PUBLIC HEALTH & PRIMARY CARE | RESEARCH ARTICLE

Joint modeling of longitudinal change in tumor cell level and time to death of breast cancer patients: In case of Ayder comprehensive specialized Hospital Tigray, Ethiopia

Brsat Tesfay¹, Tewodros Getinet² and Endeshaw Assefa Derso³*

Abstract: Breast cancer is the major public health problem throughout the world and it results in serious physical damages and death. This work proposes the use of joint model to study breast cancer in patients of Ayder Hospital. The primary motivation is to contribute to the understanding of the tumor cell progression of breast cancer, within Ayder Hospital, using a joint model that takes into account a possible existence of a serial correlation structure within a same subject observations from September 2015 till December 2018. The general aim of this study was to investigate the risk of longitudinal change in tumor cell level on time to death due to breast cancer among breast cancer patients. Hospital-based retrospective cohort study was conducted among breast cancer patients. A joint model of longitudinal and time to death model was used to determine the risk of longitudinal change in tumor cell level on time to death due to breast cancer patients. These were used by using JM package in R version. Results from joint models, showed that the longitudinal Tumor cell progression was significantly associated with the survival probability of these patients(estimated association parameter(a) in the joint model is 0.84 with corresponding (95% CI: 2.28,2.37). A comparison between parameter estimates obtained in this joint model and independent survival and longitudinal analysis lead us to conclude that independent analysis brings up bias parameter

ABOUT THE AUTHOR

The main research topic of our group is Breast cancer, a major public health problem in the developing countries. We investigate risk factors and pathophysiological mechanisms, which cause or are associated with Breast cancer. Additionally, we work on improving the present Breast cancer diagnostics. The basis for all these research work is Joint modeling of survival and longitudinal analysis of Breast cancer. Thus, we perform studies intending to investigate joint modeling of survival _ longitudinal analysis.

PUBLIC INTEREST STATEMENT

This study was about Breast cancer in Ayder referral hospital _ Ethiopia. Breast cancer destroys the lives of many young individuals in developing countries. As it is known that the main predictive factors for Breast cancer patients are more health variables for this disease, data about its pathophysiological variables are important for interpretation in disease. The study presents data of the normal cohort of the Ayder, Registry. In many cases, due to a lack of awareness for women from health workers practices, these diseases claim lots of lives of the community who should not have otherwise died. With regular research and dissemination, among other interventions, it is very possible to contain these diseases. More researches are still encouraged in this area and also in another hospital.
estimates. There is a strong association between the progression change in log(TCL) and risk of mortality due to breast cancer.

Subjects: Environmental health; Epidemiology; Health Education and Promotion

Keywords: breast cancer; tumor cell; longitudinal analysis; survival analysis; joint model

1. Background
Cancer is a group of diseases that causes by the uncontrolled growth and spread of abnormal cells anywhere in the body, if the spread is not controlled, it can result in death. Most types of cancer cells eventually from a lump or mass called a tumor, and are named after the part of the body where the tumor originates (Diabate et al., 2018).

Breast cancer is a type of malignant tumor which it starts in the cells of the breast tissue that is made up of glands for milk production, called lobules, and the ducts that connect the lobules to the nipple. The majority of breast cancer cases are classified as either invasive or noninvasive. The invasive type of breast cancer is distributed in the whole body; but not the noninvasive (Abay et al., 2018). Cancer constitutes enormous burdens in more and less economically developed countries. Studies also suggested that over the years, the burden has shifted to less developed countries, which currently account for about 57% of cases and 65% of cancer deaths worldwide (Torre et al., 2015).

Breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death among African women. In Africa from 92,600 cases, 50,000 deaths in 2008 (Jemal et al., 2012). In sub-Saharan Africa, breast cancer is one of the non-communicable diseases, and the most commonly diagnosed cancer in women (Brinton et al., 2014). It is estimated that around 10,000 Ethiopian women and men 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 (Lemlem et al., 2013).

The average age of diagnosis of breast cancers among African women tends to be young, with estimates that a majority of cancers develop among women 50 years or younger; a considerably younger age than seen in Caucasian populations (Sighoko et al., 2013). Genetic or environmental factors or interplay of the two may be additional factors involved (Brinton et al., 2014). Age has significant effect on women breast cancer patients and it indicated that 1.35 times (95% CI; 1.12, 2.22) more likely died in ages 65 years and older.

Educational status and marital status are prognostic factors for mortality of women with breast cancer (Lan et al., 2013). A study conducted from Braga’s hospital diagnosed with a malignant tumor by using linear-mixed model shows that tumor stage III and IV have an increasing of the tumor cell values as compared to tumor stage 0, I, II (Borges et al., 2015). Covariates such as time, age, tumor stage and tumor size have a positive significant effect on the linear mean progression of tumor cell level. On the other hand, tumor stage (III or IV versus 0 or I or II) have a significant effect on patients’ survival. Patients with Tumor stage III or IV were 3.52 times more likely to die than patients with Tumor stage 0 or I or II (Borges et al., 2015). Tumor size, lymph node metastasis status, and tumor extension had significant effect on breast cancer survival (Desantis et al., 2013). Stage and pathology type had significant effect on breast cancer survival (Kantelhardt et al., 2014).

High proportion of breast cancer morbidity and mortality was observed in age category of 15 years of age and above in both men and women in Tigray, Northern Ethiopia. Overall breast cancer mortality was 2.3% during the study period (Ajemu et al., 2019). Translating Raman spectroscopy technology by using the RESpect probe as a potential point-of-care screening instrument has the potential to change the paradigm of screening for cancer as an initial step to
determine when a definitive tissue biopsy would be necessary (Agsa‌‌da-Garcia et al., 2019). The profile of a cancer patient includes summarization and visualization of the results of WES and RNAseq analysis (specific variants and significantly expressed genes, respectively) and the clinical profile, integration/comparison of these results and a prediction regarding the disease trajectory (Kosvyra et al., 2019).

Joint models reduce bias in the estimates of the overall treatment effect, that is, the treatment effect on survival and the longitudinal marker (J. G. Ibrahim et al., 2010). An excellent general review article on joint modeling of longitudinal and survival data is given in Tsiatis and Davidian (Tsiatis & Davidian, 2004). JG Ibrahim et al. (2001) also give an overview of joint modeling methods in their book. Joint models for longitudinal and survival data in which the survival component of the model is a cure rate model are also useful in cancer research. Law et al. (2002), Brown (2003) consider such models for cancer clinical trials.

Even though the number of people living with breast cancer is increasing from year to year; it has emerged as one of the rapidly increasing non-communicable disease and a major public health challenge in developing countries like Ethiopia (Kantelhardt et al., 2014) with a consequence of Chronicity, disability, and death. Therefore, we formulate a method that can estimate the effect of baseline covariates on Tumor cell level and time to death jointly for breast cancer patients and its complication using suggested risk factors in literature, mainly by including Socio-demographic variables (Age, Gender, Educational status, Marital status, Residence, and Alcohol drinking) and Clinical factors (Treatment used, Pathology type, stage, Baseline Tumor cell count, stage of metastatic, Baseline tumor size, Time/month). To find the estimate, we want to use a joint model of longitudinal and time to death statistical models to determine the risk of longitudinal change in tumor cell level on time to death due to breast cancer among breast cancer patients by using JM package in R version.

The general aim of this study was to investigate the risk of longitudinal change in tumor cell level on time to death due to breast cancer among breast cancer patients from September 2015 till December 2018 at Ayder Comprehensive Specialized Hospital. Specifically 1) To estimate the effect of baseline covariates on tumor cell level and time to death jointly 2) To observe the association between tumor cell level and time to death due to breast cancer among breast cancer patients 3) To demonstrate the advantage of joint model analysis techniques over separate survival and longitudinal analysis to the data.

1.1. Significance of the study
Any concerned legal body may use the output of the research to allocate proper human and material resources for breast cancer health care institutions. On the other hand, health care institution could provide fast and effective services for those women with high-risk factors of breast cancer. This research increases the education and awareness for individual, society, the government and any legal body about the risk factor of women with breast cancer and their life expectancy.

2. Methodology

2.1. Study area and period
The study was conducted at Ayder comprehensive specialized hospital from September 2015 till December 2018. Tigray region, Northern part of Ethiopia.

Tigray region has 18 public hospitals and 170 health centers with a total population of 4,316,988.
2.2. Data source
Source of data for this study was the patient's chart from the data record office at Ayder comprehensive specialized hospital. The patients' chart was used to extract the necessary information from different BC recording formats.

2.3. Inclusion and exclusion criteria

2.3.1. Inclusion criteria
The study includes all BC patients and started Anti-cancer treatment since September 2015 and have at least two follow up until December 2018 in ACSH.

2.3.2. Exclusion criteria
This study excludes those patients who had incomplete variable registration cards, a person who died by other accidents, and those patients who started anti-cancer treatment from other healthcare institutions.

2.4. Study design
A facility-based retrospective cohort study design was conduct in ACSH. The study was focused on investigating the determinant factors that affect time to death among breast cancer patients with joint modeling of longitudinal change in tumor cell level at ACSH.

2.5. Study population
The study population consists of all BC patients who were under follow up and started the anti-cancer treatment and have at least two follow up until December 2018. From the total-registered BC patients at the hospital, only 186 of them were included in the study.

2.6. Method of data collection
The data were collected from patient charts based on those variables to be considered in this study. Both the longitudinal and survival data were extracted from the patient's chart which contains demographic and clinical variables of BC patients under treatment follow-up.

2.7. Operational definition
Censoring: When we say Observations are called censored if the information about their survival time is incomplete.

Survival time: the period that a patient Stays in life after starting the anti-cancer treatment.

Time: the period which patients visiting to the hospital during the follow up every 3 months.

Minimum follow-up time: the period that a patient who had two longitudinal measurements after started the anti-cancer treatment.

2.8. Study variables
Outcome variables:
- Survival case

The main outcome variable for survival data analysis is time to death due to breast cancer or censored (in a month).

- Longitudinal case

The main outcome variable for longitudinal data analysis is tumor cell level measured in ng/mL every 3 month.
Joint case

In this case, the two outcome variables are longitudinal change in tumor cell level and time to death due to breast cancer or censored (in month) among breast cancer patients. Figure 1

2.9. Independent variables
Several predictors were considering in this study for both survival and longitudinal cases.

2.10. Data processing and analysis
The analysis consists of exploratory data analysis and three different models namely; a linear mixed model for the longitudinal data, the Cox proportional-hazards model for the time to-death data, and a joint modeling of them altogether. The data were analyzed by using STATA12 to presents the categorical descriptive statistics, R version 3.6.1 to analyze the Cox, LMM models including the graphics. A joint model of longitudinal and time to death model was used to determine the risk of longitudinal change in tumor cell level on time to death due to breast cancer among breast cancer patients by using JM package in R version 3.6.1.

2.11. Data exploration
Exploratory data analysis (EDA) was conducted to investigate various associations, structures, and patterns exhibited in the data set. This consists of obtaining the summary statistics, individual profile plots, mean structure, correlation structure, and variance structure plots were obtained to gain some insights into the data. Since our data are unbalanced and had not equal observation time we had used smoothing techniques that highlight the typical response as a function of an explanatory variable without reliance on specific parametric models.
2.12. Statistical models
We have used survival sub models for length of time from the start date of Anti-cancer treatment until the date of death or censored (measured in months), longitudinal sub-models for tumor cell level and Joint survival longitudinal for true-unobserved Longitudinal covariate.

2.13. Data quality management
The supervisor & principal investigator had performed immediate supervision on a daily basis, & every activity of questionnaire was checked for completeness & consistency.

3. Results

3.1. Descriptive statistics
The descriptive statistics for categorical variables in described in Tables 1 and Tables 2 as. A total of 186 breast cancer patients were included in this study. Out of these, 22 (11.83%) were dead. And also, 119 (64%) patients their residence in rural. Among 186 breast cancer patients, 47 (97.92%), 42 (97.67%), 53 (82.81%), and 22 (70.97%) were censored for them in stages stage-1, stage-2, stage-3, and stage-4, respectively. When we consider based on death for educational status of the patients, 17 (18.09%) of patients are illiterate, 3 (6.38%) of them are primary educated, 1 (14.29%) of them are secondary educated and 1 (2.63%) are tertiary educated. Similarly, when we consider the covariate type of treatments, we observe that 121 (65.1%) of the patient uses surgery treatment and 33 (17.7%) chemotherapy whereas the remaining used both surgery and chemotherapy treatment. When we take the pathology type of BC patients, 94 (50.5%) of breast cancer patients are ductal carcinoma in-situ 71 (38.2%) lobular carcinoma in-situ & 8 (4.3%) were ductal invasive carcinoma. But the rest of them were lobular invasive carcinoma.

From the following Table 3 we can conclude that the median age, B tumor cell, B tumor size & B_HGL of BC patients were given as 47,246,12 &12.8, respectively. and also, the median survival time of BC patient was 34.50 months.

| Table 1. Summary statistics of socio-demographic variables for breast cancer patients |
|-------------------------------------------------|-----------------|------------------|
| **Variables**                                  | **Survival status** | **Total(%)**   |
|                                                 | **Dead**         | **Censored**    |
| Residence                                      | 5 (7.46%)        | 62 (92.54%)     | 67 (100%)   |
| Urban                                           |                  |                 |            |
| Rural                                           | 17 (14.3%)       | 102 (85.7%)     | 119 (100%) |
| Total                                           | 22 (11.83%)      | 164 (88.17%)    | 186 (100%) |
| Marital status                                 | 3 (5.56%)        | 51 (94.44%)     | 54 (100%)  |
| Single                                          |                  |                 |            |
| Married                                         | 14 (15.91%)      | 74 (84.09%)     | 88 (100%)  |
| Divorced                                        | 3 (12%)          | 22 (88%)        | 25 (100%)  |
| Widowed                                         | 2 (10.53%)       | 17 (89.47%)     | 19 (100%)  |
| Total                                           | 22 (11.83%)      | 164 (88.17%)    | 186 (100%) |
| Educational status                              | 17 (18.09%)      | 77 (81.91%)     | 94 (100%)  |
| Illiterate                                      |                  |                 |            |
| Primary                                         | 3 (6.38%)        | 44 (93.62%)     | 47 (100%)  |
| Secondary                                       | 1 (14.29%)       | 6 (85.71%)      | 7 (100%)   |
| Tertiary                                        | 1 (2.63%)        | 37 (97.37%)     | 38 (100%)  |
| Total                                           | 22 (11.83%)      | 164 (88.17%)    | 186 (100%) |
3.2. Separate analysis of survival data

3.2.1. Kaplan-Meier estimates and log-rank tests
We compared the survival time of patients by using Kaplan Meier and log-rank test. When we compared by using educational status, the patient whose educational status was illiterate category had shortest survival probability than primary, secondary, and tertiary categories.

The survival probability of the BC patients who came from urban is longer than the patients who came from rural areas.

Similarly, BC patients who used the treatment both surgery & chemotherapy has smallest survival probability than the patients, who used the treatment surgery and chemotherapy.

| Variables          | Survival status | Total(%) |
|--------------------|-----------------|----------|
| Pathology type     | Dead            | Censored |
| DCIS               | 10(10.64%)      | 84(89.36%) |
| LCIS               | 7(9.86%)        | 64(90.14%) |
| DCC                | 1(12.5%)        | 7(87.5%)  |
| LIC                | 4(30.77%)       | 9(69.22%) |
| Total              | 22(11.83%)      | 164(88.17%) |
| Stage              | 1(2.08%)        | 47(97.92%) |
| Stage-1            | 1(2.33%)        | 42(97.67%) |
| Stage-2            | 11(17.19%)      | 53(82.81%) |
| Stage-3            | 9(29.03%)       | 22(70.97%) |
| Stage-4            | 22(11.83%)      | 164(88.17%) |
| Metastatic         | 12(11.32%)      | 94(88.68%) |
| Distant            |                |          |
| Regional           | 1(10%)          | 9(90%)   |
| Locally            | 9(12.86%)       | 61(87.14%) |
| Total              | 22(11.83%)      | 164(88.17%) |
| Treatment          | 16(13.22%)      | 105(86.78%) |
| Surgery            |                |          |
| Chemotherapy       | 0(0.0%)         | 33(100%) |
| Surgery and chemotherapy | 6(18.75%) | 26(81.25%) |
| Total              | 22(11.83%)      | 164(88.17%) |

Table 2. Summary statistics of clinical variables for breast cancer patients

| Variables          | Min.  | 1st Qu. | Median | Mean   | 3rd Qu. | Max.  |
|--------------------|-------|---------|--------|--------|---------|-------|
| Age                | 23.0  | 36.0    | 47.0   | 47.6   | 58.0    | 82.0  |
| B tumor cell       | 141.0 | 217.0   | 246.0  | 247.6  | 276.0   | 418.0 |
| B tumor size       | 2.00  | 4.      | 12.00  | 13.51  | 22.0    | 37.00 |
| B_HGL              | 10.0  | 12.4    | 12.8   | 13.4   | 15.4    | 16.2  |
| Survival time      | 6.00  | 27.00   | 34.50  | 30.95  | 37.00   | 39.00 |
The BC patients whose pathology type was LCIS have longest survival probability than the patients who were DCIS, DIC, and LIC. The patients whose stage was stage-4 has smallest survival probability than the patients, whose stage was stage-1, stage-2, and stage-3.

Based on Table 4 results of long-rank test, Educational Status, Treatment, Pathology type, and Residence were statistically significant differences in the risk of mortality among BC patients at 25% level of significance. However, there is no significant difference in the risk of mortality among BC patients.

| Variable        | Chisq | Df | P—Value |
|-----------------|-------|----|---------|
| Metastatic      | 0.1   | 2  | 0.9     |
| Educational Status | 3.5   | 3  | 0.01    |
| Marital Status  | 4     | 3  | 0.3     |
| Treatment       | 7.2   | 2  | 0.03    |
| Pathology Type  | 8.3   | 3  | 0.04    |
| Stage           | 4.1   | 3  | 0.3     |
| Residence       | 2.8   | 1  | 0.01    |
between categories of Marital Status, Stage and metastatic at 25% level of significance. Because this Table was important for selection of candidate variables to pass for multivariate analysis.

Based on Table 5 variable Age, B tumor size & B tumor cell are statistically significant differences on Univariate analysis of Cox regression model at 25% level of significance. This helps for selection of variables in the final cox regression model.

3.3. Variable selection and Cox PH assumption
To determine the variables to be included in the multivariable Cox survival model we used variables whose p-value <0.25 for Tables 4 and Tables 5 on the above. Educational status, Age, B tumor size, B tumor cell, pathology type, treatment, and residence were the candidate variables in the multivariate cox regression model. However, marital status, stage, metastatic & B-HGL were not included in the multivariate cox regression model.

The proportional-hazards assumption asserts that the hazard ratios are constant overtime. That means the risk of failure must be the same no matter how long subjects have been followed. In order to test this assumption, GLOBAL test was used. The result of the GLOBAL test is displayed on Table 6. From the output, it is clear to see that the p-value of GLOBAL is insignificant. This indicates that the PH assumption is not violated. Therefore, we can assume the proportionality assumption is met.

| Variable     | Df | P—Value | HR   | 95%CI for HR |
|--------------|----|---------|------|-------------|
| Age          | 1  | 0.05    | 1.03 | (0.99,1.067) |
| B_HGL        | 1  | 0.69    | 1.04 | (0.823, 1.338) |
| B tumor size | 1  | 0.04    | 1.045| (1.002,1.009) |
| B tumor cell | 1  | 0.045   | 1.009| (1.0,1.02)   |

| Variables                | rho  | chisq   | Sig  |
|--------------------------|------|---------|------|
| Age                      | -0.0346 | 8.92e-03 | 0.9247 |
| Educational status       |      |         |      |
| Primary                  | 0.049 | 0.117   | 0.73 |
| Secondary                | 0.299 | 2.57    | 0.109|
| Tertiary                 | 0.257 | 1.47    | 0.225|
| Treatment                |      |         |      |
| Surgery                  |      | -       | -    |
| Chemotherapy             | 0.355 | 0.075   | 0.86 |
| Surgery $ chemotherapy   | -0.032 | 0.043   | 0.83 |
| Pathology type           |      |         |      |
| LCIS                     | -0.102 | 0.238   | 0.626|
| DIC                      | -0.176 | 1.15    | 0.283|
| LIC                      | -0.251 | 3.6     | 0.06 |
| Baseline tumor size      | -0.0938 | 0.136   | 0.7125|
| Baseline tumor cell      | 0.095  | 20.27   | 0.6  |
| Residence                |      |         |      |
| Urban                    | 0.0107 | 1.62e-01 | 0.6679 |
| GLOBAL                   | NA    | 15.2    | 0.81 |
3.3.1. Cox proportional hazards model
After checking the assumption of proportional hazard, the survival data were analyzed based on Cox proportional hazard model. The results are presented below. From the result of multivariable analysis presented in Table 7 show that the estimated AHR, 95% CI, and p-value of the covariates in the final cox regression model for age, educational status, residence, Baseline Tumor size & Pathology type (LIC) were found to be statistically significant effect at 5% level of significance on risk of mortality due to breast cancer. The remaining variables which were found to be non-significant at 5% of significance level in the multivariable analysis of the final cox regression model.

3.4. Separate analysis of longitudinal data
The Q-Q plots was used to check the normality of the longitudinal measures of tumor cell in ng/ml. Figure 2 represents that the Q-Q plot for tumor cell measures of original and logarithm transformed data. The normal quintile-quintile plot indicates that the log tumor cell level is normally distributed since the line concentrates around the normal line and is an approximately straight line through the origin. Thus, the analysis of this study was used the logarithm transformed of tumor cell-level data.

3.5. Exploratory Data Analysis (EDA)
Exploratory data analysis was conducted to investigate various associations, structures and patterns exhibited in the data set. In addition, the individual profile plots, mean structure plots and variance structure plots were obtained to gain some insights of the data (Verbeke and Molenberghs, 2009).

| Variables          | AHR    | 95%CI      | Sig  |
|--------------------|--------|------------|------|
| Age                | 1.06   | (1.005,1.13)| 0.032|
| Educational stat   |        |            |      |
| Illiterate         |        |            |      |
| Primary            | 0.043  | (0.00041,0.443)| 0.008|
| Secondary          | 0.033  | (0.0012,0.9) | 0.043|
| Tertiary           | 0.013  | (0.00046,0.371)| 0.011|
| Pathology type     |        |            |      |
| DCIS               |        |            |      |
| LCIS               | 7.96   | (0.245,25.76)| 0.24 |
| DCIS               | 3.188  | (0.98, 7.8) | 0.06 |
| LIC                | 6.925  | (1.82, 8.5) | 0.02 |
| Residence          |        |            |      |
| Rural              |        |            |      |
| Urban              | 0.213  | (0.05,0.92) | 0.038|
| B tumor size       | 1.172  | (1.06,1.29) | 0.041|
| B tumor cell       | 1.005  | (0.995,1.016)| 0.294|
| Treatment          |        |            |      |
| Surgery            |        |            |      |
| Chemotherapy       | 0.0038 | (0.00045,12.2)| 0.61 |
| Surgery + chemotherapy | 0.132 | (0.01577,1.16)| 0.062|
| AIC                | 178.05 |            |      |
Figure 3 visualizes the pattern change of the overall individual plots log tumor cell measurements of patients over time. It demonstrates the variability (within and between patients) in log tumor cell measurements of breast cancer patients. Since the measurements were not equally spaced across the different subjects. It shows that there is large variability between log tumor cell of patients that their measure is too extreme in some extent and deviated through time. This figure 4 indicates that there is variability in tumor cells within & between BC patients in our LMM.

3.5.1. Exploring the mean structure
We explore the mean plot of log tumor cell over time to describe the longitudinal change in log tumor cell level of breast cancer patients. Figure revealed that there seems slight increase in average of log tumor cell and whereas

3.5.2. Exploring the variance structure
Similarly, Figure 5 assures that there was increase in the mean of log tumor cell-expecting constant increase at follow-up time. The line plot of log tumor cell over time is used to observe the variability of the data. Indicated that there is log tumor cell variability at the first visit than at visit-2. The log tumor cell variability increases slightly from visit-3 to visit-4 and decreases very fast from visit-4 to visit-5, whereas slightly increases to visit-6 and slightly decreases to visit-7. Then, the pattern shows that it increases fastly except decreases for the last follow up time figure 6.
3.6. Variance-covariance structure
Table 8 below indicates that unstructured variance-covariance has smaller AIC (−3633.268) & BIC (−3133.898) value which is to mean that it was better in explaining the model under study. Therefore, unstructured variance covariance was used in identifying the correlation structure.

3.7. Linear mixed model
After exploring, the data examine whether the assumption of heterogeneous within-subject variance for the log tumor cell is supported and also identify the random effects (random intercepts, random slope, and random intercept random slope) to be included in the model.

Here from the following three model examine the great reduction in the AIC of RI RS for the model incorporating subject-specific variances is an evident that subject-specific log tumor cell variances must be considered in the analysis. Also, the random effect of table tells us there is
subject-specific variation. Hence, it supports the assumption of heterogeneous variance for the repeated log tumor cell measurements.
3.8. Univariate analysis of longitudinal data
From the result of the univariate analysis of linear-mixed models, Time, Age, residence, educational level, baseline Tumor cell level, pathology type, stage, metastatic, and treatment are important factors that can affect Log (TCL) of BC patients and should be included in the multivariable analysis. However, B-HGL, baseline Tumor size, and marital status of the disease were not statistically significant for Log (TCL) at 25% level of significance. Therefore, these variables are not included in the multivariable analysis of LMM.

3.9. Multi-variate analysis of longitudinal data
After made a transformation of the response variable tumor cell level, the most important variables were selected using purpose full variable selection method. The variables which were significant in the univariate analysis at 25% level of significance is all included in the multivariable analysis for the response variable of log (TCL). Therefore, from the result of multivariable analysis (Table 11) age, educational status, stage-3, stage-4, baseline Tumor cell level, time*age, time*stage-3 and time*stage-4 were statistically significant factors that affect change in log (TCL) of BC.

| Variables | Values | 95%CI         | p-values |
|-----------|--------|---------------|----------|
| Intercept | 2.29   | (2.18,2.4)    | 0.0000   |
| Time      | 0.067  | (-0.0147,0.28)| 0.54     |
| Age       | 0.009  | (0.000058,0.022)| 0.0405  |
| Educational status | | | |
| Illiterate | -      | -             | -        |
| Primary   | -0.058 | (-0.075,-0.026) | 0.024 |
| Secondary | -0.047 | (-0.1048,-0.01) | 0.0213 |
| Tertiary  | -0.079 | (-0.0947,0.05)  | 0.038   |
| Residence | | | |
| Rural     | -      | -             | -        |
| Urban     | -0.016 | (-0.0417,0.0106)| 0.245 |
| Baseline tumor cell | 0.089 | (0.0339,0.13608) | 0.0073 |
| Pathology type | | | |
| DCIS      | -      | -             | -        |
| LCIS      | 0.019  | (-0.048,0.079) | 0.64     |
| DIC       | -0.018 | (-0.093,0.056) | 0.62     |
| LIC       | -0.03  | (-1.1,0.043)  | 0.39     |
| Metastatic | | | |
| Distant   | 0.039  | -0.043,0.116  | 0.37     |
| Regionally | 0.003 | -0.091,0.084  | 0.94     |
| Locally   | -0.003 | -0.091,0.084  | 0.94     |
| Treatment | | | |
| Surgery   | -      | -             | -        |
| Chemotherapy | 0.41  | (-0.005,0.542) | 0.056    |
| Chemo & surgery | 0.371 | (-0.020,0.47)  | 0.08     |
| Stage     | | | |
| Stage-1   | -      | -             | -        |

(Continued)
3.10. Data analysis for joint longitudinal _ Survival analysis

The result of the joint model of longitudinal change of log Tumor cell level and time to death (Table 12) shows that age, educational status, baseline Tumor cell level, stage-3, stage-4, Time*age, Time*stage-3 and Time*stage-4 were significant factors of log (Tumor cell level) and age, educational status, pathology (DIC), pathology (LIC) and Residence were significant factors that affect the mortality rate of BC patients. The association of longitudinal change in log Tumor cell level and the mortality rate was significant, and this shows that there is a strong positive relationship (Association (HR) = 2.321 with 95% CI (2.26,3.091), p-value = <0.0001) between the two sub models. It indicated that as the number of tumor cells increases the mortality rate also increases.

3.11. Interpretation of joint model

From the result of separate and joint models, the joint models had narrow confidence interval as compared with the separate models. This indicates that the joint model is more precise than the separate models. Also, the statistical significance of the association parameter is evidence that the joint model is better than the separate models (Borges et al., 2015). The estimates of the parameters of the separate and joint models are not identical. The estimates of the association parameters in the joint models are significantly different from zero, providing strong evidence of association between the two sub-models for the joint model. The estimate of the association parameter due to trend of log Tumor cell level is positive (HR = 2.32) with 95% CI (2.28,2.37), p-value = <0.0001), which indicates that the log Tumor cell level is positively associated with the hazard of death of BC patients on follow up anti-cancer treatment. This implies that an increase of the log Tumor cell level in BC patients on follow-up treatment significantly increases the risk of death of those BC patients. Hence, the joint model of longitudinal and time to death model is better.

3.12. Discussion

We had one longitudinal and one survival sub-models. The fitted models were expressed as follows.

**Longitudinal sub-model of Tumor cell:**
### Table 12. Joint modeling of longitudinal and survival data analysis

#### Longitudinal process

| Fixed effect             | EXP(β) | 95%CI       | p-value |
|--------------------------|--------|-------------|---------|
| Intercept                | 11.13  | (2.33, 2.54)| <0.0001|
| Time                     | 1.02   | (<0.0034, 0.03)| 0.42   |
| Age                      | 1.012  | (0.0036, 0.02)| 0.031  |
| Educational status       |        |             |         |
| Illiterate               |        |             |         |
| Primary                  | 0.953  | (<0.055, 0.036)| 0.024  |
| Secondary                | 0.964  | (<0.1048, 0.01)| 0.0213 |
| Tertiary                 | 0.934  | (<0.0847, 0.04)| 0.038  |
| Baseline tumor cell      | 1.111  | (0.0928, 0.164)| 0.0031 |
| Stage                    |        |             |         |
| Stage-1                  |        |             |         |
| Stage-2                  | 0.995  | (<0.0057, 0.12)| 0.24   |
| Stage-3                  | 1.54   | (0.320, 0.562)| 0.02   |
| Stage-4                  | 2.1    | (0.5, 0.81)| 0.015  |
| Age∗time                 | 0.994  | (<0.008, 0.005)| 0.024  |
| Time∗stage               |        |             |         |
| Time∗stage-1             |        |             |         |
| Time∗stage-2             | 0.67   | (<0.65, 0.16)| 0.003  |
| Time∗stage-3             | 0.575  | (<0.799, 0.31)| <0.0001|
| Random effects           | StdDev | Corr        |         |
| Intercept                | 0.0423 |             |         |
| Time                     | 0.0102 |             |         |
| Residual                 | 0.062  |             |         |
| (intercept, Time)        |        | 0.108       |         |

#### Event process

| Event process | AHR  | 95%CI       | p-value |
|---------------|------|-------------|---------|
| Age           | 1.093| (1.015, 1.16)| 0.025  |
| Educational status |      |             |         |
| Illiterate    |      |             |         |
| Primary       | 0.064| (0.051, 0.086)| 0.0031 |
| Secondary     | 0.083| (0.064, 0.089)| 0.033  |
| Tertiary      | 0.028| (0.016, 0.029)| 0.001  |
| Pathology type|      |             |         |
| DCIS          |      |             |         |
| LCIS          | 8.43 | (1.852, 12.62)| 0.12   |
| DIC           | 4.01 | (3.8, 5.7)| 0.034  |
| LIC           | 7.422| (5.74, 8.2)| 0.001  |
| Residence     |      |             |         |
| Rural         |      |             |         |
| Urban         | 0.361| (0.12, 0.47)| 0.013  |
| Association   | 2.32 | (2.28, 2.37)| <0.0001|
| AIC           |      | -3812.910   |         |
\[ E(\log \text{Tumor}(b)) = 2.41 + 0.012 \times \text{age} - 0.048 \times \text{primary} - 0.036 \times \text{secondary} - 0.068 \times \text{tertiary} + 0.105 \times \text{baseline Tumor cell} + 0.43 \times \text{stage} - 3 + 0.741 \times \text{stage} - 4 - 0.0057 \times \text{Time} \times \text{age} - 0.405 \times \text{Time} \times \text{stage} - 3 - 0.554 \times \text{Time} \times \text{stage} - 4 \]

**Survival sub model:**

\[
\log(h(t)) = 0.089 \times \text{age} - 2.75 \times \text{primary} - 2.49 \times \text{secondary} - 3.58 \times \text{tertiary} + 1.4 \times \text{DIC} + 2 \times \text{LIC} - 1.02 \times \text{Urban} + 0.84 \times \alpha
\]

Where \( \alpha \) measures the effect of \( m(t) \) in the survival process.

Baseline age had a significant positive association with Tumor cell level (Borges et al., 2015). Similarly, the average of the log TCL increases as the age patient increases. From the result of longitudinal sub-model of log(TCL), as the baseline age of patients increases by 1 year the average log (TCL) also increases by 1.012 (p-value < 0.031) by adjusting other variables. The progression changes in log tumor cell level were decreased by 5%, 4%, and 7% for Female BC patients whose educational status was primary, secondary, and tertiary, respectively, as compared to whose educational status was illiterate by controlling other independent variables.

Clinical stage has a significant effect on the progression change of log TCL of breast cancer patients such that the more advanced stage is a more aggressive of log tumor cell level. The progression changes in log tumor level for stage III and stage IV patients had increased by 1.54 and 2.1 as compared to stage I patients and this difference is statistically significant since the 95% confidence interval did not include zero by controlling the other variables constant. This indicates that the latter stage at diagnosis is associated with the higher mean change of log tumor cell level. Such finding was similar to the study conducted from South Africa by (Dickens et al., 2014). And also another studies were conducted by (Borges et al., 2015) and (Borges et al., 2015) show that the late stage at diagnosis contributes to increase in the value of log TCL.

Baseline tumor cell level has been found to have a significant impact on the progression change of log tumor cell level. The progression change of log tumor cell level has increased by 1.11 when baseline tumor cell level is increased by one unit by controlling other independent variables. The interaction effect stage by time had a significant progression change on patients’ log tumor cell level. That is patients with stage III by visit time and stage IV by time interaction decreased the progression change of log tumor cell by 33% and 43%, respectively, as compared to patients with stage I by time interaction. Age by time interaction have a significant progression change on log TCL such that increasing the age by a visit time interaction had a decreasing progression change about 0.6% by controlling the other covariates.

Baseline age had a significant positive association with the hazard of death (Rezaianzadeh et al., 2009). Our finding also indicates that mortality rate of breast cancer patient increases 1.09 times when age increased by 1 year by controlling other independent variables. The other significantly important variable for the survival rate of the patient is educational status. The hazard rate of patient whose educational status primary, secondary & tertiary education was decreased by 93.6, 91.7, and 97.2%, respectively, when we compared with those patients, whose educational status was illiterate. Similarly, this result agreed with the research from Vietnam on Survival probability and prognostic factors for breast cancer by (Lan et al., 2013).

In this study, we found that pathology type has positive association with the risk of mortality, the Invasive pathology type at diagnosis is associated with higher risk of mortality than those in-situ. The hazard rate of patients who had pathology DIC and pathology LIC were 4.01 and 7.4 times more than those patients who had pathology DCIS, respectively. Residence is an important predictor for the survival time of breast cancer patients. The mortality rate of breast cancer patient...
whose residence was from rural side had shorter probability of surviving as compared to those patients from urban side. The hazard rate of patients whose residence was from urban is reduced by 63.9% than those patients from rural side, which shows that in Tigray regional state breast cancer patients were highly affected than their urban counterparts; this survival difference is because they reside further from treatment centers, they presented at later stages and were less compliant with treatment. This result is similar to the study conducted from South Africa by (Dickens et al., 2014).

Joint models for longitudinal and survival data are particularly relevant to breast cancer and observational studies in which longitudinal biomarkers (eg, circulating tumor cells, immune response to a vaccine, and quality-of-life measurements) may be highly associated with time to event. This result is similar to the study conducted at the University of North Carolina by J. G. Ibrahim et al. (2010) and study conducted from Portugal by Borges et al. (2015). This research also shows that there is strong positive association between the progression change in log(TCL) and risk of mortality due to breast cancer. In particular, a unit increase in the value of log(TCL) corresponding to exp (0.84) = 2.32-fold increase (95% CI: (2.28, 2.37)) in the risk for death at the same time point.

4. Conclusion
The result indicates that there is strong positive association between log(TCL) and survival time of BC patients. According to our finding, Age, educational status, pathology(DIC), pathology(LIC) and Urban were important baseline covariates that have a significant effect on time to death due to breast cancer among breast cancer patients. On the other hand, Age, educational status, baseline tumor cell level, stage-3, stage-4, time*age, Time*stage-3 and Time*stage-4 were important baseline covariates which have a significant association with progression change of TCL.

From our finding, we concluded that the relationship between log (tumor cell level) and the hazard for death was positively significant. Thus, a patient with higher tumor trend is less likely to survive. The results of both the separate and joint analyses are consistent. However, the use of a joint analysis compared to independent models adjusted for the correlation between the responses, which indicates that more adequate and efficient inferences can be made using joint model estimates. This means that joint modeling can benefit for the analyses of correlated data and an assumption of association between the two processes in a joint model of breast cancer data is necessary.

Abbreviations
AIC: Akaike Information Criterion; BC: Breast Cancer; BIC: Bayesian Information Criterion; B-HGL: Baseline hemoglobin level; CI: Confidence interval; DCIS: Ductal carcinoma in-situ; DIC: Ductal invasive carcinoma; HLM: health management information system; AHR: Adjusted hazard ratio; LIC: Lobular carcinoma in-situ; LIC: Lobular invasive carcinoma; LMM: Linear Mixed Model; PH: Proportional Hazard; TCL: Tumor cell level; TS: Tumor size; WHO: World Health Organization.

Acknowledgements
We wish to extend our thanks for all individuals involved in this study from

Funding
The authors received no direct funding for this research.

Author details
Bsrat Tesfay1
E-mail: bsrattesfaygu@gmail.com
Tewodros Getinet2
E-mail: tedomanchu@gmail.com
Endeshaw Assefa Derse1
E-mail: endussi@gmail.com

1 College of Natural and Computational Science, Department of Statistics, Mekelle University, Mek’ele, Ethiopia.
2 St. Paul’s Hospital Millennium Medical College, Addis Ababa, Ethiopia.
3 College of Natural and Computational Science, Department of Statistics, University of Gondar, Gonder, Ethiopia.

Data availability statement
The datasets analyzed during the current study were available from the corresponding author on reasonable request.

Ethical considerations
All procedures performed in studies involving human participants were in accordance with the ethical standards of University of Gondar institution and the University research committee declarations with comparable ethical standards for confidentiality.

Competing interests
The authors declare they have no competing interests.

Citation information
Cite this article as: Joint modeling of longitudinal change in tumor cell level and time to death of breast cancer patients: In case of Ayder comprehensive specialized Hospital Tigray, Ethiopia, Bsrat Tesfay, Tewodros Getinet & Endeshaw Assefa Derse, Cogent Medicine (2021), 8: 1874090.
associated with reproductive life. The Breast, 22, 828–835. https://doi.org/10.1016/j.breast.2013.02.011
Simsek, F. (2000). Five year survival analysis of patients with clinical stages I and IIA breast cancer who received initial treatment at North Carolina hospitals. Center for Health Informatics and Statistics.
Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians, 65(2), 87–108. https://doi.org/10.3322/caac.21262
Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. Stat Sin, 14, 809–834.