Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.12, No.1, 2022

A Novel Three Parameter Modified Cox Frailty Model For Noncompeting Risk Patients with Breast Cancer Malignality

Senyefia Bosson-Amedenu [1,*], Joseph Acquah [1], Christiana Cynthia Nyarko [1], Noureddine Ouerfelli[2]

[1]*Department of Mathematical Sciences, University of Mines and Technology, P.O. Box 237, Tarkwa, Ghana;*

[2]*Institut Supérieur des Technologies Médicales de Tunis, LR13SE07, Laboratoire de Biophysique et Technologies Médicales, Université de Tunis El Manar, Tunis, Tunisia.*

**Email of Corresponding author:*
sbosson-amedenu@st.umat.edu.gh ;senyefia@yahoo.com;(Bosson-Amedenu S.)

Abstract

The typical Cox proportional hazard (PH) model will provide erroneous estimates if the PH assumption is broken, which is quite prevalent in medical research. We have developed an expanded version of the basic Cox Model that includes a time-lag function and a frailty parameter to account for time-variant covariates, heterogeneity and unobserved components in this study. Secondary data from 558 Breast cancer (BC) patients diagnosed at Korle Bu teaching hospital were analyzed. The dataset was divided into two parts: training (which had 70% of the data) and validation (30 percent). Tests for the functional form of continuous covariates and outliers were included in the model diagnostics. The Shoenfeld residual test and the graphical test served as the foundation for the PH assumption test. In a noncompeting risk environment, the PH assumption was violated by progesteron receptor status, molecular subtype, and tumor grade at diagnosis. Frailty component was revealed to be a significant contributor to the developed model, accounting for around 15% of all fatalities attributable to heterogeneity and unobserved variables. Our model outperformed current models such as the Exponential AFT model, stratified Cox (interaction) model, the standard Cox PH Model and Park and Qiu (2017) model in terms of AIC, BIC, likelihood ratio test and area under the ROC curve. Breast cancer survival in Ghana is influenced by stage at diagnosis, metastatic status, lymph node involvement, and HER2 overexpression, according to our model. Among other findings, BC patients who develop metastasis are 41.264 times more likely to die from the disease than individuals who do not develop metastasis. Individuals with higher stages of BC (III and IV) are 6.89 times more likely to die from the disease than patients with lower stages (I and II). To improve BC care and prognosis, it was suggested that medical officers and diagnosticians take into account the identified significant determinants regulating survival as well as the estimated risk and survival probability.

Keywords: Breast Cancer, Frailty, Heterogeneity, Cox Proportional Hazard Model, Stratified Cox (interaction), Extended Cox.

1. Introduction

Breast cancer (BC) is a hereditary disorder caused by the uncontrolled multiplication of aberrant breast cells as a result of DNA damage [1]. Although around 75% of the causes of BC are unclear, 25% can be attributed to modifiable and non-modifiable factors [2,3].

Although the Cox PH Model (homogenous and semi-parametric) has been the most robust in comparison to parametric and nonparametric models for the last four decades, it has several drawbacks. It is based on proportional hazards, which posits that hazard ratios for certain fixed factors are stable across time. In real-world medical practice, patient factors are mainly heterogeneous and time-varying; in this instance, the Cox PH model in its original form will not be appropriate [4]. It is common in clinical trials for effects to fluctuate over time, and because standard models do not provide for a natural explanation of such effects, there is a need to extend these models [5].The Cox PH model, which is a pure version of survival, produces approximation relationships that largely disregard nonlinear factors. An accurate or true survival model reflects both the linear and nonlinear connection between variables and the response variable in practice [6].Unobserved heterogeneity, unobserved variables, and time-variant covariates for noncompeting risk age groups are not addressed by the Cox Model in its original form.

Frailty is a latent unobserved random variable that accounts for unobserved heterogeneity and unobserved variables in survival modeling [7,8].Heterogeneity emerges as a result of differences in patients, provider care, and other hidden factors [9]. In survival studies, when unobserved heterogeneity (frailty) is overlooked, bias is

introduced into the estimations, leading to misleading results. Frailty can be included in survival models to account for unobserved heterogeneity in addition to observable factors [10]. In multi-center research, frailty variance is considerable and should not be overlooked when modeling [8].

In order to make educated decisions on breast cancer prognosis, recent research have underlined the need of evaluating competing causes of death. Traditional approaches for describing survival processes, such as the Kaplan Meier product-limit method and Cox Model, are not designed to account for the competing nature of numerous causes for the same event, and hence yield inaccurate estimates when examining the marginal probability for cause-specific occurrences [11]. The occurrence of the event of interests is ruled out due to competing risk [12]. To evaluate competing risk in different age groups, noncompeting risk models are compared to competing risk models to see if the latter is overestimated [13].[14,15] discovered that the risk of competing events rises with age, from 19 percent in patients less than 75 years to 54 percent in patients older than 75 years.

According to a recent evaluation of cancer-related papers published in the last decade that used the Cox PH model, 81 percent of them failed to account for the proportionate hazard assumption [16]. The proportional hazards (PH) assumption has never been rigorously tested in a large clinical trial dataset, and the effects of time-varying treatments have never been fully understood [17].

2. Method

Secondary data from 558 BC patients diagnosed at Korle Bu teaching hospital between 2010 and 2015 and followed up on (right censored) until the end of 2015 were analyzed. Event history analysis was used to identify these individuals' survival status, demographics, and tumor features. The Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), and Reciever Operation Characteristic (ROC) curves were used to compare various survival models. The data was analyzed using the R programming language. The participants in the study ranged in age from 13 to 97 years old. Prior to the testing of the PH assumption, test of influential and nonlinear covariates was conducted. The Cox PH assumption was tested with the shoenfeld residual test and the graphical method. Comparing estimates of cumulative incidence functions and the complement of the Kaplan Meier Product limit estimator, as well as estimates of the Cox PH model and the Fine Gray model to check any overestimation of probabilities, were utilized to account for competing risk in the survival data used. Competing risk ages were identified (about 7 percent) and analysed separately. The remaining 93 percent of the survival data was used in main analysis the model. The dataset was divided into two parts: training (which had 70% of the data) and validation (which held 30% of the data) (30 percent).

The following relationships was used to develop the modified version of the standard Cox Model:

The Cox PH Model is given by :

$$\psi(t|\mathbf{X}) = \psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i\right] \quad (1)$$

The frailty parameter , ℓ in model building is assumed to have a multiplicative effect on the baseline hazard $\psi_0(t)$ given by :

$$\psi(t|\ell) = \ell\,\psi_0(t) \quad (2)$$

The Gamma distribution is given by;

$$f(x) = \frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}, x>0, \quad \alpha>0 \quad (3)$$

Partial Maximum Likelihood Estimator for the standard Cox Model is given by;

$$\log PL = \sum_{i}^{n}\delta_i\prod_{i=1}^{n}\left[\beta x_i - \log\left(\sum_{j=1}^{n}V_{ij}\exp(\beta x_i)\right)\right] \quad (4)$$

V_{ij} modifies the risk set over time.

The area under the ROC curve (AUC), AIC and BIC were the basis for comparing models. Likelihood ratio test was also employed in comparing nested models.

$$Z = \frac{A_1 - A_2}{\sqrt{\left(se\left(A_1\right)\right)^2 + \left(se\left(A_2\right)^2 - 2 \cdot r \cdot se\left(A_1\right) \cdot se\left(A_2\right)\right.}} \tag{5}$$

$$\begin{cases} AIC = -2\log\left(\text{maximum likelihood}\right) + 2p & (6) \\ BIC = -2\log\left(\text{maximum likelihood}\right) + plogn & (7) \end{cases}$$

Where p is the number of model parameters.

3. Model Formulation

Equation 1 is a one parameter |Cox Proportional Hazard Model with unspecified shape parameter (absorbed by the baseline hazard $\psi_0(t)$ and scale parameter $\boldsymbol{\beta}_i$.

Consider the Cox regression model with observed and unobserved covariates with respective vectors to be:

$$\psi(t|\mathbf{X}, \mathbf{X}^u) = \psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i + \sum_{i=1}^{k}\omega_i X_i^u\right] \tag{8}$$

Generating (8) as a two parameter distribution with the unobserved covariates factored out as random effect (ℓ) assumed to account for heterogeneity and unobserved covariates, we obtain,

$$\psi(t|\ell, X) = \ell\,\psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i\right] \tag{9}$$

Introducing a nonnegative function into equation (9) given by:

$$\tau(x) = \exp\left(\sum_{i=1}^{n}\mathbf{X}\,d(t)\right) \tag{10}$$

We obtain equation (11) as follows;

$$\psi(t|\ell, X) = \ell\,\psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i d_i(t)\right] \tag{11}$$

For the special case where the null hypothesis is true, (11) reverts to the modified frailty Cox Proportional Hazard Model (9)

$$H_0 : \theta_1 = \theta_2 = \ldots\theta_n = 0$$

$$LR = -2\ln L_{Frailty\ CoxModel} = -\left(-2InL_{Modified\ \ frailty\ Cox}\right) \square\ \chi_n^2$$

If the time function in (11) is replaced with a lagged time effect we obtain:

$$\psi(t|\ell, X) = \ell\,\psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-\phi)\right] \tag{12}$$

In a special case where we specify the lag time as $\phi = 1$ equation 12 becomes:

$$\psi(t|\ell, X) = \ell\,\psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right] \tag{13}$$

We now compute the expectation of the gamma distribution with scale and shape parameters of α and λ

respectively,

$$f(x)=\frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)},\ x>0,\quad \alpha>0 \qquad (a)$$

$$E(x)=\int_0^{\infty} x\frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{\alpha}e^{-\lambda x}dx$$

$$But\ \int_0^{\infty}x^{\alpha}e^{-\lambda x}dx \Rightarrow gamma(\alpha+1,\lambda)$$

$$\Rightarrow\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{\alpha}e^{-\lambda x}dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}x^{\alpha}e^{-\lambda x}dx$$

$$=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}\left[\int_0^{\infty}\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}e^{-\lambda x}dx\right]^{1}$$

$$E(x)=\frac{\lambda\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}=\frac{\lambda\alpha}{\lambda^{\alpha+1}}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}=\frac{\lambda^{\alpha}}{\lambda^{\alpha+1}}\frac{\alpha!}{(\alpha-1)!}$$

$$E(x)=\frac{\lambda\alpha}{\lambda^{\alpha+1}}\frac{\alpha(\alpha-1)!}{(\alpha-1)!}=\frac{\alpha}{\lambda} \qquad (b)$$

Gamma distribution has no closed forms, and to ensure that our model is identifiable, we consider the restriction $\alpha=\lambda$

$$\Rightarrow \mathrm{E}(x)=\frac{\alpha}{\lambda}=\frac{\alpha}{\alpha}=1 \qquad (c)$$

With variance

$$Var(x)=\frac{\alpha}{\lambda^{2}}=\frac{\alpha}{\alpha^{2}}=\frac{1}{\alpha} \qquad (d)$$

When $\alpha=\lambda$ equation (a) simplifies to an exponential distribution (with a constant hazard function) yielding equation (e) as shown below:

$$Now Gamma\ (1,\lambda)\ \ becomes$$

$$\frac{\lambda e^{-\lambda x}}{\Gamma(1)}=\lambda e^{-\lambda x} \qquad \Rightarrow G(1,\lambda)=Exp(\lambda) \qquad (e)$$

Substituting equation (d) into (13) yields:

$$\psi(t|\lambda,X)=\psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i+\theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)-\lambda\ln\lambda\sum_{i=1}^{k}X_i\right] \qquad (14)$$

For ease of simplicity, we let $\gamma=\lambda\ln\lambda$ and obtain

$$\psi(t|\gamma, X) = \psi_0(t)\exp\left[\sum_{i=1}^{k}\boldsymbol{\beta}_i X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1) - \gamma_i\sum_{i=1}^{k}X_i\right] \quad (15)$$

We simplify (15) further to obtain

$$\psi(t|\gamma, X) = \psi_0(t)\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right] \quad (16)$$

Equation 16 represents our 3 parameter Extended Cox Model. The scale parameters $\boldsymbol{\beta}_i$ and ϕ_i controlling the dispersion of the probability distribution and the location parameter γ_i which moves the curve of the distribution from left to right. However, because our updated model is still semi-parametric with an undetermined baseline hazard, the shape parameter is absorbed by the baseline hazard; the model has no shape parameter.

4. Model Building Assumptions

The following assumptions were considered in model development:

i. Random effect parameter , ℓ was considered to have multiplicative effect on the baseline hazard.

ii. Nonnegative function , $\tau(x)$ was considered to have multiplicative effect on the baseline hazard.

iii. Frailty parameter γ was considered to follow a Gamma distribution.

iv. The expectation of the γ frailty is assumed to be 1 with a finite variance.

5. Statistical Properties of modified Cox Model

The modified Survival frailty model developed from equation (16) is

$$S(t|, \mathbf{X}, \lambda) = \exp\left(-\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]H(t)\right) \quad (17)$$

Applying the Laplace Transformation to (17) yields

$$S(t|, \mathbf{X}, \lambda) = \int_0^{\infty}\exp\left(-\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]H(t)\right)dF_{\ell}(x) \quad (18)$$

Equation (18) can be simplified as follows

$$S(t|, \mathbf{X}, \lambda) = \exp\left(-G(-\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_i + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]H(t)\right) \quad (19)$$

$$G(y) = -\ln\left(\int_0^{\infty}\exp(-uy)\,dF_{\ell}(\mathrm{u})\right)$$

Where ,

The developed modified Hazard ratio is given by:

$$\ln\frac{\psi(t|\lambda, X_1)}{\psi(t|\lambda, X_2)} = \ln\frac{\psi_0(t)\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_1 + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_1(t-1)\right]}{\psi_0(t)\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i - \gamma_i)X_2 + \theta_i\sum_{i=1}^{k}\phi_i\mathbf{X}_2(t-1)\right]} \quad (20)$$

Simplifying yields

$$\ln\frac{\psi(t|\lambda,X_1)}{\psi(t|\lambda,X_2)}=\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i-\gamma_i)(X_1-X_2)+\theta_i\sum_{i=1}^{k}\phi_i(t-1)(\mathbf{X}_1-\mathbf{X}_2)\right] \quad (21)$$

The derived hazard ratio permits the hazard ratio to change over time which grants greater flexibility than proportional hazards assumption.

Also, we can now ascertain whether differs significantly from zero or otherwise. This will enable us evaluate the proportional hazards assumption.

6. Parameter Estimation Of The Modified Cox PH Model

In general, for time-invariant covariates the partial likelihood from a Cox regression model is given as:

$$L=\prod_{i=1}^{k}\left[\frac{\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i-\gamma_i)X_i+\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]}{\sum_{j=1}^{n}Y_{ij}\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i-\gamma_i)X_i+_i\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]}\right]^{\delta_i} \quad (22)$$

We wish to maximize the log partial likelihood by taking log of both sides to obtain

$$\log PL=\log\left(\prod_{i=1}^{k}\left[\frac{\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i-\gamma_i)X_i+\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]}{\sum_{j=1}^{n}Y_{ij}\exp\left[\sum_{i=1}^{k}(\boldsymbol{\beta}_i-\gamma_i)X_i+\sum_{i=1}^{k}\phi_i\mathbf{X}_i(t-1)\right]}\right]^{\delta_i}\right) \quad (23)$$

$$\log PL=\sum_{i=1}^{n}\delta_i\left[\sum_{i=1}^{k}\log\exp[(\boldsymbol{\beta}_i-\gamma_i)X_i+\phi_i\mathbf{X}_i(t-1)]-\log\sum_{j=1}^{n}Y_{ij}\left[\left(\exp\left[(\boldsymbol{\beta}_i-\gamma_i)X_i+\phi_i\mathbf{X}_i(t-1)\right]\right)\right]\right] \quad (24)$$

Equation (24) represents the actual maximized function. The score equations for partial likelihood are written as:

The maximum partial likelihood estimates are found using the equation,

$$U(\beta)=\frac{\partial}{\partial\beta}l(\beta)\ ;\ U(\phi)=\frac{\partial}{\partial\phi}l(\phi)\ ;\ U(\gamma)=\frac{\partial}{\partial\gamma}l(\gamma)\ ;$$

$$U(\beta)=0\ ;U(\phi)=0\ ;U(\gamma)=0.$$

7. Comparing Developed Model with Existing Models

Table 1. Comparing Models by AIC, BIC and ROC

Model	AIC	BIC	Loglik (model)	OC (AUC)
Accelerated Failure Time (Exponential) Model	1965.032	2024.213	-967.516	0.642
Park et al., (2017) Model	1923.572	1933.329	-958.7861	0.663
Stratified Cox (Interraction) Model	1745.186	1864.272	-835.977	0.945
Modified Cox Frailty Model with lag effect	**1742.8**	**1765.566**	**-864.4002**	**0.964**
Modified Cox Frailty Model without lag effect	**1751.605**	**1790.633**	**-863.8027**	**0.812**

Test of no-interaction Assumption

To know which of the Stratified Cox Model is more appropriate statistically; that is, the no-interaction model or the interaction model, we must first look at the hazard function model for the interaction situation. The test

is a likelihood ratio (LR) test which compares log-likelihood statistics for the interaction model and the no-interaction model.

Null hypothesis: $\beta_{11} = \beta_{12} = \ldots = \beta_{17} = 0$ and $\beta_{21} = \beta_{22} = \ldots = \beta_{27} = 0$

$$LR \sim \chi^2_{14df} \text{ under } H_0 : \text{no interaction}$$

Chsq	df	p-value	LR
266.68	1	< 2.2e-16	133.35

Conclusion: Reject null hypothesis; interaction model is preferred.

The results of comparing our created models (Modified Cox and Frailty Cox) to current models such as Exponential AFT, Stratified Cox model, and [18]models are shown in Table 1. Our constructed models outperformed existing models on the basis of lowest AIC, BIC values, and greatest area under the ROC curve, as seen by the findings. Our enhanced model, which added a time function with lag effect and a frailty parameter, was clearly the best overall.

Table 2. Comparing Nested Models by Likelihood Ratio Test

Null Model	Larger Model	Likelihood Ratio	df	P-Value
Cox PH Model	Modified Cox with lag effect	101	1	<0.01
Park et a., (2017) Model	Modified Cox with lag effect	96.8	1	<0.01

We employed the likelihood ratio test as an additional performance test because the formulated models are nested models with the normal Cox model. Because the difference is statistically significant, our constructed model fits the data far better than the Null Model and so optimizes the likelihood function. The null hypothesis, which asserts that the Null model fits better, is rejected.

8. Testing for Influential Covariates

The index plots below the comparison of the magnitudes of the largest dfbeta values to the regression coefficients of the 14 covariates. The output below suggests that none of the observations is terribly influential individually, even though some of the dfbeta values for some covariates are large compared with the others. If the removal of the outliers from the data did not cause a change in the coefficient of the respective covariate

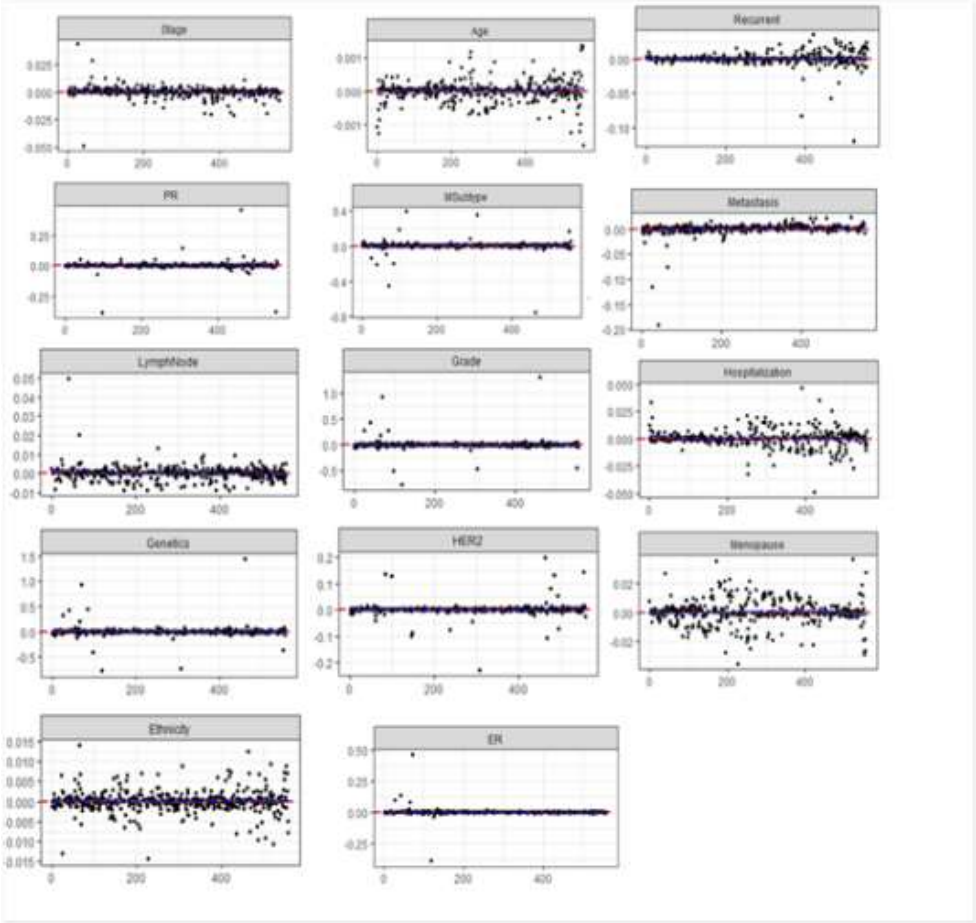

Figure1. dfbeta Test For Influential Covariates

9. Test Nonlinear Covariates

Nonlinearity is not an issue for categorical variables, so we only examine plots of martingale residuals and partial residuals against the continuous variable, Age. The Figure shows the plot of martingale residual on the y axis and the Age continuous covariate on the x-axis on the range of (-INF, +1). By inspection of the plots we can see after application of transformations of log and square root to the covariate that generally, linear model is a good fit for age ; an indication that the hazard ratios of the prospective model will be reliable when we assume that continuous covariate (Age) have a linear functional form. Expressed in another way, the LOESS curve is most reasonably linear (roughly flat) for Age variable with no transformation; hence satisfying the Cox Proportional hazard assumption. This is important to help know the functional form of Age covariate that should be included in the model, i.e either linearly by direct inclusion in the model or its transformation. We therefore assume that the continuous covariate (Age) have a linear form.

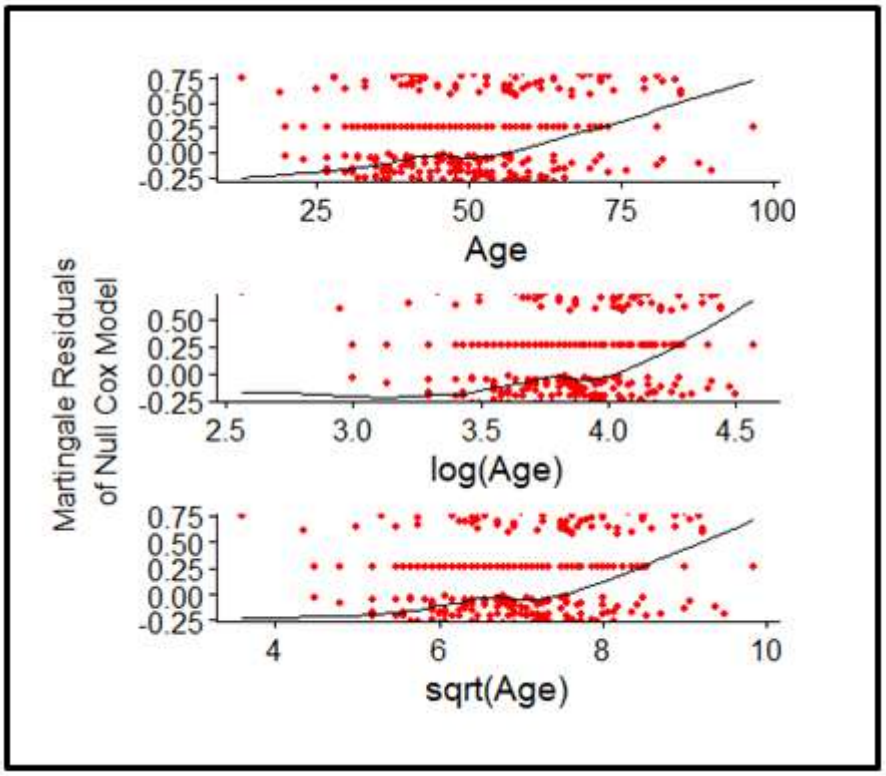

Figure2. Plot of Martingale Residuals by Age, with Loess Line (Age not in model)

Table3. Goodness of Fit Test of Proportional Hazard Assumption (No Competing Risk)

Covariate	chisq	df	p
Age	0.11	1	0.741
Recurrent	0.753	1	0.386
HER2	0.832	1	0.362
PR	4.031	1	**0.044**
ER	3.251	1	0.071
MSubtype	6.53	1	**0.011**
Grade	7.267	1	**0.007**
Stage	2.486	1	0.115
Metastasis	0.41	1	0.522
LymphNode	0.542	1	0.462
Menopause	0.319	1	0.572
Ethnicity	0.259	1	0.61
Hospitalization	1.922	1	0.166
Genetics	1.731	1	0.188
GLOBAL	22.358	14	0.072

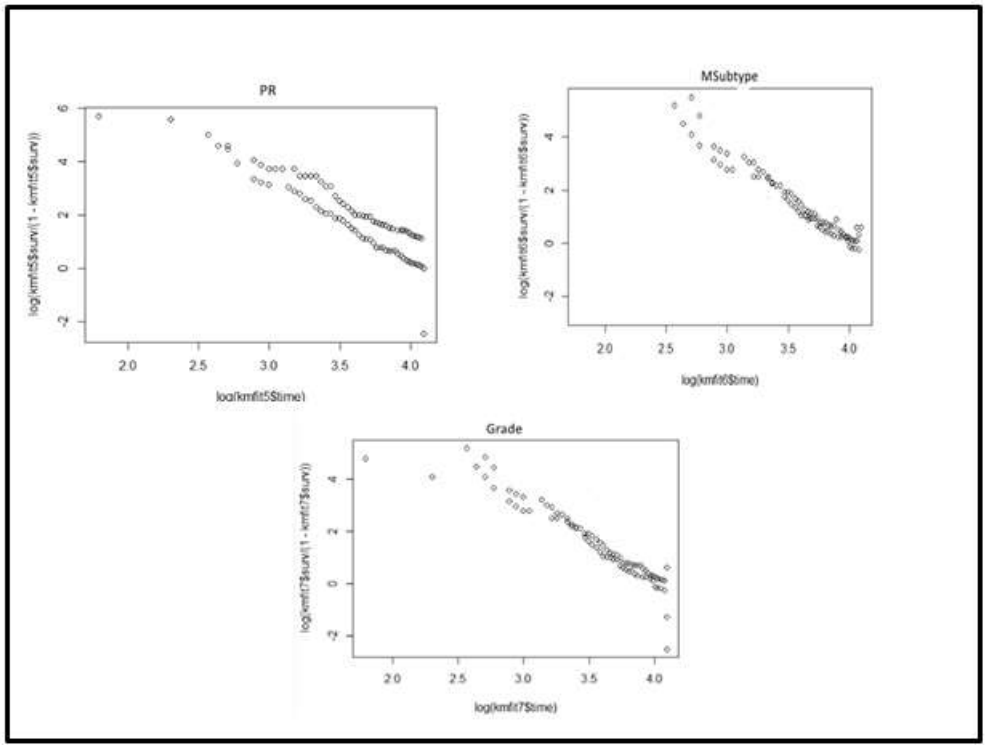

Figure 3. Graphical Test for Cox Proportional Hazard Assumption

Table4. Goodness of Fit Test of Proportional Hazard Assumption (Under Competing Risk)

Covariate	chisq	df	p
Age	0.00169	1	0.967
Recurrent	0.32053	1	0.571
HER2	0.00176	1	0.966
PR	5.16717	1	**0.023**
ER	3.77155	1	0.052
MSubtype	2.00089	1	0.157
Grade	5.63717	1	**0.018**
Stage	2.35698	1	0.125
Metastasis	0.28847	1	0.591
LymphNode	0.57144	1	0.45
Menopause	0.59849	1	0.439
Ethnicity	0.77946	1	0.377
Hospitalization	2.07502	1	0.15
Genetics	2.72288	1	0.099
GLOBAL	22.62211	14	0.067

Goodness of Fit (GOF) test for validating a proportional hazard assumption test between the observed and estimated survival function values; the resulting P value is a more objective method than a graphical method and Shoenfeld residual test.

It is evident from tables 3 and 4 that GOF test is affected by competing risk. If this assumption is ignored, from table 3 where competing risk in the data set is unaccounted for, it can be seen that there were only two significant time-dependent covariates compared to 3 significant time-dependent covariates under noncompeting risk data set. This means that when competing risk in the data set is ignored, a number of time-dependent covariates are wrongly identified as time-invariant covariates to be included in the model; which can lead to erroneous estimates and conclusions.

Table5. Test of Significance of for Competing and noncompeting Risk Ages

Groups	Statistic	P-value	df
1	0.899017	0.343	1
2	52.26348	<0.01	1

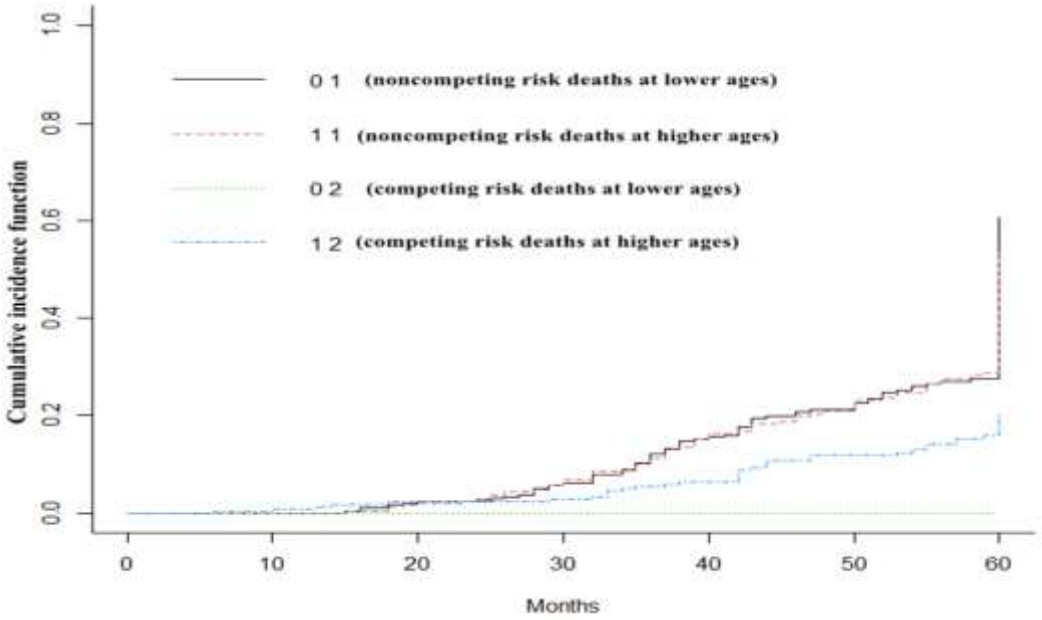

Figure4. Cumulative Incidence comparison of Competing and noncompeting Risk Ages

It's critical to evaluate the prognosis of breast cancer as well as the risk of dying from reasons other than breast cancer when analysing survival. From Figure 4 and Table 5, it is inferable that there is no significant difference between the mortalities of patients resulting from the BC disease with respect to lower and higher ages. Again, Figure4 shows that increased age and comorbidity are both linked to reduced survival. However, there was a significant difference in mortalities resulting from comorbidities (competing risk) in the lower and higher ages. This is a reasonable finding, since majority (about 77%) of BC mortality have been ascribed to postmenopausal age groups of women attributable to intense comorbidities at higher ages **[19].**

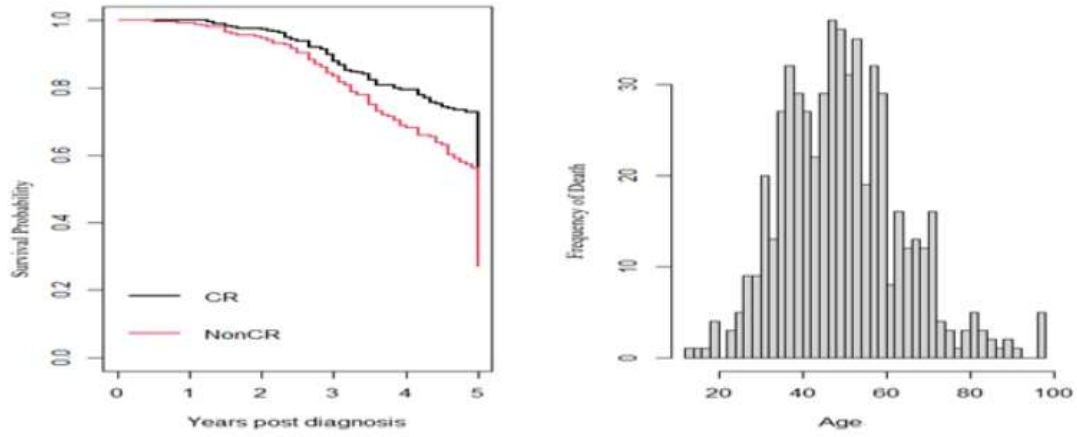

Figure5. Comparison of Survival Probabilities for Competing and Noncompeting Risk Patients

Noncompeting risk BC patients had poorer survival rates than competing risk BC patients, according to our research. While comorbidities in patients are often associated with higher risks than patients without comorbidities, this is conditional on certain conditions, according to **[20].**With regard to our study, we can explain that this conditional factor was identified in the noncompeting risk age groups over a five-year follow-up period because roughly 67 percent of BRCA 1, which causes the most aggressive Triple Negative molecular subtype, was found in the noncompeting risk age categories. Among the BC subtypes, triple negative has been linked to a poor prognosis and the greatest mortality rates, notably in Sub-Saharan African nations. Among our study, the competing risk was low (about 7%), but significant in those over the age of 57. According to our findings, anemia was a common source of comorbidity. Cancer patients frequently develop anemia, which has a substantial negative influence on their quality of life and overall prognosis **[21].**

Table6. Parameter Estimates of the Formulated Model

Predictors	Estimates	CI	P
tt(Grade)	0.98	0.94 –1.01	0.215
Age	1.01	0.98 –1.03	0.572
Recurrent	0.92	0.54 –1.57	0.759
HER2	2.84	1.86 –4.34	**<0.001**
tt(PR)	0.98	0.93 –1.02	0.342
ER	0.45	0.10 –2.00	0.293
tt(MSubtype)	1.00	0.99 –1.02	0.772
Stage	1.93	1.52 –2.46	**<0.001**
Metastasis	3.72	1.37 – 10.07	**<0.001**
LymphNode	1.46	1.20 –1.78	**<0.001**
Menopause	1.37	0.81 –2.32	0.245
Ethnicity	1.12	0.94 –1.33	0.222
Hospitalization	1.00	0.65 –1.54	0.998
Genetics	0.42	0.08 –2.30	0.315

Variance of random effect = 0.145, gamma (0.145) = 1, p<0.01

Our model identified stage at diagnosis, Metastasis status, lymph node involvement and HER2 overexpression as the significant factors that influence breast cancer survival in Ghana. The variance of random effect explains the quantum of variability among individuals across all treatments, and not within each group. Hence about 15% of all deaths are due to random effects. The estimated hazard ratio (HR) comparing risk of dying due to BC by low

and high staging of the disease is exp (1.93) = 6.89. Meaning patients at higher staging of BC (III and IV) are about 6.89 times at risk of dying from the disease than patients at lower staging (I and II) of the disease. Metastasis tells whether the cancer has spread to other parts of the body. The estimated HR with respect to metastasis is 3.72. To this end, BC patients who experience metastasis are a whopping exp (3.72) = 41.264 times at risk of dying from the disease than patients experiencing no metastasis. This is supported by [**22]** who concluded that metastasis is the most important undermining prognostic factor in breast cancer survival. Lymph nodes are classified as N1 (1 to 3 lymph nodes involved), N2 (4 to 9 lymph nodes involved), and N3 (more than 10 lymph nodes involved). From the table patients with Lymph nodes at N2 and beyond are exp(1.46) = 4.31 times at risk of dying from the disease than patients below this level. The HR comparing risk of dying due to BC by negative and positive HER2 status of the disease is exp (2.84) = 17.116. Meaning patients with positive HER2 status are about 17.116 times at risk of dying from the disease those of negative status. This finding is supported by [23] who found that HER2 amplification/overexpression is a marker of poor prognosis in breast cancer.

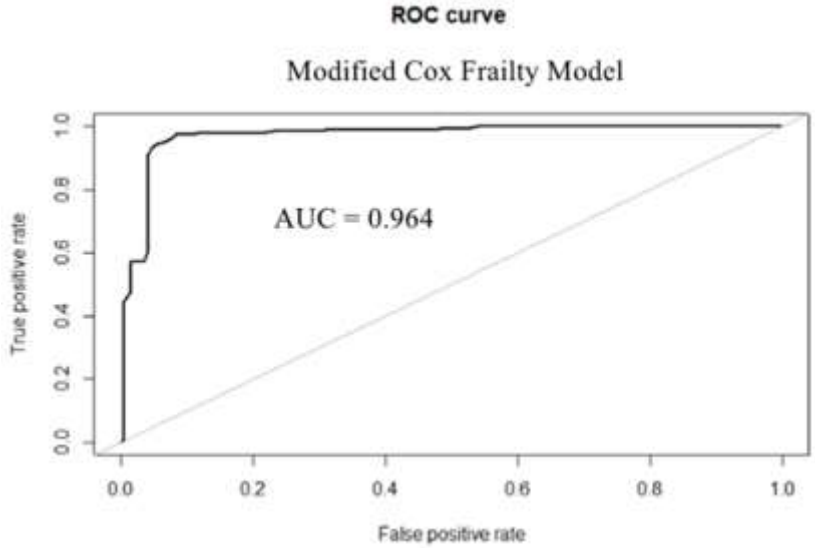

Figure 6. Reciever Operation Characteristic Curve for the Modified Cox Frailty Model

Our constructed model displayed strong prediction power (AUC = 0.964) after partitioning the survival data into 70% Training and 30% Validation sets, dichotomizing into thresholds, and computing the resulting sensitivity and specificity to predict clinical risk by stratifying BC patients into higher or lower risk categories of clinical importance.

10. Conclusion

In this paper, we focused on a joint frailty Cox Model with a time lagged function. In its basic form, the standard Cox Model is a homogenous model that is influenced by competing risk in the survival data. Again, the proportional hazard assumption precludes the inclusion of time-varying components in the model, resulting in erroneous and biased conclusions. We added a function of time with a time-lag effect and a frailty parameter to the existing Cox model to address these problems. We discovered that the extended Cox Model, which only included the Frailty parameter and did not include a time function, also outperformed the normal Cox Model. Frailty component was revealed to be a significant determinant of survival, accounting for around 15% of all fatalities attributable to heterogeneity and unobserved variables and confounders. As a result, we conclude that our suggested models captures heterogeneity and time-dependent covariates in a way that outperforms several existing models, notably Stratified Cox, Exponential AFT model and [18]model.

11. Ethical considerations

The authors have paid close attention to ethical considerations.

12. Future Research

The assumption that the Gamma distribution reduces to an exponential distribution provided a critical basis for modeling the frailty parameter. Future research could focus on modeling the scale and shape factors to determine the ideal combination for a more accurate estimation. While such a strategy may produce a better model, parameter estimate using Maximum Likelihood estimation may be difficult, if not impossible, due to the Gamma distribution's lack of a closed form. Alternative ways of estimating the parameters of such a future model could be considered in future studies.

Reference

[1] Berg M, Nielsen MH, Pedersen AN, Andersen K, Glavicic V, Jakobsen EH, Jensen I, Josipovic M, Lorenzen EL, Nielsen HM, Stenbygaard L, Thomsen MS, Vallentin S, Zimmermann S, Offersen BV (2013), Danish Breast Cancer Cooperative Group Radiotherapy Committee. Delineation of target volumes and organs at risk in adjuvant radiotherapy of early breast cancer: national guidelines and contouring atlas by the Danish Breast Cancer Cooperative Group. Acta Oncol. 2013 May;52(4):703-10. doi: 10.3109/0284186X.2013.765064. Epub 2013 Feb 19. PMID: 23421926.

[2] Qua yson, S.E., Wiredu, E.K (2014), Breast Cancer in Accra, Ghana, Journal of Medical and Biomedical Sciences, 3(3):21-26.

[3] Gnant M, Harbeck N, Thomssen C. St. Gallen/Vienna 2017: A Brief Summary of the Consensus Discussion about Escalation and De-Escalation of Primary Breast Cancer Treatment. Breast Care (Basel). 2017 May;12(2):102-107. doi: 10.1159/000475698. Epub 2017 Apr 26. PMID: 28559767; PMCID: PMC5447163.

[4] Omae K, Komori O, Eguchi S.(2020), Quasi-linear Cox proportional hazards model with cross-L1 penalty. BMC Medical Research Methodology (2020) 20:182

[5] Scheike, T.H. (2004), Time-Varying Effects in Survival Analysis, Handbook of Statistics, Vol.23, Elsevier

[6] Wu L. (2010), Mixed Effects Models for Complex Datahapman & Hall/CRC, Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742.

[7] Balan T.A, Putter H. A tutorial on frailty models. *Statistical Methods in Medical Research*. 2020;29(11):3424-3454.

[8] Yazdani, A., Yaseri, M. (2019), Investigation of Prognostic Factors of Survival in Breast Cancer Using a Frailty Model: A multicentre study, Greast Cancer Basic and Clinical Research ,Vol. 13:1-10.

[9] Haerting J. (2007). Frailty Models in Survival Analysis, Aus dem Institut f¨ur Medizinische Epidemiologie, Biometrie und Informatik.

[10] Sarkar,K.Chowdhury,R.Dasgupta,A.(2017), *Analysis of Survival Data: Challenges and Algorithm-Based Model Selection*,J Clin of Diagn Res. *11*(6), LC14-LC20. https://www.doi.org/10.7860/JCDR/2017/21903/10019

[11] Marlies Noordzij, Karen Leffondré, Karlijn J. van Stralen, Carmine Zoccali, Friedo W. Dekker, Kitty J. Jager(2013), When do we need competing risks methods for survival analysis in nephrology?, *Nephrology Dialysis Transplantation*, Volume 28, Issue 11, November 2013, Pages 2670–2677,

[12] Pham H.M., Kafle R.C. (2017), Competing Risks analysis of African American Breast Cancer Patients, Advances in Breast Cancer research, Vol 6, 28-41.

[13] Scheike, Thomas H., and Mei-Jie Zhang. “Analyzing competing risk data using the R timereg package.” Journal of statistical software 38.2 (2011).

[14] Glas, N. A. Mandy, K. Vandenbroucke, J.P (2016), Performing Survival Analyses in the presence of

Competing Risks: A clinical Example in Older Breast Cancer patients, JNCI Natl Cancer Inst, 2016,vol. 108, No.5.

[15] Arani M.D, Abadi A, Yavari A, Bashiri Y, Liley M abd Bajdik C (2018), Evaluation of risk factors in patients with breast cancer III and IV: Comparison of Cox and Fine-Grey competing risk regression models, Biomedical Research and Therapy ,Vol 5 (2), 2022-2033.

[16] Rulli E, Ghilotti F, Biagioli E, Porcu L, Marabese M, D'Incalci M, Bellocco R, Torri V. Assessment of proportional hazard assumption in aggregate data: a systematic review on statistical methodology in clinical trials using time-to-event endpoint. Br J Cancer. 2018 Dec;119(12):1456-1463. doi: 10.1038/s41416-018-0302-8. Epub 2018 Nov 13. PMID: 30420618; PMCID: PMC6288087.

[17]Jatoi I., Bandos H., Jeong J, Anderson W. (2015), Time-Varying Effects of Breast Cancer Adjuvant Systemic Therapy, *JNCI J Natl Cancer Inst* (2015) 108(1): djv304.

[18] Park, K., Qiu, P. (2017). *Evaluation of the treatment time-lag effect for survival data. Lifetime Data Analysis, 24(2), 310–327.* doi:10.1007/s10985-017-9390-7

[19] Yancik R, Wesley MN, Ries LAG, Havlik RJ, Edwards BK, Yates JW. Effect of Age and Comorbidity in Postmenopausal Breast Cancer Patients Aged 55 Years and Older. *JAMA.* 2001;285(7):885–892. doi:10.1001/jama.285.7.885

[20] Derks MGM, van de Velde CJH, Giardiello D, Seynaeve C, Putter H, Nortier JWR, Dirix LY, Bastiaannet E, Portielje JEA, Liefers GJ. Impact of Comorbidities and Age on Cause-Specific Mortality in Postmenopausal Patients with Breast Cancer. Oncologist. 2019 Jul;24(7):e467-e474. doi: 10.1634/theoncologist.2018-0010. Epub 2019 Jan 3. PMID: 30606886; PMCID: PMC6656441

[21] Busti, F., Marchi, G., Ugolini, S., Castagna, A., & Girelli, D. (2018). Anemia and Iron Deficiency in Cancer Patients: Role of Iron Replacement Therapy. *Pharmaceuticals (Basel, Switzerland)*, *11*(4), 94. https://doi.org/10.3390/ph11040094

[22] Karimi A, Delpisheh A, Sayehmiri K. (2016), Application of accelerated failure time models for breast cancer patients' survival in Kurdistan Province of Iran. J Cancer Res Ther. 2016 Jul-Sep;12(3):1184-1188. doi: 10.4103/0973-1482.168966. PMID: 28054533.

[23] Ménard S, Tagliabue E, Campiglio M, Pupa SM. Role of HER2 gene overexpression in breast carcinoma. *Journal of Cellular Physiology.* 2000;182(2):150–162