were considered for multivariable model [24]. Then, the full multivariable Cox proportional hazard model was fitted including all the potential covariates that were significant at 25% level. For multivariable analysis, variables with P-value less than or equal to 5% were selected as significant covariates. The result from the standard Cox PH model is presented in the appendix (Table 3) show that treatment taken, age, stage, tumor size, histological grade, breast feeding and oral contraceptive were significantly associated with time to recurrence of breast cancer patients. The Standard Cox PH model considers different line of data contributed by the same subjects as independent contributions from different subjects.

The PH assumption of all variables included in the model was checked using the Schoenfeld residuals. The results show that the covariates are not statistically significant implying that the covariates are time independent because all the p-values are greater than 5%. The overall proportionality test is also not statistically significant implying that the proportionality assumption was not violated (see Table 4 of the appendix).

#### **Shared Gamma Frailty Model**

In recurrent event data, subjects may have more than one events of interest. Thus, patients with the same id are considered as correlated. An extension of the Cox model can be considered by taking into account the clustered structure of the data. Thus clustering can be considering as a random effect. Here the main interest is rather in the heterogeneity between subjects. In the shared gamma frailty model, first uni-variable analysis was conducted and significant variables at 25% level of significance were taken to the multiple shared gamma frailty model. Treatment taken, age, stage, tumor size, histology grade, breast feeding and oral contraceptive were significant covariates selected from the saturated multiple shared gamma frailty model.

**Table 5:-** Parameter Estimates for Shared Gamma Frailty Model uses penalized likelihood

Variable names	Category	Coef	exp(coef)	SE coef (H)	Z	p-value	95% CI
Treatment taken	Chem	0.3678	1.4445	0.1978	1.8597963	0.0629	[0.98 , 2.13]
	Rad	0.3250	1.3840	0.2365	1.3741657	0.1694	[0.87 , 2.20]
	Horm	0.1762	1.1927	0.2467	0.7141271	0.4752	[0.74 , 1.93]
	Comb	0.4343	1.5438	0.1961	2.2143393	0.0268	[1.05 , 2.27]
Age	36-49	-0.0061	0.9939	0.0980	-0.0619469	0.9506	[0.82 , 1.20]
	>=50	-0.3192	0.7267	0.1178	-2.7091665	0.0067	[0.58 , 0.92]
Stage	II	0.3104	1.3640	0.1126	2.7552245	0.0059	[1.09 , 1.70]
	III	0.3084	1.3612	0.1079	2.8587071	0.0043	[1.10 , 1.68]
	IV	0.66744	1.9492	0.1592	4.1913276	≤\$0.0001	[1.43 , 2.66]
Tumor size	3cm-5cm	0.5216	1.6847	0.1195	4.3630154	≤\$0.0001	[1.33 , 2.13]
	>5cm	0.6798	1.9734	0.1216	5.5887134	≤\$0.0001	[1.55 , 2.50]
Hgrade	II	0.1387	1.1488	0.1135	1.2219717	0.2217	[0.92 , 1.44]
	III	0.2565	1.2923	0.1118	2.2936376	0.0218	[1.04 , 1.61]
Breast feeding	Yes	-0.3672	0.6927	0.0860	-4.2717911	≤\$0.0001	[0.59 , 0.82]
o.cont	Used	0.2236	1.2506	0.0853	2.6216171	0.0088	[1.06 , 1.48]
Frailty parameter, Theta: 6.49157e-14				SE (H): 2.53472e-08		p = 0.5	
<p><i>penalized marginal log-likelihood = -2207.38</i>  <i>Convergence criteria: parameters = 1.08e-05 likelihood = 7.1e-05 gradient = 5.63e-09</i>  <i>LCV = the approximate likelihood cross-validation criterion in the semi parametrical case = 2.23908</i></p>							
<p><i>Coef: estimated coefficient for each covariates; exp (coef): exponentiation value of coefficient, se coef (H) and se coef (HIH): estimator of standard error direct and the Hessian ("sandwich estimator") respectively; the First Category of covariates' is Reference.</i></p>							

Parameter estimates of the shared gamma frailty model are presented in table 5 above interpreted as following after controlling for others prognostic factors and accounting for frailty. The estimated hazard rate for age implied those patients who are grouped in age interval 50 or above 50 is  $\widehat{HR} = \exp(\hat{\beta}) = \exp(-0.31921178) = 0.727$  with [95% CI: 0.58, 0.92] and the p-value are small (p-value = 0.007) which implies that age has been increase the recurrence of the diseases. Thus, the result revealed that recurrence of breast cancer for age interval 50 or above 50 of breast cancer patients at hazard rate about 0.727 times less than patients who were grouped to age interval  $\leq 35$ .

Looking for stages of women with breast cancer, the estimated hazard rate of breast cancer with stage II, III and IV were 1.364 with [95% CI: 1.09, 1.70], 1.361 with [95% CI: 1.10, 1.68] and 1.950 with [95% CI: 1.43, 2.66] respectively and p-value (0.006, 0.004, <0.0001) this indicates that stage II, III and IV of breast cancer has a significant effect on the increasing recurrence of women with breast cancer and the risk of recurrence breast cancer increased by 1.364, 1.361 and 1.950 for stage II, III and IV of women with breast cancer respectively as compared to stage I.

Regarding tumor size of women breast cancer, the estimated hazard rate of breast cancer patients for tumor size 3 to 5 and above 5 centimeters are estimated to be 1.685 with [95% CI: 1.33, 2.13] and 1.973 with [95% CI: 1.55, 2.50] respectively and p-value are small (<0.0001) which implies that both tumor size of women with breast cancer has a significant effect on the increasing recurrence of breast cancer and the expected hazard rate of breast cancer patients increased by 1.685 for tumor size 3 to 5 centimeters and 1.973 for tumor size above 5 centimeters as compared to women with breast cancer of tumor size 2 or below 2 centimeters.

Moreover, for those patients having histology grade III, the estimated hazard rate is to be 1.292 with [95% CI: 1.04, 1.61] in which the expected hazard rate is 1.292 times more than women having histology grade I and p-value is small (0.022) this implies that women having histology grade III has a significant effect on the increasing recurrence of breast cancer.

By observing breast feeding of women with breast cancer, the estimated hazard rate of breast cancer recurrence for women in breastfeed a child is estimated to be 0.693 with [95% CI: 0.59, 0.82] in which the expected hazard rate is 0.693 times less than women did not breastfeed child and p-value is small (<0.0001) which implies that women with breast feeding were significantly decreases the risk of recurrence of breast cancer.

Finally, observing for women oral contraceptive use, the estimated hazard rate of breast cancer patients using oral contraceptive is 1.251 with [95% CI: 1.06, 1.48] which indicate the expected hazard rate is 1.251 times more than women did not use oral contraceptive and the p-value is small (0.0088) this indicated that a women using oral contraceptive were significantly increases the risk of recurrence of breast cancer.

### Shared Log-Normal Frailty Model

Similarly, we conducted uni-variable analysis for the shared log-normal frailty model. The result of uni-variable analysis indicate treatment taken, age, smoking habit, stage, tumor size, histology grade, breast feeding and oral contraceptive were statistically significant at 25% level of significance. The result from the shared log-normal frailty model is presented on table 6 below. It is observed that age, stage, tumor size, histology grade, breast feeding and oral contraceptive were the only significant covariates selected from the saturated multiple shared log-normal frailty models.

**Table 6:-** Parameter Estimates of Shared Lognormal Frailty Model using penalized likelihood.

Variable names	Categ ory	Coef	exp(co ef)	SE coef (H)	Z	p-value	95% CI
Age	36-49	-0.0124	0.9877	0.1052	-0.117856	0.9062	0.80 - 1.21
	$\geq$ \$50	-0.3346	0.7156	0.1245	-2.687229	0.0072	0.56 - 0.91
Stage	II	0.3428	1.4090	0.1187	2.887342	0.0039	1.12 - 1.78
	III	0.3137	1.3685	0.1126	2.786767	0.0053	1.10 - 1.71
	IV	0.7286	2.0722	0.1696	4.295193	$<$ \$0.0001	1.49 - 2.89
Tumor size	3cm-5cm	0.4598	1.5838	0.1234	3.726685	0.00019	1.24 - 2.02
	$>$ \$5 cm	0.6640	1.9425	0.1253	5.298460	$<$ \$0.0001	1.52 - 2.48

Hgrade	II	0.1160	1.1231	0.1183	0.98092 6	3.266 3e-01	0.89 - 1.42
	III	0.2492	1.2831	0.1162	2.14524 5	0.031 9	1.02 - 1.61
Breast feeding	Yes	-0.3514	0.7037	0.0894	- 3.93204 7	\$.<\$0. 0001	0.59 - 0.84
o.cont	Used	0.2233	1.2503	0.0900	2.48287 9	0.013 0	1.05 - 1.49
Frailty parameter, Sigma Square: 0.075618				SE (H): 0.00864336		p = < 1e-16	
<p><i>penalized marginal log-likelihood = -2162.87</i>  <i>Convergence criteria: parameters = 5.95e-07 likelihood = 0.000894 gradient = 6.81e-09</i>  <i>LCV = the approximate likelihood cross-validation criterion in the semi parametrical case = 2.19644</i></p>							
<p><i>Coef: estimated coefficient for each covariates; exp (coef): exponentiation value of coefficient, se coef (H) and se coef (HIH): estimator of standard error direct and the Hessian ("sandwich estimator") respectively; the First Category of covariates' is Reference.</i></p>							

Parameter estimates of the shared Log-normal frailty model are presented in table 6 above are interpreted as following after controlling for others prognostic factors and accounting for frailty. The estimated hazard rate for age implied those patients who are grouped in age interval 50 or above 50 years old is  $\widehat{HR} = \exp(\hat{\beta}) = \exp(-0.3345833) = 0.716$  with [95% CI: 0.56 – 0.91] and p-value are small (p-value= 0.0072) which implies that younger women, particularly those under the age of 35, have a higher risk of recurrence of breast cancer and the result revealed that age interval 50 or above 50 of breast cancer patients has rate of recurrence 0.716 times less than patients who were grouped to age interval  $\leq 35$ .

Looking for stages of women with breast cancer, the estimated hazard rate of breast cancer for stage II, III and IV were 1.409 with [95% CI: 1.12 - 1.78], 1.368 with [95% CI: 1.10 - 1.71] and 2.0722 with [95% CI: 1.49 - 2.89] and the p-value are small (0.0039, 0.0053, <0.0001) respectively and this indicates that women with advanced stages have a greater risk for recurrence of breast cancer. Thus, the expected hazard rate of women with recurrence of breast cancer increased by 1.409, 1.368 and 2.0722 for stage II, III and IV respectively as compared to stage I.

Regarding tumor size of women breast cancer, the estimated hazard rate of recurrence for tumor size 3 to 5 and above 5 centimeters are estimated to be 1.584 with [95% CI: 1.24 - 2.02] and 1.942 with [95% CI: 1.52 - 2.48] respectively and p-value are small ( $<0.0001$ ,  $<0.0001$ ) which implies that both tumor size of women with breast cancer has a significant effect on the increasing recurrence of women breast cancer and women with larger tumors have a greater risk of recurrent breast cancer and also the expected hazard rate of women with breast cancer increased by 1.584 for tumor size 3 to 5 centimeters and 1.942 for tumor size above 5 centimeters as compared to tumor size 2 or below 2 centimeters.

Moreover, for those patients having histology grade three, the estimated hazard rate is estimated to be 1.283 with [95% CI: 1.02 - 1.61] in which the expected hazard rate is 1.283 times more than women having histology grade one and the p-value is small (0.032) which implies that women having histology grade three has a significant effect on the increasing recurrence of breast cancer.

By observing breast feeding of women with breast cancer, the estimated hazard rate of breast cancer recurrence for women in breastfeed a child is estimated to be 0.7037 with [95% CI: 0.59 - 0.84] which is the expected hazard rate is 0.7037 times less than women did not breastfeed child and the p-value is small ( $<0.0001$ ) which implies that women breast feeding a child is significantly decreases the risk of recurrence of breast cancer.

Finally, observing for women oral contraceptive use, the estimated hazard rate of breast cancer patients using oral contraceptive is estimated to be 1.250 with [95% CI: 1.05 - 1.49] which implies the expected hazard rate is 1.250 times more than women did not use oral contraceptive and the p-value is small (0.013) this implies that women use oral contraceptive is significantly increases the risk of recurrence of breast cancer.

Test hypothesis for the variance term of shared log-normal Frailty term is given by:

$$H_0 : \sigma^2 = 0 \text{ vs } H_1 : \sigma^2 > 0$$

The variance of frailty term (**Sigma Square**):  $\sigma^2 = 0.0756$ (SE (H):0.0086) is significantly different from zero, meaning that there is heterogeneity between the subjects explained by non-observed covariates. We can deduce this by using a modified Wald test:  $W_m(\sigma^2) = 0.075618 / 0.00864336 = 8.75$ , with the critical value for a normal one-sided test. The modified Wald test ( $W_m$ ) is a significance test for the variance of the random effects distribution occurring on the boundary of the parameter space. The usual squared Wald statistic is simplified to a mixture of two distributions and hence the critical values must be derived from this mixture (Molenberghs and Verbeke (2007)). In the case our result have a p-value that is less than 5% for shared log-normal Frailty but not for shared gamma Frailty distribution. This mean that there is a significant Frailty

effect, that within subject correlation cannot be ignored for shared log-normal frailty but not for shared gamma frailty.

### Comparison of the Cox-PH and Shared Frailty Models

Efficiency of the fitted models was compared using Penalized marginal log-likelihood and LCV (likelihood cross-validation) criterion. The likelihood cross-validation criterion assesses the goodness of fit of a statistical model [21]. In case of parametric approach, LCV is approximately equivalent to Akaike Information criterion (AIC). Lower values of LCV indicate a better fitting model. Table 7 depicts the LCV results of Cox PH, shared gamma Frailty and shared log-normal Frailty models. The shared log-normal Frailty model was chosen as the best fit for our recurrent events data on breast cancer based on the residual analysis and minimum value of LCV. Although the difference in LCV value of the fitted models was negligible, the Cox-Snell residual plot suggested that the shared log-normal Frailty model fits the data better. The Wald test results indicated that the heterogeneity parameter was significant, implying that there is a significant frailty effect, or that within subject correlation cannot be ignored.

**Table 7:-** Comparison of Cox PH, Shared Gamma and Log-normal Frailty Models.

Model	Frailty Parameter	SE	PM logL	LCV
Cox-PH	-	-	-2217.48	2.24721
Shared Gamma Frailty	6.49157e-14	2.53472e-08	-2207.38	2.23908
Shared Log Normal Frailty	0.075618	0.00864336	-2162.87	2.19644

## 4. Discussion

The main aim of this study is to identify factors affecting the recurrence of women breast cancer, which was obtained from TASH. The most frequently applied analysis method for recurrent time-to-event data is the model by Andersen and Gill was used to analyze breast cancer data set. The Andersen-Gill model assumes independence between all observed event times [12]. In addition, frailty models used to account for the dependence among the recurrent event times based on Andersen- Gill (A-G) survival model [25, 26].

From the total of 997 case of recurrent observation, 609 (61.1%) them experienced the event. From the total of 421 patients 42(10%),153(36.3%), 40(9.5%), 35(8.3%), 151(35.9%) breast cancer patients treated in the hospital with surgery, chemotherapy, radiotherapy, hormone-therapy and combination of two or more treatments respectively and out of this 6(1.4%), 77(18.3%), 14(3.3%), 13(3.1%), 56(13.3%) of them experienced the events respectively. The minimum and maximum times of recurrence of women breast cancer were 5 and 32 months respectively. The median recurrence time of the breast cancer patients was 10 months in the recurrent events of women breast cancer.

Efficiency of the fitted models was compared using penalized marginal log-likelihood and LCV (likelihood cross-validation) criterion and shared log-normal frailty model found to be the best fit from the Cox PH and shared gamma frailty model [21].

Under uni-variable analysis the shared log-normal Frailty model shows that treatments taken, age, smoking, tumor size, stages of breast cancer, obesity, histology grade, alcohol use, family history of breast cancer, breast feeding and oral contraceptives were significantly associated with recurrence of women breast cancer at 25% level of significance [24].

From result of multivariable analysis of shared log-normal Frailty model the recurrence of women breast cancer were significantly affected by age, tumor size, stages of breast cancer, histology grade, breast feeding and oral contraceptives.

From the result of this study the younger women experienced the greatest hazard rate for recurrence of breast cancer. The study by Dignam in 2009 shows the same results. In addition, the histology grade at diagnosis was significantly affected the recurrence of breast cancer and the hazard rate was high for women in histology grade III as compared to women in histology grade I. This was also indicated by study done [27, 28].

Similarly the hazard rate of women with breast cancer for tumor size 3 to 5 centimeters and above 5 centimeters were increase as compared to women with breast cancer of tumor size 2 or below 2 centimeters this implies that tumor size has a significant effect on increasing the risk of breast cancer recurrences [28, 29]

The stages of breast cancer have significant effect on the recurrence of women breast cancer. The study done by Demicheli in 2010 and Dignam in 2009 also shows that the stages of breast cancer at diagnosis have been significantly affect the recurrence of women breast cancer. From the results of these two studies the hazard rate of recurrence of breast cancer was greatest as the stage increases [27, 30].

Oral contraceptives were associated with increasing recurrence of women breast cancer. This implies that oral contraceptives facade a higher risk of breast cancer recurrence. The current study was consistent with the study [31, 32]. Furthermore, Women treated for breast cancer that previously breast fed their babies have lower risk of recurrence than those who did not. These findings are consistent with a study conducted by [33].

The model adequacy of shared log-normal frailty model checked by plotting Cox-Snell residuals against cumulative hazard function of residuals and its result shows that the log-normal frailty model fit women breast cancer data well, since the plot makes approximately straight lines through the origin for women breast cancer data set.

### **Conclusion**

The data consisted of 997 observations from a total of 421 patients. About (61.1%) of them experienced the event of interest while the remaining (38.9%) of observations did not experience the event of interest throughout the study period.

The main goal of this study was to investigate determinants of recurrence of women breast cancer and to fit best model for recurrence of breast cancer. The standard Cox PH shared gamma frailty and shared log-normal Frailty models were fitted. The value of the LCV was used to identify the best model. Accordingly, the shared log-normal frailty models provide suitable choice for the recurrent event time of the breast cancer as compared to standard Cox proportional hazard and shared gamma frailty models.

From the result of shared log-normal frailty model age, stage, tumor size, histology grade, breast feeding and oral contraceptive were found to be statistically significant factors for recurrence of women breast cancer. Of all this significant covariates stage (II, III, IV), tumor size ((3-5) cm, >5 cm), histology grade (poorly differentiated) and oral contraceptive (use) were significantly increases the risk of recurrence of women breast cancer. While, breast feeding was significantly decreases the risk of recurrence of women breast cancer.

## **Declaration**

During conducting the study, the investigators have included the following declarations.

### **List of abbreviations**

Cox-PH: Cox Proportional Hazard model, EM: Expectation Maximization, FMOH: Federal Ministry of Healthy, LCV: Likelihood cross-validation criterion, PH: Proportional Hazard, TASH: Tikur Anbessa Specialized Hospital, WHO: World Health Organization.

### **Availability of Data and Materials**

The data set was taken from Tikur Anbessa Specialized Hospital, Ethiopia, from 1<sup>st</sup> January 2013 and 30<sup>th</sup> January 2019. There are no restrictions on the availability of data and R-code, and the investigators are willing to provide the code as well.

### **Competing interests**

The authors declare that they have no competing interests

**Funding:** The study was fully funded by Jimma University and the Ministry of science and higher education of Ethiopia. The collaborative fund obtained from the two institutions covered all costs associated with data access and related activities.

### **Authors' contribution**

This study was designed and compiled by Tashome Fenta (MSc in Biostatistics) as the principal investigator. The development of the basic research questions, identifying the problems and selecting appropriate statistical models, data collection, data analysis, interpretation, and critical review of the paper have been done by him. The edition of the overall progress of the work was supported by Geremew Muleta, Yasin Negash, Selamawit Endale.

### **Authors' Information**

<sup>1</sup>Department of Statistics, College of Natural Science, Assosa University, Assosa, Ethiopia

<sup>2</sup>Department of Statistics, College of Natural Science, Jimma University, Jimma, Ethiopia

<sup>3</sup>Department of Statistics, College of Natural Science, Assosa University, Assosa, Ethiopia

### **Consent for publication**

Not applicable

### **Acknowledgments**

We want to express a sincere acknowledgement to Jimma University and the Ethiopian ministry of science and higher education for financing this study and guidance that contributed to the successful realization of the study during data collection and analysis.

### **Ethical Consideration**

The research ethics review board of Jimma University would provide an ethical clearance for the study. The data has been collected after written permeation is obtained from oncology department of Tikur Anbessa Specialized Hospital and department of statistics write an official cooperation letter to the Hospital for the permeation. The data has been carefully reviewed from the registration log book and patients registration card. Confidentiality of any information related to the patients and their clinical history has been maintained by keeping both the hard-copy and soft-copy of every collected data in a locked cabinet and password secured computer. Only the researcher would access to the de-identified data that has been kept in a secure place. All data has been coded with numbers and hospital numbers and without personal identifiers. All analysis has been on de-identified and coded data. During the study, there is no contact between the patients and the researcher. The study is noninvasive and without any harm to the patients. Then, the data obtained from the hospital has been secured.

## Appendix

**Table 2:-** Descriptive Statistics summary

No.	Variable Names	Category	Recurrent event of breast cancer patients status		
			Censoring	Recurrence	Total
1.	treatment taken	Surgery	51 (63.0%)	30 (37.0%)	81 (8.1%)
		Chemo	126 (36.1%)	223 (63.9%)	349 (35.0%)
		Rad	40(45.5%)	48(54.5%)	88 (8.8%)
		Horm	40(51.3%)	38(48.7%)	78 (7.8%)
		Comb	131(32.7%)	270 (67.3%)	401 (40.1%)
		Total	388(38.9%)	609(61.1%)	997(100.0%)
2.	Age	<=35	85 (33.9%)	166 (66.1%)	251 (25.2%)
		36-49	131(30.2%)	303 (69.8%)	434 (43.5%)
		>=50	172 (55.1%)	140 (44.9%)	312 (31.3%)
		Total	388(38.9%)	609(61.1%)	997(100.0%)
3.	Smoking	No	317 (38.1%)	516 (61.9%)	833 (83.6%)
		Yes	71(43.3%)	93(56.7%)	164(16.4%)
		Total	388(38.9%)	609(61.1%)	997(100.0%)
4.	Number of children	<= 2	144(38.0%)	235(62.0%)	379(38.0%)
		3-5	162(39.9%)	244(60.1%)	406(40.7%)
		>5	82(38.7%)	130(61.3%)	212(21.3%)
5.	Region	Addis Ababa	195(39.2%)	302(60.8%)	497 (49.8%)
		Snnp	32(40.0%)	48(60.0%)	80(8.0%)
		Amhara	49(38.6%)	78(61.4%)	127(12.7%)
		Oromia	87(37.7%)	144(62.3%)	231(23.2%)
		Others	25(40.3%)	37(59.7%)	62(6.2%)
6.	Residence	Urban	256 (38.8%)	403(61.2%)	659(66.1%)

		Rural	132(39.1%)	206(60.9%)	338(33.9%)
7.	Tumor size	<=2cm	195 (65.0%)	105 (35.0%)	300 (30.1%)
		3cm-5cm	137 (36.5%)	238 (63.5%)	375 (37.6%)
		>5cm	56 (17.4%)	266 (82.6%)	322 (32.3%)
8.	Stage	I	56(83.6%)	11 (16.4%)	67(6.7%)
		II	86(31.9%)	184(68.1%)	270(27.1%)
		III	127 (36.5%)	221 (63.5%)	348 (34.9%)
		IV	148 (47.4%)	164 (52.6%)	312 (31.3%)
9.	Obesity	Normal	102 (41.0%)	147 (59.0%)	249 (25.0%)
		Underweight	119 (37.1%)	202 (62.9%)	321 (32.2%)
		Overweight	167 (39.1%)	260 (60.9%)	427 (42.8%)
10.	Histological grade	Well differentiated	163 (56.4%)	126(43.6%)	289(29.0%)
		modretly differentiated	130 (36.3%)	228(63.7%)	358(35.9%)
		poorly differentiated	95 (27.1%)	255(72.9%)	350(35.1%)
11.	Alcohol	No	182 (38.3%)	293 (61.7%)	475 (47.6%)
		Yes	206 (39.5%)	316 (60.5%)	522 (52.4%)
12.	Family history	No	195(38.5%)	311(61.5%)	506(50.8%)
		With	193(39.3%)	298(60.7%)	491(49.2%)
13.	Breast feeding	No	124 (25.6%)	360 (74.4%)	484 (48.5%)
		Yes	264 (51.5%)	249 (48.5%)	513 (51.5%)
14.	Menopausal status	pre-menopausal	187(38.6%)	297(61.4%)	484(48.5%)
		post-menopausal	201(39.2%)	312(60.8%)	513(51.5%)
15.	Oral contraceptive	Not used	250 (48.3%)	268 (51.7%)	518 (52.0%)
		Used	138 (28.8%)	341 (71.2%)	479 (48.0%)

16.	Marital status	Single	37(36.3%)	65(63.7%)	102(10.2%)
		Married	294 (39.6%)	448(60.4%)	742(74.4%)
		Widowed	37(37.0%)	63(63.0%)	100(10.0%)
		Divorced	20(37.7%)	33(62.3%)	53(5.3%)
Total			388(38.9%)	609(61.1%)	997

Variable names	category	Coef	exp(coef)	SE coef (H)	p-value	95% CI
Treatment taken	Chem	0.3334	1.3957	0.1969	0.0900	[0.95 , 2.05]
	Rad	0.2959	1.3443	0.2357	0.2091	[0.85 , 2.13]
	Horm	0.1567	1.1696	0.2460	0.5241	[0.72 , 1.89]
	Comb	0.3690	1.4462	0.1954	0.0590	[0.99 , 2.12]
Age	36-49	-0.0022	0.9978	0.0979	0.98225	[0.82 , 1.21]
	>=50	-0.3117	0.7322	0.1175	0.0080	[0.58 , 0.92]
Stage	II	0.2976	1.3466	0.1123	0.0081	[1.08 , 1.68]
	III	0.2865	1.3318	0.1077	0.0078	[1.08 , 1.64]
	IV	0.6330	1.8833	0.1586	≤\$0.0001	[1.38 , 2.57]
Tumor size	3cm-5cm	0.4786	1.6138	0.1197	≤\$0.0001	[1.28 , 2.04]
	>5cm	0.6448	1.9055	0.1215	≤\$0.0001	[1.50 , 2.42]
Hgrade	II	0.1455	1.1566	0.1134	0.1993	[0.93 , 1.44]
	III	0.2364	1.2667	0.1117	0.0342	[1.02 , 1.58]
Breast feeding	Yes	-0.3301	0.7188	0.0859	0.00012	[0.61 , 0.85]
Oral.Cont	Used	0.2083	1.2316	0.0850	0.0142	[1.04 , 1.45]

**Table 3:-** Parameter Estimates of Cox PH Model using penalized likelihood.

**Table 4** Cox PH assumption checking test statistics based on Schoenfeld residuals.

Variable	Category	Rho	Chisq	p-value
Treatment taken	Chem	0.05218	-1.58611	0.2079
	Rad	0.03978	0.90082	0.3426
	Horm	0.01692	0.17448	0.6762
	Comb	0.05864	2.03430	0.1538
Age	36-49	-0.00549	0.00973	0.9214
	>=50	0.03850	0.35782	0.5497
Smoking	Yes	-0.09713	2.00309	0.1570
Stage	II	-0.02969	0.15248	0.6962
	III	-0.05021	0.46661	0.4946
	IV	-0.07048	1.05617	0.3041
Tumor size	2cm-5cm	0.10263	3.42868	0.0641
	>5cm	0.10073	3.79636	0.0514
Obesity	Underweight	-0.05153	0.59943	0.4388
	Overweight	-0.01558	0.05379	0.8166
Histologic grade	Moderately differentiated	-0.04014	0.53773	0.4634
	Poorly differentiated	-0.07704	1.91115	0.1668
Alcohol	Yes	-0.00560	0.00883	0.9251
Breast feeding	Yes	-0.05511	0.66005	0.4165
Oral contraceptive	Used	-0.07222	1.44803	0.2288
GLOBAL		NA	18.66570	0.4785

## References

1. Adesina A, Chumba D, Nelson AM, Orem J, Roberts D, Wabinga H, *et al.*, (2013) 'Improvement of pathology in sub-Saharan Africa'. *The Lancet Oncology*. 14(4).
2. International Agency for Research on cancer (2013) Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed. WHO: Press release N 223.
3. World Health Organization (2015) Cancer a Growing Public Health Concern for Ethiopia. Available at [www.who.int](http://www.who.int)
4. Federal Ministry of Health (2015) National cancer control plan, 2016-2020 FMOH; 2015
5. Parkin DM, Bray F, Ferlay J, Jemal A (2012) 'Cancer in Africa'. *Cancer Epidemiology and Prevention Biomarkers* 2014, 23(6):953-966.
6. Brinton L, Figueroa J, Awuah B, Yarney J, Wiafe S, Wood S, *et al.*, (2014) 'Breast cancer in Sub-Saharan Africa: opportunities for prevention'. *Breast cancer research and treatment*.144 (3):467-78.
7. Global Burden of Disease Cancer Collaboration (2017) Fitzmaurice C, Allen C, Barber R, Barregard L, Bhutta Z, *et al.*, Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: 'A systematic analysis for the Global Burden of Disease Study'. *JAMA Oncol*. 2017; 3:524-48.
8. Da Costa Vieira, R., Biller, G., Uemura, G., Ruiz, C., & Curado, M. (2017) Breast cancer screening in developing countries. *Clinics (Sao Paulo, Brazil)*, 72(4), 244-253. doi:10.6061/clinics/2017 (04)09.
9. Abate S, Yilma Z, Assefa M, Tigeneh W (2016) Trends of Breast Cancer in Ethiopia. *Int J Cancer Res Mol Mech* 2(1): <http://dx.doi.org/10.16966/2381-3318.121>
10. Kaplan EL. & Meier P. (1958) Non-parametric estimation from incomplete observations. *J am Stat Assoc* 1958(53):457-81
11. Aalen (1980) Non-parametric estimation of partial transition probabilities in multiple decrements models. *Ann Stat*.
12. Andersen P, Gill R (1982) Cox regression model for counting processes: A large sample study. *Ann Stat*; 10(4):1100-20.

13. Kleinbaum and Klein, J.P. (2005). *Survival Analysis: A Self-Learning Text*. Second Edition, Springer.
14. Cox DR (1972) Regression models and life-tables. *J R Stat Soc Series B Methodological*. 34(2):187-220.
15. Clayton, D. (1978) Model for Association in Bi-variate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence.
16. Petersen, P. K. Andersen, and R. D. Gill, "Variance components models for survival data," *Statistical Neerlandica* vol. 50 pp. 193-211, 1996.
17. Therneau TM, Grambsch PM. (2010) *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer-Verlag.
18. Therneau T, Grambsch P. (2000) *Modeling survival data*. New York: Springer
19. Duchateau, L. and P. Janssen (2008). *The Frailty Model*. New York: Springer-Verlag.
20. Joly P., Letenneur D. L., Alioum A., Commenges D., (1999): a computer program for hazard estimation using a penalized likelihood method with interval-censored and left-truncated data, *Computer Method and Program*.
21. Gray R, (1992) Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis. *JASA*, 87 942-951.
22. Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* 11 (2),431-441
23. Cox D. R. and Snell E. J. (1968). A general definition of residuals with discussion. *Journal of the Royal Statistical Society. Series B* 30 (1968), 248-275.
24. Bursac Z, Gauss C, Williams D, Hosmer D.(2008) Purposeful selection of variables in logistic regression. *Source Code Biol Med*. 3:17. Doi10.1186/1751-0473-3-17
25. Pickels, A., Crouchley, R., Simonoff, E.L., Meyer, J., Rutter, M., Hewitt, J. and Silbery J., (1994) Survival models for developmental genetic data: Age at onset of puberty and antisocial behavior in twins. *Genetic Epidemiology*, 11: 155-170.
26. Yashin, A., Vaupel, J., and Jachine, I. (1995) Correlated individual Frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical population studies*, 5:145-149.
27. Dignam J, Dukic V, Anderson S, Mamounas E, Wickerham D, Wolmark N (2009). Hazard of recurrence and adjuvant treatment effects over time in lymph node negative breast cancer. *Breast Cancer Res Treat* 116:595-602. doi:10.1016/j.mcm.2010.02.002.
28. Mauguen A, Rachet B, Mathoulin-Plissier S, *et al.*, (2013) Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Stat Med*.2013; 32(30):5366-80.

29. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P (2007) Joint frailty models for recurring events and death using maximum penalized likelihood estimation: Application on cancer events. *Biostatistics*. 2007; 8(4):708-21.
30. Demicheli, R., Ardoino, I., Boracchi, P. *et al.*, (2010) 'Recurrence and mortality according to Estrogen Receptor status for breast cancer patients undergoing conservative surgery'. Ipsilateral breast tumour recurrence dynamics provides clues for tumor biology within the residual breast. *BMC Cancer* 10, 656 (2010) doi:10.1186/1471-2407-10-656
31. Lu Y, Ma H, Malone KE, Norman SA, Sullivan-Halley J, *et al.*,(2011) Oral contraceptive use and survival in women with invasive breast cancer. 'Cancer Epidemiology Biomarkers' *Prev* 20: 1391-1397
32. Saxe G, Rock C, Wicha M Schottenfeld D (1999) Diet and risk for breast cancer recurrence and survival. *Breast Cancer Res Treat* 53: 241-253.
33. Anderson KN, Schwab RB, Martinez ME (2014) 'Reproductive risk factors and breast cancer subtypes: a review of the literature'. *Breast Cancer Res Treat*. 144(1):1-10. [PMC free article] [PubMed] [Google Scholar].