

Risk Factors Affecting Survival Time to First Recurrence of Breast Cancer Patients: The Case of South West Ethiopia

Reta Habtamu Bacha (✉ habtamuretab@yahoo.com)

Jimma University

Yasin Negash Jabir

Jimma University

Yohannes Yebabe Tesfay

Cleveland, ohio operation

Kenenisa Tadesse Dame

Jimma University

Research

Keywords: Acceleration factor, parametric shared frailty survival models, Breast cancer, Ethiopia

Posted Date: August 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-53957/v1>

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

[Read Full License](#)

Risk Factors Affecting Survival Time to First Recurrence of Breast Cancer Patients: The case of South West Ethiopia

Reta Habtamu Bacha^{1*}, Yasin Negash Jabir², Yohannes Yebabe Tesfay³, Kenenisa Tadesse Dame⁴

***Corresponding Author:** Reta Habtamu Bacha: habtamuretab@yahoo.com

^{1, 2,4}Department of statistics, College of Natural Sciences, Jimma University, Jimma, Ethiopia

³CREE Inc., Professor of Quantitative Methods and Process Engineering,

4600 Silicon Drive Durham, North Carolina, 27703, USA

Abstract

Background: Among the cancers, next to cervical cancer, in Ethiopia, breast cancer is the second common cancer health problem for women. This article aimed to identify the potential risk factors affecting of time-to-first-recurrence of breast cancer patients in southwest Ethiopia: Jimma Medical Center, Bedelle hospital, Mizan-Aman hospital, and Mettu hospital and Karl hospital.

Methods: The data were taken from the patient's medical record that registered from January 2012 to January 2019. The study considered a sample size of 642 breast cancer patients. Different non-parametric and parametric shared frailty survival models employed.

Results: Out of 642 breast cancer patients, about 447 (69.6 %) recovered/cured of the disease. The median cure time from breast cancer found 13 months. The lognormal parametric shared frailty survival model predicted that treatment, stage of breast cancer, smoking habit, and marital status significantly affects the time to the first recurrence of breast cancer.

Conclusion: Treatment, stages of breast cancer, and marital status were improved while smoking habits worsen the time to first cure of breast cancer. To mitigate breast cancer diseases awareness should be given for the community on identified factors

Keywords: Acceleration factor, parametric shared frailty survival models, Breast cancer, Ethiopia

1. Background

Breast cancer is amongst the category of non-communicable diseases. Breast cancer is one of the root causes of death among women followed by cardiovascular and cervical cancer diseases. Both in the developed and the developing world, breast cancer causes the loss of life in women (WHO, 2017).

Worldwide, breast cancer is the utmost frequently diagnosed cancer and the leading cause of cancer death in women. In 2013, the International Agency for Research on Cancer (IARC) report indicated that there were 6.7 million new breast cancer cases and 5.5 million death among females worldwide (Ferlay, 2013). By 2030, the projection showed that the new cases could increase to 8.9 million cases and 5.5 million deaths among females annually as a result of the growth and aging of the population (Ferlay, 2013).

In Africa, breast cancer is held responsible for 28 % of all cancer patients and 20 % of all utmost cancers deaths in women (Brinton et. al., 2014). Correspondingly, it is increasing as the major public health problem in Sub-Saharan Africa (SSA) for the reason that the population mature and development, as well as the factors related to social and economic variables. Nowadays, numerous low- and middle-income countries of African face a twofold burden of cervical and breast cancer, which embodies the top cancer killers in women for the last 30 years old. Generally, in developed countries, breast cancer is the leading cause of death among women (Ferlay, 2013). However, the mortality rates of breast cancer range from 6 % to 20 % in western Africa (WHO, 2012). The incidence rate of breast cancer is 30.4 in eastern Africa, 26.8 in central Africa, 38.6 in western Africa, and 38.9 in southern Africa (Brinton et al., 2014).

Including Ethiopia and Sub-Saharan countries, breast cancer is the furthermost common cause of cancer death problems in Africa. According to a study done in "Tikur Anbessa" specialized hospital (Tigeneh.et al., 2015), in Ethiopia, breast cancer found to be the second predominant cancer following gynecological malignancy. Breast cancer becomes incurable due to late medical presentation, lack of awareness, limited medical resources, and hospitals. Especially, the symptoms of breast cancer have some strong traditional beliefs that delay biomedical care (Timotewos, 2018). Because of confined access to well-timed and popular treatments, breast

cancer survival tends to be poor in most of the Ethiopian regions. Notwithstanding the government's concern on a cancer issue: reduce the recurrence, incidence, mortality, and the survival of women, breast cancer is still emergence in Ethiopia.

It is estimated that about 9,900 Ethiopian women have breast cancer. Thousands of more cases unreported as women living in rural areas regularly seek remedy from traditional healers earlier than seeking help from the government health system. A retrospective study conducted in TASH from 1997-2012 indicates that of a total of 16,622 new cancer cases registered 3460 were new cases of breast cancer with a prevalence of 20.8% and approximately 216 cases per annual (Abate et al.,2016).

Traditionally, studies on the survival analysis of breast cancer patients performed using the cox proportional hazards model. However, to fill the research gaps, in this article we advance our methods on analysis using various parametric shared frailty models. The study will expect to predict the status of patients by identifying determinants of risk factors of time-to-first-recurrence due to breast cancer in southwest Ethiopia.

The term frailty was first suggested by (Vaupel, 1979) in the context of mortality studies, and (Lancaster, 1979) incorporated the frailty concept into a study of duration of unemployment. For a random time-to-data, T , we define the probability density function (pdf) of T as $f(t)$ and the cumulative distribution function as $F(t) = p(T \leq t)$. Two other functions that prove useful in this context are the survivor function. $S(t) = p(T > t) = 1 - F(t)$, and the hazard function $h(t) = \frac{f(t)}{S(t)}$, which can be integrated as the instantaneous rate of data given survival up until time t .

Frailty models have been used frequently for modeling dependence in multivariate time-to-event data (Weinke, 2003). Suggested the variability of lifetimes is fundamental as arising from two different sources. The first one is natural variability, which includes the baseline hazard function, while the second one is explained by the frailty. Lifetimes are conditionally independent given the frailty (as an individual random effect), the frailty term represents unobserved covariates. It is assumed that, given unobserved frailty, the hazard for each survival time follows a proportional hazards model with the frailty variable and the covariate effect acting multiplicatively on the baseline hazard.

2. The Data and the Methodology

2.1. The Data

In this study, the data is obtained from hospital databases in southwest Ethiopia: Jimma Medical Center, Bedelle Hospital, Mizan-Aman Hospital, and Mettu Karl Hospital. The data are convenient summarizing through estimates of the survival and hazard function. A total of 642 breast cancer patients included in this survival analysis from 2012 to 2019.

Variable Definition: Stages of breast cancer

Stage I

Stage I refer to invasive breast cancer that the cancer cells invade the breast tissue. In this stage, the tumor may measures up to 2cm in length and the cancer cell has not spread the exterior part of the breast. Furthermore, no lymph nodes are involved.

Stage II

In stage II, the invasive breast cancer shows one or more of the following characteristics:

- The tumor may not be found in the breast. Nonetheless, cancer can be larger than 2mm in length, which found in 1 to 3 axillary lymph nodes (the lymph nodes underside the arm) or in the lymph nodes adjacent to the breast bone (found throughout a sentinel node biopsy).
- The tumor measures less than 2 cm and has spread to the axillary lymph nodes.
- The tumor is between 2cm to 5 cm and has not spread to the axillary lymph nodes has not spread to the lymph nodes with HER2-negative, estrogen-receptor-positive, progesterone-receptor-negative and has an Oncotype DX Recurrence Score of 9.

Stage III

In stage III, the invasive breast cancer shows one or more of the following characteristics:

- The tumor can be any size or it may not be found in the breast. But, cancer is found in 4 to 9 axillary lymph nodes or in the lymph nodes adjacent the breastbone (found during imaging tests or a physical exam) or
- The tumor is larger than 5cm, small groups of breast cancer cells are found in the lymph nodes or

- the tumor is larger than 5 cm; cancer has spread to 1 to 3 axillary lymph nodes or to the lymph nodes near the breastbone (found during a sentinel lymph node biopsy)
- Still, if the cancer tumor measures more than 5 cm across and is grade 2.
- The cancer is found in 4 to 9 axillary lymph nodes is estrogen-receptor-positive, progesterone-receptor-positive and HER2-positive.
- The tumor may be any size and has spread to the chest wall and/or skin of the breast and caused swelling or an ulcer and may have spread to up to 9 axillary lymph nodes or may have spread to lymph nodes near the breastbone
- Still, if the cancer tumor measures more than 5 cm across and is grade 3.
- The cancer is found in 4 to 9 axillary lymph nodes is estrogen-receptor-positive, progesterone-receptor-positive and HER2-positive.
- Inflammatory breast cancer typically features reddening of a large portion of the breast skin, the breast feels warm and may be swollen. The cancer cells have spread to the lymph nodes and may be found in the skin

Stage IV

Stage IV defines invasive breast cancer, which spreads beyond the breast and neighboring lymph nodes to other organs of the body: lungs, chest wall, distant lymph nodes, skin, bones, liver, or brain.

The monthly response variable, i.e., the time-to-first-recurrent from breast cancer obtained by taking the difference of times from the start of registration at the hospital and until the patient recovered or censored.

For time-to-first-event (in our case time-to-first-recurrent) different survival models were used. In survival analysis, the data are convenient summarizing through estimates of the survival and hazard function.

2.2. Accelerated Failure Time (AFT) Model

Models of survival analysis are used to study data in which the “time until the event” is of interest. The time to an event that can be measured yearly, monthly, weekly, daily, etc. can be failure time or survival time (Wogi et al. 2018, Chere et al. 2015, Collet, 1994).

In the AFT models, we measure the direct effect of the explanatory variables on the survival time instead of a hazard. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the median survival time. The model works to measure the effect of a covariate to "accelerate" or "decelerate" survival time. When the proportional hazard assumptions are not satisfied, the parameters AFT model can be used instead of the Cox proportional hazard model. The AFT model states the survival function of an individual with covariate x at a time t is the same as the survival function of an individual with a baseline survival function at a time $(t \exp(\alpha'x))$, where $\alpha'x = (\alpha_1, \dots, \alpha_p)$ is a vector of regression coefficients (Klein, 2003).

We can consider on a log scale of the AFT model concerning time is given analogous for the classical linear regression approach. In this approach, the natural logarithm of the survival time $y = \log(T)$ is will be modeled. Let us denote S_0 the survival function by setting then we have the following relationship.

$$p\left(T > \frac{t}{x}\right) = p\left(y > \log\left(\frac{t}{x}\right)\right) = S_0\{t^* \exp(-\alpha'x)\}$$

The effects of covariates on the survival function are the time scale is changed by a factor $(-\alpha'x)$ and we call this an acceleration factor. We note that, $\exp(-\alpha'x) > 1$, when $\alpha < 1$, the survival process accelerates. $\exp(-\alpha'x) < 1$, when, the survival process decelerates.

The survival function of T_i can be expressed by ε_i (Klein, 2003). For each distribution of ε_i , there is a corresponding distribution for T . In this case, the AFT models can be interpreted in terms of the speed of the progression of a disease. The effect of the covariates is an accelerated time model is to change the scale and not the location of the baseline distribution of survival times.

Estimation of AFT model

AFT models are fitted using the maximum likelihood estimation technique. The likelihood of the n observed survival time t_1, \dots, t_n is given by:

$$L(\alpha, \mu, \delta) = \prod_{i=1}^n \{f(t_i)\}^{\sigma_i} \{\sigma_i(t_i)\}^{1-\sigma_i}$$

Where $f(t_i)$ and $S_i(t_i)$ are the density and survival functions for the i^{th} individual at t_i and δ_i is the event indicator for the i^{th} observation. The logarithm of the above equation yields:

$$\log L(\alpha, \mu, \delta) = \sum_{j=1}^n \{-\delta_i \log(\sigma t_i + \delta_i \log f_i(x_i) + (1-\delta_i) \log S_i(w_j))\}$$

where $w_j = \{log t_i - \frac{\mu + \alpha_1 x_1 + \dots + \alpha_p x_p}{\sigma}\}, Z = \{z_{ij}\}$ is a vector of covariates for the j^{th} subject.

The maximum likelihood parameter estimates are calculated by using Newton-Raphson procedure.

Weibull AFT Model

Then the Weibull distribution is a very flexible model for time-to-event data. It has a hazard rate which is monotonic increasing, decreasing, and constant. The survival and hazard function of Weibull model with scale parameter and shape parameter is given by (Klein, 2003):

$$f(t, \mu, \alpha) = \frac{\alpha}{\mu} \left(\frac{t}{\mu} \right)^{\alpha-1} \exp \left(-\frac{t}{\mu} \right)^\alpha, \text{ where } \mu > 0 \text{ and } \alpha > 0.$$

$$S_{ei}(t) = \exp \left(-\exp \left(\frac{\log t - (\mu + \alpha' x)}{\sigma} \right) \right) = \exp \left(-\exp \left(\frac{-(\mu + \alpha' x)}{\sigma} t^{\frac{1}{\sigma}} \right) \right)$$

Using the above equation, the proportional hazard representation of the survival function of the Weibull model is given by:

$$S_i(t) = \exp \{ -\exp(\beta_1 x_{1i} + \dots + \beta_p x_{pi}) \lambda t^\gamma \}$$

Comparing the above two formulas the parameter λ, η, β_j in the proportional hazard model

can be expressed by the parameter μ, σ, α_j in the AFT model: $\lambda = \exp(-\frac{\mu}{\sigma}), \eta = \frac{1}{\sigma}, \beta_j = \frac{\alpha_j}{\sigma}$

The AFT representation of hazard function of the Weibull model is given by:

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp \left(\frac{-\mu - \alpha' x_{1i}}{\sigma} \right)$$

Log-Logistic AFT Model

The log-logistic distribution has a frailty flexible functional form. It is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions. The log-logistic model has two parameters λ and ρ , where λ the scale parameter is and ρ is the shape of the parameter. Its probability density function (pdf) given by (Bennet, 1983) and (Cox, 1972). (Bennet, 1983):

$$f(t) = \frac{\lambda \rho t^{\rho-1}}{(1 + \lambda t^\rho)^2}$$

The corresponding survival and hazard functions are given respectively as follows:

$$S(t) = \frac{1}{1 + \lambda t^\rho}$$

$$h(t) = \frac{\lambda \rho t^{\rho-1}}{1 + \lambda t^\rho}$$

Where; $\lambda > 0$, $\rho > 0$. When $K \leq 1$, the hazard rate decreases monotonically and $K > 1$, it increases from zero to a maximum and then decreases to zero.

The AFT model the hazard function for the i^{th} individual in this case is given as:

$$h_i\left(\frac{t}{x}\right) = h_0(t \exp(-\alpha' x_i)) \exp(-\alpha' x_i) = \frac{\rho \exp(\lambda) t \exp(-\alpha' x_i)}{1 + \exp(\lambda) \{t \exp(-\alpha' x_i)\}^\rho}$$

The log-logistic AFT model with a covariate x may be written as: $Y = \log T = \mu + \alpha' x_i + \delta_\varepsilon$, where : $\alpha' = (\alpha_1, \dots, \alpha_p)$; ε has standard logistic distribution. The survival and the hazard functions, respectively, with covariate x are given as follows:

$$S_T\left(\frac{t}{x}\right) = \frac{1}{1 + \lambda \exp(\alpha' x) t^\rho} = \frac{1}{1 + \exp(\log \lambda + \beta' x)}$$

$$h_T\left(\frac{t}{x}\right) = \frac{\rho t^{\rho-1} \lambda \exp(\alpha' x)}{1 + \lambda \exp(\alpha' x) t^\rho} = \frac{\rho t^{\rho-1} \lambda \exp(\alpha' x)}{1 + \exp(\log \lambda + \alpha' x)}$$

To interpret the factor $\exp(\beta' x)$ for the log-logistic model, one can notice that the odds of survival beyond time t for log-logistic model is given by $\frac{S_T(t)}{1-S_T(t)}$. Therefore, We can see that the log-logistic distribution has the proportional odds property. So, this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as:

$$\frac{S_T(t)}{1-S_T(t)} = \exp(-\alpha' x) \frac{S_0(t)}{1-S_0(t)}$$

The factor $\exp(-\alpha' x)$ is an estimate of how much the baseline odds of survival at any time changes when individual has covariate x . And $\exp(-\alpha' x)$ is the relative odds of the experiencing the event for an individual with covariate relative to an individual's baseline characteristics. At this representation of log-logistic regression is the AFT model with a log-logistic baseline survival function, then the log-logistic model is the only parametric model with both a proportional odds and an accelerated failure time representation. The survival function of logistic distribution is given by (Collet, 2003).

$$S_{ei}(\varepsilon) = \frac{1}{1 + \exp(\varepsilon)}$$

Then, the AFT representation of log- logistic survival function is given by:

$$h_i(t) = \frac{1}{\sigma_t} \left[1 + t^{-\frac{1}{\sigma}} \exp\left(\frac{-\mu - \alpha \text{nx}}{\sigma} \right) \right]^{-1}$$

Log-Normal AFT Model

Assuming the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by (Collet, 2003):

$$h_0(t) = \frac{\phi\left(\frac{\log t}{\sigma}\right)}{\left[1 - \phi\left(\frac{\log t}{\sigma}\right)\right]\sigma'^t}$$

Where μ and σ are parameters, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ is the normal probability density function and

$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du$ is the cumulative normal distribution function of the hazard normal distribution. The survival function of the i^{th} individual is:

$$S_i(t) = S_0\left(\frac{t}{\eta_i}\right) = S_0(t * \exp(\mu + \alpha' x_i)) \\ = 1 - \phi(\log t - \alpha' x_i - \mu)$$

Where $\eta_i = \exp(\alpha_1 x_1, \dots, \alpha_p x_p)$, the log survival time for the i^{th} individual has normal $\mu + \alpha' x_i$, σ then the log-normal distribution of AFT property to be

$$\phi^{-1}[1 - S(t)] = \frac{1}{\phi}(\log t - \alpha' x_i - \mu),$$

Where x_i is the value of a categorical variable which takes the value 0 in one group and 1 in the other group. This implies that the plot $\phi^{-1}[1 - S(t)]$ against $\log(t)$ will be linear if the log-normal is appropriate for the given data set.

Multivariate Frailty Models

Frailty models are random effect models for time variables. Thus, the base baseline hazard has a multiplicative effect due to the random effects in the model. The shared frailty model is used with multivariable survival data where unobserved frailty is shared among groups of individuals, and thus a shared frailty model may be thought of as a random effect model for survival data. The frailty model for univariate data has long been used to account for heterogeneous times-to-failure. Consider a parametric survival model characterized by its hazard function $h(t)$. Implicit in the definitions of all these functions are the effects of any covariate, whether we parameterize the model as having PH with respect to changes in covariate values, or accelerated failure time (AFT) due to the covariate. Suppose we have the k^{th} observations and i subgroups. Each subgroups consists of n_i

observation and $\sum_{i=1}^G n_i = n$, where n is the total sample size. The hazard rate for the j^{th} individual in the i^{th} subgroup is given by:

$$h_{ij} = h_0(t)u^i \exp(Z_{ij}^t \beta), i = 1, \dots, n_i$$

Where u_i frailty terms for subgroups and their distribution is again assumed to be independent with a mean of 0 and a variance of 1. If the number of subjects n_i is 1 for all groups, the univariate frailty models obtained (Weinke, 2010); otherwise the model is called the shared frailty model (Dauchateau, 2002) and (Klein, 1992) because all subjects in the same cluster share the same frailty value. In general, we use the notation $h_i(t) = h\left(\frac{t}{x_i}\right)$. The shape parameter p and regression coefficients β are estimated from the data. A frailty model in the univariate case introduces an unobservable multiplicative effect α on the hazard, so that conditional on the frailty, $h\left(\frac{t}{\alpha}\right) = \alpha h(t)$, α is some positive quantity assumed mean of one with variance θ , where (Klein, 1992) discuss the ramifications of the assumed distribution of the frailty, whether gamma or inverse Gaussian.

In this analysis we apply the Gamma Frailty Distribution (Wienke, 2010, Duchateau 2002, Guitierrez, 2002, Hougaard, 1986) and Inverse Gaussian Frailty Distribution (Hougaard, 1984).

Parameter Estimation for shared frailty model

Parameter estimation for right censored clustered survival data, the observation for subject $j \in j_i = \{1, \dots, n\}$ from cluster $i \in I = \{1, \dots, s\}$ is the couple (y_{ij}, δ_{ij}) , where $y_{ij} = \min(t_{ij, c_{ij}})$ is the minimum between survival time t_{ij} and the censoring time c_{ij} and where $\delta_{ij} = I(t_{ij} \leq c_{ij})$ is the event indicator. When covariate information are been collected the observation will be $y_{ij}, \delta_{ij}, x_{ij}$, where x_{ij} denote the vector of covariate for the ij^{th} observation. In the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution. Under assumptions of non-informative right-censoring and independence between the censoring time and survival time

random variables, given the covariate information, the marginal log likelihood of the observed data can be written as:

$$\begin{aligned}
 l_{marg}(\psi, \beta, \theta, Z, X) &= \prod_{i=1}^s \left(\prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(x_{ij}^T \beta))^{\delta_{ij}} \right) \int_0^\infty Z_i^{d_i} \exp(Z_i \sum_{j=1}^{n_i} h_0(y_{ij} \exp(x_{ij}^T \beta))) f(z_i) dZ_i \\
 &= \prod_{i=1}^s \left(\prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(x_{ij}^T \beta)) \right)^{\delta_{ij}} (-1)^{d_i} L^{d_i} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(X_{ij}^T) \right)
 \end{aligned}$$

Where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of events in the i^{th} cluster and $L^q(\cdot)$ the q^{th} derivative of the

Laplace transform of the frailty distribution defined as:

$$L(s) = E(\exp(Zs) \int_0^\infty \exp(-Z_{is}) f(z_i) dZ_i, s \geq 0.)$$

Where ϕ represents a vector parameters of the baseline hazard function, β the vector of regression coefficients and θ the variance of the random effect. Estimates of ϕ , β , θ are obtained by maximizing the marginal log-likelihood above. This can be done if one is able to compute high order derivatives $L^q(\cdot)$ of the Laplace transform up to $q = \max\{d_1, \dots, d_s\}$.

Model building

In model building, to select the best subset of the covariates we were used purposeful selection. Purposeful selection of covariates conducts with the following step: formally, it begins by including covariates that are statistically significant in the univariate analysis. Secondly, include covariates that are considered as more important. Thirdly, use these covariates (those selected out in steps 1 and 2) to fit the multivariable model. Fourthly, select covariates that are statistically significant ($p\text{-value} < 0.05$).

Comparison of Models

In this article, the AIC (Akaike Information Criterion) were used to compare various candidates of parametric shared frailty models (Tesfay, 2014).

Model Diagnostics

After the model have been compared, it is important to check the effectiveness of the model in describing the outcome. The identified accelerated failure model should be linear and goes through the origin with the baseline distribution (Klein, 1992).

3. The Results

The outcome was time to first recurrent breast cancer. As we observe from Table 1, a total of 642 events, about 447(69.6 %) of breast cancer patients recovered/cured. The median cure time from breast cancer was 13 months and the remaining 195 (30.4 %) were censored.

From the total of the events, 45 (7.0%) have taken surgery treatment of which 17 (37.8%) experienced recurrence, and 28 (62.2%) are censored. In the Breast cancer events, 214 (33.3%) are those who took chemotherapy from these events 165 (77.1%) of them recovered/cured the events whereas 49(22.9%) are censored. And also, About 162 (22.5%) women breast cancer patients in the hospital treated with radiotherapy out of this event 102(63.0%)of them cured the events whereas 60(37.0%) are censored. About 51 (7.9%) women breast cancer patients in the hospital treated with a hormone therapy out of this events 32(62.7%) of them recovered/cured the events whereas 19(37.3%) are censored and 170 (26.5%) women breast cancer patients in the hospital treated with a combination of two or more treatments out of this event 131 (77.1%) of them recovered/cured the events whereas 39(22.9%) are censored.

Non-smoker incorporates 603 (93.9%) of the total events where the disease recovered/cured in 426 (70.6%) of them and 177 (29.4%) are censored. About 237(70.1%), 95(69.3%), 54(65.9%), 33(68.8%) and 28(75.7%) women live in Gambela, SNNP, Amhara, Oromia and as well as all other regions together recovered/cured breast cancer respectively. Looking for tumor size of women with breast cancer patients, about 23 (88.5%), 161 (76.7%) and 157 (68.6%) have a size of $\leq 2\text{cm}$, 3 - 5cm and $> 5\text{cm}$ tumor recovered/cured breast cancer respectively, in which the recurrence of women with breast cancer seems higher for higher tumor.

Most of the patients come to the hospital at the second stage (II) and stage four (IV), which is an advanced stage of breast cancer and with this stages out of 195 (30.4%) and 177 (27.6%) events 106 (59.9%) recovered/cured of breast cancer whereas 71 (40.1%) are censored. Even if almost half of women with breast cancer patients, about 355 (55.3%), did not use oral contraceptives, 247 (69.6%) recovered/ cured breast cancer whereas 108 (30.4%) are censored. In the marital

status of the patients, 481(74.9%) events are married and the disease recovered/cured in 329(68.4%) of the events, and 152 (31.6%) are censored.

Table 1: Descriptive summary of the breast cancer patients in the southeast Ethiopia

Variables		No. patients	of		95%	
			Recovered	Censored	Median	CI
Age of patients	<=30	73	58(79.5%)	15(29.4%)	11	(10,13)
	31-40	364	242(66.5%)	122(33.5%)	13	(11,14)
	> 40	205	147(68.9.6%)	58(28.3%)	13	(11,14)
Residence	Urban	446	316(70.9%)	130(29.5%)	12	(11,14)
	Rural	196	131(66.8%)	65(33.2%)	13	(11,14)
Treatment taken	Chemotherapy	214	165(77.1%)	49(22.9%)	13	(10,14)
	Radiotherapy	162	102(63.0%)	60(37.0%)	13	(11,15)
	Surgery	45	17(37.8%)	28(62.2%)	18	(14,15)
	Hormone.	51	32(62.7%)	19(37.3%)	15	(10,19)
	Combination.	170	131(77.1%)	39(22.9%)	11	(10,12)
Stage	I	26	23(88.5%)	3(11.5%)	11	(9,16)
	II	210	161(76.7%)	49(23.3%)	11	(11,13)
	III	229	157(68.6%)	72(31.4%)	12	(11,14)
	IV	177	106(59.9%)	71(40.1%)	14	(12,16)
Obesity	Underweight	162	110(67.9%)	52(32.1%)	11	(10,14)
	Normal	132	132(72.5%)	50(27.5%)	12	(11,14)
	Overweight	298	205(68.8%)	93(31.2%)	13	(11,14)
Histological grade	Well diff	159	114(71.7%)	45(28.3%)	11	(11,14)
	Moderately	239	169(70.7%)	70(29.3%)	12	(11,14)
	Poorly diff	244	164(67.2%)	80(32.8%)	14	(12,14)
Alcohol consumption	No	312	219(70.2%)	93(29.8%)	13	(11,14)
	Yes	330	228(69.1%)	102(30.9%)	12	(11,14)
Family history	No	312	215(68.9%)	97(31.1%)	12	(11,14)
	With	330	232(70.3%)	98(29.7%)	13	(11,14)
Breast feed	No	297	203(68.4%)	94(31.6%)	12	(11,14)
	Yes	345	244(70.7%)	101(29.3%)	13	(11,14)
Tumor size	< 2cm	271	187(69.0%)	84(31.0%)	11	(10,14)
	2cm-5cm	256	176(68.8%)	80(31.2%)	13	(11,14)
	> 5cm	115	84(73.0%)	31(27.0%)	13	(11,14)
Smoking habit	No	603	426(70.6%)	177(29.4%)	13	(11,14)
	Yes	39	21(53.8%)	18(46.2%)	9	(8,10)
Menopause status	pre-meno.	345	239(69.3%)	108(30.4%)	13	(11,14)
	post-meno.	297	208(70.0%)	89(30.0%)	12	(11,14)
Oral contraceptive	Not used	355	247(69.6%)	108(30.4%)	13	(11,14)
	Used	287	200(69.7%)	87(30.3%)	13	(11,14)

Marital status	Single	57	42(73.7%)	15(26.3%)	11	(10,14)
	Married	481	329(68.4%)	152(31.6%)	13	(11,14)
	Widowed	69	49(71.0%)	20(29.0%)	11	(10,15)
	Divorced	35	27(77.1%)	8(22.9%)	13	(10,14)
Total		642	447(69.6%)	195(30.4%)	13	(11,14)

Univariate Analysis

The univariate analysis is applied to select covariate variables that predict the time-to-first recurrence of breast cancer patients before proceeding to the multivariate analysis. The univariate analyses were fitted for each covariate by using different baseline distributions like Weibull, log-logistic, exponential, and lognormal and frailty distribution (gamma and inverse Gaussian) for the data set.

Table 2: Univariable analysis results

Variables		coef	se(coef)	se2	P-value
Age of patients	(Intercept)	2.598	0.0527	0.05	0.001
	<=30	Ref			
	31-40	0.113	0.0527	0.0555	0.043
Residence	> 40	0.157	0.0592	0.059	0.0079
	(Intercept)	2.6961	0.0216	0.0216	0.0001
Treatment taken	Urban	Ref			
	Rural	0.0189	0.0398	0.0398	0.63
Stage	(Intercept)	2.682	0.0345	0.0299	0.0001
	Chemotherapy	Ref			
	Radiotherapy	0.085	0.0474	0.0474	0.073
	Surgery	0.264	0.0963	0.0962	0.0061
	Hormone.	0.185	0.0731	0.0729	0.012
Obesity	Combination.	-0.071	0.0441	0.044	0.011
	(Intercept)	2.5846	0.0813	0.0793	7.20E-22
	I	Ref			
	II	0.0675	0.0846	0.0846	4.20E-01
Histological Grade	III	0.121	0.085	0.0848	1.50E-01
	IV	0.2265	0.0873	0.0873	9.50E-03
	(Intercept)	2.6713	0.0417	0.0373	0.00E+00
Underweight	Underweight	Ref			
	Normal	0.0504	0.0373	0.0494	0.031
	Overweight	0.0497	0.0373	0.0453	0.027
Well diff	(Intercept)	2.70212	0.0401	0.0362	0.002
	Well diff	Ref			
	Moderately	-	0.0467	0.0465	0.89

					0.00672
	Poorly diff	0.03052	0.0478	0.0472	0.52
	(Intercept)	2.71264	0.0319	0.0266	0.0001
Alcohol	No	Ref	-		
Consumption	Yes	0.00476	0.0365	0.0363	0.9
	(Intercept)	2.7358	0.0323	0.0269	0.0012
Family History	No	Ref			
	With	-0.0501	0.0361	0.0361	0.17
	(Intercept)	2.6786	0.033	0.0277	0.00001
Breast feed	No	Ref			
	Yes	0.0553	0.0363	0.0363	0.13
	(Intercept)	2.6955	0.0464	0.0427	0.003
Tumor size	< 2cm	Ref			
	2cm-5cm	-0.012	0.0515	0.0514	0.0082
	> 5cm	0.0427	0.0502	0.0501	0.04
	(Intercept)	2.719	0.0263	0.0195	0.0001
Smoking Habit	No	Ref			
	Yes	-0.281	0.0849	0.0848	0.00094
	(Intercept)	2.6902	0.0309	0.0254	0.0025
Menopause Status	pre-meno.	Ref			
	post-meno.	0.0434	0.0366	0.0364	0.24
	(Intercept)	2.714	0.031	0.0253	0.00013
Oral	Not used	Ref	-		
Contraceptive	Used	0.00829	0.0365	0.0365	0.0082
	(Intercept)	2.556	0.0619	0.0591	0.0002
Marital Status	Single	Ref			
	Married	2.0176	0.0622	0.0622	0.0046
	Widowed	0.113	0.0802	0.0798	0.01
	Divorced	0.137	0.0938	0.0936	0.04

From Table 2, the results of a univariate analysis using Weibull with Gamma and Weibull with inverse Gaussian shared frailty distribution shows each covariate variables like age of patients, treatment is taken, stages of diseases, smoking habit, obesity, alcoholic consumption, tumor size, oral contraceptive, and marital status were significantly associated with time-to-first recurrence of breast cancer at 5% level of significance. These significant variables were considered in multivariable analyses.

Multivariable Analysis and Comparison of Models

Six multivariable survival models were fitted with exponential, Weibull, and log-normal as baseline hazard function with gamma and the inverse Gaussian frailty distributions using all the covariates that are significant in the corresponding univariate analysis at 5% level of significance. To compare the efficiency of different models, the AIC was used. It is the most common applicable criterion to select a model. A model having the minimum AIC value was preferred.

Table 3: Comparison of model by AIC

Baseline hazard function	Frailty distribution	AIC
Exponential	Gamma	3417.55
	Inverse-Gaussian	3228.85
Weibull	Gamma	2904.83
	Inverse-Gaussian	2913.14
Lognormal	Gamma	2953.33
	Inverse-Gaussian	2843.43

Table 3 shows summaries of the AIC (Akaike Information Criterion) results of the two shared frailty models with three baseline hazard functions. Among those models, the inverse-Gaussian shared frailty model with log-normal baseline hazard function has the smallest AIC (Akaike Information Criterion). That indicates that the inverse-Gaussian shared frailty model with lognormal was relatively the most appropriate model to fit the time-to-first recurrence of breast cancer patients in southwest Ethiopia under the selected covariate variables.

Table 4: Multivariable analysis using the log-logistic-inverse Gaussian frailty model

Covariates		Coefficients	S.E.	P-value	ϕ	95% CI of ϕ
Age of patients	≤ 30	Ref				
	31-40	-0.258	0.167	0.123	0.77	(0.557,1.071)
	> 40	-0.269	0.188	0.153	0.76	(0.528,1.104)
Treatment taken	Chemotherapy	Ref				
	Radiotherapy	-0.261	0.131	0.047 *	0.77	(0.595,0.995)
	Surgery	-0.631	0.267	0.018 *	0.53	(0.315,0.897)
	Hormone.	-0.346	0.22	0.115	0.71	(0.459,1.088)
	Combination.	0.163	0.121	0.178	1.18	(0.928,1.492)
Stage	I	Ref				
	II	-0.241	0.22	0.273	0.79	(0.510,1.209)
	III	-0.309	0.22	0.16	0.73	(0.477,1.129)
	IV	-0.635	0.222	0.004 **	0.53	(0.343,0.818)

Obesity	Underweight	Ref				
	Normal	-0.149	0.132	0.26	0.86	(0.665,1.115)
	Overweight	-0.126	0.127	0.321	0.88	(0.687,1.130)
Alcohol consumption	No	Ref				
	Yes	-0.036	0.107	0.736	0.96	(0.805,1.189)
Tumor size	< 2cm	Ref				
	2cm-5cm	0.034	0.143	0.813	1.03	(0.781,1.369)
	> 5cm	-0.002	0.134	0.991	1	0.767,1.297)
Smoking habit	No	Ref				
	Yes	1.122	0.246	<.001 ***	3.07	(1.896,4.973)
Oral contraceptive	Not used	Ref				
	Used	0.066	0.101	0.516	1.07	(0.876,1.302)
Marital status	Single	Ref				
	Married	-0.427	0.184	0.02 *	0.65	(0.454,0.935)
	Widowed	-0.254	0.24	0.291	0.78	(0.484,1.241)
	Divorced	-0.085	0.269	0.753	0.92	(0.542,1.556)
	$\theta = 0.11(S.E=0.019)^*$					$\tau=0.05$
	$\rho=2.168(S.E=0.076)^*$					AIC=2843.43

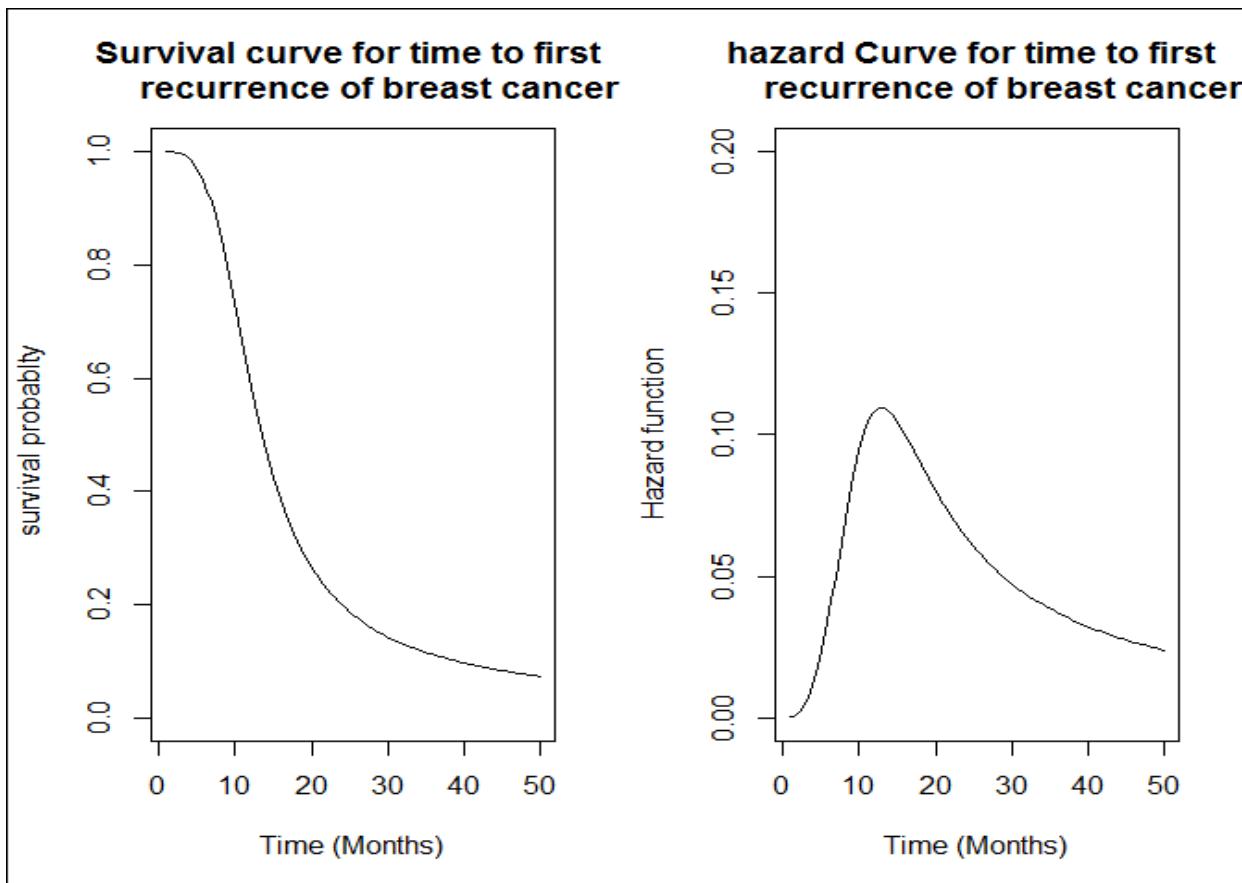
S.E= standard error, ϕ =acceleration factor; 95% CI= 95% confidence interval for acceleration factor,

Ref. =Reference, θ = variance of random effect, ρ =shape parameter, τ =Kendall's tau, AIC= Akaike's Information Criteria, the empty cell indicates the covariate variable is not significant with respect to the data.

Analysis based on log-normal with inverse Gaussian frailty model showed that treatment is taken, stage, smoking habit, and marital status were significant. Women who took radiotherapy and surgery had significantly different curing than the reference group (chemotherapy) with the acceleration factors of (ϕ = 0.77, ϕ = 0.53) and 95% confidence intervals of acceleration factor (0.595, 0.995) and (0.315, 0.897) respectively, did not include 1. The estimated coefficient of the parameters for women who took radiotherapy and surgery was - 0.261 and -0.631 respectively. The sign of the coefficients is negative which implies that decreasing the log-logistic of survival time and hence, the shorter expected duration of time to cure breast cancer. Therefore, women who took radiotherapy and surgery had shrunk cure time by factors of 0.77 and 0.53 then the reference group (chemotherapy). But the other treatment taken group was not significantly different from the baseline at a 5% level of significance. The time rate and 95% confidence interval of the acceleration factor of stage IV were 0.53(0.343, 0.818) respectively when compared to the category of stage I as a reference category. This indicates patients with stage IV their survival was reduced by a factor of ϕ = 0.53 compared with patients who were in stage I. The confidence interval of the acceleration factor of smoking habit was (1.896, 4.973), did not include 1, indicating smoking habit is a significant prognostic factor for time cure from breast cancer. Accordingly, women

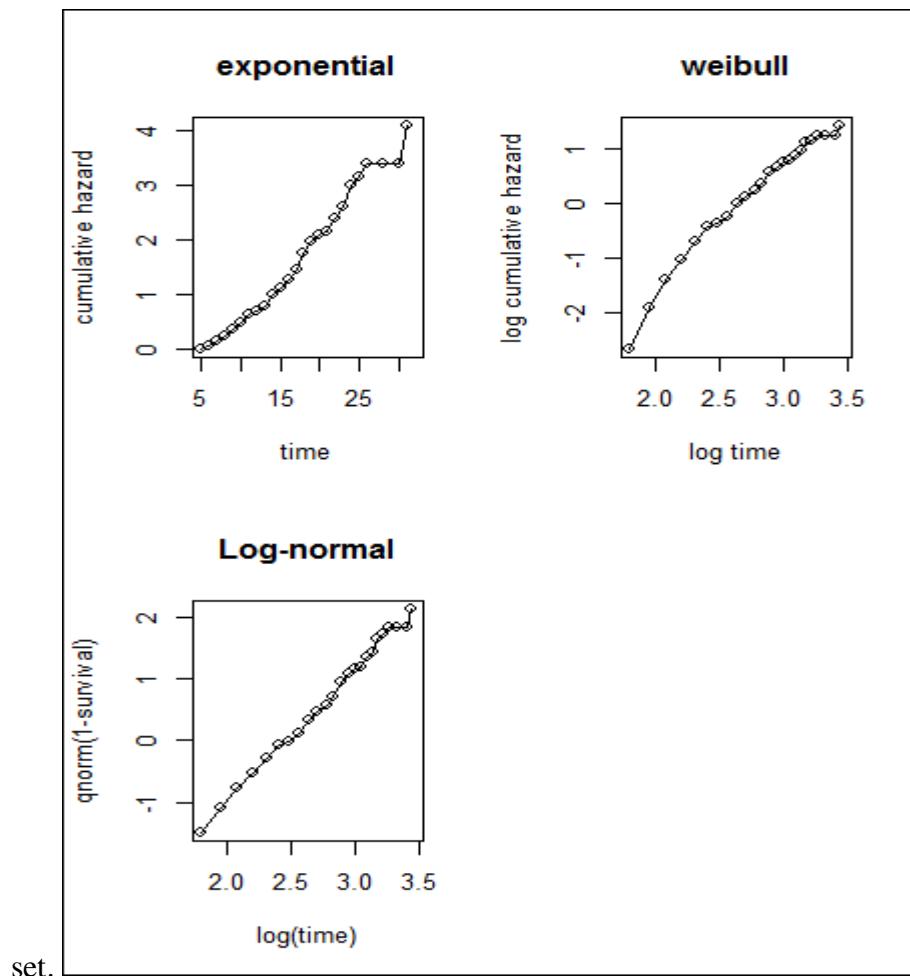
who had a smoking habit had prolonged cure time by a factor of 3.07 than the women who had not smoked (reference group). The confidence interval of the acceleration factor of the marital status of the married group is (0.454, 0.935), did not include 1, at a 5% level of significance. This indicates patients of married group survival were shrunk by a factor of $\phi = 0.65$ compared with the reference category (single). The value of the shape parameter in our working model (log-normal model) is $\rho = 2.168$. Since this value is greater than unity the hazard function is unimodal i.e. it increases up to some time and then decreases and The Km curves to the survival and hazard function of time- to- first recurrent of breast cancer is shown in Figure 1. The curve shows that the probability that time to the first recurrence is decreasing with increasing time. The hazard function increases at the beginning a few months and the curve decrease slightly after some months (Figure 1). The heterogeneity in the population of clusters (region) estimated by log-normal was $\theta = 0.11$ and dependence within clusters was about $\tau = 5\%$ (Table 4).

Figure 1: The Kaplan Meier plots of Survival and hazard functions of time to first recurrent of breast cancer



After the model has been fitted, it is desirable to determine whether a fitted parametric model adequately describes the data or not. Therefore, the appropriateness of the model is measured based on the linearity of the plot of given baseline distribution for the given dataset. Accordingly, their respective plots are given in Figure 2 and the plot for the log-normal baseline distribution looks straight line than exponential and Weibull baseline distribution. This evidence also strengthens the decision made by AIC value that log-normal baseline distribution is appropriate for the data.

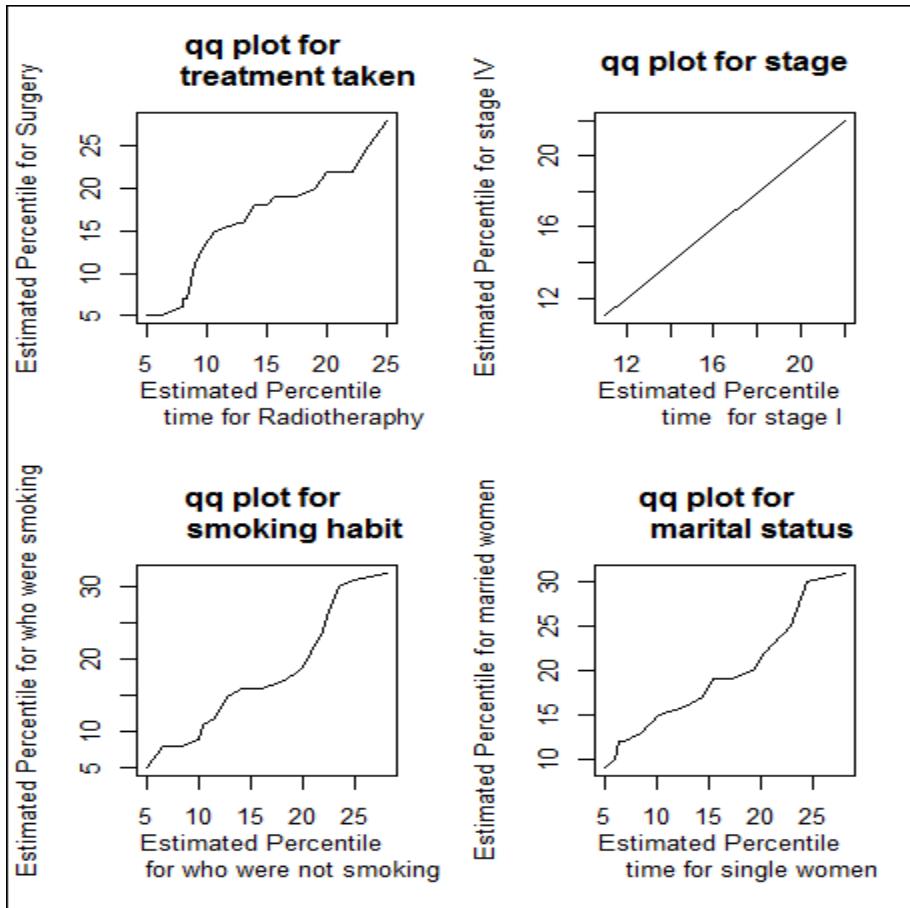
Figure 2: Graphs of Exponential, Weibull, and Log-normal baseline distributions for time-to-first recurrent of breast cancer data set



A quantile-quantile or q-q plot is made to check if the AFT provided an adequate fit to the data using by two different groups of the population. We shall graphically check the adequacy of the model by comparing the significantly different groups of the covariate variables. All most all of

the plot appears to be approximately linear as shown in Figure 3. Therefore, the accelerated failure time appears to be the best to describe the time-to-first recurrences of the data set.

Figure 3: Quantile- Quantile plot to check the adequacy of the accelerated failure time model



4. Discussion

To select prognosticators that affect time to the first recurrence of breast cancer patients, we used univariate analysis with different baseline distribution like Weibull, log-logistic, exponential, and lognormal and frailty distribution (gamma and inverse Gaussian) for the data set. All significant variables in the univariate were considered in multivariate analysis. The comparison of different models with their baseline hazard function was performed using AIC criteria, where a model with minimum AIC is the most appropriate model to fit the time-to-first recurrence of breast cancer patients data (Munda et.al, 2012). Therefore, Inverse-Gaussian shared frailty model with lognormal has the smallest AIC. The result of this study showed that there was a clustering effect and heterogeneity in the region in southwest Ethiopia from which

women with breast cancer came and here we assumed that women coming from the same region share the same factors relating time to the first recurrence of breast cancer. Accordingly, we considered the clustering effect in hazard function.

The hazard function indicates increases more sharply at the early stage and failure occurs at random and failure happened (Duchateau, 2008). The heterogeneity and dependence in the region were estimated to be $\theta = 0.11$ and $\tau = 5\%$. These values are the minimum among other variance effects and Kendall's tau of all the models, which related to the idea of the better the model, the less observed heterogeneity (Banbeta et al. 2015). The time to the first recurrence of breast cancer data set was most appropriate fitted with log-normal depending on model diagnostic (assumption of linearity) which was compared with exponential and Weibull hazard function.

The results obtained from multivariable analysis with the appropriate model Inverse-Gaussian shared frailty model with lognormal showed that treatment is taken, stage, smoking habit, and marital status were the factors that determine the time to first cure of breast cancer.

The estimated median survival time to the first recurrence from breast cancer of women in southwest Ethiopia is found to be 13 months with a 95% confidence interval of [11,14]. This finding is almost similar to (Dye et.al.2012).

Women who took radiotherapy and surgery had shrunk curing time than the reference group (chemotherapy). This finding is supported by (Abadi et.al 2014). But the results obtained by (Kuru, 2008, Hu. 2006) showed that radiotherapy increases the overall survival rate of patients with stage III cancer. Stage III was not significant in our study.

The stages of breast cancer have a significant effect on the time to the first recurrence of breast cancer patients. Patients with stage IV their survival was shrunk by a factor of $\phi=0.53$ compared with patients who were in stage I. This is done by (Demicheli et al., 2010, Dignam et al., 2009. In general, the results of this study showed that for patients with stage IV, their time to cure breast cancer was decreased.

5. Conclusion

The main objective of this study was to identify risk factors affecting time to the first recurrence of breast cancer patients in southwest Ethiopia using different parametric shared frailty models. 642 women of breast cancer patients from southwest Ethiopia hospitals were included in the study.

The median age of women of the time to cure breast cancer was 13 months. Treatment taken, stage, smoking habit, and marital status were statistically associated with time to first cure of breast cancer.

Awareness has to be given on breast cancer for the society regarding treatment taken, stage, smoking habit, and marital status to mitigate breast cancer.

Abbreviation

AIC: Akaike;s Information Criteria

Declaration

Ethics approval and consent to participate

Ethical approval to conduct the study and human subject research approval for this study was received from Jimma University, College of Natural Sciences, Research Ethics Committee and the Medical Director of the Hospitals. As the study was retrospective, informed consent was not obtained from the study participants, but data were anonymous and kept confidential.

Consent for publication

Not applicable.

Availability of data and material

The data for this analysis will be provided based on request.

Competing interests

The authors declare that they have no competing interest.

Funding

This study project was awarded funding from Jimma University, college of natural sciences research and postgraduate office.

Authors' contribution

RHB contributed to the study concept and design. YNJ participated in the data collection and drafted the manuscript. YYT reviewed the findings of data analyses and was a major contributor in writing the manuscript. KTD performed the statistical analyses and drafted the manuscript. All authors approved the final manuscript.

Acknowledgement

First of all authors would like thank Jimma University, college of natural sciences research and post graduating office for providing financial support for this study and would like to thank Jimma medical center, Bedelle, Mizan-Aman and Mettu Karl hospitals in south west Ethiopia for providing us data on breast cancer patients.

References

- Abadi, A., Yavari, P., Dehghani-Arani, M., Alavi-Majd, H., Ghasemi, E., Amanpour, F., & Bajdik, C. (2014). Cox models survival analysis based on breast cancer treatments. Iranian journal of cancer prevention, 7(3), 124.
- Abate SM, Yilma Z, Assefa M, Tigeneh W (2016). *Trends of Breast Cancer in Ethiopia*. Int J Cancer Res Mol Mech 2(1): doi <http://dx.doi.org/10.16966/2381-3318.121>
- Banbeta, A., Seyoum, D., Belachew, T., Birlie, B., & Getachew, Y. (2015). Modeling time-to-cure from severe acute malnutrition: application of various parametric frailty models. Archives of Public Health, 73(1), 6.
- Brinton LA, Figueroa JD, Awuah B, Yarney J, Wiafe S, Wood SN, et al. (2014) 'Breast cancer in Sub-Saharan Africa: opportunities for prevention'. Breast cancer research and treatment. 144 (3):467-78.
- Chere Fikadu Zawdie, Tesfay Yohannes Yebabe, Fikre Enquoselassie (2015). *Parametric Survival Modelling of Risk Factor of Tuberculosis Patients under DOTS Program at Hawassa Town, Ethiopia*. 01/2015; 4(1):1-17., DOI:10.4018/IJBCE.2015010101

Collet, D.(1994). Modeling Survival Data in Medical Research (2nd edition). London, UK: Chapman and Hall/CRC.

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 tumour biology within the residual breast. *BMC Cancer* 10, 656 (2010) doi:10.1186/1471-2407-10-656

Dignam J.J, Dukic V, Anderson SJ, Mamounas EP, Wickerham DL, Wolmark N (2009). Hazard of recurrence and adjuvant treatment effects over time in lymph node-negative breast cancer'. *Breast Cancer Res Treat* 116:595602. doi:10.1016/j.mcm.2010.02.002.

Dye TD, Bogale S, Hobden C et al (2012) Experience of initial symptoms of breast cancer and triggers for action in Ethiopia. *Int J Breast Cancer* 2012:908547.

Duchateau L, Janssen P (2008). The Frailty Model. New York: Springer-Verlag

Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin D, (2009). Cancer incidence and mortality worldwide: IARC Cancer Base No. 10. Lyon, France: International Agency for Research on.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN (2013). Cancer Incidence and Mortality Worldwide:IARC Cancer Base No. 11 [Internet]. Lyon, France: International Agency for Research onCancer. Available from: <http://globocan.iarc.fr>, 2013

Globocan (2012) <http://globocan.iarc.fr/Default.aspx>. Accessed 16 Sep2015

Hougaard, P.(1984).Survival Models for heterogeneous population derived from stable distributions. *Biometrika*, 72(2),387-396.

Klein, J.(1992). Semi parametric estimation of random effects using Cox model based on the EM algorithm. *Biometrics*, 48, pp.795-806.

Munda M (2012). Parametric frailty models in R. *American Statistics Association*; 55:1–21

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011). Global cancer statistics. A cancer journal for clinicians American cancer society, 61, 69-90.

Kuru B, Camlibel M, Dinc S, Gulcelik MA, Gonullu D, Alagol H Singapore Med J. 2008 Nov; 49(11):904-11.

- Parkin DM, Bray F, Ferlay J, Jemal A (2012). ‘Cancer in Africa’. *Cancer Epidemiology and Prevention Biomarkers* 2014, 23(6):953-966.
- Tefera,B., Assefa, M., Abebe, B. and Rauch, D. (2016). Patterns of Cancer in Uni- versity of Gondar Hospital: North-West Ethiopia. *Journal of oncology medicine and practice*, 1,2.
- Tesfay, Y. (2013). The Fluctuation of the Prices of Crude Oil at International level explained by a Dynamic Discrete Stochastic Models. *American Journal of Economics*, 3(3), 140-148.
- Tigeneh W (2015). Pattern of Cancer in Tikur Anbessa Specialized Hospital Oncology Center in Ethiopia from 1998 to 2010. *Int J Cancer Res MolMech* Volume 1.1: doihttp://dx.doi.org/10.16966/2381-3318.103
- Timotewos G, Solomon A, Mathewos A, (2018). First data from a population based cancer registry in Ethiopia. *Cancer Epidemiology* 53:93–8
- Wienke A., Ripatti S., Palmgren J., Yashin A.I., (2010). A bivariate survival model with compound Poisson frailty. *Statistics in Medicine* 29, 275.83.
- Wogi Abebe Argaw, Wakweya Shibru Temesgen, Tesfay Yohannes Yebabe (2018). *Determinants of Time-to-Under-Five Mortality in Ethiopia: Comparison of Parametric Shared Frailty Models*. 01/2018; 7(1):1-24., DOI:10.4018/IJBCE.2018010101
- WHO (2017). World Health Organization Latest report cancer statistics. Geneva: International Agency for Research in Cancer, 2017
- WHO (2012). World Health Organization. Latest world cancer statistics. Geneva: International Agency for Research in Cancer, 2012.

Figures

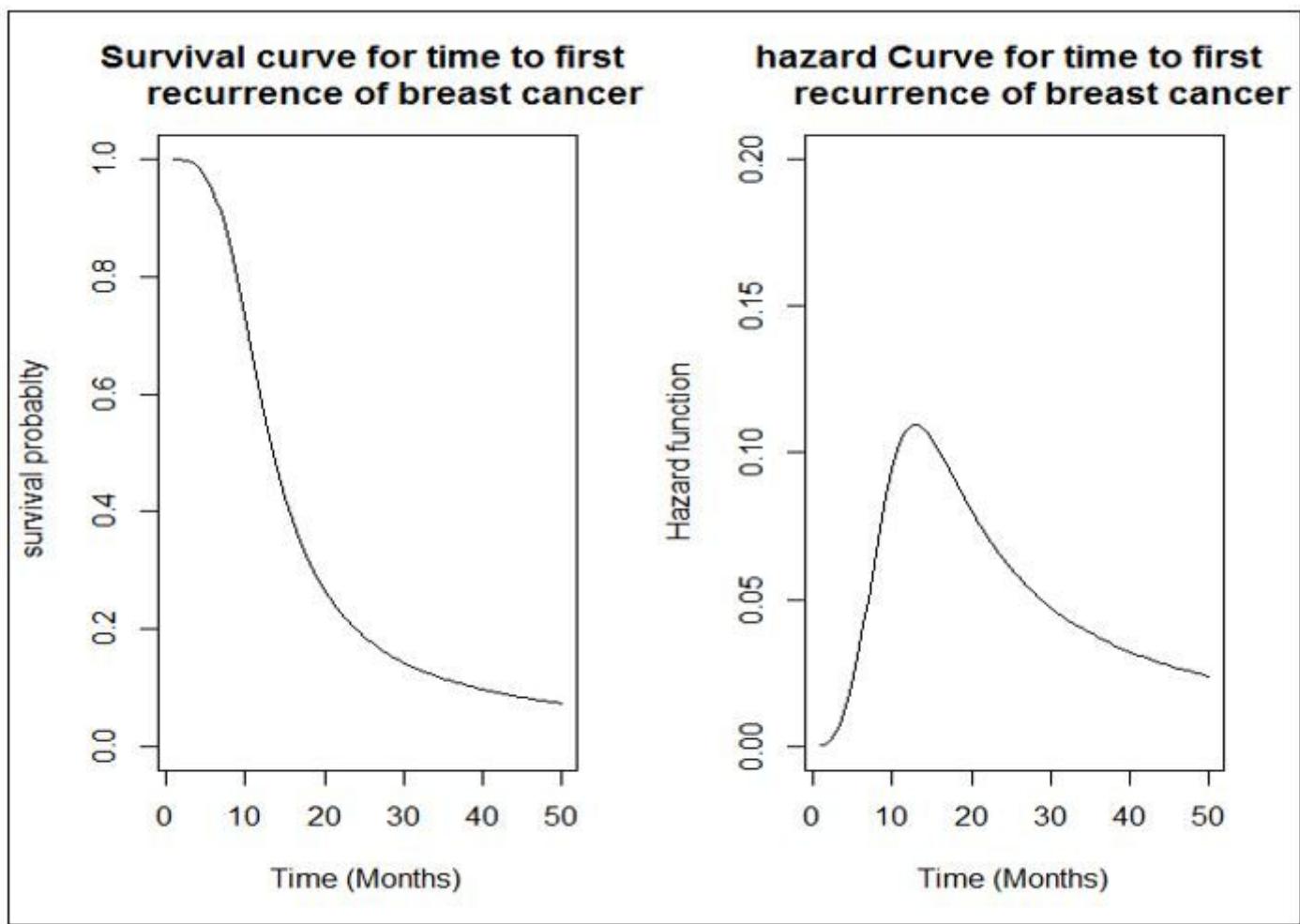


Figure 1

The Kaplan Meier plots of Survival and hazard functions of time to first recurrent of breast cancer

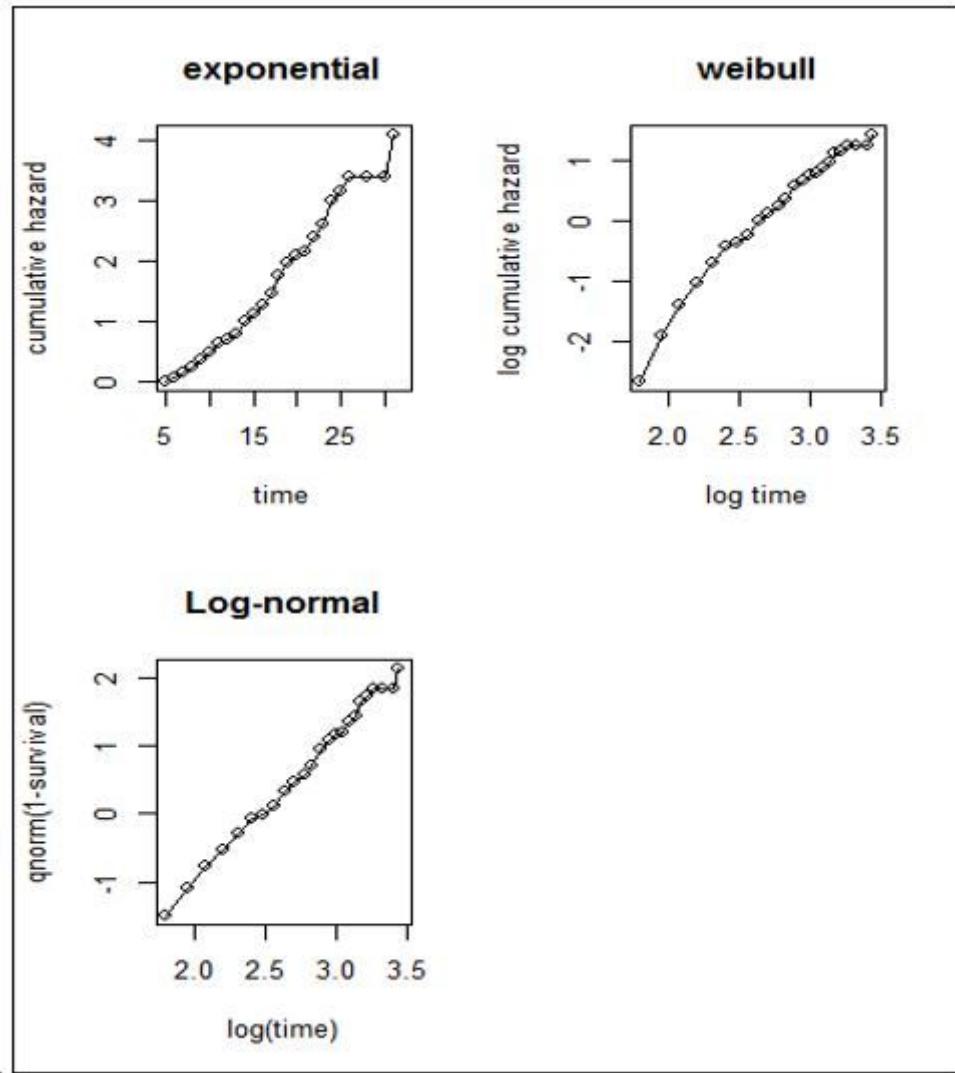


Figure 2

Graphs of Exponential, Weibull, and Log-normal baseline distributions for time-to-first recurrent of breast cancer data set

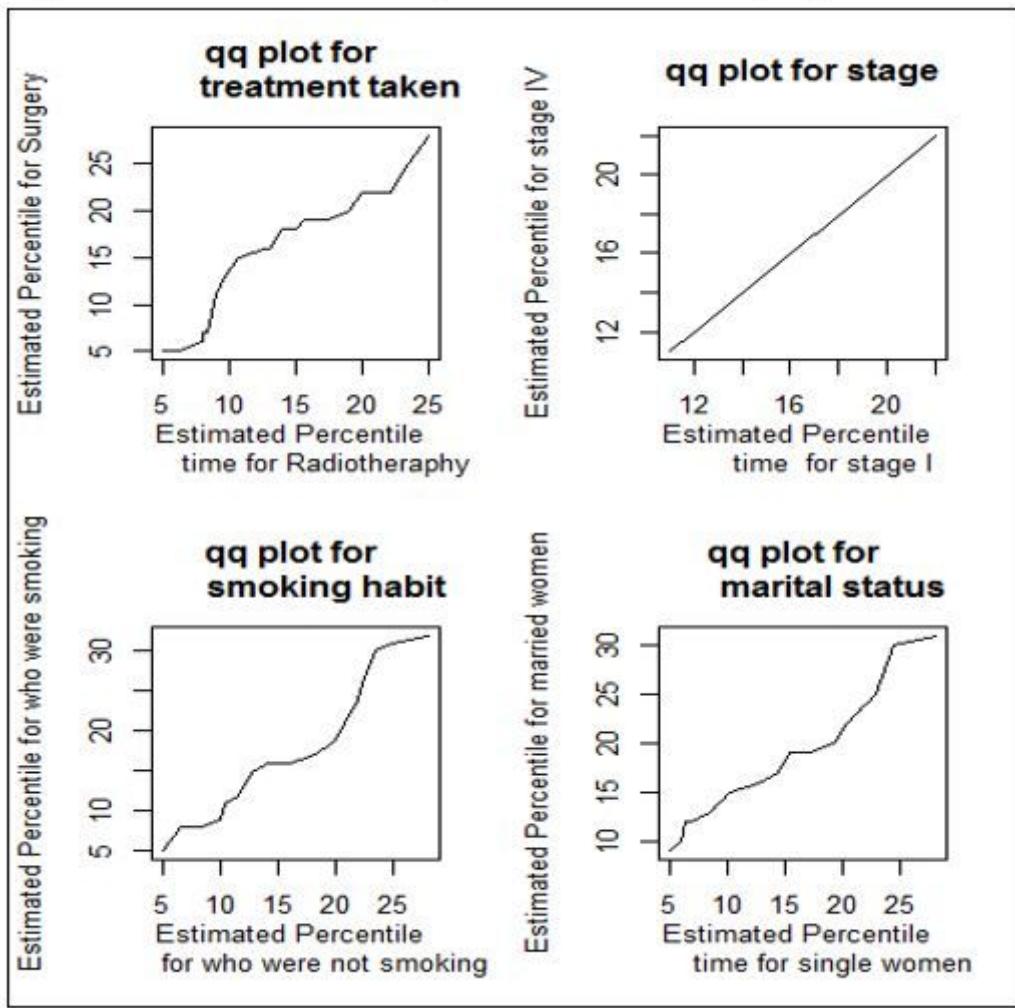


Figure 3

Quantile- Quantile plot to check the adequacy of the accelerated failure time model