Modeling Risk Factors Associated With Time to Pregnancy Termination in Ethiopia: Application of Frailty Models

AbomaTemesgen Sebu *a

aDepartment of Statistics, Haramaya University, Dire Dawa, Ethiopia, E-mail: aboma.temesgen@gmail.com

May 14, 2020

Abstract

Background: Pregnancy termination commonly known as abortion is one of the preventable causes for the maternal mortality worldwide that largely forgotten. About 45% of these pregnancy terminations are unsafe causing death of more than 22,000 women every year and remains one of the major public health problems in developing countries including Ethiopia. This study was also aimed to model and investigate risk factors associated with time to pregnancy termination in Ethiopia by applying survival model considering the clustering effects which handles unobservable heterogeneity across the administrative regions of the country.

Methods: The study considered 15,683 reproductive age group women from 2016 Ethiopian Demographic and Health Survey data. Kaplan-Meier(KM) was employed to estimate the survival curve and this estimated KM survival curve estimated for different groups were tested based on log rank test. To come up with appropriate model for the time to pregnancy termination and the associated risk factors both semi-parametric and parametric survival model with no frailty effects as wells as with shared frailty effects were employed and compared based AIC and BIC of the fitted models.

Results: The result of the study showed generalized gamma and lognormal survival models were appropriate models compared with semi-parametric and other candidate parametric models. Fitting these survival model with frailty showed the improvement of the models which was an indication for the presence of unobservable random effects in clusters. Regarding the frailty models comparison, lognormal with gamma frailty model was considered as appropriate model for fitting time to pregnancy termination model in Ethiopia compared with other candidate frailty models. Furthermore, the selected frailty model result showed that age of women, ever trying to avoid pregnancy, contraceptive method use, age at first sex, total number of children ever born and place of residence were the identified risk factors for the time to pregnancy termination at 5% level of significance.

Conclusions: Based on the finding of this study, starting sex at early age, residing urban areas, having lower number of children, being in married marital status group, chewing chat and do not using contraceptive methods were the risk factors that results pregnancy termination at early age that needs serious consideration to prevent the problem in Ethiopia.

Keywords: Pregnancy Termination, Survival Model, Frailty Model, Model Information Criteria

*Corresponding author
1 Background

Pregnancy termination is the removal of pregnancy tissue, products of conception or the fetus and placenta (afterbirth) from the uterus [1]. This happens because miscarriage and stillbirth that its incidence affect fertility levels since a sizable proportion of pregnancies, ranging between 4.9% and 52.0% in a comparative study of 20 countries, end in pregnancy termination instead of a live birth [2]. About 10.9% of pregnancies do not end in live-birth and about 63.7% of these pregnancies was terminated due to miscarriage. Based on reported it is higher among women using contraceptives, consistent with expectations and very low levels of reported in some countries, particularly in sub-Saharan Africa, suggests possible underreporting [3].

Worldwide, about 25 million unsafe abortions occurred every year between 2010 and 2014, according to a new study by WHO and the Guttmacher Institute published on the Lancet. Most this unsafe abortions accounting 97% occurred in developing countries in Africa, Asia and Latin America [4]. There where an estimated 8.2 million induced abortions occurred in Africa and this number represents an increase from 4.6 million annually during 1990 to 1994, mainly because of an increase in the number of women of reproductive age. The annual rate of abortion for the region is an estimated 34 per 1,000 women of reproductive age (15-44) and remained constant between 1990-1994 and 2010-2014. The regional abortion rate is roughly 26 per 1,000 for married women and 36 per 1,000 for unmarried women [5]. Even if the World has made significant progress on four of the five main causes of maternal mortality, unsafe abortion the only almost completely preventable cause has been largely forgotten. About 45 per cent of abortions globally are deemed unsafe, and more than 22,000 women and girls die each year after undertaking an unsafe abortion in 2018[6].

The problem of unsafe abortion is a neglected public health problem contributing for 13% of maternal death worldwide and about 99% of these pregnancy termination in Africa are unsafe resulting in one maternal death per 150 cases. Its prevalence is associated with restricted abortion law, poor quality of health service, and low community awareness [8]. It is also one of the major medical and public health problems specially developing countries including Ethiopia. However, there is a lack of up-to-date and reliable information on induced abortion distribution and its determinant factors in these countries. Most of these induced abortion is underpinned by high magnitude of unwanted pregnancy. There is requirement for widespread expansion of increased access to high quality family planning service and post-abortion care[7]. Even though abortion rate in Ethiopia is lower compared to other East Africa countries, termination of pregnancy and the consequences are still major public health problems in the country. Among women undergo abortions outside of health facilities 40% were experiencing severe complications and it was higher in women residing in urban areas which is about 92 per 1,000 pregnancies [9].

Several reported studies on the prevalence of pregnancy termination and its associated factors largely based on logistic regression with no consideration of unobservable random effects and limited study was focused on time to pregnancy termination in Ethiopia. Therefore, the current study was focused modeling time to pregnancy termination and the associated factors with consideration of unobservable heterogeneity subjects considered for the study by applying different survival and frailty models. The frailty models are also part of survival models, which accounts for heterogeneity and random effects through a latent multiplicative effect on the hazard function and is assumed to have unit mean and variance $\theta$, which is estimated along with the other model parameters. It is heterogeneity model where the frailties are assumed to be subject specific and frailties are common (or shared) among groups of subjects and are randomly distributed across groups or clusters [10]. It is first suggested by Vaupel
et al. [11] in the context of mortality studies, and Lancaster [12] incorporated the frailty concept into a study of duration of unemployment and Hougaard [13] detailed the ramications of the assumed distribution of the frailty. This study was also mainly focused on these frailty models which considers the clustering effects based on unobservable random effects to model time to pregnancy termination of women in Ethiopia.

2 Methods

2.1 Source of the Data

Data considered for this study was 2016 Ethiopian Demographic and Health Survey (EDHS) data which was the fourth survey implemented from January 18, 2016, to June 27, 2016 by the Ethiopian Central Statistical Agency (CSA). By virtue of its mandate, the CSA has conducted the survey in collaboration with the Federal Ministry of Health (FMoH) and the Ethiopian Public Health Institute (EPHI) with technical assistance from ICF International. The main aim of the survey was to provide current and reliable data on fertility and family planning behavior, child mortality, adult and maternal mortality, childrens nutritional status, use of maternal and child health services, knowledge of HIV/AIDS, and prevalence of HIV/AIDS and anemia [14].

2.1.1 Sampling Design

The survey considered two-stage cluster sampling technique in which the country was stratified into nine regions and two administrative cities, in the first stage involved selecting 645 enumeration areas (202 in urban areas and 443 in rural areas) with probability proportional to size. In the second stage, a predetermined 28 households per cluster were selected with an equal probability systematic selection from the newly created household list. The data related with pregnancy was taken from a womans survey and the survey data was collected by conducting face-to-face interviews with women who met the eligibility criteria (women aged 15-49 years). The study population in this study was all reproductive age group women (15-49 years) who responded on pregnancy termination of EDHS questionnaire in 2016.

2.2 Variables of the study

The outcome variables considered for the study was the survival time (time to pregnancy termination) women during the reproductive age while the survey was conducted. It was measured from birth to pregnancy termination during the reproductive age of women in years which helps to know the exact age at which women had the pregnancy termination and to relate this time to the associated risk factors during reproductive age. Women who had not experience the event (pregnancy termination) was considered as right censoring and the censoring time was also measured from birth to current age of women while the was survey conducted.

To identify factors associated factor of time to pregnancy termination different independent covariates were proposed based on different literatures. Accordingly, a total of 16 independent variables were proposed among eleven of these representing
region where women residing, place of residence, marital status, religion, wealth index, ever trying to avoid pregnancy, contraceptive use, internet use, frequency of smoking, getting medical help and chewing chat were categorical whereas five of these independent variables representing age of women, year of education, total number of sex partner ever had, age at first sex and total number of children ever born were quantitative variables measured numerically. The summary of these independent variables with the pregnancy termination status were detailed in Table 2 of result and discussion section.

2.3 Methods of Data Analysis

Data analysis of the study begin with exploratory data analysis to understand the nature of time to pregnancy data better and to characterize pregnancy termination status with independent covariates. This was done using proportion calculation to know the prevalence of pregnancy termination in censored and event in relation with independent variables. Furthermore, mean and standard deviation were also employed to know the average and variation of quantitative variables by their pregnancy termination status. After exploratory data analysis was done the survival modeling was followed to identify the risk factors associated with time to pregnancy termination.

2.3.1 Survival Data Analysis

Survival models seek to explain how the risk or hazard of an event occurring at a given time is affected by covariates of theoretical interest where the event of interest is defined as transition from one state to another state [15,16]. Survival data analysis requires unique statistical methods for analysis due to the data is always incomplete because of censored observations. These censored observation happens when some information about a subjects event time do not know the exact event time when event happened [17]. In survival data analysis, let the survival times of \( n \) is the number of independent subjects which are right censored, the observed time for the \( i^{th} \) subject for, \( i = 1, \ldots, n \) is \( t_i \) where \( t_i \) is random variable defined by \( t_i = \min(T_i, C_i) \), for \( T_i \) representing time for observed event time and \( C_i \) is the observed censoring time which are independent. In order to know whether the event actually happened or not an indicator \( \delta_i = I(T_i \leq C_i) \) is introduced, which equals 1 if the event happened and 0 otherwise given as:

\[
\delta_i = \begin{cases} 
\delta_i = 0, & \text{for } T_i > t_i, \\
\delta_i = 1, & \text{for } T_i < t_i
\end{cases}
\]

The survival time \( t_i \) (time to specified event) assumed to follow a continuous probability distribution and this distributions are defined over the interval \([0, \infty)\) which uniquely determined by its probability density function defined by:

\[
F(t) = P(T \leq t) = \int_0^t f(s)ds
\]

This \( F(t) \) represents the cumulative distribution function (CDF) of \( T \) and the probability that a subject survives beyond time \( t \) known to be the survival function defined by:

\[
S(t) = 1 - F(t) = \int_t^\infty f(s)ds
\]

This survival function also helps to know the hazard function which is known as the failure rate, hazard rate, or force of mortality which measures the instantaneous rate of occurrence of the event which is defined as:
This hazard rate characterizes the hazard of an event changing over time and specifies the instantaneous failure rate at time $t$, given that the subject survives until time $t$. It is also called the exit rate to stress the fact that it means exit out of the state of interest.

The sum of the risks a subject faced going from duration 0 to $t$ which gives cumulative (or integrated) hazard function is defined by:

$$
\Lambda(t) = \int_0^t \lambda(s) \, ds = \int_0^t \frac{f(s)}{1 - F(s)} \, ds = -\ln(1 - S(t))
$$

The survival and hazard functions tell alternative but equivalent characterizations of the distribution of $T$ because of survival function is used for comparing the survival progress whereas the hazard function used to describe the risk of failure at any time point [18].

### 2.3.1.1 Non-parametric Survival Methods

Non-parametric method were employed to summarize survival data through the estimation of the hazard and survival function. This method is used to produce a graphical summary of the survival times for a given subjects considered in the investigation. Kaplan-Meier based nonparametric estimator of the survival function is not based on the actual observed event and censoring times, but rather on the order in which events occur and assign probability to and only to uncensored failure times. Assuming there are $n$ observations, $t_1, ..., t_n$ with corresponding censoring indicators, $\delta_1, ..., \delta_n$. Then, the Kaplan-Meier(KM) estimator for the number of distinct event times be $r$ ($r \leq n$), with the ordered event times given by $t_{(1)} < ..., < t_{(r)}$ is given by product limit of survival function estimator is given as:

$$
\hat{S}(t) = \begin{cases} 
1, & \text{for } t < t_j, \\
\prod_{j:t \leq t_i} \left(1 - \frac{d_j}{R_j}\right), & \text{for } t \geq t_j
\end{cases}
$$

Where, $d_j$ is the observed number of events at time $t_j$ and $R_j$ is the number of individuals at risk at time $t_j$. This K-M estimator of the survival function is expected to decrease as the failures occurs and the graph of the estimated K-M survival function against time is plotted to enhance understanding of the function [19]. This estimated K-M survival curve between groups is compared by plotting them on the same graph to visualized if there is difference between the estimated curves. The difficulty with this graphical visualization procedure is that there might be a difference between the two groups, but it may not be significant. Therefore, formal way of testing this significance difference between the curves is based on Log rank test according to Fleming & Harrington [20] which is given as follows for a two-category covariate:

$$
\text{Logrank test} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{\text{Var}(O_i - E_i)}
$$

Where, $O$ is the observed survival count and $E$ is the expected survival count. This test is also approximated for comparing more than two category covariates which is approximated by:

$$
\text{Logrank test} \approx \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{\sum_{i=1}^{k} E_i^2}
$$

This test statistic has chi-square distribution with $k - 1$ degree freedom where $k$ is the number of categories of covariate.
2.3.1.2 Semi Parametric Survival Model: Cox PH

In survival analysis, to determine if the variation in subjects survival experience is partially explained by covariates or to find any possible relationship between survival times and important covariates, a popular approach is to model the hazard function rather than the mean of the survival times as in the classical regression models. That is, survival models are most often defined in terms of the hazard function. Since a hazard function may be complicated, a parametric assumption can be avoided, and the hazard function allowed being nonparametric. Such survival models with no distributional assumption of the hazard function are termed as semi-parametric model which is proposed by Cox [18] and widely used survival model which is expressed as:

$$\lambda(t|X) = \lambda_0(t) exp(X_β)$$

Where; $\lambda_0(t)$ is base line hazard function, $β$ is $p$ by 1 vector of parameters and $X_i$ is $n$ by $p$ matrix of predictors. Based on this specified model the effect of the predictors are observed through the multiplicative term $exp(X_i^T β)$.

Estimation of this model is based on the partial likelihood which starts with ordering a random sample of $n$ individuals according to the rank of survival times assuming no observed ties. Since the baseline hazard has an arbitrary form in the Cox model, intervals between successive survival times do not provide information on the multiplicative effects of covariates on the hazard function and $t$ is assumed to be a continuous function. Then, the conditional probability transforms to a continuous hazard function is given by:

$$\text{Hazard rate at } t_i \text{ for } i\text{th individual with covariates } X_i$$

This can be expressed in the form of $β$ vector of the required parameters of model as:

$$\frac{\lambda_0(t_i|X_i)}{\sum_{i \in R_{t_i}} \lambda_0(t_i|X_i)} = \frac{\lambda_0(t_i|X_i) exp(X_i^T β)}{\sum_{i \in R_{t_i}} \lambda_0(t_i|X_i) exp(X_i^T β)} = \frac{exp(X_i^T β)}{\sum_{i \in R_{t_i}} exp(X_i^T β)}$$

Then, the joint likelihood for the required parameters $β$, is simply specified as the product of the above equation over all $t_i$ values, which is given by:

$$L_p(β) = \prod_{i=1}^{n} \frac{exp(X_i^T β)}{exp(X_i^T β)} \delta_i$$

Taking the log of the partial likelihood this likelihood is given by:

$$ln(L_p(β)) = \sum_{i=1}^{n} \delta_i \{ X_i^T β \} - ln(\sum_{i \in R_{t_i}} exp(X_i^T β))$$

The corresponding $β$ values are obtained by maximizing this given loglikelihood functions with respect of the required components of $β$’s Cox (1972).

2.3.1.3 Parametric Survival Models

Even if the Cox proportional hazard model is the most commonly used technique for analyzing survival data, the model is no longer valid when the proportional hazard assumption is failed. In such conditions the alternative way of modeling survival time is using parametric models which requires specification of a probability distribution. For the parametric survival models, the survival times needs to follow a certain parametric distribution with unknown parameters. The corresponding survival and hazard functions of survival time were also determined based on specified parametric probability density function [21] and the effect of covariates on survival time is through the conditional hazard function which specified as:

$$\lambda(t|X) = \lambda_0(t)(exp(X_i^T β))$$

Unlike that of Cox PH the baseline hazard function of parametric survival models is modeled parametrical, and it represent baseline hazard when all covariates are zero. If the proportional hazard is no longer valid an alternative method modeling
survival time regression modeling is based on accelerated failure time (AFT) since it does not require the proportional hazard assumption. The corresponding AFT model which is based on log scale of time given by:

$$Log(T) = X_i \beta + \sigma \varepsilon_i$$

Where; $$\varepsilon_i \sim F$$ and F is parametric error distribution and $$\sigma$$ is scale parameter. Different distributional choices for lead to different models and the most common choice for the distribution of is the Gumbel distribution which is an extreme value distribution. Some these parametric distribution that commonly used and employed in this study is summarized in Table 1 as follows with their respective survival and hazard functions.

| Distribution       | Density function | Survival function | Hazard function | Parameters |
|--------------------|------------------|-------------------|----------------|------------|
| Exponential        | $$\lambda e^{-\lambda t}$$ | $$e^{-\lambda t}$$ | $$\lambda$$ | $$\lambda > 0$$ |
| Weibull            | $$\lambda t^\mu e^{-\lambda t^\mu}$$ | $$e^{-\lambda t^\mu}$$ | $$\lambda t^\mu^{-1}$$ | $$\mu > 0, \lambda > 0$$ |
| Loglogistic        | $$\frac{1}{(1 + \nu (\nu t)^{\nu - 1})}$$ | $$\frac{1}{1 + \nu (\nu t)^{\nu - 1}}$$ | $$\frac{1}{\nu (\nu t)^{\nu - 1}}$$ | $$\nu > 0, \kappa > 0$$ |
| Lognormal          | $$\frac{1}{\sqrt{2\pi}} e^{\frac{-(\log(t) - \mu)^2}{2\sigma^2}}$$ | $$1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$$ | $$\frac{\sigma}{\sqrt{2\pi}} e^{\frac{-(\log(t) - \mu)^2}{2\sigma^2}}$$ | $$\mu > 0, \sigma > 0$$ |
| Generalized Gamma  | $$\frac{\lambda^\mu e^{-\lambda t}}{\Gamma(\nu)}$$ | $$1 - I_\nu(\lambda t)$$ | $$\frac{\lambda^\mu e^{-\lambda t}}{\Gamma(\nu)}$$ | $$\lambda > 0, \kappa > 0$$ |
| Gompertz           | $$\lambda e^{\nu t} e^{-\lambda (e^{\nu t} - 1)}$$ | $$e^{-\lambda (e^{\nu t} - 1)}$$ | $$\lambda e^{\nu t}$$ | $$\lambda > 0, \varphi > 0$$ |

The hazard function based on AFT models were also given by:

$$\lambda(t) = \lambda_0(t \exp(-X_i^T \beta)) \exp(-X_i^T \beta)$$

The effects of the covariates can be observed through $$\exp(-X_i^T \beta)$$ that is the time scale is changed by factor of $$\exp(-\beta)$$ [22].

2.3.1.4 The Estimation of Parametric Survival Models

The parameters of parametric survival models are estimated based on full maximum likelihood estimation methods. The likelihood of $$n$$ observed survival times, $$t_1, t_2, t_3, \ldots, t_n$$, with unknown all required parameter vectors $$\vartheta$$ for right censored data is given by:

$$L(\vartheta) = \prod_{i=1}^{n} f(t_i; \vartheta; X_i)^{\delta_i} S(t_i; \vartheta; X_i)^{1-\delta_i}$$

Where; $$f(t_i; \vartheta; X_i)$$ is the density function. $$S(t_i; \vartheta; X_i)$$ is the survival function and $$\delta_i$$ is the censoring indicator associate with $$i^{th}$$ subjects. This likelihood function can be expanded in terms of a parametric regression model on the hazard rate characterized with a baseline hazard function and a vector of all parameters of the model can be expressed as:

$$L(\vartheta) = \prod_{i=1}^{n} (\lambda_0(t) \exp(X_i^T \beta) )^{\delta_i} \int_0^{t} \lambda_0(s) \exp(X_i^T \beta) ds$$

The likelihood function can be also constructed in terms of the AFT regression perspective. In the presence of right censoring, the likelihood function of an AFT regression model is also specified as:

$$L(\vartheta) = \prod_{i=1}^{n} (\lambda_0(t \exp(-X_i^T \beta)) \exp(-X_i^T \beta) )^{\delta_i} \exp(-\lambda_0(t \exp(-X_i^T \beta)) )$$

Where the parameter vector $$\vartheta$$ now contains different components of the models parameters. Taking logarithm of the above log likelihood equation the maximum likelihood parameters estimates are found based on NewtonRaphson procedure [23].
2.3.2 Frailty Models

The frailty models are survival models which accounts unobservable random effects for the existence of unmeasured attributes such as genotype that do introduce heterogeneity into a study population and the interest lies on the hazard function of subject, conditional on an unknown frailty term for the cluster containing this subjects. It is a multiplicative hazard model consisting of three components: a frailty (random effect), a baseline hazard function (parametric or nonparametric), and a term modeling the influence of observed covariates (fixed effects). The variance of the frailty that represents the heterogeneity and the frailty is greater than one indicates increased hazard to be frailer. For each observed data $y_{ij} = (T_{ij}, \delta_{ij})$, for $i = 1,...,G$ clusters and $j = 1,2,...,n_i$ observations for $i$th cluster $T_{ij}$ is the observed time and $\delta_{ij}$ is the censoring indicator the of hazard of frailty model at time $t$ for subject $j$ in cluster $i$ with covariate information $X_{ij} = (Z_{ij1},...,Z_{ijp})^T$ is defined as:

$$
\lambda_{ij}(t|X_{ij}) = \lambda_0(t)exp(X_{ij}^T\beta + w_i) = \lambda_0(t)u_iexp(X_{ij}^T\beta)
$$

Where, $u_i = exp(w_i)$ is the random effect, $\lambda_{ij}(t)$ is the hazard function for the $j$th individual from the $i$th group, $\lambda_0(t)$ is the baseline hazard at time $t$, and $X_{ij}$ is the matrix of $n$ by $p$ covariates recorded for the individual. The frailty term $u_i$, for $i = 1,...,G$ for each cluster expresses that it assumes different subjects in the same cluster behave in a similar but unknown manner. The consideration frailty term as a realization of a random variable $U$ with a given frailty distribution allow it to vary over the different clusters. These unknown frailty terms with their imposed distribution are used to express the association between the different subjects in a cluster [24,25].

2.3.2.1 Gamma Shared Frailty Distribution

Due to its mathematical convenience, the one-parameter gamma distribution is a popular choice for the distribution of frailty term. From a computational point of view, it fits very well into survival models, because it is easy to derive the formulas for any number of events. The gamma frailty distribution has been widely used in parametric modeling of intra-cluster dependency because of its simple interpretation, flexibility, and mathematical tractability [11, 26] which is specified as:

$$
f_U(u) = u^{\theta-1}exp \left(-\frac{u}{\theta}\right) \frac{\theta^\theta}{\Gamma(\theta)}
$$

where, $\Gamma(.)$ is the gamma function, it corresponds to a Gamma distribution of $Gam(1, \theta)$ with variance $\theta$ and its Laplace transformation is given by:

$$
L(z) = (1 + \frac{z}{\theta})^{-\theta}
$$

For $\theta > 0$ which measures the degree of heterogeneity among clusters and association within groups. The conditional Gamma frailty survival and hazard function is also given by:

$$
S_\theta(t) = [1 - \theta \ast ln(S(t))]^{-\frac{1}{\theta}} \text{ and }\lambda_\theta(t) = \lambda(t)[1 - \theta ln(S(t))]^{-1}
$$

where, $S(t)$ and $\lambda(t)$ are the survival and the hazard functions of the baseline distributions respectively [27].

2.3.2.2 Inverse Gaussian Frailty Model

The inverse Gaussian (inverse normal) distribution was introduced as an alternative to the gamma distribution by Hougaard (1984) and has been used for example by Manton et al.[28], Keiding et al.[29] and Price and Manatunga [30]. It has simple closed-form expressions for the unconditional survival and hazard functions that makes the model attractive and its probability
density function having mean one and variance $\theta$ is given by:

$$f(u) = \frac{1}{\sqrt{2\pi u^2}} e^{-\frac{(u-1)^2}{2u}}$$ For $\theta > 0$, $u > 0$

Where its lapse transformation is given by:

$$L(x) = \frac{1}{\frac{(1+2\theta x)}{\theta}}$$ For $\theta > 0$, $x > 0$

The conditional inverse Gaussian frailty survival and hazard function is given by:

$$S_0(t) = \exp\left(\frac{1}{\theta}[1-(1-2\theta \ln(S(t)))^{\frac{1}{2}}]\right)$$ and $\lambda_0(t) = \lambda(t)(1-2\theta \ln(S(t)))^{-\frac{3}{2}}$

Where; $S(t)$ and $\lambda(t)$ are the survival and the hazard functions of the baseline distributions respectively [27].

### 2.3.2.3 Estimation of Frailty Models

The parameters estimate of frailty models are obtained by maximizing the marginal log-likelihood. Within the context of clustered survival data, let we denotes the observed (event or censoring) times $y_{ij} = (y_{11}, \ldots, y_{Gn_G})^T$ with the censoring indicators $(\delta_{11}, \ldots, \delta_{Gn_G})^T$, the covariate information $x = (x_{11}, \ldots, x_{Gn_G})^T$, and finally the unobserved information $u = (u_1, \ldots, u_G)$ then the conditional likelihood for the $i^{th}$ cluster, with $\Lambda_0(t)$ the cumulative baseline hazard, is given by:

$$L(\phi, \beta | u_1) = \prod_{j=1}^{n_i} (\lambda_0(y_{ij})(u_i \exp(X_{ij}^T \beta)))^{\delta_{ij}} \exp(-\Lambda_0(y_{ij})u_i \exp(X_{ij}^T \beta))$$

Where, $\phi$ is containing the parameters of the baseline hazard. Denoting $\xi = (\phi, \theta, \beta)$ the joint marginal likelihood $L_{\text{marg},i}(\xi)$ for the $i^{th}$ cluster is given as:

$$L_{\text{marg},i}(\xi) = \int_0^\infty \prod_{j=1}^{n_i} (\lambda_0(y_{ij})(u_i \exp(X_{ij}^T \beta)))^{\delta_{ij}} \exp(-\Lambda_0(y_{ij})u_i \exp(X_{ij}^T \beta)) f_{U_i}(u_i) du_i$$ and making some rearrangement to this joint marginal likelihood it is given by:

$$L_{\text{marg},i}(\xi) = \prod_{i=1}^{G} \left[ \prod_{j=1}^{n_i} (\lambda_0(y_{ij})(u_i \exp(X_{ij}^T \beta)))^{\delta_{ij}} \exp(-u_i \sum_{j=1}^{n_i} \exp(X_{ij}^T \beta)) f_{U_i}(u_i) du_i \right]$$

$$L_{\text{marg},i}(\xi) = \prod_{i=1}^{G} \left[ \prod_{j=1}^{n_i} (\lambda_0(y_{ij})(u_i \exp(X_{ij}^T \beta)))^{\delta_{ij}} \exp(-1) \lambda_0^2 \sum_{j=1}^{n_i} \exp(X_{ij}^T \beta) \right]$$

Taking the logarithm, the marginal likelihood is:

$$l_{\text{marg}}(\xi) = \sum_{i=1}^{G} \left( \sum_{j=1}^{n_i} \delta_{ij} \lambda_0(y_{ij}) + X_{ij}^T \beta \right) + \ln((-1)^{\delta_i} \lambda_0^{\delta_i} \sum_{j=1}^{n_i} \exp(X_{ij}^T \beta)))$$

Where; $\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of events in the $i^{th}$ cluster, and $L^{(q)}(.)$ is the $q^{th}$ derivative of the Laplace transform of the frailty distribution defined as:

$$L(x) = E(\exp(-ux)) = \int_0^\infty \exp(-ux) f(u) du \text{ for } x > 0$$

Where; $\phi$ represents a vector of parameters of the baseline hazard function, $\beta$ is the vector of regression coefficients and $\theta$ the variance of the random effect. Accordingly, the estimates of $\phi$, $\beta$, $\delta_\theta$ are obtained by maximizing the marginal log-likelihood following Breslow and Clayton [31] approach.

### 2.3.3 Model Comparisons

To come up with an appropriate model among the candidate model considered for modeling time to pregnancy termination graphical comparison and model information criteria were employed. Graphically, the estimated survival time with considered parametric models were compared with the survival function estimate by KM and to observe how the estimated survival models by candidate models were closer(same) to the estimated survival time compared with the KM estimated survival time. In addition to the graphical techniques AIC and BIC of the fitted model information criteria were also considered to compare the fitted models where the AIC and BIC of the model is given as follows:
\[ AIC = -2 \log(L) + 2k \] 
\[ BIC = -2 \log(L) + \log(nk) \]

Where, \( k \) is the number all parameters for the estimated model and \( n \) is sample size and estimated model with minimum AIC and BIC values was considered as an appropriate model [32,33]

## 3 Results

### 3.1 Descriptive Statistics of Pregnancy Termination and Log Rank Test for Estimated Survival Difference

The study considered a total of 15,683 women in reproductive age groups from nine regions and two city administration of Ethiopia. Among these women 1236(7.88\%) of them experienced abortion whereas about 14,447(92.11\%) of women were censored (had not experienced) abortion when the survey was conducted. The average age of women who had not experienced abortion was estimated 27.43 years with standard deviation 9.07 years and they had the first sex at an average age of 17.46 years whereas the average of women who had experienced abortion was estimated 33.81 years with standard deviation of 8.13 years and these women also had their first sex at an average age 16.62 years.

Regarding the regional distribution of abortion prevalence, larger proportion (174(10.30\%)) of women in regional state had experienced abortion followed by women in Afar, Amahara an Oromia regional state women accounting 107(9.50\%), 151(8.80\%) and 147(7.80\%) among the total women considered in the region respectively. In contrast to this result, lower proportion women who had experience abortion were found in Benishangul gumuz and Gambella regional states accounting 43(4.20\%) and 49(4.40\%) of women among the total women considered in the regions respectively. Regarding place of residences, about 435(8.10\%) of women and 801(7.80\%) women had experienced abortion. Similarly, 1046(10.60\%) of married or in relationship women, 44(9.80\%) of windowed women and116(10.30) divorced/separated women experienced abortion in comparison with women who had not experienced abortion.

Furthermore, among the total women who had experienced abortion 475(8.00\%) of them, 158(7.90\%) and 603(7.80\%) of them from poor, middle income, and rich family background comparison with women who had no history of abortion. On the same way, about 661(11.20\%) every tried to avoid abortion whereas about 575(5.90\%) do not in comparison with women who had no history of abortion during the survey. Regrading use of contraceptive methods, among the women who had the history of abortion 927(7.50\%) used these methods whereas 309(9.30\%) do not used and about 299(6.89\%) of them faced the big problem on getting medical help whereas 937(8.26\%) of them do not faced the big problem in getting medical helps. Regarding the women background on frequency of smoking and chewing chat, among women who had history of abortion about 1097(7.77\%) of the women chew chat whereas 139(8.93\%) do not chew chat and about 1223(7.90\%), 8(5.59\%) and 5(9.09\%) of women were none smokers, every day smokers and sometimes smokers respectively in comparison with women who had no history abortion.

Regarding prevalence of abortion by religious categories, is was more prevalent accounting 7(9.70)in other religion followers even if the sample size is smaller for this religion categories followed by orthodox religion followers which accounts 564(8.80\%) of the cases in comparison with women who had no history of abortion. It was also less prevalent in catholic religion followers which accounts 5(5.50\%) of the total catholic religion followers followed by protestant religion followers.
| Covariates       | Categories | Pregnancy Termination Status | Log rank test |
|------------------|------------|------------------------------|---------------|
|                  |            | Censored n(%) | Event n(%) | Chi-square value (DF) | P-value |
| Region           | Tigray     | 1508(89.70) | 174(10.30) | 88.3(10) | < 0.001 |
|                  | Afar       | 1021(90.50) | 107(9.50)  |           |         |
|                  | Amhara     | 1568(91.20) | 151(8.80)  |           |         |
|                  | Oromia     | 1745(92.20) | 147(7.80)  |           |         |
|                  | Somali     | 1285(92.40) | 106(7.60)  |           |         |
|                  | Benishangul| 1077(95.60) | 49(4.40)   |           |         |
|                  | SNNPR      | 1728(93.50) | 121(6.50)  |           |         |
|                  | Gambela    | 992(95.80)  | 43(4.20)   |           |         |
|                  | Harari     | 847(93.50)  | 59(6.50)   |           |         |
|                  | Addis Ababa| 1643(90.10) | 181(9.90)  |           |         |
|                  | Dire Dawa  | 1033(91.30) | 98(8.70)   |           |         |
| Place of residence| Urban     | 4913(91.90) | 435(8.10)  | 6.5(1)    | 0.01    |
|                  | Rural      | 9534(92.20) | 801(7.80)  |           |         |
| Marital status   | Single     | 4248(99.30) | 30(0.70)   | 85.6(3)   | < 0.001 |
|                  | Married    | 8778(89.40) | 1046(10.60)|           |         |
|                  | Widowed    | 407(90.20)  | 44(9.80)   |           |         |
|                  | Divorced(separated) | 1014(97.90) | 16(2.10)  |           |         |
| Religion         | Orthodox  | 5849(91.20) | 564(8.80)  | 22.2(5)   | 0.092   |
|                  | Catholic  | 86(94.50)   | 5(5.50)    |           |         |
|                  | Protestant| 2647(94.10) | 167(5.90)  |           |         |
|                  | Muslin    | 5721(92.10) | 488(7.90)  |           |         |
|                  | Traditional| 79(94.00)  | 5(6.00)    |           |         |
|                  | Other     | 65(90.30)   | 7(9.70)    |           |         |
| Wealth index     | Poor      | 5465(92.00) | 475(8.00)  | 1.0(2)    | 0.6     |
|                  | Middle    | 1844(92.10) | 158(7.90)  |           |         |
|                  | Rich      | 7138(92.20) | 603(7.80)  |           |         |
| Trying to avoid pregnancy | Not Tried   | 9223(94.10) | 575(5.90)  | 47.30(1) | < 0.001 |
|                  | Tried     | 5224(88.80) | 661(12.20) |           |         |
| Contraceptive Use| Nonuser   | 11444(92.50) | 927(7.50)  | 1.50(1)   | 0.2     |
|                  | User      | 3003(90.70) | 309(9.30)  |           |         |
| Internet use     | Nonuser   | 13175(91.90) | 1154(8.10) | 1.8(1)    | 0.2     |
|                  | User      | 1272(93.90) | 82(6.10)   |           |         |
| Frequency of smoking | Nonsmoker | 14262(92.10) | 1223(7.90) | 5.8(2)    | 0.05    |
|                  | Every day | 135(94.41)  | 8(5.59)    |           |         |
|                  | Sometimes | 50(90.91)   | 5(9.09)    |           |         |
| Getting medical help | Big problem | 4042(93.11) | 299(6.89)  | 8.5       | 0.003   |
|                  | Not a big problem | 10405(91.74) | 937(8.26)  |           |         |
| Chewing chat     | No        | 13030(92.23) | 1097(7.77) | 4.9(1)    | 0.03    |
|                  | Yes       | 1417(91.07) | 139(8.93)  |           |         |

| Quantitative Covariates | Censored Mean (SD) | Event Mean (SD) |
|-------------------------|--------------------|-----------------|
| Age                     | 27.43(9.07)        | 33.81(8.13)     |
| Year of education       | 4.16(4.73)         | 3.21(4.64)      |
| Total number of children ever born | 2.54(2.82)   | 3.83(2.78)      |
| Total number of sex partner | 1.32(5.81)   | 2.20(7.6)       |
| Age at first sex        | 17.46(3.8)         | 16.62(3.45)     |

On average, censored women spent 4.16 years on education and women who had history of abortion spent 3.12 years. The average total number of children ever born was 2.54 in non-abortion women and the estimated total number of total children ever born for women who had history abortion was 3.83 children. This descriptive result indicated that year of education spent
by women who had history of abortion was lower than that of women who had not, and the total number of children that aborted women had was greater than that of women who do not have abortion history. Similarly, the average total number of sex partners (2.2) of women who had history of abortion was also greater than the average total number of sex partners (1.32) of women who had no history of abortion.

To identify the estimated survival function difference among the categorical covariates considered in the study, estimated Kaplan-Meier survival curves for some of these covariates were depicted in Figure 1 and 2. As can be visualized from the Figure 1, women who reside in rural areas have longer estimated time to pregnancy termination than women residing urban areas whereas women from rich family background have shorter survival estimated time to pregnancy termination compared with women from poor family background. In the same way, single women have longer estimated survival function curve than compared with married women and women who do not use internet have also longer time to pregnancy termination compared with women who do not use internet as depicted on estimated KM survival curve plot.

Figure 1: The KM estimated Survival function curve for place of residence, Wealth index, internet use and Marital status

Figure 2 of estimated KM survival curve also depicted that women who had the big problem in getting medical help have longer time to pregnancy termination than women who do not have the big problem in getting medical help whereas women who use contraceptive methods have shorter time to pregnancy termination than women who used contraceptive methods. With similar KM estimated curve comparisons women who chews chat and ever tried to avoid pregnancy have shorter time to pregnancy termination compared with women’s who do not chew chat and who haven’t ever tried to avoid pregnancy termination. But, visualizing estimated KM curves does not tell us clear significance difference among the estimated survival time by groups. Therefore, to reveal the significance difference among the estimated KM curve log rank test was employed. The log rank tested result also depicted that there was significant estimated survival curve difference by place of residences, marital status, ever
trying to avoid pregnancy, getting medical help and chewing chat whereas there was no significance difference among estimated KM curves by contraceptive use, internet use, wealth index, religion and frequency of smoking at 5% level of significance as depicted in Table 2 result.

3.2 Survival Modeling Results

To reach on an appropriate model that appropriately fits time to pregnancy termination in Ethiopia, the survival model comparison begins with graphical comparison of estimated survival time by KM of non parametric methods and survival time estimated by parametric methods. As depicted in Figure 3 of estimated survival curve coloured lines represents survival probability estimated by parametric methods and the black lines represents Kaplan-Meier estimate and the plot shows survival estimate from the generalised gamma model appears to match the Kaplan-Meier estimate most closely throughout during reproductive age followed by lognormal survival estimates. Similarly, as observed from survival estimates by the Weibull and Gompertz, the estimated survival curve by Weibull was preferable than Gompertz since its was closer to KM estimates whereas the estimated survival time by Exponential model was very far from non parametric estimates and other parametric estimates which can be considered as the worse estimate of survival time to pregnancy termination.
In addition to graphical methods, all the candidate survival models were compared based on fitted model information criteria. Before fitting the models, step wise automatic variable selection were employed to all candidate survival models in order to identify significant independent variables of time to pregnancy termination. Accordingly, the same covariates were selected for both cox and Weibull models whereas frequency of smoking and total number of sex partner were excluded variable in exponential model automatic variable selection and wealth index were included in additionally in log logistic model automatic variable selection. For the seek of consistency among models the same predictor variables were consider for all candidate models. The considered predictor variables were also age of women, religion, use of contraceptive methods, marital status, Ever trying to avoid pregnancy, age at firs sex, total children ever born, use of internet ,chewing chat ,total sex partner, getting medical help ,frequency of smoking and place of residence of women.

The result of these fitted models also showed even if the Cox PH model was estimated based on the partial likely hood function it has larger information criteria than the parametric survival model showing the considered parametric survival models were appropriate fit than Cox PH model. Regarding the estimated parametric survival modes comparison result, the generalized gamma survival model with estimated AIC value of 6691.924 and BIC value of 6845.131 has minimum information criteria compared with other models followed by Lognormal survival models which has estimated AIC value of 6708.184 and BIC value of 6853.73 as depicted in Table 3 were considered as appropriate fit of the data than the remaining candidate survival models which also confirmed with graphical comparison result.
Table 3: Estimated survival models information criteria

| Candidate models   | DF | -2Log likelihood | BIC        | AIC        |
|--------------------|----|------------------|------------|------------|
| Cox PH             | 17 | -10769.82        | 21703.86   | 21573.63   |
| Weibull            | 19 | -3368.521        | 6920.589   | 6775.043   |
| Exponential        | 18 | -4210.021        | 8593.928   | 8456.042   |
| log logistic       | 19 | -3361.723        | 6906.992   | 6761.446   |
| Log normal         | 19 | -3335.092        | 6853.73    | 6708.184   |
| Gopertez           | 19 | -3516.599        | 7216.745   | 7071.198   |
| Generalized Gamma  | 19 | -3325.962        | 6845.131   | 6691.924   |

3.2.1 The Generalized Gamma and Lognormal Survival Models Results

As depicted on Table 4 below in both model age of women, wealth index, marital status, history of trying to avoid pregnancy, age at first sex, total number of children ever born, ever tried to avoid pregnancy, use of contraceptive methods, number of total number children ever born, age at fist sex, marital status, challenge of getting medical help and place of residence have p-value less 5% level of significant and showed that they have significant effects on time to pregnancy termination of women in both cases. But when we compared the estimated coefficients and their standard error the magnitude of estimated value for the lognormal survival model was lower than that of generalized Gamma survival model. This lower estimated standard error of the estimates showed even if the Gamma survival model was a good fit of the data than the lognormal survival model, the lognormal survival model was more efficient in estimating the model parameters than the generalized Gamma survival model.
Table 4: Estimated generalized gamma and lognormal survival models

| Factors                        | Generalized gamma model | Lognormal model |
|--------------------------------|-------------------------|-----------------|
|                               | Coeff.(SE)   | Z     | P-value | Coeff.(SE)   | Z     | P-value |
| Age                           | 0.004(0.002) | 2.240 | 0.025   | 0.008(0.001) | 5.760 | 0.000   |
| Wealth index                  |             |       |         |             |       |         |
| Poor(Ref.)                    |             |       |         |             |       |         |
| Middle income                 | 0.022(0.027) | 0.790 | 0.428   | 0.025(0.026) | 0.970 | 0.332   |
| Rich                          | 0.067(0.025) | 2.670 | 0.008   | 0.058(0.024) | 2.450 | 0.014   |
| Using internet                |             |       |         |             |       |         |
| Nonuser(Ref.)                 |             |       |         |             |       |         |
| User                          | -0.052(0.039) | -1.330 | 0.183   | -0.060(0.036) | -1.650 | 0.100   |
| Trying to avoid pregnancy     |             |       |         |             |       |         |
| Tried                         | -0.096(0.023) | -4.260 | 0.000   | -0.099(0.021) | -4.820 | 0.000   |
| Contraceptive Use             |             |       |         |             |       |         |
| Nonuser                       |             |       |         |             |       |         |
| User                          | 0.103(0.025) | 4.160 | 0.000   | 0.105(0.023) | 4.610 | 0.000   |
| Total children ever born      | 0.035(0.005) | 7.260 | 0.000   | 0.032(0.004) | 7.250 | 0.000   |
| Age at first sex              | 0.028(0.003) | 9.100 | 0.000   | 0.022(0.003) | 8.590 | 0.000   |
| Marital status                |             |       |         |             |       |         |
| Single                        |             |       |         |             |       |         |
| married                       | -0.361(0.042) | -8.620 | 0.000   | -0.412(0.043) | -9.480 | 0.000   |
| widowed                       | -0.236(0.058) | -4.040 | 0.000   | -0.273(0.059) | -4.640 | 0.000   |
| Divorced(separated)           | -0.269(0.049) | -5.530 | 0.000   | -0.320(0.049) | -6.500 | 0.000   |
| Total sex partner             | -0.001(0.001) | -0.920 | 0.356   | -0.001(0.001) | -1.290 | 0.198   |
| Chewing chat                  |             |       |         |             |       |         |
| No                            | -0.045(0.028) | -1.600 | 0.109   | -0.049(0.026) | -1.900 | 0.057   |
| Yes                           |             |       |         |             |       |         |
| Place of residence            |             |       |         |             |       |         |
| Urban                         | 0.063(0.027) | 2.350 | 0.019   | 0.063(0.025) | 2.550 | 0.011   |
| Rural                         |             |       |         |             |       |         |
| Frequency of smoking          |             |       |         |             |       |         |
| Sometimes                     |             |       |         |             |       |         |
| Do not smokes                 | -0.107(0.134) | -0.800 | 0.424   | -0.071(0.125) | -0.570 | 0.572   |
| Every Smoker                  | 0.108(0.159) | 0.680 | 0.499   | 0.140(0.152) | 0.920 | 0.358   |
| Getting medical help          |             |       |         |             |       |         |
| not big problem               |             |       |         |             |       |         |
| Big problem                   | 0.059(0.020) | 2.960 | 0.003   | 0.050(0.019) | 2.670 | 0.007   |
| Constant                      | 3.721(0.157) | 23.710 | 0.000   | 3.756(0.147) | 25.560 | 0.000   |
| \(\ln(\hat{\sigma})\)       | -0.281(0.067) | -4.190 | 0.000   | -0.662(0.024) | -27.690 | 0.000   |
| \(\ln(\hat{\kappa})\)       | -0.979(0.254) | -3.860 | 0.000   | - -            | - -      | - -     |
| \(\hat{\sigma}\)            | 0.755(0.051) | -      | -       | 0.516(0.012) | -      | -       |

**Likelihood ratio value = 378.55, p-value = < 0.0001**

**Likelihood ratio value = 446.01, P-value = < 0.0001**

### 3.3 Frailty Models Comparison Result

To select an appropriate frailty model among ten candidates models fitted with Gamma and inverse Gaussian frailty distributions the models were compared based on their information criteria. Even if the generalized Gamma survival model with none frailty was considered as an appropriate fit as depicted in Table 3 result due the complexity of model it did not achieved convergency and was excluded from the frailty model candidates. As depicted in Table 3 of estimated none frailty survival models information criteria and Table 5 of frailty models information criteria, compared with none frailty models the frailty models counterpart has
minimum AIC and BIC values showing there was an improvement of the survival model due to consideration of the random effects in the frailty models which handles unobservable regional heterogeneity of time to pregnancy termination.

Regarding the frailty models comparison, as depicted Table 5 of the fitted frailty models information criteria result, exponential with frailty model has larger AIC and BIC value showing that it was the worst fit than the other parametric frailty models whereas lognormal frailty model which has minimum information criteria compared to other candidate models was more appropriate fit of time to pregnancy termination. Furthermore, among the fitted two lognormal frailty model with two different frailty distributions, Gamma frailty lognormal model with AIC value of 6643.63 and BIC value of 6796.837 was minimum compared to Inverse Gaussian frailty lognormal model. Therefore, the lognormal survival model with Gamma frailty distribution was considered as an appropriate fit for the time to pregnancy termination in Ethiopia.

Table 5: Fitted frailty models information criteria comparison

| Model          | Gamma Frailty | Inverse Guassian frailty |
|----------------|---------------|--------------------------|
| Exponential    | -4183.32      | 8404.633                 |
| Weibull        | -3337.52      | 6715.034                 |
| lognormal      | -3301.82      | 6643.63                  |
| loglogistic    | -3330.16      | 6700.317                 |
| Gopertez       | -3486.62      | 7013.247                 |
|                | AIC           | BIC                      |
|                | 8550.179      | 8550.205                 |
|                | 8686.241      | 8686.837                 |
|                | 6796.377      | 6796.837                 |
|                | 6853.524      | NC                       |
|                | 7166.454      | NC                       |
|                | 8550.751      | 8550.751                 |
|                | 6868.837      | 6868.837                 |
|                | 6797.3        | 6797.3                   |
|                | NC            | NC                       |
|                | 7167.01       | 7167.01                  |

3.3.1 Risk Factors Associated to Time to Pregnancy Termination Based on Lognormal Gamma Frailty Model

As showed in estimated lognormal frailty survival model in Table 6 age, total number of children ever born and age at fist sex have positive effects on time to pregnancy termination of women. The estimated random effect (the frailty term) was also significant showed that there was time to pregnancy termination differences across the Ethiopian regions and which was an indication for the existence of unobservable heterogeneity in the data considered for the study due to the clustering effects by region. The estimated coefficients of log normal with gamma frailty model can be also interpreted in different ways and the simplest of interpreting it is by exponentiating the coefficient which represents the time ration. The time ration for the case of log normal survival model survival model represents the relative risk which was interpreted as follow for the time to pregnancy termination in Ethiopia.

The estimated time ratio for the age 1.008 showed, the relative risk time to pregnancy termination will increases by 0.8% with unit increment of age of women in years holding other covariates constant and showing age of women have contribution for women pregnancy termination at early age. Regarding every trying to avoid pregnancy on time to pregnancy terminations, for ever tried to avoid pregnancy women the risk to time to pregnancy termination was 10.10% earlier than women who havent tried to avoid pregnancy termination and this result showed that the risk of time to pregnancy termination was higher in women who have tried to avoid pregnancy. Similarly, using contraceptive methods has significant effects and its estimated time ration showed time to pregnancy termination for women who use contraceptive method prolongs 10.40% in comparison with non-contraceptive user women. Furthermore, regarding the effects of getting medical help women who had big problem to access medical help has 5.7% longer time to get pregnancy termination compared with women with no big problem of getting medical help. In Ethiopian context most of the time women who do not get medical help is women’s who resides in rural areas.
and the result of this study also revealed that residence has significant effects which showed that women’s who resides in rural areas has 0.1% longer time to pregnancy termination in comparison with women residing in urban areas.

Another significant risk factors showed in this model result was age at first sex and total number of children ever born have positive effects on time to pregnancy termination and their estimated time ration showed that the time to pregnancy termination will increases by 2.3% and 3.4% with unit increment of age at fist sex in years and total number one children respectively holding the effects of other factors constant. Chewing chat and wealth index is also another significant risk factors at 10% level of significance that showed women’s who Chews chat women and from poor family has early time to pregnancy termination compared with women who do not chews chat and from rich family background respectively. Regarding the significant effects of marital status categories on time to pregnancy termination, time to pregnancy termination were 33.6%, 24.00% and 26.5% earlier in married, widowed, divorced(separated) women categories respectively in comparison with single or women which were not in any form of relation ship women holding the effects of other variables constant.
Table 6: Lognormal gamma frailty model result

| Factors                        | Coeff.(SE)   | Z     | P-value | 95% CI       | RR=exp(coeff.) |
|-------------------------------|--------------|-------|---------|--------------|----------------|
| Age                           | 0.008(0.001) | 5.660 | 0.000   | [0.005,0.011] | 1.008          |
| Wealth index                  |              |       |         |              |                |
| Poor                          |              |       |         |              |                |
| Middle income                 | 0.008(0.027) | 0.280 | 0.778   | [-0.045,0.060] | 1.008          |
| Rich                          | 0.045(0.025) | 1.790 | 0.073   | [-0.004,0.093] | 1.046          |
| Using internet                |              |       |         |              |                |
| Nonuser                       |              |       |         |              |                |
| User                          | -0.032(0.035)| -0.920| 0.360   | [-0.101, 0.037]| 0.969          |
| Tried to avoid pregnancy      |              |       |         |              |                |
| Never                         |              |       |         |              |                |
| Tried                         | -0.107(0.022)| -4.920| 0.000   | [-0.149,-0.064]| 0.899          |
| Contraceptive use             |              |       |         |              |                |
| Nonuser                       |              |       |         |              |                |
| User                          | 0.099(0.023) | 4.350 | 0.000   | [0.054,0.144] | 1.104          |
| Total children ever born      | 0.033(0.005) | 7.310 | 0.000   | [0.024,0.042] | 1.034          |
| Age at first sex              | 0.023(0.003) | 9.060 | 0.000   | [0.018,0.029] | 1.023          |
| Marital status                |              |       |         |              |                |
| Single                        |              |       |         |              |                |
| married                       | -0.409(0.043)| -9.560| 0.000   | [-0.493,-0.325]| 0.664          |
| widowed                       | -0.274(0.058)| -4.710| 0.000   | [-0.388,-0.160]| 0.760          |
| Divorced(separated)           | -0.308(0.048)| -6.390| 0.000   | [-0.403,-0.214]| 0.735          |
| Total sex partner             | -0.002(0.001)| -1.470| 0.140   | [-0.004,0.001] | 0.998          |
| Chewing chat                  |              |       |         |              |                |
| No                            | -0.048(0.026)| -1.860| 0.063   | [-0.099, 0.003]| 0.953          |
| Place of residence            |              |       |         |              |                |
| Urban                         |              |       |         |              |                |
| Rural                         | 0.001(0.028) | 0.050 | 0.960   | [-0.054,0.057] | 1.001          |
| Frequency of smoking          |              |       |         |              |                |
| Sometimes                     |              |       |         |              |                |
| Do not smokes                 | -0.012(0.130)| -0.090| 0.926   | [-0.267,0.242] | 0.988          |
| Every Smoker                  | 0.116(0.159) | 0.730 | 0.465   | [-0.195,0.427] | 1.123          |
| Getting medical help          |              |       |         |              |                |
| not big problem               |              |       |         |              |                |
| Big problem                   | 0.055(0.020) | 2.760 | 0.006   | [0.016,0.094] | 1.057          |
| Constant                      | 3.688(0.157) | 23.530| 0.000   | [3.381,3.995] | 39.965         |
| $ln(\hat{\sigma})$           | -0.675(0.030)| -22.170| 0.000   | [-0.734,-0.615] |                |
| $ln(\hat{\theta})$           | -2.209(0.490)| -4.510| 0.000   | [-3.170,-1.249] |                |
| $\hat{\sigma}$               | 0.509(0.016) |       |         | [0.480,0.541] |                |
| $\hat{\theta}$               | 0.110(0.054) |       |         | [0.042, 0.287] |                |

LRT chi-square= 454.71, p-value= 0.00

4 Discussion

The main of this study was to model time to pregnancy termination and the associated risk factors by applying an appropriate survival model which considers the nature of the data considered for the study. Before directly proceeding to modeling part preliminary analysis was done using exploratory data analysis which helps to know the prevalence of pregnancy termination in Ethiopia and none parametric methods based on Kaplan-Meier (KM) estimates of survival curve was also employed to compare
the estimated survival curve of time to pregnancy termination by different categorical covariates to look for the existence time to pregnancy termination difference between the groups and their significance difference was also tested based on log rank test. The test log rank test result also revealed that there were a significant estimate survival curve difference of time to pregnancy termination by region, place of residence, ever trying to avoid pregnancy, getting medical help and chewing chat at 5% level of significance and significant estimated survival curve difference by religion and frequency of smoking of women at 10% level of significance.

The selection of risk factors for the time to pregnancy termination was done based on automatic variable selection for all models and the same predictor variables based on variable selection techniques were considered to fit all candidate models for comparison reasons based on AIC and BIC of the fitted model. The fitted survival models comparison result revealed that generalized gamma and log normal survival models was selected based on their minimum values of AIC and BIC compared with other candidate models for the non-frailty survival model. To check for the improvement of the models all the parametric survival models were refitted with inclusion of frailty term in the model with consideration region as the clustering effects which handle unobservable regional heterogeneity of time to pregnancy termination in Ethiopia. The result these estimated frailty models also showed that improvements of the refitted models due to consideration of frailty effects in the model and the estimated frailty parameters also has significant effect showing there was observable regional heterogeneity in the data due to the clustering nature by administrative regions of Ethiopia. The studies conducted by Gutierrez[10], Vaupel et al.[11], Lancaster[13] and Duchateau and Janssen [25] recommended the importance and performance of frailty models to handle unobservable heterogeneity when there is clustering nature in the data considered for given study.

The final estimated model results also depicted that an increase in current women age was one the risk factor which prolongs time to pregnancy termination. The study conducted by dos Santos et al. [34] in Brazil; Woldeamanuel, B.T.[35] in Ethiopian; Yogi, A., et al. [36] in Nepal and Tesfaye et al.[37] in Ethiopia on associated of factors abortion based on odds ratio showed age of women has significant effect on the pregnancy termination and the risk of pregnancy termination becomes lowered as women goes older.

The study conducted by Tesfaye, G., et al. [37] on induced abortion and associated factors in health facilities of Guraghe Zone of Southern Ethiopia based on logistic regression model showed that unwanted pregnancy increased the risk of induced pregnancy termination. This study also revealed the significant effects of ever trying to avoid pregnancy showing women who tried to avoid pregnancy experience pregnancy termination at early age in comparison with women who never tried to avoid pregnancy. Similar, the result of this study showed that using contraceptive methods and residing in rural areas in comparison with women who do not use contraceptive and women who resides in urban areas have significant effects which prolongs time to pregnancy termination. The study conducted by Senbeto, E.[38] based logistic regression revealed that place of residence and using contraceptive methods were significantly associated with pregnancy termination and similar result by Woldeamanuel [35] and Bago BJ et al.[39] found that use of contraceptive methods reduces the risk of pregnancy termination which confirms the effects of factors which have effects on pregnancy termination.

Furthermore, the result of this study revealed that starting sex at early age, having less total number of children, not having a big problem for getting medical help and chewing chat were significant risk factors that results in pregnancy termination at early age of women. The study conducted by Behulu et al. [40] on repeat induced abortion and associated factors among
reproductive age women who seek abortion services in Debre Berhan town health institutions, Central Ethiopia, 2019 based on the odds ratio, found age of the first sexual intercourse less than 18 years positively associated with the repeatedly induced abortion among women who sought abortion services.

5 Conclusion

The study identified there were the estimated time to pregnancy termination curve difference by place of residence, marital status, ever trying to avoid pregnancy, frequency of smoking, getting medical help and chewing chat chewing chat based on the result of log rank test of Kaplan-Meier estimated survival curve.

With no consideration shared frailty, generalize gamma and lognormal survival models were appropriate models and lognormal survival model was more efficient than the generalized gamma based on the estimate model results. Compared to the none frailty survival models the estimated frailty models which accounts unobservable regional heterogeneity were more appropriate and improves the none frailty survival models showing the existence of time to pregnancy termination difference among the administrative regions of Ethiopia. Among the fitted frailty models, lognormal survival model with gamma frailty distribution was an appropriate fit of time to pregnancy termination.

The fitted model result also showed that, women current age, trying to avoid pregnancy, use of contraceptive methods, total number of children ever born, and age at first sex, marital status, chewing chat, place of residence and getting medical help were the risk factors associated with time to pregnancy termination of women in Ethiopia. Furthermore, the model depicts starting sex at early age, not using contraceptive methods compared with contraceptive method user group, chewing chat in comparison with not chewing chat group, residing in urban area comparison with rural group, married marital status in comparison with single marital status group and where there no big problem of getting medical help in comparison with where there is a big problem of getting medical help women accelerates time to pregnant termination in Ethiopia that needs great attention to lower the risk of pregnancy termination.

List of abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
CSA: Central Statistical Agency
EDHS: Ethiopia Demographic and Health Survey
KM: Kaplan-Meier
PH: Proportional hazard
Compliance with ethical standards

Acknowledgements I acknowledge the Ethiopian central statistical agency for providing me the data.

Conflict of interest The authors declare that they have no competing interests.

Availability of data and materials I can provide the data set that has been used during the current study on reasonable request.

References

1. Harvard Publishing, 2019. Abortion (Termination Of Pregnancy) - Harvard Health [Internet]. Available from: https://www.health.harvard.edu/medical-tests-and-procedures/abortion-termination-of-pregnancy-a-to-z

2. Bradley, S.E., Croft, T. and Rutstein, S.O., 2011. The impact of contraceptive failure on unintended births and induced abortions: estimates and strategies for reduction. Calverton eMaryland Maryland: ICF Macro.

3. Sanchez-Pez, D.A. and Ortega, J.A., 2019. Reported patterns of pregnancy termination from Demographic and Health Surveys. PloS one, 14(8).

4. WHO, 2017. Worldwide, an estimated 25 million unsafe abortions occur each year [Internet]. Available from: https://www.who.int/news-room/detail/28-09-2017-worldwide-an-estimated-25-million-unsafe-abortions-occur-each-year

5. Guttmacher Institute, 2019. Abortion in Africa [Internet]. Available from: https://www.guttmacher.org/factsheet/abortion-africa

6. Mdecins Sans Frontires (MSF) International, 2019. Unsafe abortion: a forgotten emergency — Women's health — MSF [Internet]. Available from: https://www.msf.org/unsafe-abortion-forgotten-emergency-womens-health

7. Tesfaye, G., Hambisa, M.T. and Semahegn, A., 2014. Induced abortion and associated factors in health facilities of Guraghe zone, southern Ethiopia. Journal of pregnancy, 2014.

8. Gebremedhin, M., Semahegn, A., Usmael, T. and Tesfaye, G., 2018. Unsafe abortion and associated factors among reproductive aged women in Sub-Saharan Africa: a protocol for a systematic review and meta-analysis. Systematic reviews, 7(1), p.130.

9. Moore, A.M., Gebrehiwot, Y., Fetters, T., Wado, Y.D., Bankole, A., Singh, S., Gebreselassie, H. and Getachew, Y., 2016. The estimated incidence of induced abortion in Ethiopia, 2014: changes in the provision of services since 2008. International perspectives on sexual and reproductive health, 42(3), p.111.

10. Gutierrez, R.G., 2002. Parametric frailty and shared frailty survival models. The Stata Journal, 2(1), pp.22-44.

11. Vaupel, J.W., Manton, K.G. and Stallard, E., 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, 16(3), pp.439-454.
12. Lancaster, T., 1979. Econometric methods for the duration of unemployment. Econometrica: Journal of the Econometric Society, pp.939-956.

13. Hougaard, P., 1984. Life table methods for heterogeneous populations: distributions describing the heterogeneity. Biometrika, 71(1), pp.75-83.

14. Central Statistical Agency (CSA)[Ethiopia] and ICF, 2016. Ethiopia Demographic and Health Survey 2016: Key Indicators Report. Addis Ababa, Ethiopia, and Rockville, Maryland, USA. CSA and ICF.

15. Stevenson, M. and EpiCentre, I.V.A.B.S., 2009. An introduction to survival analysis. EpiCentre, IVABS, Massey University.

16. Wintrebert, C.M., Putter, H., Zwinderman, A.H. and Van Houwelingen, J.C., 2004. Centreffect on survival after bone marrow transplantation: application of timedependent frailty models. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 46(5), pp.512-525.

17. Aalen, O., Borgan, O. and Gjessing, H., 2008. Survival and event history analysis: a process point of view. Springer Science & Business Media.

18. Cox, D.R., 1972. Regression models and lifetables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), pp.187-202.

19. Hosmer, D.W., Lemeshow, S. and May, S., 2008. Model development. Applied Survival Analysis: Regression Modeling of Time-to-Event Data, pp.132-168.

20. Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), pp.457-481.

21. Fleming, T.R. and Harrington, D.P., 2011. Counting processes and survival analysis (Vol. 169). John Wiley & Sons.

22. Collett, D., 2003. Modeling Survival Data in Medical Research. Chapman & Hall, London

23. Vaupel, J.W. and Yashin, A.I., 1985. Heterogeneity’s ruses: some surprising effects of selection on population dynamics. The American Statistician, 39(3), pp.176-185.

24. Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(1), pp.141-151.

25. Duchateau, L. and Janssen, P., 2007. The frailty model. Springer Science & Business Media.

26. Clayton, D. and Cuzick, J., 1985. Multivariate generalizations of the proportional hazards model. Journal of the Royal Statistical Society: Series A (General), 148(2), pp.82-108.

27. Gutierrez, R.G., 2002. Parametric frailty and shared frailty survival models. The Stata Journal, 2(1), pp.22-44.

28. Manton, K.G., Stallard, E. and Vaupel, J.W., 1986. Alternative models for the heterogeneity of mortality risks among the aged. Journal of the American Statistical Association, 81(395), pp.635-644.

23
29. Keiding, N., Andersen, P.K. and Klein, J.P., 1997. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. Statistics in medicine, 16(2), pp.215-224.

30. Price, D.L. and Manatunga, A.K., 2001. Modelling survival data with a cured fraction using frailty models. Statistics in medicine, 20(910), pp.1515-1527.

31. Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American statistical Association, 88(421), pp.9-25.

32. Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), pp.716-723.

33. Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics, 6(2), pp.461-464.

34. dos Santos, A.P.V., Coelho, E.D.A.C., Gusmo, M.E.N., da Silva, D.O., Marques, P.F. and Almeida, M.S., 2016. Factors associated with abortion in women of reproductive age. Revista Brasileira de Ginecologia e Obstetricia/RBGO Gynecology and Obstetrics, 38(06), pp.273-279.

35. Woldeamanuel, B.T., 2019. Assessment of determinant factors of pregnancy termination among women of reproductive age group in Ethiopia: Evidence from 2016 Ethiopian Demographic and Health Survey. Int J Sex Reprod Health Care, 2(1), pp.010-015.

36. Yogi, A., Prakash, K.C. and Neupane, S., 2018. Prevalence and factors associated with abortion and unsafe abortion in Nepal: a nationwide cross-sectional study. BMC pregnancy and childbirth, 18(1), p.376.

37. Tesfaye, G., Hambisa, M.T. and Semahegn, A., 2014. Induced abortion and associated factors in health facilities of Guraghe zone, southern Ethiopia. Journal of pregnancy, 2014.

38. Senbeto, E., Alene, G.D., Abesno, N. and Yeneneh, H., 2005. Prevalence and associated risk factors of Induced Abortion in Northwest Ethiopia. Ethiopian Journal of health development, 19(1), pp.37-44.

39. Bago, B.J., Hibstu, D.T. and Woldemariam, S.H., 2017. Prevalence of Pregnancy Termination and Its Associated Factors Among Women of Reproductive Age Group in Ethiopia Using 2011 Ethiopian Demography and Health Survey, 2016. J Preg Child Health, 4(359), p.2.

40. Behulu, G.K., Fenta, E.A. and Aynalem, G.L., 2019. Repeat induced abortion and associated factors among reproductive age women who seek abortion services in Debre Berhan town health institutions, Central Ethiopia, 2019. BMC research notes, 12(1), p.499.