scientific reports

OPEN

Modelling of the time to death of breast cancer patients at Hiwot Fana Specialized University Hospital

Malkitu Tasfa, Kasahun Takele[✉] & Million Wesenu

Breast cancer is the most common cause of cancer death and is a frequently diagnosed cancer among women worldwide. It is becoming a challenging health condition in Ethiopia with a high rate of morbidity and mortality. The main aim of this study was to model the time to death in breast cancer patients at Hiwot Fana Specialized University Hospital. A retrospective cohort study was carried out from April 1st, 2020, to April 1st, 2023, and 296 women were included in the study. We used nonparametric methods and Bayesian accelerated failure time models (with Laplace approximation) to identify risk factors and choose a model fitting breast cancer patient data. Model comparison was performed using the marginal likelihood, deviance information criterion and Watanabe Akaike information criterion. From the total of 296 patients in the study, 56 (18.9%) died. The estimated median survival time was 33 months. The log-rank test showed that age group, stage, alcohol consumption, smoking habit, and comorbidity were potential risk factors associated with the time to death in breast cancer patients at the 5% level of significance. The Bayesian Weibull accelerated failure time model was found to be the best fitted model for predicting the survival time of patients with minimum DIC (520.39) and WAIC (521.59) values. The final Bayesian Weibull AFT model with the integrated nested Laplace approximation estimation technique revealed that age group, stage, alcohol consumption, smoking habit, and comorbidity were significantly associated with the time to death in breast cancer patients. Individuals older than 65 years, with stage IV disease, drinking alcohol, smoking cigarettes and having comorbidities had shortened survival times in patients with breast cancer. Hence, Hiwot Fana Specialized University Hospital and related bodies should work on awareness creation to reduce smoking habits and alcohol use as well as give due attention to elderly and stage IV breast cancer patients during intervention.

Cancer is a noncommunicable disease and a major cause of death worldwide[1]. Breast cancer is the most common cause of cancer death and the most frequently diagnosed cancer among women worldwide[2]. In 2020, 2.3 million women were diagnosed with breast cancer, and 685,000 deaths were reported globally. As of the end of 2020, 7.8 million women were alive and diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer[3]. According to the Global Burden of Cancer (GLOBOCAN) report in 2021, an estimated 2,261,419 new cases were recorded, and approximately 684,996 deaths were registered worldwide due to breast cancer, which is ranked next to lung cancer[4]. Despite the fact that there are more incidences of breast cancer worldwide, late-stage presentation and delayed diagnosis are frequent issues, especially in low- and middle-income countries[5].

Furthermore, in Africa, breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death among African women. The mortality rate was 17 per 100,000 people, 168,690 new cases were diagnosed, 74,072 deaths were diagnosed, and the age-standardized incidence rate was 37.9 per hundred thousand people in 2018[6]. According to the 2020 GLOBOCAN data, 186,598 breast cancer patients were diagnosed in Africa, with 85,787 related deaths. The mortality and incidence rate of breast cancer are highest in Africa, especially in sub-Saharan African countries. In sub-Saharan Africa, breast cancer is a noncommunicable disease and is the most commonly diagnosed cancer in women. Approximately 627,000 breast cancer-related

Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dhawa, Ethiopia. ✉email: kastake10@gmail.com

deaths were recorded in 2018, with the majority occurring in sub-Saharan Africa, where approximately 15% of all cancer-related deaths occurred[7]. In East Africa, especially Ethiopia, there is a lack of knowledge and resources available, making it difficult to detect breast cancer and treat it, which has increased breast cancer mortality[8] .

In Ethiopia, breast cancer is the major cause of death. According to the World Health Organization (WHO) estimation, the age-standardized incidence of breast cancer was 12,956, and the mortality rate was 25 per 100,000 women[9]. In 2018, the estimated incidence of breast cancer in Ethiopia was 13,987, with a crude incidence rate of 28.2 per 100,000 people, and breast cancer accounts for 33% of all cancer cases among women[10]. According to[11], Ethiopia is one of the sub-Saharan African countries with the highest incidences of new cases and deaths from breast cancer. Breast cancer is the most common cancer and constitutes 33% of all cancers in women and 23% of all cancers in Ethiopia; moreover, the expected national age-standardized incidence rate for females is approximately 43 per 100,000 people[12]. Approximately 10,000 Ethiopian women are estimated to have breast cancer, with thousands of additional cases unreported, as women living in rural areas often seek treatment from traditional healers before seeking help from the government health system[13].

Despite rising breast cancer rates, understanding factors affecting survival is crucial for treatment and risk awareness. While prior studies used non-parametric and classical methods[10,11,14–22], these have limitations in predicting survival time or comparing survival functions. This study addresses these limitations by employing a Bayesian accelerated failure time model with integrated nested Laplace approximation. This approach offers more accurate estimates, faster computation, and allows for better interpretation of survival data in breast cancer patients at Hiwot Fana Specialized University Hospital, Eastern Ethiopia.

Methods
Data source
The data were collected from patient medical records at the Hiwot Fana Specialized University Hospital. All female breast cancer patients who provided complete information, including study variables of interest on the registration card, were eligible for the study. The study period was between 1st April 2020 and 1st April 2023, 2020, with a three-year follow-up time. The starting time is the time at which the breast cancer patients are diagnosed. The end time was the time (in months) at which the event occurred, when the breast cancer patients died, were lost to follow-up before the completion of the study, or completed the study duration without any events (censored observations). Breast cancer patients who were transferred out, lost to follow-up, died from another cause, or did not develop the disease at the end of the follow-up period were considered censored. This study was conducted following the relevant guidelines and regulations of the Hospital. Retrospective data collection was approved by the College of Health and Medical Science, Haramaya University.

Outcome variable
The response variable of this study was the time to death for breast cancer patients. Time is measured in months and is the difference between the time of diagnosis and the time of death (event of interest) or censoring. The status variable is coded as 0 for censored data and 1 for death.

Risk factors
The risk factors were age, marital status, residence, alcohol consumption, body mass index, stage, tumor size, treatment taken, oral contraception, breastfeeding, smoking habits, and comorbidity.

Methods of data analysis
In this study survival analysis techniques were employed to examine the data. The Kaplan-Meier method, a nonparametric approach, was utilized to estimate survival curves. To compare survival experiences between groups, the log-rank test was conducted. The median survival time used as a central tendency measure for the data. The Kaplan−Meier (KM) method, proposed by[23], is a nonparametric estimator of survival function that is used to describe the survival of patients both graphically and numerically. It uses information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. The log-rank test is the most well-known and widely used test statistic. The log-rank test compares the survival time between groups and can be thought of as a test of whether the survival curves are identical (overlapping). The log-rank test is most powerful for the case where the hazard ratio remains constant over time[24]. This is called the proportional hazards case, and this test is used to check the proportionality of the categorical covariates in addition to other methods.

Bayesian survival analysis
Survival analysis is normally carried out with the help of nonparametric methods and semiparametric and parametric methods (the parametric PH model and accelerated failure time model). The parametric AFT model provides an alternative to the PH model for the statistical modelling of survival data. Under AFT models, we measure the direct effect of the explanatory variables on survival time instead of hazard, as we do in the PH model[25]. In the presence of complex censoring schemes, survival models are generally quite difficult to fit. The Bayesian approach to survival analysis may overcome this by using MCMC techniques and other numerical integration methods, such as INLA[26].

In the frequentist paradigm, the calculation of variance estimates can be complicated to derive (or even impracticable) since it is based on asymptotic arguments and therefore requires large sample sizes. Bayesian analysis, however, does not require large samples and can typically be used in smaller datasets without losing power while maintaining accuracy[27]. We can also highlight that the Bayesian approach incorporates prior knowledge in a natural way, whereas the frequentist approach does not. The advantages of the Bayesian approach include the ability to produce more accurate parameter estimates and greater convergence[28].

Bayesian accelerated failure time models

The accelerated failure time model is an alternative to the Cox-PH model for the analysis of survival time data. The AFT model is obtained by regressing the logarithm of the survival time over the covariates, and the effect of the explanatory variables on the survival time is directly measured. This characteristic allows for easier interpretation of the results because the parameters measure the effect of the covariates on survival time. The survival function of an individual with covariate X at time t, in the accelerated failure time models, is the same as the baseline survival function at time $t * \exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}\right)$

where β_1, β_2, β_p are the coefficients of the regression models. Thus, the survival function of time t is

$$S(t|X) = S_0\left(t * \exp\left(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}\right)\right), \ for \ all \ t \geq 0 \tag{1}$$

The effect of the covariates on the survival function is that the time scale is changed by a factor $exp\left(\beta' X\right)$, called the accelerated factor. The AFT model treats the logarithm of survival time as the response variable and includes an error term that is assumed to follow a particular distribution. The AFT model can be written as follows:

$$logT_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \sigma \epsilon_i \tag{2}$$

This model shows the log-linear representation of the AFT model for the ith individual, where $logT_i$ is the log-transformed survival time, X_1, X_2, ..., X_p are explanatory variables with coefficients β_1, β_2, ..., β_p, ϵ_i represents residual or unexplained variation in the log-transformed survival times, and μ and σ are the intercept and scale parameters, respectively.

Since many Bayesian studies in reality are conducted using parametric survival models, parametric AFT survival models play a crucial role in Bayesian survival analysis. Parametric modelling provides simple modelling and analysis approaches[26]. Some of the standard parametric AFT models are exponential, Weibull, log-normal, and log-logistic[18]. The parametric AFT models considered in this study were exponential, Weibull, log-logistic, and lognormal models. The exponential distribution is a special case of the Weibull distribution. If the scale parameter of the exponential model is estimated rather than set to one, the Weibull model can be fitted.

Bayesian exponential AFT model

The exponential model is the most fundamental parametric model in survival analysis[26]. Suppose we have independent identically distributed survival times $t = (t_1, t_2, \ldots t_n)$', each having an exponential distribution with parameter ; denote the censoring indicators by $\delta = (\delta_1, \delta_2, \ldots \delta_n)$', where $\delta_i = 0$ if t_i is right censored and $\delta_i = 1$ if t_i is a failure time. Let $f(t_i/\lambda) = \lambda e^{-\lambda t_i}$ denote the density for t_i and $S(t_i/\lambda) = e^{-\lambda t_i}$ denote the survival function. We build a regression model by introducing covariates through and write $\lambda_i = \phi(x_i'\beta)$, where x' is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of regression coefficients, $\phi(.)$ is a known function, and D = (n, t, X, δ) denotes the observed data for the regression model.

$$L(\beta/D) = \prod_{i=1}^{n} f\left(\frac{t_i}{\lambda}\right)^{\delta_i} S\left(\frac{t_i}{\lambda}\right)^{1-\delta_i}$$

$$= exp\left\{\sum_{i=1}^{n} \delta_i x_i'\beta\right\} exp\left\{-\sum_{i=1}^{n} t_i exp(x_i'\beta)\right\}$$

Suppose we specify a normal prior for β with mean μ_0 and variance σ_0^2. Then, the prior distribution is $\pi(\beta/\mu_0, \sigma_0) = \frac{1}{2\pi\sigma_0^2} exp\left\{-\frac{1}{2\sigma_0^2}(\beta - \mu_0)^2\right\}$

Then, the posterior distribution of β is given by

$$\pi(\beta/D) \propto L(\beta/D)\pi(\beta/\mu_0, \sigma_0), \tag{3}$$

where $\pi(\beta/\mu_0, \sigma_0)$ is the multivariate normal density with mean μ_0 and variance σ_0^2. The posterior in Eq. (3) does not have a closed form in general.

Bayesian Weibull AFT model

The Weibull model is perhaps the most widely used parametric survival model and a popular generalization of the exponential model with two parameters. Suppose we have independent identically distributed survival times $t = (t_1, t_2, \ldots t_n)$', each having a Weibull distribution, denoted by $\omega(\alpha, \gamma)$. It is often more convenient to write the model in terms of the parameterization $\lambda = log(\gamma)$, leading to $f(t_i/\alpha, \lambda) = \alpha t_i^{\alpha-1} e^{\left(\lambda - e^\lambda t_i^\alpha\right)}$, the survival function being given by $S(t_i/\alpha, \lambda) = e^{\left(-e^\lambda t_i^\alpha\right)}$. We can write the likelihood function of (α, λ) as

$$L(\alpha, \lambda/D) = \prod_{i=1}^{n} f(t_i/\alpha, \lambda)^{\delta_i} S(t_i/\alpha, \lambda)^{1-\delta_i}$$

$$= \alpha^d exp\left\{d\lambda + \sum_{i=1}^{n} (\delta_i(\alpha - 1))log(t_i) - exp(\lambda) t_i^\alpha\right\}$$

where $d = \sum_{i=1}^{n}\delta_i$ and δ_i are indicator variables taking the value 1 if t_i is the failure time and 0 if t_i is right censored. To build the Weibull regression model, we introduce covariates through λ and write $\lambda_i = x_i'\beta$. Here, x_i' is a px1 vector of covariates, and β is a px1 vector of regression coefficients. A normal prior to parameter (μ_0, σ_0^2) is assumed for β and gamma prior to parameter (α_0, k_0)[26]. The distribution of the normal prior for β is $\pi(\beta/\mu_0, \sigma_0^2) = \frac{1}{2\pi\sigma_0^2}exp\left\{-\frac{1}{2\sigma_0^2}(\beta - \mu_0)^2\right\}$, and the distribution of the gamma prior for α is $\pi(\alpha/\alpha_0, k_0) = \alpha^{\alpha_0-1}exp(-k_0\alpha)$.

The joint posterior distribution of (α, λ) is given by

$$\pi(\beta, \alpha/D) \propto \alpha^{\alpha_0+d-1}exp\left\{\sum_{i=1}^{n}\delta_i x_i'\beta + \delta_i(\alpha-1)\log(t_i)) - (t_i^{\alpha}\exp(x_i'\beta))\right.$$
$$\left. -k_0\alpha - \frac{1}{2}(\beta-\mu_0)\frac{1}{\sigma_0^2}(\beta-\mu_0)\right\} \quad (4)$$

where D = (n, t, X, δ) denotes the observed data for the regression model.

Bayesian log-logistic AFT model
The log-logistic distribution is among the parametric survival models where the hazard rate initially increases and then decreases. Suppose we have independent identically distributed survival times $t = (t_1, t_{2,...} t_n)'$, each having a log-logistic distribution, denoted by $L(\alpha, \lambda)$, with density $f(t_i/\alpha, \lambda) = \frac{\alpha\lambda^{\alpha}t_i^{\alpha-1}}{(t_i^{\alpha}+\lambda^{\alpha})^2}$, for $\alpha > 0, \lambda > 0$ and t > 0; the survival function is given by

$S(t_i/\alpha, \lambda) = \frac{\lambda^{\alpha}}{(t_i^{\alpha}+\lambda^{\alpha})}$ for t > 0. We can write the likelihood function of (α, λ) as

$$L(\alpha, \lambda/D) = \prod_{i=1}^{n} f(t_i/\alpha, \lambda)^{\delta_i} S(t_i/\alpha, \lambda)^{1-\delta_i}$$
$$= \alpha^d \lambda^{n\alpha} t_i^{(\alpha-1)d}(t_i^{\alpha}+\lambda^{\alpha})^{-d}$$

where $d = \sum_{i=1}^{n}\delta_i$ and δ_i are indicator variables taking the value 1 if t_i is the failure time and 0 if t_i is right censored. To build the log-logistic regression model, we introduce covariates through λ and write $\lambda_i = x_i'\beta$. Here, x_i' is a px1 vector of covariates, and β is a px1 vector of regression coefficients. We assume a normal prior for β and a gamma prior with parameters (α_0, k_0), for[26]. The distribution of the normal prior for β is $\pi(\beta/\mu_0, \sigma_0^2) = \frac{1}{2\pi\sigma_0^2}exp\left\{-\frac{1}{2\sigma_0^2}(\beta - \mu_0)^2\right\}$, and the distribution of the gamma prior for α is $\pi(\alpha/\alpha_0, k_0) = \alpha^{\alpha_0-1}exp(-k_0\alpha)$.

We will have the following joint posterior

$$\pi(\beta, \alpha/D) \propto \alpha^d(n\alpha + \alpha_0 - 1)\left\{\exp(x_i\beta) + d(\alpha-1)exp(t_i) - dexp(t_i^{\alpha} + \lambda^{\alpha} + log(k_0 x_i\beta)\right\}, \quad (5)$$

Bayesian log-normal AFT model
Another commonly used parametric survival model is the log-normal model. For this model, we assumed that the logarithms of the survival times were normally distributed. If t_i has a log-normal distribution with parameters (μ, σ^2), then the density function is given by $f(t_i/\mu, \sigma) = 2\pi^{-1/2}(t_i\sigma)^{-1}exp\left(\frac{-1}{2\sigma^2}(\log(t_i) - \mu)^2\right)$, and the survival function is given by

$$L(\mu, \sigma/D) = \prod_{i=1}^{n} f(t_i/\mu, \sigma)^{\delta_i} S(t_i/\mu, \sigma)^{1-\delta_i}$$
$$= 2\pi^{-1/2}exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\delta_i(\log(t_i) - \mu)^2\right\} * \prod_{i=1}^{n} t_i^{-\delta_i}\left(1 - \phi\left[\frac{\log(t_i) - \mu}{\sigma}\right]\right)^{1-\delta_i}$$

Then, the likelihood function of (μ, σ) is

$$L(\mu, \sigma/D) = \prod_{i=1}^{n} f(t_i/\mu, \sigma)^{\delta_i} S(t_i/\mu, \sigma)^{1-\delta_i}$$
$$= 2\pi^{-1/2}exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\delta_i(\log(t_i) - \mu)^2\right\} * \prod_{i=1}^{n} t_i^{-\delta_i}\left(1 - \phi\left[\frac{\log(t_i) - \mu}{\sigma}\right]\right)^{1-\delta_i}$$

Let $\tau = 1/\sigma^2$ and $\mu_i = x_i'\beta$, where x_i is a covariate and β is a parameter. Assume that the prior for β and the gamma prior for τ are normal[26]. The distribution of the normal prior for β is

$\pi\left(\beta / \mu_0, \sigma_0{}^2\right)=\frac{1}{2\pi\sigma_0{}^2}exp\left\{-\frac{1}{2\sigma_0{}^2}(\beta-\mu_0)^2\right\}$, and the distribution of the gamma prior for τ is $\pi\left(\tau / \alpha_0, k_0\right)=\tau^{\alpha_0-1}exp(-k_0\tau)$.

Then, the posterior distribution of (β, τ) is given by

$$
\begin{aligned}
\pi\left(\beta, \tau / D\right) \propto \alpha^{\frac{\alpha_0+d}{2}-1}exp & \left\{\frac{-\tau^2}{2}\left(\sum_{i=1}^{n}\delta_i(\log(t_i)-x_i'\beta)^2+(\beta-\mu_0)'\frac{1}{\sigma_0^2}(\beta-\mu_0)+\lambda_0\right)\right\} \\
& *\prod_{i=1}^{n}t_i^{-\delta_i}\left(1-\phi\left(\tau^{\frac{1}{2}}(\log(t_i)-x_i'\beta)\right)\right)^{1-\delta_i},
\end{aligned}
\tag{6}
$$

Parameter estimation
Bayesian parameter estimation is an alternative framework for parameter estimation. This study employed the integrated nested Laplace approximation method to estimate the parameters of the Bayesian AFT models.

Model selection criterion
The marginal likelihood, deviance information criterion (DIC), and Watanabe Akaike Information Criterion (WAIC) were used for Bayesian survival model comparisons.

Model diagnostics
In models for survival data, checking model adequacy is a crucial factor. The predictive distribution can be used both to validate and compare models[29]. To evaluate the goodness of fit of the model and whether there are any outliers, the conditional predictive ordinates (CPO) and probability integral transform (PIT) values can be examined. When the expected failure is 0, the computed value of CPO seems to be reliable; when the expected failure is 1, the computed value of CPO is known to be completely unreliable. When the expected failure is 0, the computed value of the PIT seems to be reliable; when the expected failure is 1, the computed value of the PIT is known to be completely unreliable. The Kullback-Leibler divergence (kld) is the value that describes the difference between the normal approximation and the simplified Laplace approximation. Small values indicate that the posterior distribution is well approximated by a normal distribution[30].

Results
Descriptive statistics
In this study, 296 women who were followed up for breast cancer treatment and had at least one follow-up at Hiwot Fana Specialized University Hospital during the period 1st April 2020 to 1st April 2023 were considered. The minimum and maximum event times observed in follow-up were 2 and 36 months, respectively. Among those breast cancer patients, approximately 56 (18.9%) died, and the remaining 240 (81.1%) were censored (Table 1).

Out of breast cancer patients, one-third (33.4%) were younger than 40 years old, while just over half (58.5%) were between 40 and 65 years old. The remaining 8.1% were older than 65. It's important to note that the event of interest happened more often in the middle age group (40–65-year-olds) at 10.5%, compared to 4% for those under 40 and 4.4% for those over 65. The study also investigated factors like residence and tumour size. Nearly 40% (39.5%) of the patients lived in rural areas. Interestingly, the event of interest occurred slightly more often in this group (11.1%) compared to the overall rate. Looking at tumor size, the results showed a small percentage (1.4%) of patients with tumours 2 cm or smaller experienced the event. This rate increased to over 12.8% for those with tumours larger than 5 cm. Furthermore, the study considered BMI and breastfeeding history. The patients fell into three BMI categories: underweight (13.2%), normal weight (58.4%), and overweight (28.4%). Interestingly, the event rate increased with BMI, affecting 2.4% of underweight patients, 7.4% of normal weight patients, and rising to 9.1% of overweight patients. And, 26.4% of the patients had breastfed, and 3.4% of these women experienced the event of interest.

The distribution of breast cancer stages among the patients was: 2 (0.7%) with stage I, 5 (1.7%) with stage II, 16 (5.4%) with stage III, and 33 (11.1%) with stage IV disease. The majority of the breast cancer patients (255, or 86.1%) were non-smokers. Of the total number of breast cancer patients who experienced these events, 32 (10.8%) were not alcohol users, and 24 (8.1%) were alcohol users. Additionally, 26 (8.8%) of the patients had comorbidities, and 14 (4.7%) of them experienced these events.

Likewise, with 58.5% receiving chemotherapy only, 18.9% undergoing surgery only, and 22.6% receiving both. The percentages of patients who died from chemotherapy, surgery or both chemotherapy and surgery were 28 (9.5%), 9 (3.0%), and 19 (6.4%), respectively. The study also investigated marital status, finding that 51 (17.3%) were single, 209 (70.6%) married, 22 (7.4%) divorced, and 14 (4.7%) widowed. Interestingly, the event rate varied by marital status, with the highest rate among divorced patients (18.2%) and the lowest among married patients (6.7%). Additionally, 95 (32.1%) of the patients were oral contraceptive users, and 19 (6.4%) of them experienced these events.

Survival analysis
Median survival time
According to Table 2, the estimated median survival time of the breast cancer patients in this study was 33 months. Overall, 50% or half of the breast cancer patients who received treatment were expected to survive 33 months or more in this study. Because of the existence of censoring and the skewed nature of the survival time, the median was used to describe the time to death in breast cancer patients.

Risk factors	Categories	Patient status		
		Censored	Death	Total
Age	< 40	87 (29.4%)	12 (4.0%)	99 (33.4%)
	40–65	142 (48.0%)	31(10.5%)	173 (58.5%)
	> 65	11 (3.7%)	13 (4.4%)	24 (8.1%)
Marital status	Single	36 (12.2%)	15 (5.1%)	51 (17.3%)
	Married	174 (58.8%)	35 (11.8%)	209 (70.6%)
	Divorced	18 (6.1%)	4 (1.3%)	22 (7.4%)
	Widowed	12 (4.0%)	2 (0.7%)	14 (4.7%)
Residence	Rural	84 (28.4%)	33(11.1%)	117 (39.5%)
	Urban	156 (52.7%)	23 (7.8%)	179 (60.5%)
Tumor size	≤ 2 cm	49 (16.5%)	4(1.4%)	53 (17.9%)
	3–5 cm	93 (31.5%)	14 (4.7%)	107 (36.2%)
	> 5 cm	98 (33.1%)	38 (12.8%)	136 (45.9%)
Breastfeeding	No	172 (58.1%)	46(15.5%)	218 (73.6%)
	Yes	68 (23.0%)	10 (3.4%)	78 (26.4%)
BMI	Underweight	32 (10.8%)	7 (2.4%)	39 (13.2%)
	Normal	151 (51.0%)	22 (7.4%)	173 (58.4%)
	Overweight	57 (19.3%)	27 (9.1%)	84 (28.4%)
Stage	Stage I	36 (12.1%)	2 (0.7%)	38 (12.8%)
	Stage II	63 (21.3%)	5 (1.7%)	68 (23.0%)
	Stage III	81 (27.4%)	16 (5.4%)	97 (32.8%)
	Stage IV	60 (20.3%)	33(11.1%)	93 (31.4)
Treatment taken	Chemotherapy	145 (49.0%)	28 (9.5%)	173 (58.5%)
	Surgery	47 (15.9%)	9 (3.0%)	56 (18.9%)
	Chemo and surgery	48 (16.2%)	19 (6.4%)	67 (22.6%)
Alcohol consumption	No	201 (67.9%)	32(10.8%)	233 (78.7%)
	Yes	39 (13.2%)	24 (8.1%)	63 (21.3%)
Oral contraceptive	No	164 (55.4%)	37(12.5%)	201 (67.9%)
	Yes	76 (25.7%)	19 (6.4%)	95 (32.1%)
Smoking habits	No	218 (73.6%)	37(12.5%)	255 (86.1%)
	Yes	22 (7.5%)	19 (6.4%)	41 (13.9%)
Comorbidity	No	228 (77.0%)	42(14.2%)	270 (91.2%)
	Yes	12 (4.1%)	14 (4.7%)	26 (8.8%)

Table 1. Descriptive results of demographic and clinical variables of breast cancer patients.

N	Events	Median	0.95LCL	0.95UCL
296	56	33	30	NA

**Table 2**. The estimated median survival time (in months) for breast cancer patients.

Comparison of survival functions
Figure 1a shows that breast cancer patients aged younger than 40 years had a greater probability of surviving than patients aged 40 to 65 years and patients aged older than 65 years. This finding indicates that patients older than 65 years have a lower survival rate than do the other patients. On the other hand, the survival curves of breast cancer patients with a tumor size greater than 5 cm were lower than those of patients with a tumor size 3–5 cm and less than or equal to 2 cm (Fig. 1b).

The Kaplan–Meier survival curves in Fig. 2a show that patients whose stage was stage IV had a smaller survival probability than patients whose stage was stage I, stage II, or stage III. Similarly, Fig. 2b indicates that breast cancer patients who did not consume alcohol had a longer survival probability than patients who consumed alcohol. Figure 3a below shows that breast cancer patients who did not smoke had a better survival probability than patients who smoked. Figure 3b shows that breast cancer patients who had comorbidities had a shorter survival rate than patients who did not have comorbidities.

Significance tests
Table 3 below shows the log-rank test results for the different covariate categories. The results showed that there were significant differences in the survival time of breast cancer patients according to age, residence, tumor
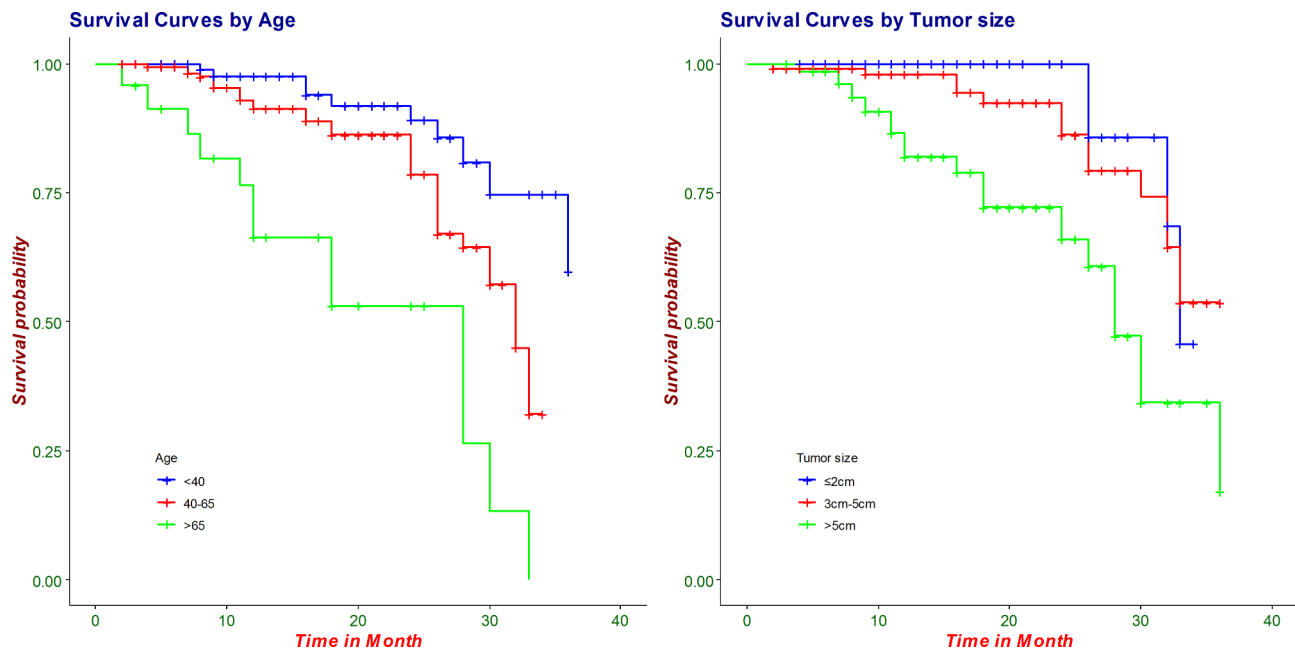
Fig. 1. K–M survival curves for breast cancer patients according to age and tumor size.
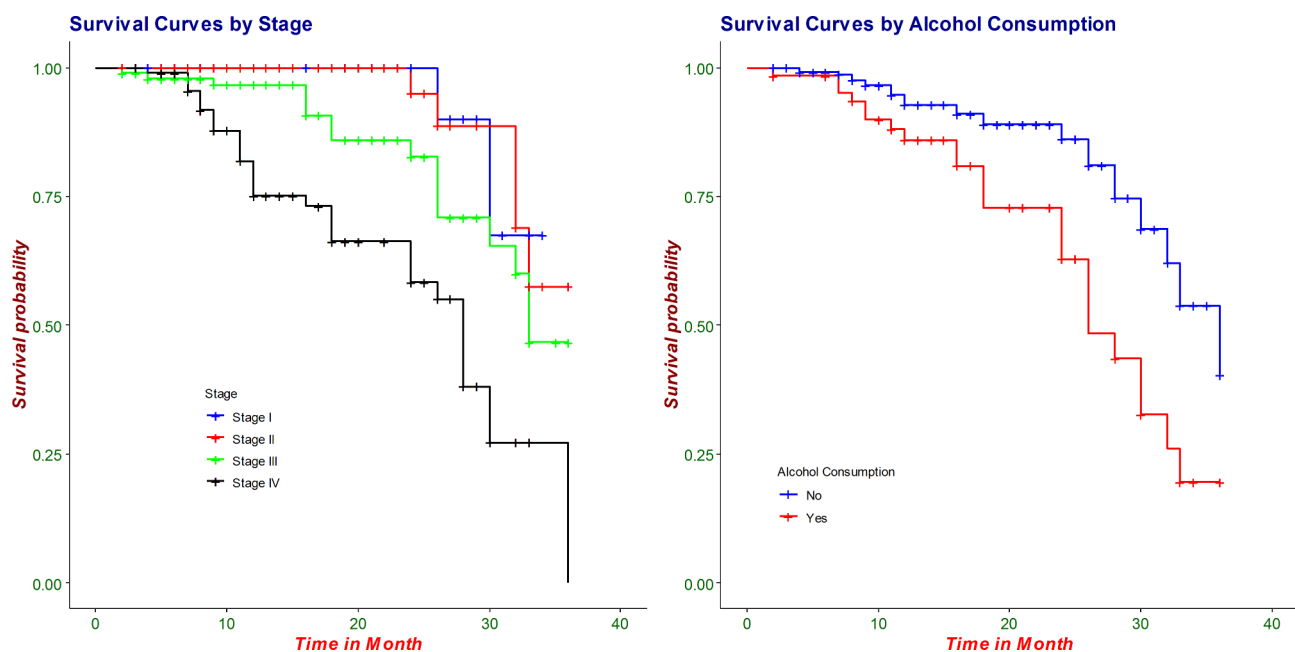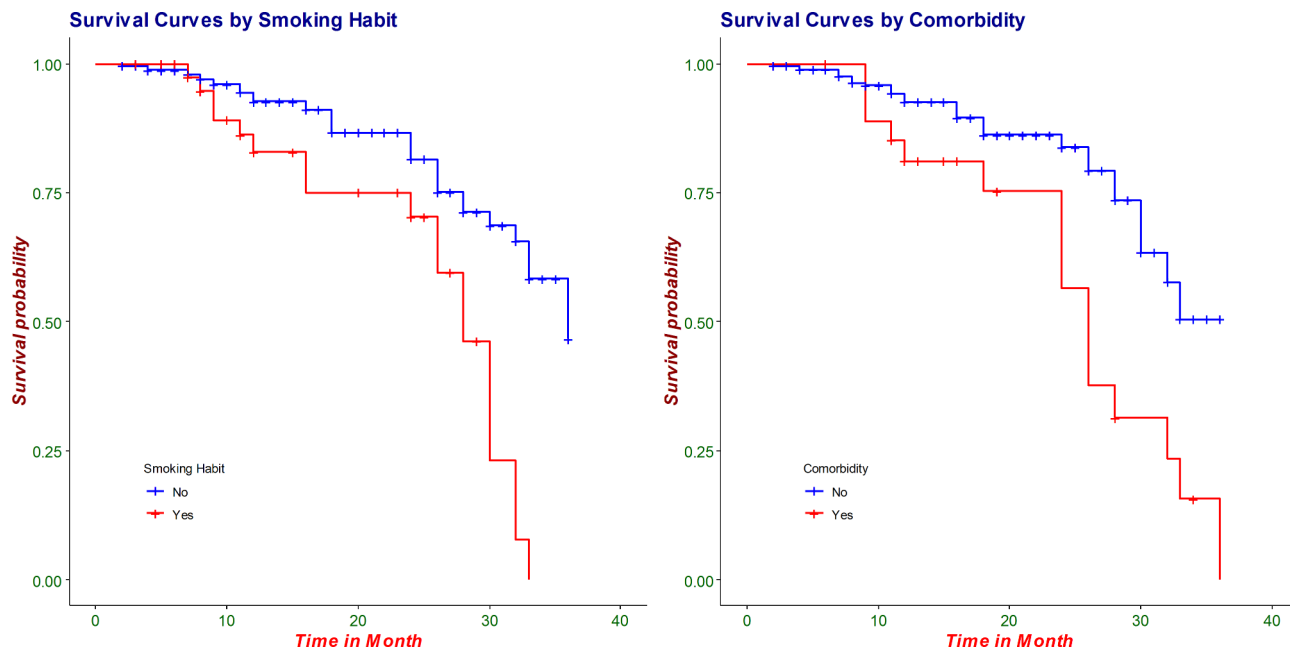

Fig. 2. K–M survival curves for breast cancer patients by stage and alcohol consumption status.

size, breast feeding status, BMI, stage, alcohol consumption, smoking habit and comorbidity at the 5% level of significance. However, the remaining covariates, such as marital status, treatment taken, and oral contraceptives, were not significantly different.

Bayesian accelerated failure time models
A Bayesian Accelerated Failure Time (AFT) model was used in this study to understand how various factors influence breast cancer patient survival. AFT offers a distinct advantage: it explains the impact of these factors in real-world terms, such as the number of years survival time might change, instead of just relative terms. The study configured the model using specific probability distributions.

Before building a complex model, the study performed a univariate analysis, a process that examines individual factors to identify which significantly affect survival time. This analysis revealed significant associations between

Fig. 3. K–M survival curves for breast cancer patients according to smoking habit and comorbidity status.

Risk factors	df	Log-rank test X^2-square	P value
Age	2	29.9	0.0001
Marital status	3	2.3	0.5
Residence	1	5.2	0.02
Tumour size	2	21.9	0.0001
Breastfeeding	1	7.5	0.006
BMI	2	8.9	0.01
Clinical stage	3	35.9	0.0001
Alcohol consumption	1	13.6	0.0002
Treatment taken	2	2.2	0.3
Oral contraceptive	1	0.3	0.6
Smoking habit	1	15.9	0.0001
Comorbidity	1	13.7	0.0002

Table 3. Significance tests.

survival time and factors like age, residence, tumor size, breastfeeding history, BMI, cancer stage, alcohol consumption, smoking habits, and the presence of other illnesses (comorbidities).

Then, to select factors for the final model predicting patient survival, the study looked at two approaches. A nonparametric log-rank test identified relevant factors, while the univariate analysis provided more detailed information about each factor's influence. Combining these insights, they chose age group, residence, tumor size, breastfeeding history, BMI, cancer stage, alcohol consumption, smoking habits, and comorbidities as candidate factors for the final Bayesian AFT model.

Multivariate analysis of the bayesian accelerated failure time model

For the survival time of the patients in the breast cancer dataset, a Bayesian AFT model with baseline exponential, Weibull, log-logistic and log-normal distributions was fitted by including all the covariates that were significant in the univariate analysis at the 5% level of significance, and the backwards variable selection method was applied to select the significant covariates. From Table 4, we compared the models by using DIC, WAIC, and the Marginal log-likelihood. The models with the smallest DIC and WAIC and a large marginal log-likelihood were the best preferable Bayesian AFT models for analysing the time to death in breast cancer patients. The Bayesian Weibull AFT model (DIC = 520.39, WAIC = 521.59, and ML = -289.02) was found to be the best fitting model for our dataset, as it has the smallest values of DIC and WAIC and has the largest marginal log-likelihood among the models.

Models	DIC	WAIC	Marginal log-likelihood
Exponential	566.29	565.42	−311.20
Weibull	**520.39**	**521.59**	**−289.02**
Log-logistic	527.63	526.34	−297.77
Log-normal	522.44	523.14	−303.53

Table 4. The values of DIC, WAIC and the marginal log-likelihood of bayesian AFT models.

Risk factors	Categories	Mean	Exp (β)	SD	95% CI	Kld
Intercept		−10.821	0.0001	1.087	[−12.962, −8.705]	0
Age	< 40 (Ref.)					
	40–65	0.377	1.457	0.358	[−0.324, 1.077]	0
	> 65	1.410	4.095	0.418	[0.589, 2.228]	0
Stage	Stage I (Ref.)					
	Stage II	0.224	1.251	0.840	[−1.424, 1.869]	0
	Stage III	1.376	3.959	0.754	[−0.102, 2.852]	0
	Stage IV	1.976	7.213	0.738	[0.527, 3.421]	0
Alcohol cons	No (Ref.)					
	Yes	0.624	1.866	0.298	[0.040, 1.207]	0
Smoking habits	No (Ref.)					
	Yes	0.984	2.675	0.303	[0.389, 1.578]	0
Comorbidity	No (Ref.)					
	Yes	0.666	1.946	0.316	[0.046, 1.286]	0

Table 5. Summary results of the bayesian Weibull AFT model.

The results from the bayesian Weibull AFT model
The final results for the Bayesian Weibull AFT model using the INLA method are shown in Table 5, and the decision about the significance of the variables is based on the 95% credible interval for the posterior mean of the coefficients. The posterior summary results of the Bayesian Weibull AFT model showed that the survival time of breast cancer patients was significantly affected by age, stage, alcohol consumption, smoking habits, and comorbidities. The final model was interpreted using the acceleration factor and 95% credible interval of the Bayesian accelerated failure time estimated values. The estimated acceleration factor is defined as exp (posterior mean). To determine the significance of the covariates in the model, the 95% credible interval was used. The factors whose credible intervals for the posterior mean of parameters were 0 or whose credible intervals for the acceleration factor were 1 implied that these factors were not significant.

The Bayesian Weibull AFT model estimated posterior coefficients and showed increased survival time when the mean posterior coefficient was negative and decreased (delayed) survival time when the estimated posterior coefficient was positive due to its parameterization. The Kullback–Libler divergence values for all the significant parameters in the Bayesian Weibull AFT model were 0; thus, small values indicate that the posterior distribution was well approximated by a normal distribution (Table 5). This finding implies that a simplified Laplace approximation is the most efficient algorithm with improved efficiency and higher computation speed.

The estimated acceleration factor for breast cancer patients aged > 65 years was estimated to be exp (1.410) = 4.095, with a 95% credible interval of acceleration [1.802, 9.28]. Thus, the expected survival time of breast cancer patients older than 65 years was delayed by a factor of approximately 4.095 compared to that of patients younger than 40 years, keeping the effects of other factors constant. The 95% confidence interval for the acceleration factor for patients older than 65 years was not included, which implies that having more than 65 years has a significant effect on the time to death in breast cancer patients.

By examining the stage of the patients, the estimated acceleration factor for stage IV breast cancer patients was estimated to be 7.214, with a 95% credible interval of acceleration [1.694, 30.599]. Thus, the expected survival time of stage IV breast cancer patients was shortened by a factor of approximately 7.214 compared to that of stage I patients, keeping the effects of other factors constant. These findings revealed that stage IV breast cancer significantly affects the survival time of breast cancer patients.

Regarding alcohol consumption, the estimated acceleration factor for breast cancer patients who consumed alcohol was estimated to be 1.866, with a 95% credible interval [1.041, 3.343]. This finding indicated that when all the other factors were held constant, the expected survival time of breast cancer patients who consumed alcohol decreased by 86.6% compared to that of those who did not consume alcohol. This finding suggested that alcohol consumption has a significant effect on the survival of breast cancer patients.

By observing smoking habits, the estimated acceleration factor for breast cancer patients who smoke was estimated to be 2.675, with a 95% credible interval [1.475, 4.845]. This finding showed that when all the other factors were held constant, the expected survival time of breast cancer patients who had a smoking habit was

delayed by a factor of approximately 2.675 compared to that of those who did not have smoking habits. Thus, smoking habits had a significant effect on the time to death in breast cancer patients.

The estimated acceleration factor for breast cancer patients who had comorbidities was estimated to be 1.946, with a 95% credible interval of acceleration [1.047, 3.618] obtained by keeping all the other factors constant. Thus, the expected survival time of breast cancer patients who had comorbid diseases was 94.6% shorter than that of patients who had no comorbid disease. These findings showed that comorbidities had a significant effect on the survival time of breast cancer patients.

Discussion

The aim of this study was to model the time to death of breast cancer patients at Hiwot Fana Specialized University Hospital by using Bayesian accelerated failure time models. Among the study participants who fulfilled the inclusion criteria, approximately 18.9% died due to breast cancer. In the present study, the estimated median survival time for breast cancer patients was 33 months; the minimum and maximum event times were 2 and 36 months, respectively.

Previous research conducted in the rural Ethiopia, and Northwest Amhara, Ethiopia has reported median survival times for women with breast cancer of 28 months, and 45 months, respectively[15,16]. Our finding is consistent with these reported outcomes. Classical logistic regression and survival analysis, while valuable tools, face limitations in handling time-to-event data and complex censoring schemes. To address these challenges, this study employed a Bayesian accelerated failure time (AFT) model with integrated nested Laplace approximation. This approach provides a more comprehensive analysis of breast cancer risk factors, offering superior insights into patient outcomes[10,11,15,22,31]. The results of the Bayesian Weibull AFT model using the INLA method showed that the survival time of breast cancer patients was significantly affected by age, stage, alcohol consumption, smoking habits, and comorbidities.

Accordingly, age group had a significant effect on the time to death in breast cancer patients. This finding implies that the survival time of breast cancer patients is highest for the youngest women and decreases with increasing age. This revealed that the expected survival time of women with breast cancer decreases as they age. Moreover, patients older than 65 years had a low survival rate. When a patient ages, she cannot survive the disease because of the influence of age, which may be due to the expected limit of survival caused by age. The findings of this study conform with previous study, which has consistently indicated the significant impact of age on breast cancer results. Older age over 65 years, was identified as a key risk factor associated with shorter survival times among breast cancer patients. For instance, the hazards of mortality were more than two times higher for breast cancer patients aged 65 years and above compared to those under 40 years[17,32]. These consistent findings across different studies highlight the critical need to prioritize the care and management of elderly breast cancer patients, who face a disproportionately higher risk of mortality[14,17,32].

The stage of breast cancer also has a significant effect on the survival time of breast cancer patients, and the expected survival time of patients with stage IV breast cancer was shorter than that of patients with stage I breast cancer. The findings of this study emphasize the critical importance of disease stage as a key predictor of survival for breast cancer patients. Consistent with prior study, the finding show that advanced stages of breast cancer, particularly stages III and IV, are associated with significantly poorer survival outcomes[17,33,34]. For instance, the retrospective cohort study conducted in northwest Ethiopia reported that the hazards of mortality were 1.82 times higher for patients diagnosed with stage IV breast cancer compared to those with stage I disease[17]. Similarly, research from South Africa has emphasized that later-stage diagnosis is a well-established factor linked to diminished breast cancer survival[33]. This indicates that the critical need for early detection and screening programs to identify breast cancer cases at earlier, more treatable stages.

The findings of this study also revealed that breast cancer patients who consumed alcohol had shorter survival times than those who did not consume alcohol; thus, alcohol consumption had a significant effect on the time to death. This result is consistent with those of[35,36], who suggested that alcohol consumption increases the risk of mortality due to breast cancer. Similarly, the study on the relationship between alcohol consumption and breast cancer risk clearly reveals a concerning link. Studies have consistently found that increased alcohol intake is associated with a heightened risk of developing and dying from breast cancer. For example, a meta-analysis presented to the American Cancer Society found that each additional 10 g of alcohol consumed per day can increase a woman's risk of breast cancer by 7–12%[35]. Similarly, a recent study by[36] reported that moderate alcohol consumption of more than 35–44 g per day was linked to a 46% increased risk of breast cancer. The study highlights the importance of promoting moderation and limiting alcohol intake, particularly among women, as a potential strategy for breast cancer prevention.

Similarly, smoking is an important risk factor that significantly influences breast cancer patients. The expected time to death was lower for breast cancer patients who smoked than for those who did not smoke. This finding implies that nonsmoking breast cancer patients have longer survival times than smokers. Likewise, multiple studies across diverse geographic regions have consistently shown that tobacco use is a key determinant of breast cancer development and survival. A case-control study conducted in Saudi Arabia found that women who smoked had a six-fold increased risk of breast cancer compared to their non-smoking counterparts[37]. Furthermore, the study by[10] reported that patients without a smoking habit had significantly longer survival times than those who smoked. These findings underscore the crucial need to prioritize tobacco cessation and prevention efforts as part of comprehensive breast cancer control strategies[10,37,38].

This study revealed that the survival time of breast cancer patients who had comorbidities was shorter than that of patients who did not have comorbidities. This finding suggested that patients who have no associated diseases, such as anaemia, diabetes, or hypertension, are more likely to have a survival rate than are breast cancer patients who have comorbidities. This result is similar to the findings of the study performed by[39], in which the survival time of patients with comorbidities increased the risk of breast cancer mortality compared

to that of patients without comorbidities. These results are also consistent with those of previous studies[31,40]. Their results suggested that breast cancer patients with comorbidities had shorter survival times than patients without comorbidities. The shorter survival times observed among patients with comorbidities highlight the need for targeted interventions and multidisciplinary care approaches. Strategies to optimize the management of coexisting conditions, such as improving glycaemic control, managing hypertension, and addressing anaemia, may contribute to improved overall outcomes for breast cancer patients.

Conclusion
The estimated median survival time of the breast cancer patients in this study was 33 months. The Bayesian Weibull accelerated failure time model was found to be the best fitted model for predicting the survival time of patients with minimum DIC and WAIC as well as large marginal likelihood values among the proposed Bayesian AFT models.

The results of the Bayesian Weibull AFT model with INLA estimation techniques revealed that age, stage, alcohol consumption, smoking habit, and comorbidity were the most important risk factors significantly associated with the time to death. Thus, from the findings of this study, we conclude that age older than 65 years, stage IV breast cancer, alcohol consumption, smoking, and comorbidity are factors that increase the risk of mortality due to breast cancer by decreasing survival time. Therefore, Hiwot Fana Specialized University Hospital should work on creating awareness to reduce smoking habits and alcohol use, as well as paying attention to elderly patients and stage IV breast cancer patients during intervention.

Limitation of the study and areas for further research
The study was conducted based on secondary data which might have incomplete information. Besides, potential risk factors such as educational level, family history of breast cancer, and the presence of positive lymph nodes, are thought to have an impact on the survival of breast cancer patients, as various pieces of literature have noted. However, these variables were not included in this study since data on them could not be found in hospital records. Furthermore, incorporating additional potential risk factors using machine learning techniques could help provide a more comprehensive understanding of the problem.

Practical implications of the study

1. Importance of early detection and screening:

   - The study highlights that breast cancer is a major health concern in Ethiopia, with high morbidity and mortality rates.
   - Emphasis should be placed on early detection and screening programs to identify breast cancer cases at earlier, more treatable stages.

2. Risk factor identification and targeted interventions:

   - The research identified several key risk factors associated with shorter survival times, including older age, advanced disease stage, alcohol consumption, smoking, and comorbidities.
   - Targeted interventions should be developed to address these modifiable risk factors, such as:
   - Promoting awareness campaigns to reduce alcohol use and smoking among women.
   - Providing additional support and care for elderly breast cancer patients and those with advanced disease.
   - Integrating comorbidity management into breast cancer treatment protocols.

3. Promoting knowledge sharing:

   - The study's insights can be shared with other healthcare facilities and policymakers in Ethiopia to inform the development of national breast cancer control strategies.
   - Collaborative efforts between the hospital, research institutions, and public health authorities could help to scale up successful interventions and optimize breast cancer care across the country.

Data availability
Access to the datasets used for this study are available from the corresponding author on reasonable request.

References
1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 3 (2018).
2. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clinicia* (2021).
3. Organization, W. H. (2021). https://www.who.int/news-room/fact-sheets/detail/breast-cancer
4. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. *Cancer J. Clin.* **72**(1), 7–33. https://doi.org/10.3322/caac.21708 (2022).
5. Adamu, A. *Application of Copula on Time to Last Stage of Left and Right Breast Cancer of Patients at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar, Ethiopia* (Bahir Dar University, 2022).
6. Sharma, R. Breast cancer burden in Africa: Evidence from GLOBOCAN 2018. *J. Public Health.* **43**(4), 763–771 (2021).
7. Anyigba, C. A., Awandare, G. A. & Paemka, L. Breast cancer in sub-saharan Africa: The current state and uncertain future. *Exp. Biol. Med.* https://doi.org/10.1177/15353702211006047 (2021).

8. Hadgu, E. et al. Breast cancer in Ethiopia: Evidence for geographic difference in the distribution of molecular subtypes in Africa. (2018).
9. World Health Organization. Cancer country profiles. (2017). http://www.who.int/cancer/country-profiles/idn_en.pdf. Accessed April 20, 2023 (2017).
10. Reta, B., Yasin, J., Anberbir, A. & Abebe, L. Risk factors affecting survival time of breast cancer patients: The case of southwest Ethiopia. *J. Res. Health Sci.* **21**(4), e00532 (2021).
11. Deressa, B., Cihoric, N., Badra, E. V., Tsikkinis, A. & Rauch, D. Breast cancer care in northern Ethiopia—Crosssectional analysis. *BMC Cancer.* **19**(1), 393 (2019).
12. Solomon, M. et al. Estimates of cancer incidence in Ethiopia in 2015 using population-based registry data. *J. Glob Oncol.* **4**, 1–11 (2018).
13. Sefinew, A., Zelalem, Y., Mathewos, A. & Wondmagegnehu, T. Trends of breast cancer in Ethiopia. *Int. J. Cancer Res. Mol. Mech.* **2**(1), 1 (2016).
14. Bsrat, T., Tewodros, G. & Endeshaw, D. Joint modelling of longitudinal change in tumor cell level and time to death of breast cancer patients: In case of Ayder Comprehensive Specialized Hospital Tigray, Ethiopia. *Cogent Med.* **8**(1), (2021).
15. Eber, P., Wakuma, S., Christian, T., Adamu, R. & Claudia, A. Survival of breast cancer patients in rural Ethiopia. *Breast Cancer Res. Treat.* **170**(1), 111–118. https://doi.org/10.1007/s10549-018-4724-z (2018).
16. Bereket, F., Lijalem, T. & Aweke, M. Survival analysis of women breast cancer patients in Northwest Amhara, Ethiopia. *Front. Oncol.* https://doi.org/10.3389/fonc.2022.1041245 (2022).
17. Mulu, T., Aragaw, T. & Desalegn, T. Survival and predictors of mortality among breast cancer patients in Northwest Ethiopia: A retrospective cohort study. *Cancer Manag Res.* **13**, 9225–9234. https://doi.org/10.2147/CMAR.S3399 (2021).
18. Datwyler, C., Stucki, T. & Parametric survival model. (2011). https://stat.ethz.ch/education/semesters/ss2011/seminarhandout-9.pdf
19. Royston, P. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Stat. Neerl.* **55**(1), 89–104 (2001).
20. Tolessa, L., Endalew, G., Negalign, D. & Assefa, D. Risk factors associated with breast cancer among women in Addis Ababa, Ethiopia: Unmatched case–control study. *Int. J. Women's Health.* **13**, 101–110 (2021).
21. Babiker, S. et al. Prospective breast cancer risk factors prediction in Saudi women. *Saudi J. Biol. Sci.* **27**(6), 1624–1631 (2020).
22. Nikbakht, R. & Bahrampour, A. Determining factors influencing survival of breast cancer by fuzzy logistic regression model. *J. Res. Med. Sci. Off J. Isfahan Uni Med. Sci.* **22** (2017).
23. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
24. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**(4), 719–748 (1959).
25. Kleinbaum, D. G. & Klein, M. *Survival Analysis: A Self-Learning Text* (Springer, 1996).
26. Ibrahim, J. G., Chen, M-H., Sinha, D., Ibrahim, J. G. & Chen, M. H. *Bayesian Survival Analysis*, Vol. 2. (Springer, 2001).
27. Soodejani, M. T., Tabatabaei, S. M. & Mahmoudimanesh, M. Bayesian statistics versus classical statistics in survival analysis: An applicable example. *Am. J. Cardiovasc. Dis.* **11**(4), 484–488 (2021).
28. Ibrahim, J. G., Zhu, H. & Tang, N. Bayesian local influence for survival models. *Lifetime Data Anal.* **17**(1), 43–70 (2021).
29. Gelman, A. Prior distribution. *Encyclopedia Environmetr.* **3**, 1627–1628 (2002).
30. Bivand, R., Gómez-Rubio, V. & Rue, H. Spatial data analysis with R-INLA with some extensions. *Am. Stat. Assoc.* (2015).
31. Wondimeneh, S. *Survival Status and Predictor of Mortality among Breast Cancer Patients in Black Lion Specialized Hospital* (Addis Ababa University, 2018).
32. Pham, M. H. Survival analysis-breast cancer. *Undergr. J. Math. Model. One + Two.* **6**(1), 4 (2014).
33. Dickens, C. et al. Stage at breast cancer diagnosis and distance from diagnostic hospital in a Periurban setting: A South African public hospital case series of over 1000 wome. (2014).
34. Lijalem, T., Teshale, T. & Muluneh, E. Assessing the incidence, epidemiological description and associated risk factors for breast cancer in western Amhara, Ethiopia. *Breast Cancer Manag.* **9**(3), BMT47 (2020).
35. DeSantis, C. E. et al. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA Cancer J. Clinicians.* **66**(1), 31–42 (2016). (2016).
36. Kashyap, D. et al. Global increase in breast cancer incidence: Risk factors and preventive measures. *BioMed. Res. Int.* (2022).
37. Alsolami, F. J. et al. Determinants of breast cancer in Saudi women from Makkah region: A case–control St. (2019).
38. Catsburg, C., Miller, A. B. & Rohan, T. E. Active cigarette smoking and risk of breast cancer. *Int. J. Cancer.* **136**(9), 2204–2209 (2015).
39. Nelson, S. H. et al. Impact of very low physical activity, BMI, and comorbidities on mortality among breast cancer survivors. *Breast Cancer Res.* (2016).
40. Misganaw, M., Zeleke, H., Mulugeta, H. & Assefa, B. Mortality rate and predictors among patients with breast cancer at a referral hospital in northwest Ethiopia: A retrospective follow-up study. *PLoS One* **18**(1), e0279656 (2023).

Author contributions

MT was involved in data collection, cleaning, data analysis, interpretation and manuscript preparation. KT and MS were helped with data analysis and manuscript preparation. All the authors read and approved the manuscript.

Funding

Declarations

Ethics approval and consent to participate

Informed consent was waived due to patients' data were anonymized by College of Health and Medical Science, Haramaya University Hiwot Fana Comprehensive Specialized Hospital (Institutional Health Research Ethics Review Committee (IHRERC)) (Ref: D/R/G/P/05/37/2023).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.