ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

SURVIVAL ANALYSIS OF TIME TO FIRST BIRTH AFTER MARRIAGE

BY

EPHRATA HAILU

A Thesis submitted to the office of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics (Applied Statistics)

July, 2015

ACKNOWLEDGEMENTS

First, and foremost, I thank almighty GOD as all is and has been possible through him.

My special and heartfelt gratitude goes to Professor M. K. Sharma, my advisor and instructor, for his guidance, support, continual help, and comments from proposal development up to the final work of this thesis and will always be grateful for that.

Then I would like to thank my family especially my mother Tsege Debela, my father Hailu Beyene, my husband Daniel Alemayehu and my best friend Mehiret Tena for their help, appreciation and encouragement.

Finally, I am highly grateful to all my friends for their continual support and encouragements towards success of my work.

LIST OF ABBREVIATIONS (ACRONYMS)

BDHS	Bangladesh Demographic and Health Survey
CI	Confidence Interval
CSA	Central Statistics Agency
EAs	Enumeration Areas
EDHS	Ethiopian Demographic and Health Survey
GDHS	Ghana Demographic and Health Survey
HIV/AIDS	Human Immune deficiency Virus/ Acquired Immune deficiency Syndrome
HR	Hazard Ratio
KM	Kaplan-Meier estimate
PM	Proportional Hazards
SE	Standard Error

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Survival analysis of time to first birth after marriage

By: Ephrata Hailu

ADDIS ABABA, ETHIOPIA July, 2015

The first birth describes a woman's transition into motherhood. It has a significant role in the future life of each individual woman and a direct relationship with fertility. The study aimed to model the determinant of the waiting time to first birth after marriage of Ethiopian women. The data source for the analysis was the 2011 EDHS data collected during September 2010 through January 2011. Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event. Non-parametric survival analysis techniques such as the life table, log-rank test and semi-parametric Cox proportional hazard regression analysis were used to investigate the determinants of waiting time to first birth after marriage. The Cox Proportional Hazard Model is a multiple regression method used to evaluate the effect of multiple covariates on the survival. Out of the total (7594) about 89% of women were give birth and the rest of 11% were not gave birth in marriage at all. The median of waiting time of a woman to first birth after marriage was 2 years. The minimum and maximum waiting time of first birth in the data was 1 years and 35 years respectively. The skewness of waiting time of a woman to first birth after marriage is 3.480. This shows that a data is skewed to the right distribution. The 25th and 75th percentile of waiting time of a woman to first birth after marriage was 2 and 4 years respectively. The result of Cox proportional hazard model showed that the factors that determine the waiting time to first birth after marriage which are Age at first marriage, Women educational level, Region, Place of residence, Educational level of husband and Employment status of the respondents are statistically significant.

Key words: First birth interval, Marriage, Survival

CHAPTER ONE
INTRODUCTION

1.1 Background of the study

Child bearing as a crucial period of human development might have serious consequences on young mothers. It can limit educational attainment and their capacity to support themselves financially. It can also restrict the skills young women acquire for the work and negatively affect the health and quality of their life (Ngalinda, 2000).

The birth of a child is an event of great social and individual significance and its importance is recognized in all human societies. It is of special importance as it signifies the transition of a couple into a new social status i.e. parenthood with its related expectations and responsibilities. It marks the sexual and social maturity of the mother and the visible consummation of sexual intercourse.

The first birth describes a woman's transition into motherhood. It has a significant role in the future life of each individual woman and a direct relationship with fertility. It is believed that the age at which child bearing begins influences the number of children a woman bears throughout her reproductive period in the absence of any active fertility control. Therefore, the timing of the first birth has important demographic implication as both the timing of subsequently births and completed family sizes are related to the age at first birth.

For developing country like Ethiopia where contraceptive use is relatively low giving first birth at early age tends to increase the number of children a woman will have in her future reproductive period. Even in a widespread family planning, the timing of first birth can affect the completed family size if contraception is used for spacing rather than for limiting fertility.

The decline in fertility in every society often proceeds in two ways. The first way is the decline in fertility due to fertility control and the second way involves the adoption of contraception and a change in the fertility behavior. However, much attention in stimulating fertility decline has been devoted to the provision of family planning and thus the first way received less policy attention.

Generally, fundamental social, economic, and cultural transformations, which gave changed norms relating to family and reproduction as well as personal values and practices, have influenced fertility and thus time to first birth.

1.2 Statement of the problem

Ethiopia is the second most populous country in Africa next to Nigeria. Its population has increased nearly seven times from 11.8 million at the beginning of the 20^{th} century to about 80 million in 2007. The estimated annual rate of growth and doubling period was 2.7% and 26 years respectively (MOH, 2007). Women of reproductive age make up one-fifth of the total population of the country and about 45% of the female population.

Fertility is one of the elements in population dynamics that has significantly contributed to changing the population size and structure over time (EDHS, 2011). In order to reduce fertility and control population growth of the country, the factors that influence fertility should be clearly identified (Zhang, 2007).

The fertility level of Ethiopia especially in the rural area is unacceptably high (Bekele, 2011). The higher the fertility of women, the more the risk associated with each birth. The National Population Policy of Ethiopia targeted the reduction of fertility (National Office of Population, 1993).

The first birth plays a significant role in the future life of each individual woman and her family; her overall fertility depends on it (Ida, 2013). The marriage to first birth interval in terms of the age of mother has strong impact on fertility. Along with other variables, it determines the observed reproductive behavior of the women. In developed societies with a higher use of contraception, first birth interval may be less relevance to the study of fertility. However, in the population where contraceptive practice is quite low like in Ethiopia, marriage to the first birth interval could be one of the prime determinants of fertility (Mukhlesur et. al, 2013). In general it is proved that early child bearing has direct impact on high total fertility and young population, short biographical distance between generations and short doubling time of the population at all (Ngalinda, 2000).

Numerous studies (Kinfu, 2000; Pathfinder Ethiopia, 2006) have been done in investigating factors affecting the timing of marriage in reason of controlling fertility. However, little has been done on the relationship between the timing of first marriage and first birth (Eshetu and Dula, 2014).

The purpose of this paper is, therefore, to examine factors that determine the interval between first marriage and first birth.

1.3 Objectives of the study

General objective

- The general objective of this study is to investigate the effect of some selected variables on marriage to first birth interval in Ethiopia.

Specific objectives

- To identify significant factors or covariates found to have significant effect on first birth interval for a case of Ethiopian women.
- To find out which women are most likely to have short and long marriage to first birth interval

1.4 Limitations

In this study, the researcher considers only women of child-bearing age (15-49 years). It considers only women who went into marriage without a child or a pregnancy.

1.5 Significance of the study

This study is significant since it provides information on first birth interval in Ethiopian women by analyzing the impact of different variables on survival of first birth. Some of it is significance are the following.

- It is expected to give some knowledge about the determinants or risk factors of first birth interval in Ethiopian women
- It will provide information to the government and other concerned bodies for the future usage.
- It can also be used as a basis for further studies related to fertility and others in the future.

CHAPTER TWO
LITERATURE REVIEW

2.1 General Overview about First Birth Interval

The waiting time to first birth from marriage, which is defined to be the time interval between the time of marriage and the time of occurrence of the first birth, is said to be highly influenced by several socio-economic and demographic factors (MacInnes, 2005).

Fundamental social, economic and cultural transformations, which have changed norms relating to family and reproduction, as well as personal values and practices, are shown to influence fertility and thus time to first birth (Gillespie, 2001; MacInnes, 2005; Murphy, 1992; Murphy, 1993; Sobotka, 2004).

In a study they conducted among young Chinese couples, Feng and Quanhe (1996) argued that the decline in arranged marriages, rapid increase in formal education and non-familial employment, changes in norms related to sexuality due to exposure to media and government's family planning program are the four modernization related social transformations which have changed couple's sexual behavior and timings of first marriage and first birth.

The study conducted in Nepal also proved that temporal change in the first birth interval is strongly influenced by social transformation which encourages quicker intimacy between couples (Shrestha, 1998).

As per Shrestha (1998), the time interval between the time of marriage and the time of occurrence of the first birth is becoming short now-a-days because couples take less time to become near. For example, in China the shortening of first birth interval has resulted due to the shift from arranged marriage to love marriage, which Hong (2006) attributed to increased intimacy and coital frequency (Eshetu and Dula, 2014).

First birth interval is also associated with couple's personal characteristics like age at marriage, education, occupation, and place of residence. Age at first marriage is often used as a proxy for the onset of women's exposure to the risk of pregnancy by some researches. Others also identified an inverse relation between waiting time and the age at first marriage of a mother (Ngalinda, 2000).

Several studies have also identified that high education strongly reduces the likelihood of first birth (Blossfeld, 1995; Hoem, 2000; Joshi 2002; Kohler et al, 2002; Martin, 2000; Morgan, 1999).

Khan and Reside (1998) also reveals that a women's education has a greater influence on her fertility than the education of her husband.

In a study conducted in Bangladesh, it is found that the median birth interval increases in all education groups, with better -educated women having consistently longer birth interval (Salway et al., 1993). However, as per the study conducted in Taiwan education of both spouses had not shown any significant effect on the first birth interval (Asifa and Muhammad, 2013).

Place of residence (rural or urban), region of residence, level of education of mother as well as the wealth index, are key factors that influence first birth and overall fertility as per the report of the Ghana Demographic and Health Survey (Ida, 2013).

Rodriguez and Hobcraft (1980) have compared results of identical structural for nine countries and found that a woman's education and age, time period, and the length of the previous birth interval all had substantial effects on birth-interval length.

Rindruss et al., (1987) analyzed the determinants of birth intervals for five countries. For all countries, they found significant and important differences in child spacing for the following variables: country and ethnicity, age at first birth and urban experience. They also discovered that education had relatively little effect on interval lengths except at the higher birth orders.

The waiting time of first marriage to first birth has different effects on the mother, children and family at all. First birth interval not only affects the length of rest of birth intervals but also affect reproductive pattern of women (Millman and Hendershott, 1980; Trussel and Menken, 1978; Yamaguchi and Ferguson, 1995).

Majumder (1991) observed that children born after a longer birth interval had a lower mortality rate than those born within a short interval.

It is found that infants who were conceived less than 6 months after a preceding birth have higher mortality than other infants, particularly in the neonatal period (Yerushalmy et al, 1956; Nortman et al, 1976).

Martin et al. (1964) reported that children born at a longer birth intervals of 3 -4 years will be heavier, taller and are not likely to suffer from portion, irrespective of family income and other factors energy malnutrition.

2.2. Predictors of first birth interval

Proportional hazards modeling is the most frequently used type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (Stevens and Hollis, 1989), affective disorders (Shapiro et al. 1989), childhood family breakdown (Fergusson et al, 1985); interruptions in conversation (Dress, 1986), employee turnover (Morita, et al. 1989), and in medical areas for identification of important covariates that have significant impact on the response of the interested variables(Abreham, 2009).

Empirically, many studies have shown that the interval of first birth is influenced by a number of socio-economic and demographic factors. For instance, Ida (2013) used data from the Ghana Demographic and Health Survey (GDHS) conducted in 2008 to investigate the effect of education, place of residence and religion on the timing of marriage among Ghana's women and the relative effects of education across generations of women. He used Cox proportional hazard model to analyze the data. The results revealed that level of education had a statistically significant ($p<0.031$) and strong positive effect on a woman's first birth interval. As per this study specifically Secondary and Higher level of education shortens the timing of first birth after marriage. This is because at the time of entry to marital life, they are emotionally prepared, biologically mature, and financially secured to have a child.

Eshetu and Dula (2014) used data from the 2005 EDHS to examine the effects of socioeconomic and demographic factors on age at first marriage and First Birth Interval in Ethiopia. The analysis was also made using Cox proportional hazard model and Kaplan Meier plot to assess the relationship between age at first marriage and first birth interval in Ethiopia. The study showed that place of residence, religion, region, wealth index, education, and occupation of women were significantly important factors for determining age at first birth. As per this study the transition of women with no education to motherhood took longer time compared to the literates. Uneducated women on average got married early at age 15 whereas those who had secondary and higher education got married later after 19. First birth interval was longer for those who got married before age 15 but shorter for those who engaged in marital life after legal age of 18 years. Also, the findings of this study showed that the likelihood of having first birth immediately after first marriage was significantly higher in Oromia and SNNP regions compared to Amhara region. For women in Oromia region, the likelihood of first birth soon after marriage had increased by 65% (p<0.001) while it was 63% higher for women in SNNP region (p<0.001).

The study by Choe et al. (2005) in Nepal used a proportional hazards model in order to examine the effect of covariates on age at early marriage. The result of this study revealed that age at marriage was varied by the ecological zones of the Hills, Mountains and Terrain regions i.e., region of residence significantly determines age at the marriage. In addition to region of residence, they also found that education plays an important role and the study found that children of parents with higher education were less likely to get married at an early age.

The study by Mukhlesur et al. (2005) in Bangladesh used a proportional hazards model in order to examine the effect of covariates on marriage to first birth interval. The result of their study revealed that region of residence was found to be statistically significant to determine the waiting time of first birth. They also found education to be statistically significant.

Mosammat et al. (2013) used data from the 2007 Bangladesh Demographic and Health Survey (BDHS) to examine the effects of socioeconomic and demographic factors on age at first marriage linkage to reproductive behavior of women. Binary logistic regression is used to assess the relationship between age at first marriage and its relation to fertility and risk factors. The study showed that place of residence, religion, region, wealth index, education, and occupation of parents were found to be statistically important factors to determine age at first marriage. Also, the findings of this study showed that if the age at first marriage of adolescents was increased by 1 year, the age at first birth was increased by 0.728 years. With the increase in age at first marriage, the fecund ability of women sharply rises, whereas the proportion of temporary sterility decreases. Also, the study indicated that there were a positive association between age at first marriage and age-specific marital fertility rates of the women.

CHAPTER THREE
METHODOLOGY

3.1 Data source

The study used data from the survey entitled "Ethiopia Demographic and Health survey" which was conducted by Central Statistical Agency (CSA/2011 EDHS). The data provide in-depth information on marriage, fertility, family planning, infant, child, adult and maternal mortality, maternal and child health, gender, nutrition, malaria, knowledge of HIV/AIDS and other sexually transmitted diseases.

3.2 Sample Design

The sample for the 2011 EDHS was designed to provide population and health indicators at the national and regional levels. The 2007 Population and Housing Census, conducted by the CSA, provided the sampling frame from which the 2011 EDHS sample was drawn. Administratively, regions in Ethiopia are divided into zones, and zones into administrative units called woredas. Each woreda is further subdivided into the lowest administrative unit, called kebele. During the 2007 census each kebele was subdivided into census enumeration areas (EAs), which were convenient for the implementation of the census.

The 2011 EDHS sample was selected using a stratified, two-stage cluster design and EAs were the sampling units for the first stage. The sample included 624 EAs, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling. A complete listing of households was carried out in each of the 624 selected EAs from September 2010 through January 2011. A representative sample of 17,817 households was selected for the 2011 EDHS, of these, 16,702 were successfully interviewed. In the interviewed households 17,385 eligible women were identified for individual interview; complete interviews were conducted for 16,515. Women whose current ages are 15-49 years are included in the survey. After a certain rearrangement, reorganization and removal of missing values, the total number of women with complete information became 7594. In this analysis, the marital-conception-first birth sequence was used as a dependent variable to define the first birth interval. Though premarital conception and premarital birth could lead to first birth, these sequences were not considered as their chances of occurrence were rare in Ethiopia, where birth out of-wedlock has no societal acceptance (Kinfu 2000; Molla et al., 2008; Ayalu and Lindstrom, 2014). Due to lack of detailed information on marriage history, it was not possible to run separate analysis for women married only once, as divorced women without children cannot be censored upon dissolving the first marriage.

3.3 Variables in the study

3.3.1 The response variable

In this study, the dependent variable is the time to first birth interval after marriage. It is measured as the length of time from marriage up to the first birth of child which is measured in years. During the survey all women were asked a series of questions regarding to time at first birth. The response to this question comprises the women waiting time to first birth from marriage and women who had not yet experienced the events resulting in right censoring of the data.

3.3.2 Predictor variables

Several predictors are considered in this study to investigate the determinant factors for the waiting time of first birth after first marriage. The variables were selected based on the findings of previous works (Mukhahlesur et. al, 2013) and they play potential roles in influencing the timing of first marriage and first birth after marriage. The predictor (independent) variables included in the model are:

- Age at first marriage (0 = $\leq$ 14 years, 1= 15-20 years, 2= $\geq$21 years)
- Women level of education (0=Illiterate, 1=Primary, 2=Secondary and 3= Higher)
- Place of residence for women (0= Urban, 1= Rural)
- Region (0=Tigray, 1=Affar, 2=Amhara, 3=Oromia, 4=Somalia, 5=Benishangul-Gumuz, 6=SNNP, 7=Gambela, 8=Harari, 9=Addis Ababa,10= Dire Dawa)
- Wealth index (0= Poor, 1=Medium, 2=Rich)
- Religion (0= Orthodox, 1= Protestant, 2= Muslim, 3= Others)
- Employment status of the respondent (0= No ; 1= Yes) and

- Husband education level (0= Illiterate, 1=Primary, 2=Secondary and 3= Higher)

3.4 The methodology: Survival analysis

3.4.1 The Survival Model

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. The term survival analysis applies to techniques in which the data being analyzed is the time the process takes for a certain event of interest to occur. It is most important when there are censoring data as opposed to the use of different statistical methods (Aalen, 2008). It involves the modeling and analysis of data that has a principal end point the time until an event occurs (time-to-event data). By time, it is to mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs. Survival data, or time-to-event data, measure the time elapsed from a given origin to the occurrence of an event of interest.

In Survival analysis, it is usually referred to the time variable as survival time because it gives the time that an individual has 'survived' over some follow-up period. Also the term ‘failure’ is used to define the occurrence of the event of interest even though the event may actually be a 'success' such as recovery from therapy (Kleinbaum and Klein, 2005).

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. Such summaries may be of interest in their own right or as a precursor of a more detailed analysis of the data. Routine applications of standard formulas for measures of central tendency and variability will not yield estimates of the desired parameters when the data include censored observations.

In summarizing survival data there are two functions of central interest namely the survivor function and the hazard function. They are discussed briefly as follow.

3.4.1.1 The survivor function S(t)

Let 'T' be a random variable associated with the survival times (*t*), survival times (*t*) be the realization of the random variable T and *f* (*t*) be the underlying probability density function of the survival time *t*. Then the cumulative distribution function *F* (*t*), which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given by:

$$F(t) = P(T < t) = \int_0^t f(u)du, \quad t \geq 0 \qquad [1]$$

where *F (t)* = P (that a women has her first child before time t)

The survivor function *S (t)* is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time t and so

$$S(t) = P(T \geq t) = 1 - F(t), \quad t \geq 0 \qquad \ldots\ldots\ldots\ldots [2]$$

where *S(t)*=1-P(a woman has her first child before time t)

The survivor function can, therefore, be used to represent the probability that, a randomly selected subject survives from the time origin to some specified time beyond *t*.

Then from equations [1] and [2] the relationship between *f* (*t*) and *S* (*t*) can be derived as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - s(t)) = \frac{-d}{dt}s(t), \quad t \geq 0 \ldots\ldots\ldots\ldots [3]$$

3.4.1.2 The hazard function h(t)

The hazard function is widely used to express the risk of hazard of death at time t. It is obtained from the probability that an individual dies at time *t*, given that the individual has survived up to time *t*. It is also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic process, the age specific failure rate in epidemiology, the inverse of the Mill's ratio in economics or simply the hazard rate (Hosmer and Lemeshow, 1999). It gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time *t*.

The hazard function $h(t) \geq 0$ is given as

$$h(t) = \lim_{\Delta t \to 0} \frac{p[(t \leq T < t + \Delta t)/T \geq t]}{\Delta t} \qquad \ldots\ldots\ldots\ldots [4]$$

By applying the theory of conditional probability and the relationship in equation [4], the hazard function can be expressed in terms of the underlying probability density function and the survival function as follows

$$h(t) = \frac{f(t)}{s(t)} = \frac{-d}{dt} \ln S(t) \quad [5]$$

A related quantity is the cumulative hazard function H (t) which is defined by

$$H(t) = \int_0^t h(u)du = -\ln S(t) \quad [6]$$

Thus, S(t) = exp(-H(t)) consequently f(t) = h(t)exp(-H(t))……………………...[7]

3.4.2 Non-parametric methods

In survival analysis, it is always a good idea to present numerical or graphical summaries of the survival times for the individuals. Survival data are conveniently summarized through estimates of the survival and hazard functions. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. The Kaplan-Meier, Nelson-Aalen and Life Tables are used most widely to estimate the survival and hazard functions. These methods are said to be non-parametric or distribution-free since they do not require the specific assumption to be made about the underlying distribution of the survival times.

3.4.2.1 Kaplan-Meier (KM)

The Kaplan-Meier (KM) estimator, or Product Limit estimator, is the standard non-parametric estimator of the survival function s(t), proposed by Kaplan and Meier (1958). It incorporates 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. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated, but still manageable when censored times are included.

Suppose $t_1, t_2, \ldots, t_n$ be the survival times of n independent observations and $t_{(1)} \le t_{(2)} \le \text{---} \le t_{(m)}$, $m \le n$ be the m distinct ordered marriage times. The Kaplan-Meier estimator of the survivorship function (or survival probability) at time t, $S(t) = P(T \ge t)$ is defined as:

$$\hat{S}(t) = \prod_{t_s < t}\left(1 - \frac{d_j}{n_j}\right) \qquad \text{[8]}$$

where: d_j is the number of women bears a child in the jth interval and

n_j is the number of women exposed to bear a child in the jth interval, with the convention that $\hat{S}(t) = 1$ for $t < t_{(1)}$.

The Cumulative hazard function of the KM estimator can be estimated as:

$\hat{H}(t) = -\ln[\hat{s}(t)]$, where $\hat{S}(t)$ is KM estimator

3.4.3 Comparison of Survivorship Functions

After obtaining statistics which provide a description of the overall survival experience, one is expected to proceed with a comparison of the survivorship experience of subgroups in the data. These groups might be defined by the values of a covariate which are thought to be related to survival times. When comparing groups of subjects, it is always preferable to begin with a graphical display of the data in each group.

In studies of survival time, the Kaplan-Meier estimator of the survival function for each of the groups should be graphed. That is to plot the corresponding estimates of the two survivor functions on the same axes of Kaplan-Meier estimator. Then if the plot shows the pattern of one survivorship function lying above another, it means that the group defined by the upper curve lived longer, or had a more favorable survival experience than the group defined by the lower curve. But, the statistical question is whether the observed difference seen on the plot is significant. This needs to be answered using appropriate statistical test (Hosmer and Lemeshow, 1999).

$$Q = \frac{\left[\sum_{i=1}^{m} w_i \; |(d_{1i} - \hat{e}_{1i})\right]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \qquad \text{[9]}$$

where:

$e_{1i} = \frac{n_{1i} \times d_i}{n_i}$ is expected number of failures corresponding in group 1 at time t_i

$\hat{v}_{1i} = \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$ is the variance of the number of failures in group 1 at time t_i

n_{1i} is the number at risk at observed survival time $t_{(i)}$ in group 1

n_{2i} is the number at risk at observed survival time $t_{(i)}$ in group 2

d_{1i} is the number of observed births in group 1

d_{2i} is the number of observed births in group 2

n_i is the total number at risk

di is the total number of births at time $t_{(i)}$

The contribution to test statistic depends on which of the various tests is used. But each may be expressed in the form of a ratio of weighted sums over the observed survival times. Under the null hypothesis assuming that the two survivorship functions are the same, and that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, Q follows a chi-square distribution with one degree of freedom. The above can be used to compare more than two groups (Hosmer & Lemeshow, 1999). The log rank test which is a special case of Q is used in this study.

3.4.3.1 The Log Rank test

The log rank test, developed by Mantel and Haenszel, is a non-parametric test for comparing two or more independent survival curves. Since it is a non-parametric test, no assumption about the distributional form of the data is required. This test is however most powerful when used for non-overlapping survival curves.

The log rank test is based on weight equal to one, i.e w_i =1. Its statistic becomes

$$Q_{LR} = \frac{\left[\sum_{i=1}^{m} \left(d_{1i} - \hat{e}_{1i} \right) \right]^2}{\sum_{i=1}^{m} \hat{v}_{1i}} \qquad \text{[10]}$$

3.4.4 Regression Models for Survival Data

In considering regression modeling of survival data, the first question has to be answered is; what is going to be modeled? Specifically, what will play the role of the systematic component in a regression model? The natural aging process that is present when subjects are followed over time is what distinguishes survival research approach. A natural place to begin is to put a regression model type structure on the hazard function.

In general, the hazard function as a function of time and covariates is specified. A regression model for the hazard function is:

$$h(t, x, \beta) = h_0(t) r(x, \beta). \qquad [11]$$

The hazard function, as expressed in [11], is the product of two functions. The function, $h_0(t)$ characterizes how the hazarded function changes as a function of survival time. The other function, $r(x, \beta)$, characterizes how the hazard function changes as a function of subject covariates. The function must be chosen such that $h(t, x, \beta) > 0$. $h_0(t)$ is the hazard function when $r(x, \beta) = 1$.When the function $r(x = 0, \beta) = 1$, $h_0(t)$ is frequently referred to as the baseline hazard function.

Under the model in [11] the ratio of the hazard functions for two subjects with covariate values denoted x_1 and x_2 is

$$H(t, x_1, x_2) = \frac{h(t, x_2, \beta_1)}{h(t, x_1, \beta_1)}$$

$$= \frac{h_0(t) r(x_2, \beta)}{h_0(t) r(x_1, \beta)}$$

$$= \frac{r(x_2, \beta)}{r(x_1, \beta)}$$

The hazard ratio (H) depends only on the function $r(x, \beta)$.

One of the most popular types of regression models used in survival analysis is the Cox proportional hazard model (Cox, 1972).

3.4.4.1 The Cox Proportional Hazards Model

David Cox's 1972 paper took a different approach to standard parametric survival analysis and extended the methods of the non-parametric Kaplan-Meier estimates to regression type arguments for life-table analyses. Cox advanced prediction of survival time in individual subjects by only utilizing variables co-varying with survival and ignoring the baseline hazard of individuals. He did this by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time.

When there are several explanatory variables, and in particular when some of these are continuous, it is much more useful to use a regression method such as Cox rather than a KM approach. The Cox Proportional Hazard Model is a multiple regression method used to evaluate the effect of multiple covariates on the survival. Cox was the first to propose the model in [11] when he suggested using $r(x,\beta) = e^{\beta x}$.

With this parameterization, the hazard function is:

$$h(t, x, \beta) = h_0(t) e^{\beta' \chi}$$

where

$h_0(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time,

$h(t, x, \beta)$ represents the hazard function at time t with covariates $x=(x_1, \ldots, x_p)$

$\beta = (\beta_1 \ldots \beta_p)$ is a column vector of p regression parameters,

$e^{\beta' \chi}$ characterizes how the hazard function changes as a function of subject covariates.

t is the failure time

The survival time of each member of the sample is assumed to follow its own hazard function. In such a case, the above model can equivalently be written as:

$$h(t, x_i, \beta) = h_0(t) \exp(\beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \quad i=1, \ldots, n,$$

where: n is total number of observations in the study

$x_i = (x_{i1}, \ldots, x_{ip})$ is a column vector of measured covariates for the i^{th} individual which are expected to affect the survival probability

3.4.4.2 Assumption of Cox proportional hazard model

1. The baseline hazard $h_0(t)$ depends on t, but not on covariates $x_1, x_2, \ldots, x_p$, the hazard ratio $e^{\beta\chi}$ depends on the covariates $x = (x_1, x_2, \ldots, x_p)'$ but not on time t.
2. The covariate x_i does not depend on time t.

To express this mathematically, consider two distinct values of continuous covariate x, say, x_1 and x_2.

$$h(t, x_i, \beta) = h_0(t) e^{\beta'\chi}$$

Then the hazard ratio becomes

$$H(t, x_2, x_1) = \frac{h(t, x_2, \beta)}{h(t, x_1, \beta)} = \frac{h_0(t) e^{\beta\chi_2}}{h_0(t) e^{\beta'\chi_1}} = \frac{e^{\beta'\chi_2}}{e^{\beta'\chi_1}} = e^{\beta'(x_2 - x_1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [12]$$

Since $e^{\beta'(x_1 - x_2)}$ is independent of time, it shows that the ratio of the hazard functions for two individuals with different covariate values does not vary with time.

3.4.4.3 Parameter Estimation

In Cox proportional hazards model the vector of parameters β can be estimated without having any assumptions about the baseline hazard $h_o(t)$. Instead of the full likelihood, the partial likelihood proposed by Cox (1972) is used.

Consider the conditional probability that an individual birth at time $t_{(i)}$, given that $t_{(i)}$ is one of the r observed birth times $\{t_{(1)}, t_{(2)}, \ldots, t_{(m)}\}$.

By definition, $h_i(t) = \frac{f_i(t)}{S_i(t)}$

The contribution to the likelihood for an observed failure at time t is

$$\begin{aligned} f_i(t) &= h_i(t) S_i(t) \\ &= h_0(t) \exp(\beta' x_i) [S_0(t)]^{\exp(\beta' x_i)} \end{aligned}$$

The contribution to the likelihood for a right censored observation at time t is

$$S_i(t) = \left[S_0(t)\right]^{\exp(\beta' x_i)}$$

Let $\delta_i = 1$ if t_i is a failure time, and $\delta_i = 0$ if t_i is a censoring time.

Then the joint likelihood is

$$\prod_{i=1}^{m} [f_i(t_{(i)})]^{\delta_i} [S_i(t_{(i)})]^{1-\delta_i} = \prod_{i=1}^{m} [h_i(t_{(i)}) S_i(t_{(i)})]^{\delta_i} [S_i(t_{(i)})]^{1-\delta_i}$$

$$= \prod_{i=1}^{m} [h_i(t_{(i)})]^{\delta_i} S_i(t_{(i)})$$

$$= \prod_{i=1}^{m} \left[\frac{h_i(t_{(i)})}{\sum_{j\in R(t)} h_j(t_{(i)})}\right]^{\delta_i} \times \left[\sum_{j\in R(t)} h_j(t_{(i)})\right]^{\delta_i} S_i(t_{(i)})$$

$$= \prod_{i=1}^{m} \left[\frac{\exp(\beta' x_i)}{\sum_{j\in R(t)} \exp(\beta' x_j)}\right]^{\delta_i} \times \prod_{i=1}^{n} \left[\sum_{j\in R(t)} h_j(t_{(i)})\right]^{\delta_i} S_i(t_{(i)})$$

where R (t_i) represents the risk set just prior to time t_i.

The partial likelihood is

$$L(\beta) = \prod_{i=1}^{m} \left[\frac{\exp(\beta' x_i)}{\sum_{j\in R(t)} \exp(\beta' x_j)}\right]^{\delta_i} \quad \text{.. [13]}$$

The expression [13] assumes that there are no tied times, and it is often modified to exclude terms when δ_i=0, yielding

$$l_p(\beta) = \prod_{i=1}^{m} \frac{\exp(\beta' x_i)}{\sum_{j\in R(t)} \exp(\beta' x_j)} \quad \text{... [14]}$$

where the product is over the 'm' distinct ordered survival times and $x_{(i)}$ denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$.

The log partial likelihood function is

$$\mathrm{L_p}(\beta) = \sum \left\{ x_{(i)}\beta - \ln\left[\sum_{j\in R(t)} \exp(\beta' x_j)\right]\right\} \text{................................ [15]}$$

By taking the log of [13] and differentiating with respect to β, yielding

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left\{ x_i - \frac{\sum_{j\in R(t)} x_j \exp(\beta' x_j)}{\sum_{j\in R(t)} \exp(\beta' x_j)} \right\} \text{.............................. [16]}$$

The estimator of the variance estimator is obtained in most maximum likelihood estimation applications. The estimator is the inverse of the negative of the second derivative of the log partial likelihood at the value of the estimator. In particular, taking the derivative of [16] with respect to β, the following expression is obtained:

$$\frac{\partial^2 L_p(\beta)}{\partial \beta^2} = -\sum_{i=1}^{m} \left\{ \frac{\left[\sum_{j\in R(t_{(i)})} \exp(\beta' x_j)\right]\left[\sum_{j\in R(t_{(i)})} x_j^2 \exp(\beta' x_j)\right] - \left[\sum_{j\in R(t_{(i)})} x_j \exp(\beta' x_j)\right]^2}{\left[\sum_{j\in R(t_{(i)})} \exp(\beta' x_j)\right]^2} \right\} \ldots\ldots [17]$$

The negative of the second derivative of the log partial likelihood in [17] is called the observed information, and it will be denoted as

$$I(\beta) = -E(\frac{\partial^2 L(\beta)}{\partial \beta^2}) \ldots\ldots\ldots\ldots [18]$$

The estimator of the variance of the estimated coefficient is the inverse of [18] evaluated at $\hat{\beta}$ and is

$$\text{Var}(\hat{\beta}) = \text{I}(\hat{\beta})^{-1} \ldots\ldots\ldots\ldots [19]$$

The estimator of the standard error, denoted as S.E ($\hat{\beta}$), is the positive square root of the variance estimator in [19]

3.4.5 Assessment of Regression Coefficients

It is started by presenting two different tests to assess the significance of the coefficient: the Partial likelihood ratio test and the Wald test.

3.4.5.1 The Partial Likelihood Ratio Test (LR)

The partial likelihood ratio test, denoted as G, is calculated as twice the difference between the log partial likelihood of the model containing the covariate and the log partial likelihood for the model not containing the covariate. Specifically,

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\} \ldots\ldots\ldots\ldots [20]$$

$$L_p(0) = -\sum_{i=1}^{m} \ln(n_i), \ldots\ldots\ldots\ldots [21]$$

and n_i denotes the number of subjects in the risk set at observed survival time $t_{(i)}$.

Under the null hypothesis that the coefficient is equal to zero, this statistic will follow a chi-square distribution with 1 degree-of-freedom. This distribution can be used to obtain p-values to test the significance of the coefficient. In practice, the "sufficiently" large sample size cited for likelihood ratio test translates in this case to have the number of observed non censored survival times be large.

3.4.5.2 The Wald Test

To test H_0: $\beta_q = (0, 0, \ldots, 0)$, the multivariable Wald statistic is used.

$$Q_W = \hat{\beta}_q^{'}[I_q(\hat{\beta})]^{-1}\hat{\beta}_q$$

$\hat{\beta}_q$ and $I_q(\hat{\beta})$ are the corresponding estimates of $\hat{\beta}_q$ and sub matrix of the inverse of the observed information matrix from the full model. Under H_0 and for large sample the statistics $Q_w \sim X^2(q)$ at α level of significance.

The Wald test can also be used to test the significance of individual variables. The test statistic then becomes

$$Z = \frac{\hat{\beta}_j}{S\hat{E}(\hat{\beta}_j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{[22]}$$

Under the null hypothesis H_0: $\beta_j=0$ the statistic Z~N (0, 1) at a significance level α, consequently the 100 (1- α) % Wald statistic based confidence interval for β_j is

$$\hat{\beta}_j \pm Z_{\alpha/2} se(\hat{\beta}_j)$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution.

3.4.6 Selection of covariates

Step 1: It begins by including covariates that are statistically significant (P-value < 25% modest level of significant).

Step 2: Include covariates that are considered more important.

Step 3: Use these covariates (those selected out in step1 and step2).

Step 4: Select covariates that are statistically significant (p-value<0.05).

Step 5: Retain some covariates that are "important" even though they are not significant and check also for potential confounders.

Step 6: A final check is made to ensure that neither significant variable is eliminated from the model nor non-significant variable is included in the model. At this stage the interactions between any of the main effects currently in the model can be considered for inclusion if the inclusion significantly modifies the model.

3.4.7 Assessment of Model Adequacy

The methods for assessment of a flitted proportional hazards model are essentially the same as for other regression models. In general, requirements for model assessment are:

1. Methods for testing the assumption of proportional hazards;
2. Subject-specific diagnostic statistics that extend the notations of leverage and influence to the proportional hazards model; and
3. Overall summary measures of goodness of fit.

3.4.8 Residual analysis

Many model checking procedures are based on quantities known as residuals. The central to the evaluation of model adequacy in any setting is an appropriate definition of a residual. A residual is the difference between the observed value of the outcome variable and that value predicted by the model. The two key assumptions in the definition of a residual are the value of the outcome is known and the fitted model provides an estimate of the mean of the dependent variable or systematic component of the model.

However, the two assumptions are not valid when using partial likelihood to fit the proportional hazards model to censored survival data. The absence of an obvious residual has led to the development of several different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazard model. These include the Cox-Snell, martingale, Deviance and Schoenfeld residuals (Collett, 2003 and Hosmer & Lemeshow, 1999). Here Deviance and Schoenfeld residuals are used.

3.4.8.1 Deviance residuals

The deviance residuals (rD_i) can be regarded as an attempt to make the martingale residuals symmetrically distributed about zero, and are defined by

$$rD_i = \text{sgn}(rM_i)[-2\{rM_i + \delta_i \log(\delta_i - rM_i)\}]^{\frac{1}{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [23]$$

where rM_i is the martingale residual and

$\text{sgn}(rM_i)$ is the sign of the martingale residual for the i^{th} observation.

The deviance residuals are expected to be symmetrically distributed about zero if the fitted model is appropriate although they do not necessarily sum to zero.

3.4.8.2 Schoenfeld residuals

The above residuals have the disadvantages that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. However, Schoenfeld proposed residuals overcome these disadvantages. These residuals are calculated for each individual and for each covariate. Thus, the Schoenfeld residual for the i^{th} observation in the k^{th} covariate, the i^{th} component in the k^{th} score vector, is given by

$$rS_{ik} = c_i \left\{ x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} \exp(\hat{\beta}' x_j)}{\sum_{j \in R(t_i)} \exp(\hat{\beta}' x_j)} \right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [24]$$

where R (t_i) represents the risk set just prior to time t_i.

The sum of these residuals is zero and they have a large sample property that, their expected value is zero and are uncorrelated with one another. The vector of these residuals for i^{th} observation can be written as rS_i= (rS_{i1} , rS_{i2}, rS_{i3},….., rS_{ip}) and the convention is that it is set to be missing for censored observations.

Scaling a vector of schoenfeld residuals by an estimator of its variance is more effective in detecting departures from the assumed model. The vector of the scaled Schoenfeld residuals is then given by:

$$rs_i^* = [\text{var}(rs_i)]^{-1} rs_i \approx m\,\text{var}(\hat{\beta}) rs_i$$

where, m is the number of events.

3.4.9 Checking the assumption of proportional hazards

The next step of any model building procedure is the assessment of the specified model's adherence to the model assumptions. The critical assumption of the Cox model is the extent to which the effect of a covariate on the outcome variable has been the same over the study time. That is, the extent to which the coefficients of a covariate vary over the study time will be examined. Although there are various ways to check the proportionality of hazard ratio over time, Grambsch and Therneau (1994) and simulation comparisons by Ngandu (1997) have shown that one numerical test and an association graph yield a powerful and effective method to check this assumption.

As an alternative to the proportional hazards regression model, $h(t,x,\beta)=h_0(t)e^{\beta x}$, consider a model with time varying coefficient β as $h(t,x,\beta)=h(t,x,\beta)e^{\beta_{(t)}X}$. Let the j^{th} covariate coefficient β_j vary over time as $\beta_j(t)=\beta_j+\gamma_j g_j(t)$ where j=1,2,3,….,p. β_j is constant, $g_j(t)$ is some specified function of time (usually $g_j(t)$ is specified as ln(t)) and γ_j is coefficient of $g_j(t)$. Then the hazard function becomes: $h(t,x,\beta)=h_0(t)e^{(\beta+\gamma\ln t)X}=h_0(t)e^{(\beta X+\gamma X\,\text{int})}$.

The rationale behind this model is that the effect of a covariate may change over the period of follow up. Here whether the coefficient γ_j is zero or different from zero will be tested and if it is different from zero it indicates that the proportional hazard assumption fails to hold but if it is zero then the model is reduced to the proportional hazard model with satisfied assumption. In line with this the plots of the scaled Schoenfeld residuals against the log of time can be used and, if this plot shows random distribution around the reference line through zero then the assumption is satisfied, however, if it looks to have some systematic pattern, then the assumption of proportional hazard is violated.

The test procedure to check whether the coefficient is really time varying or not is

H_0: $\gamma_j=0$ Vs $H_1=\gamma_j\neq 0$ j=1, 2,…., p

The test statistic to be used is the Wald test statistic which is given as:

$$Z=\frac{\gamma_j}{S.E(\gamma_j)},$$ which follows a standard normal distribution.

The test rule used is:

The coefficient of the jth covariate is not time dependent if $Z < Z_{\left(\alpha/2\right)}$, then don't reject H_0, that is model assumption is satisfied.

The coefficient of the jth covariate is time dependent if $Z > Z_{\left(\alpha/2\right)}$, then reject H_0, that is model assumption is not met.

3.4.10 Checking for Outliers

Another important aspect of the model evaluation is a thorough examination of regression diagnostic statistic to identify which, if any, observation have an undue influence on the fit of the model. It is important to determine whether the hazard ratio will be affected to a large extent by any one individual observation's data in the model. This is done by fitting the model for all *n* observations in the data set and then refits the same model to the sets of *n-1* observations obtained by omitting each of the *n* observations one at a time. Whenever an observation can exert unusual influence on the hazard ratio, this will be noticeable in significant change in the parameter estimate(s) while removing the observation's data from the data set and refit the same model. That is $\Delta_i \hat{\beta}_j = (\hat{\beta}_j - \hat{\beta}_{j(n-i)})$ can be used, which is known as delta-beta (Collett, 2003), as the statistic to detect an outlier observation of the data.

The change in the parameter estimate is large for an outlier observation while the delta-beta is not large for others. The influence of each observation was measured on the estimated regression coefficients.

To examine the influence of a j^{th} covariate value of the i^{th} individual on the j^{th} regression coefficient estimate delta-beta statistics will be used which is:

$$\Delta_i \hat{\beta}_j = (\hat{\beta}_j - \hat{\beta}_{j(n-i)})$$

where $\hat{\beta}_j$ and $\hat{\beta}_{j(-i)})$ are the jth coefficient based on n and *n-1* observations respectively (i^{th} observation is deleted)

The overall summary statistic of the influence of a subject on the estimator of all the coefficients may be approximated using the likelihood displacement statistic ld_i. The statistics ld_i is an approximation to the amount of change in log partial likelihood when the i^{th} subject is deleted. In this context the statistic is called the likelihood displacement statistic. It can be shown that:

$$ld_i = 2[l_p(\hat{\beta}) - l_p(\hat{\beta}_{(n-1)})]$$

The next step in the modeling process is to identify explicitly the subjects with the extreme values, and refit the model by deleting these subjects, then calculate the percentage change in the individual coefficients as

$$\Delta\hat{\beta} = 100\left(\frac{\hat{\beta}_{reduced} - \hat{\beta}_{all}}{\hat{\beta}_{all}}\right)$$

The final decision on the continued use of a subject's data to fit the model will depend on the observed percentage change in the coefficients that result from deleting the subject's data and, more importantly, the clinical plausibility of that subject's data (Hosmer and Lemeshow 1998).

3.4.11 Overall goodness of fit

As in regression analysis, some measure analogous to R^2 may be of interest as a measure of model performance. In particular, all measures depend on the proportion of values that are censored. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to a high percent of censored data. R^2 is used as it is the easiest and best one to use, and it is defined as

$$R_p^{\ 2} = 1 - \exp\left(\frac{2}{N}\left(LL_0 - LL_{\hat{\beta}}\right)\right) \quad \text{..} [27]$$

where N is the total number of observation in the model.

LL_o is the log partial likelihood for model zero.

$LL_{\hat{\beta}}$ is the log partial likelihood for the fitted model with p covariates.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Introduction

In this section results of data analysis, discussion and interpretation are presented as follows. The first part presented summary statistics of factors and the second part compares the survival time in different groups. Descriptive survival analysis is also used in comparing the survival time in different groups. The third part is about fitting the model by which the adequacy of the model is investigated. The statistical packages SPSS and STATA are employed to analyze the data throughout the study. Finally, the results are discussed and interpreted.

4.2 Results of Descriptive Statistics

Descriptive statistics were used to get some information about the distribution of the variables. A total of 7594 women were included in the study during the data collection. From the total, 6758(89%) of them gave birth in marriage and the rest 836(11%) were not gave birth in marriage at all (censored).

The minimum waiting time of a woman to first birth after marriage in this data is 1 year and the maximum is 35 years. The median of waiting time of a woman to first birth after marriage is 2 years. The skewness of waiting time of a woman to first birth after marriage is 3.480. This shows that a data is skewed to the right distribution. The 25^{th} and 75^{th} percentile of waiting time of a woman to first birth after marriage were 2 and 4 years respectively.

Table 4.1 showed the educational attainments of women; about 64.6% had no education while 26.7% had primary education and the remaining 8.6% had attended secondary and higher education. From the total number of respondents, 32.1 were employed and the rest 67.55% were unemployed. With regard to their residence, 22.2% were lived in urban where as 77.8% lived in rural areas.

Table 4.1 Descriptive Summary of covariates for the waiting time of a woman to first birth after marriage dataset, EDHS, 2011

Covariate/factor variable	Category	Women who not give birth		Women who give birth		Total
		In number	In Percentage	In number	In Percentage	
Women education level	No education	354	7.21	4557	92.79	4911
	Primary	337	16.5	1695	83.42	2032
	Secondary	72	818.75	312	81.25	384
	Higher	73	27.34	194	72.66	267
Age at first marriage	≤ 14	120	5.42	2092	94.58	2212
	15-20	519	11.72	3908	88.28	4427
	≥ 21	197	20.63	758	79.37	955
Residence	Urban	239	14.18	1447	85.82	1686
	Rural	597	10.1	5311	89.9	5908
Region	Tigray	82	12.01	601	87.99	683
	Affar	116	15.36	639	84.64	755
	Amhara	136	12.02	995	87.98	1131
	Oromia	87	7.67	1047	92.33	1134
	Somalia	47	10.96	382	89.04	429
	Benishangul-Gumuz	75	10.92	612	89.08	687
	SNNP	61	6.77	840	93.23	901
	Gambela	70	12.61	485	87.39	555
	Harari	46	10.18	406	89.82	452
	Addis Ababa	69	15.54	375	84.46	444
	Dire Dawa	47	11.11	376	88.89	423
Wealth index	Poor	324	9.74	3001	90.26	3325
	Medium	119	9.75	1102	90.25	1221
	Rich	393	12.89	2655	87.11	3048
	Orthodox	342	12.33	2431	87.67	2773

Religion	Protestant	125	9.46	1196	90.54	1321
	Muslim	354	10.73	2946	89.27	3300
	Others	15	7.58	183	92.42	198
Employment status	Yes	273	11.08	2191	88.92	2464
	No	563	10.97	4567	89.03	5130
Husband educational level	No education	319	8.38	3488	91.62	3807
	Primary	314	11.69	2371	88.31	2685
	Secondary	111	18.08	503	81.92	614
	Higher	95	19.35	396	80.65	491

4.3 Descriptive survival analysis

Before proceeding to analysis of survival models, a descriptive analysis will be used as initiation to the subsequent findings.

4.3.1 Life Table

The life table method was applied to include censored observation with the assumption that women who did not give birth to a child were not systematically different from those who experienced the event. The validity of this assumption was thus proved.

Table 4.2: Survival Distribution Table of Marriage to First Birth Interval

Time	S(t)	F(t)	SE F(t)	Time	H(t)	SE for H(t)
1	0.8448	0.1552	0.0042	0.5	0.1683	0.0049
2	0.5301	0.4699	0.0059	1.5	0.4577	0.0094
3	0.3513	0.6487	0.0057	2.5	0.4057	0.0113
4	0.2463	0.7537	0.0052	3.5	0.3513	0.0130
5	0.1792	0.8208	0.0047	4.5	0.3155	0.0148
6	0.1316	0.8684	0.0041	5.5	0.3064	0.0173
7	0.1006	0.8994	0.0037	6.5	0.2668	0.0190
8	0.0763	0.9237	0.0033	7.5	0.2753	0.0224
9	0.0602	0.9398	0.0030	8.5	0.2351	0.0240

10	0.0479	0.9521	0.0027	9.5	0.2279	0.0269
11	0.0385	0.9615	0.0025	10.5	0.2186	0.0298
12	0.0322	0.9678	0.0023	11.5	0.1775	0.0303
13	0.0264	0.9736	0.0021	12.5	0.1987	0.0361
14	0.0233	0.9767	0.0020	13.5	0.1224	0.0316
15	0.0204	0.9796	0.0019	14.5	0.1353	0.0361
16	0.0181	0.9819	0.0018	15.5	0.1156	0.0365
17	0.0167	0.9833	0.0018	16.5	0.0833	0.0340
18	0.0149	0.9851	0.0017	17.5	0.1157	0.0437
19	0.0146	0.9854	0.0017	18.5	0.0185	0.0185
20	0.0137	0.9863	0.0017	19.5	0.0625	0.0361
21	0.0137	0.9863	0.0017	20.5	0.0000	-
22	0.0133	0.9867	0.0017	21.5	0.0278	0.0278
23	0.0125	0.9875	0.0017	22.5	0.0667	0.0471
24	0.0120	0.9880	0.0017	23.5	0.0408	0.0408
25	0.0120	0.9880	0.0017	24.5	0.0000	-
26	0.0120	0.9880	0.0017	25.5	0.0000	-
27	0.0113	0.9887	0.0017	26.5	0.0625	0.0625
28	0.0113	0.9887	0.0017	27.5	0.0000	-
29	0.0113	0.9887	0.0017	28.5	0.0000	-
30	0.0113	0.9887	0.0017	29.5	0.0000	-
31	0.0113	0.9887	0.0017	30.5	0.0000	-
32	0.0113	0.9887	0.0017	31.5	0.0000	-
33	0.0113	0.9887	0.0017	32.5	0.0000	-
34	0.0113	0.9887	0.0017	33.5	0.0000	-
35	0.0113	0.9887	0.0017	34.5	0.0000	-

The result showed that about 16.8% of married women gave birth within the first year of marriage. This represents the proportion of women who get pregnant within the first few months of marriage. Also, by the end of the second year of marriage, 45.8% of the married women had given birth. Thereafter, their probability generally keeps declining about 20 years in marriage; about 3.6% (0.036) of married women gave birth for the first time after 20 years of marriage. This provides hope for couples who have not yet given birth, especially those who are within the range of dates considered in this study.

4.3.2 Log Rank test

By using the Log Rank test, test of equality was done along the probabilities across the different groups. The null hypothesis to be tested is that there is no difference between the probabilities of an event occurring at any time point for each population. It is considered that the predictor will be included in a model if the Log Rank test has a p-value of < 0.25. However, if the predictor has a p-value greater than 0.25, it is highly unlikely that it will contribute anything to a model which includes other predictors.

Table 4.3: Results of the log-rank test for the categorical variables for waiting time of a woman to first birth after marriage in EDHS, 2011

Covariates	Ch-square	DF	Pr>Chi-Square
Age at first marriage	319.6	2	<0.001
Women education level	330.27	3	<0.001
Employment status	0.29	1	0.5881
Religion	3.55	3	0.314
Residence	20.25	1	<0.001
Region	51.87	10	<0.001
Wealth index	12.11	2	0.0023
Husband education level	155.81	3	<0.001

The log-rank test of equality for the predictors: Age at first marriage, women education level, residence, region, wealth index and husband education level has p-value <0.05. Thus, they are included as a potential candidate for the final model. It also indicates that statistically there is a significant difference of survival experience among groups. On the other hand, the log-rank test of equality across strata for the predictors ‘religion’ and ‘employment Status’ has a p-value >0.05 and these variables will not be included as a potential candidate for the final model or there are no significant difference, but further test is considered.

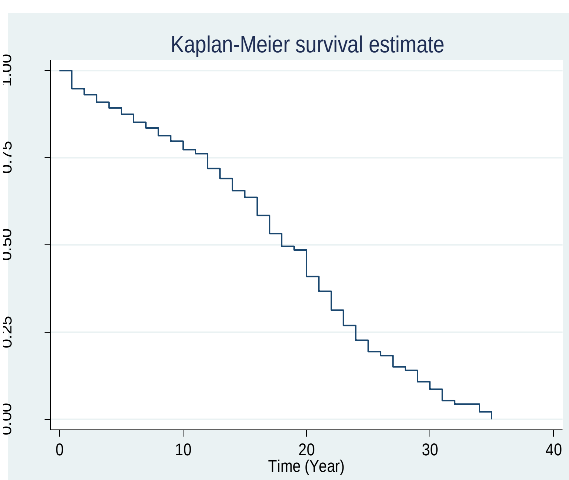


Figure 4.1: The plot of the overall estimate of Kaplan-Meier survivor function of waiting time of a woman to first birth after marriage in Ethiopia, EDHS, 2011

The Kaplan-Meier estimator survival curve gives the estimate of survivor function among different strata or groups of covariates to make comparisons. Separate graphs of the estimates of the Kaplan-Meier survivor functions are constructed for different categorical covariates. In general, the pattern that one survivorship function lying above another means the group defined by the upper curve has a better survival than the group defined by the lower curve. There are clear differences among the various groups. However, the differences are not clear among religion and employment Status of women (see Appendix D).

4.4 Results of the Cox proportional hazards model

In statistical modeling, when the number of variables is relatively large it can be computationally expensive to fit all possible models. Thus, one of the options is fitting a multivariable model containing the variables that are significant at a modest level of significance in a univariable analysis.

The model will be constructed by first identifying factors which are significant at p-value 0.25 in univariate Cox proportional hazard analysis. Wald chi-square test is used to test the significance of the probability. Accordingly, Table 4.4 is the summary of univariate analysis used to select potential predictors for further analysis.

Table 4.4 Results of the univariable proportional hazards Cox regression model of the waiting time to first birth from marriage in Ethiopia.

Variable	D.F	$\hat{\beta}$	$S.e(\hat{\beta})$	χ^2	-2LL	Pr > χ^2	$\widehat{HR}$	95% CI
Age at first marriage	2		0.006	262.48	12975.16	0.000		
15-20		-1.934	0.119	262.46		0.000	0.145	(0.114,0.183)
21-42		-0.721	0.084	73.522		0.000	0.486	(0.412,0.573)
Women Education	3			280.91	13003.20	0.000		
Primary	1	-1.580	0.129	149.58		0.000	0.206	(0.160, 0.265)
Secondary	1	-0.549	0.130	17.95		0.000	0.578	(0.448, 0.745)
Higher	1	-0.344	0.166	4.271		0.039	0.709	(0.512, 0.982)
Employment status Yes	1	0.039	0.074	0.281	13276.26	0.596	1.040	(0.899, 1.203)
Religion	3			3.362	13272.73	0.339		
Protestant	1	0.440	0.264	2.782		0.095	1.553	(0.926, 2.606)

Muslim	1	0.388	0.274	2.008		0.156	1.475	(0.826, 2.518)
Other	1	0.461	0.264	3.054		0.081	1.586	(0.946, 2.660)
Region	10			48.397	13230.09	0.000		
Affar	1	0.078	0.183	0.183		0.669	1.083	(0.755, 1.549)
Amhara	1	0.476	0.173	7.532		0.006	1.610	(1.146, 2.261)
Oromia	1	-0.144	0.170	0.716		0.397	0.866	(0.621, 1.208)
Somalia	1	-0.154	0.182	0.714		0.398	0.858	(0.601, 1.225)
Benishangul-Gumuz	1	-0.026	0.206	0.015		0.902	0.975	(0.650, 1.461)
SNNP	1	-0.024	0.187	0.016		0.898	0.976	(0.677, 1.407)
Gambela	1	-0.346	0.195	3.167		0.075	0.707	(0.483, 1.036)
Harari	1	0.177	0.189	0.880		0.348	1.194	(0.824, 1.730)
Addis Ababa	1	0.02	0.208	0.009		0.923	1.020	(0.679, 1.534)
Dire Dawa	1	0.378	0.189	3.978		0.046	1.459	(1.007, 2.115)
Residence Rural	1	0.336	0.077	19.202	13258.29	0.000	1.4	(1.204, 1.627)
wealth	2			11.88	13264.73	0.003		
Medium	1	-0.251	0.075	11.063		0.001	0.778	(0.671, 0.902)
Rich	1	0.205	0.105	3.628		0.051	0.816	(0.663, 1.001)
Husband Education	3			139.44	13143.68	0.000		
Primary	1	-1.134	0.118	92.692		0.000	0.322	(0.255, 405)
Secondary	1	-0.605	0.117	26.680		0.000	0.546	(0.434, 0.687)
Higher	1	-0.135	0.141	0.922		0.337	0.874	(0.663, 1.151)
The value of -2LL for the null model is 13276.55								

The most important subset of these predictors to be included in the multivariable model will be selected based on their contribution to the maximized log partial likelihood of the model (-2LL). The highest reduction in -2LL ($\hat{\beta}$) is observed for age at first marriage that reduced the value for the null/empty model, from 13276.55 to 12975.16, the difference is 301.39 and found to be statistically significant (p-value <0.0001). The next highest change is obtained for women education level of 273.35 followed by husband educational level 132.87.

Proceeding in this manner covariates will be eliminated in order of the magnitude in which they increase the -2LL ($\hat{\beta}$). Thus, using the Wald chi-square test, the predictors that are found to be significant were considered for the next multivariate analysis at p-value 0.25; Age at first marriage, education level of women, place of residence, region, wealth index and husband education level. Age at first marriage, educational level of women, place of residence, region, wealth index and husband educational level with p-value less than 0.05 (standard level of significance), have statistically significant positive influence on the waiting time to first birth from marriage. Religions and women Employment status were not significant.

4.5 Multivariable Analysis

From Table 4.4 above, variables which are significant in relation to the waiting time to first birth interval from marriage were included in multivariable analysis at 25% level of significance. The result of univariable analysis showed that six covariates are statistically important to be included in the multivariable analysis. Then, another multivariable Cox proportional regression was fitted by eliminating those covariates that are not significant at 10% level of significance. Accordingly, from the total of six covariates, wealth index (p-value 0.450) is eliminated from the model. Hence, at this point there is a multivariable model which includes the five covariates; Age at first marriage, women educational level, region, residence and husband educational level (Appendix-A, Table-1 and Table-2).

The next step was assessing the importance of the variables which were not significant in the univariable analysis as predictors or useful confounders of survival time of the waiting time to first birth and their effects. The effect of those variables that are not significant in multivariable analysis is also investigated. These variables are added one at a time in the model containing the variables age at first marriage, women educational level, region, residence and husband educational level which are significant at 5% significance level. The result of the analysis reveals that employment status was found to be significant and therefore it can be recalled in the model (Appendix B, Table 2).

The final step in model development strategy is to consider interaction terms that may be useful in the improvement of the model. Thus, all two-way possible interactions of the variables under the hierarchic principle are formed and the significance of adding each of the interactions in the main effects model, one at a time, is checked using the Wald test. The Wald test p-value indicated that age at first marriage * employment status, educational level of women * region and educational level of women * residence were found to be statistically significant (Appendix C, Table 1). Then by adding all significant interactions with the main effect, highly insignificant variables are eliminated one at a time and the process continued until only significant variables are obtained (Appendix C, Table 2-5).

Thus, the final model will contain the six variables which are: age at first marriage, women educational level, region, residence, husband educational level and employment status, and the two interaction variables which are: age at first marriage * employment status and educational level of women * region (out of the two interactions educational level of women * region is highly significant). However, the interaction variable Educational level of women * Region in the final model have thirty groups. It is difficult to write equation and analysis for each group. Therefore, the following table showed the fitted Cox-Proportional hazards model for covariates age at first marriage, women educational level, region, residence, employment status and husband educational level.

Table 4.5 Variables in the Final Model

Covariates	DF	$\hat{\beta}$	S.E ($\hat{\beta}$)	Chi-Square	Pr> ChiSq	Hazard Ratio	95% CI
Age at first marriage	2			187.872	0.000		
15-20	1	-1.752	0.129	184.184	0.000	0.173	(0.135,0.223)
20-42	1	-0.581	0.092	39.447	0.000	0.560	(0.467,0.671)
Women Education level	3			144.105	0.000		
Primary	1	-1.361	0.189	51.887	0.000	0.256	(0.177,0.371)
Secondary	1	-0.315	0.173	3.317	0.069	0.730	(0.520,1.024)
Higher	1	-0.188	0.177	1.126	0.289	0.829	(0.586,1.172)
Region	10			92.677	<0.001		
Affar	1	0.282	0.189	2.219	0.136	1.326	(0.915, 1.922)
Amhara	1	0.901	0.179	25.325	0.000	2.462	(1.734,3.498)
Oromia	1	0.452	0.179	6.392	0.011	1.572	(1.107,2.231)
Somalia	1	-0.114	0.189	0.364	0.546	0.892	(0.616,1.292)
Benishangul-Gumuz	1	0.243	0.21	1.322	0.250	1.275	(0.843,1.928)
SNNP	1	0.274	0.19	2.027	0.155	1.315	(0.902,1.919)
Gambela	1	-0.407	0.20	4.032	0.045	0.666	(0.447,.990)
Harari	1	-0.044	0.19	0.051	0.822	0.957	(0.650,1.408)
Addis Ababa	1	-0.058	0.20	0.078	0.780	0.943	(0.627,1.420)
Dire Dawa	1	0.078	0.19	0.161	0.688	1.081	(0.740,1.578)
Residence	1	-0.57	0.11	26.762	0.000	0.561	(0.451,0.698)
Employment status	1	0.182	0.08	5.104	0.024	1.199	(1.024, 1.404)

Husband Education	3			19.518	0.000		
Primary	1	-0.43	0.16	6.611	0.010	.648	(0.465,0.902)
Secondary	1	-0.07	0.15	0.241	0.624	.928	(0.689, 1.250)
Higher	1	0.096	0.14	0.410	0.522	1.100	(0.821,1.474)

Table 4.5 above shows the result of the fitted hazards model; based on the result predictors having statistical significant relationship with the hazards are observed.

4.6 Model Diagnostics

In survival regression analysis assessment of model adequacy, it is must to:

i) Test the assumption of proportional hazards

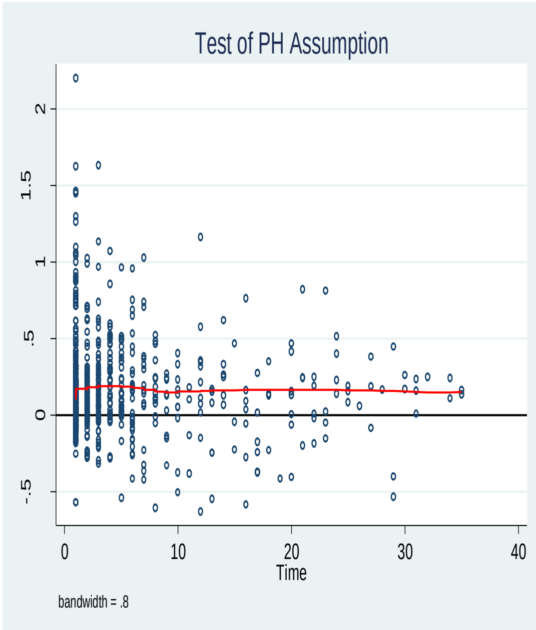
ii) Check influence and poorly fit subjects and

iii) Asses nonlinearity in the relationship between the log hazard and the covariates

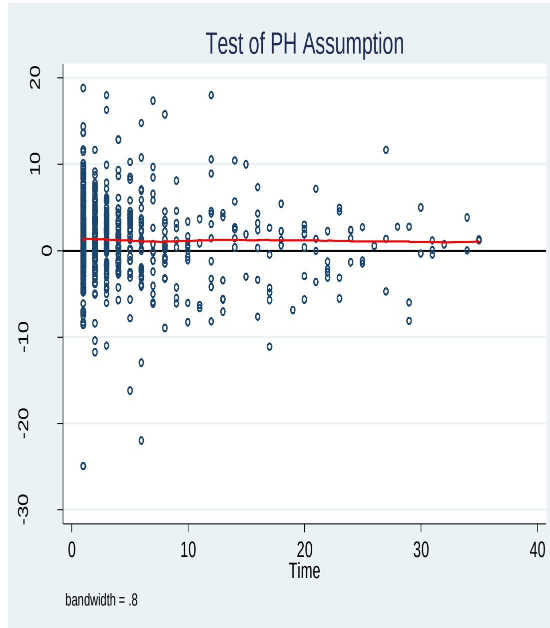
4.6.1 The assumption of proportional hazards

A proportional hazard is one of the very important assumptions in the Cox model. Cox proportional hazards models assume that the hazard ratio is constant over time. In order to test this assumption, the extended Cox model is employed and a graphical display is used to substantiate the same. Thus, in this study, a test based on the interaction of the covariates with the log of time and the plot of the scaled Schoenfeld residuals are used to see if the assumption of proportionality is violated or not.

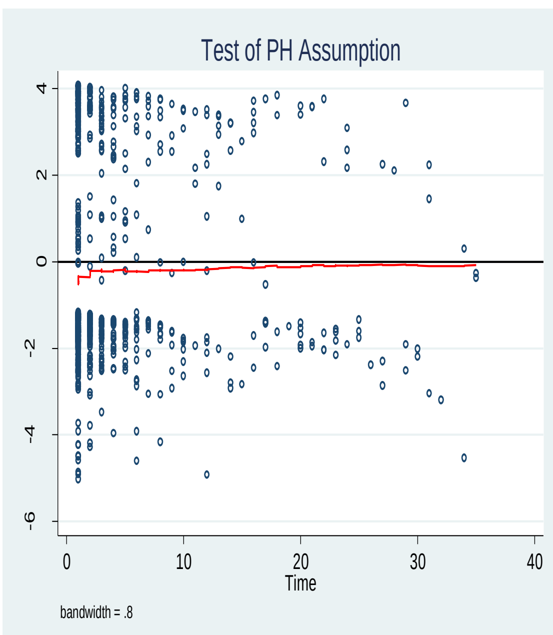
Under the assumption of proportionality of the proportional hazards model, the distribution of residuals overtime is random and LOWESS smoothing line should be a straight line around zero. Here below are the Graphical assessments of the proportional hazard assumptions.


Test of PH Assumption
-.5
0
.5
1
1.5
2
0
10
20
30
40
Time
bandwidth = .8

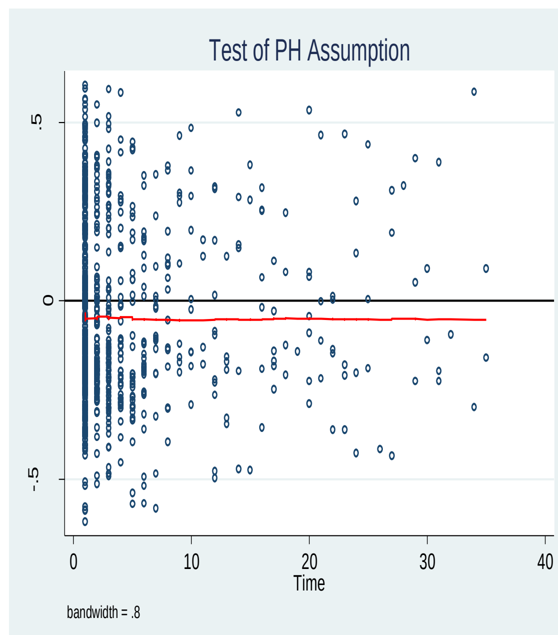
(1)


Test of PH Assumption
-30
-20
-10
0
10
20
0
10
20
30
40
Time
bandwidth = .8

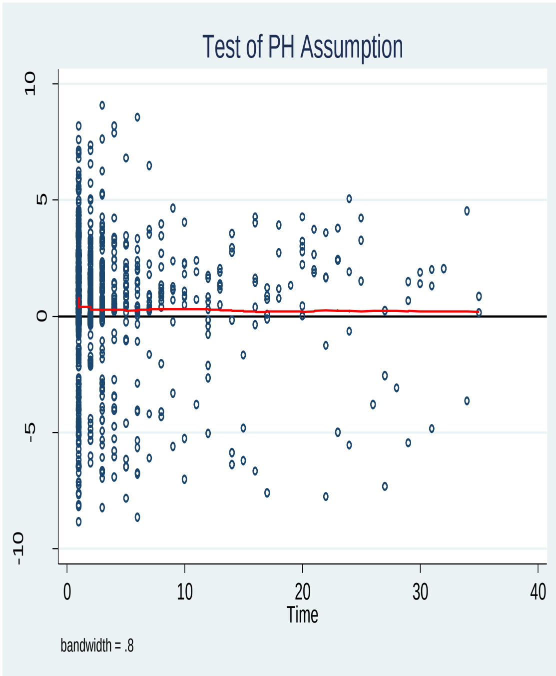
(2)


Test of PH Assumption
-6
-4
-2
0
2
4
0
10
20
30
40
Time
bandwidth = .8

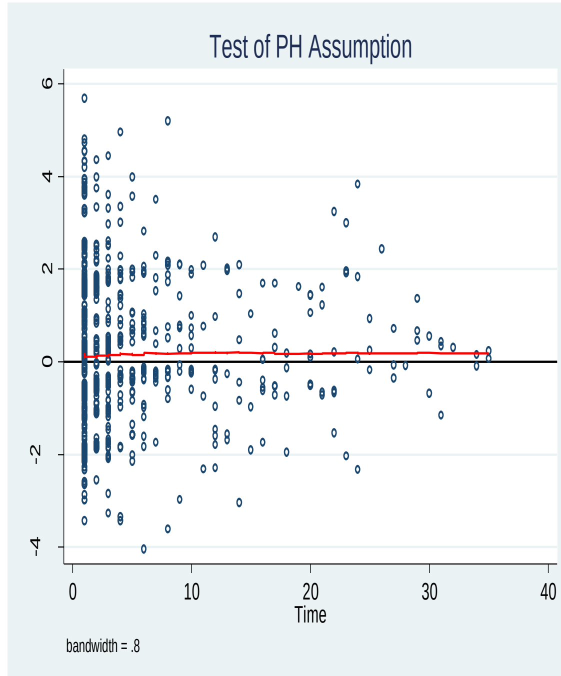
(3)


Test of PH Assumption
-.5
0
.5
0
10
20
30
40
Time
bandwidth = .8

(4)



(5)



(6)

Figure 4.2 plots of Scaled Schoenfeld residuals for 1) Age at first marriage, 2) Educational level of women, 3) Employment status of women, 4) Region 5) Residence and 6) Educational level of husband

The graphical display showed plots of the scaled Schonfield residuals against the survival time for each covariate:-Age at first marriage, Educational level of women, Employment status of women, Region, Residence and Educational level of husband. As it shown in the figure, the plots of scaled Schoenfeld residuals show randomness. Moreover, the smoothed curve is an approximate horizontal line. So the above six covariates have satisfied the assumption of proportional hazards.

The graphical test is not enough to be certain of the proportionality assumption of the model.
The reason is that it is subjected to the interpretation of different persons. That is, graphical method of assessing the proportionality assumption is more or less subjective but it can be used as a supportive argument for proportionality test.

The other statistical tests for proportional hazards assumption is to generate time varying covariates by creating interactions of the predictors and a function of survival times, usually product of covariate and log of time, and including them in the model. If any of the time dependent covariates are significant then those predictors do not show a proportional effect over the study period. That is the proportional hazard assumption fails to hold.

The result of the assumption of proportionality test showed that the chi-square value and corresponding p-values for each covariate. The p-values for the seven covariates are greater than 0.05, implying that the proportionality assumption is satisfied (Appendix D-Table 1).

4.6.2 Assessing for influential observations

In the assessment of model adequacy, it is important to determine whether any particular observation, if any, has an undue impact (leverage) on inferences made on the basis of model fitted to an observed set of survival data. It is therefore of particular interest to examine the influence of each particular observation on these estimates. This is done by examining the extent to which the estimated parameters and the maximized likelihood in the fitted model are affected by omitting the data record for each individual in the study. The DFBETA and the likelihood displacement statistics are used to examine the untoward effect of each observation on the j^{th} parameter estimate and the maximized log partial likelihood in the fitted Cox regression model respectively. Also Belsley, Kuh, and Welsch (1980) suggest that cutoff value for DFBETAs is $2/\sqrt{n}$, where n is the number of observations. However, another cut-off is to look for observations with a value greater than 1.00. Here cutoff means, this observation could be overly influential on the estimated coefficient. That is, if $|DFBETAS_{j,i}| > 2/\sqrt{n}$, then the i^{th}observation warrants examination. This is particularly made sense for only the large samples.

But, DFBETAs is mainly used for continuous variable and in this study there is no continuous variable used.

4.7 Overall Goodness of Fit

R^2 is used as a measure of overall goodness of model fit. As it is defined in chapter three, it is given as

$$R_p^2 = 1 - \exp\left(\frac{2}{N}(LL_o - LL_{\hat{\beta}})\right)$$

where N= 7594 is the total number of observation in the model.

LL_o = - 6638.275is the Log partial likelihood for model zero.

$LL_{\hat{\beta}}$ = - 6314.6015 is the Log partial likelihood for the fitted model with p covariates.

For the fitted model in Table 4.5, the value is

$$R_p^2 = 1 - \exp\left(\frac{2}{7594}(-6638.275 - (-6314.6015))\right)$$

$$= 0.1825$$

The model shown in Table 4.5 above has passed all the tests for a good fitting model.

4.8 Discussion of the results

This study used survival analysis techniques to study the waiting time to first birth after marriage of women using the 2011 EDHS data which contained both censored and uncensored responses. Non-parametric survival analysis techniques such as the life table, log-rank test and semi-parametric Cox proportional hazard regression analysis were used to investigate determinants of waiting time to first birth after marriage.

Factors that are considered in this study were age at first marriage, educational attainment of women, women's religion, place of residence, region, employment status of women, and husband educational level. The univariable analysis given in Table 4.4 revealed that two variable (religion and employment status of women) were not significantly related to the waiting time to first birth after marriage of women. Whereas, age at first marriage, educational attainment of women, place of residence, region and husband educational level were found to be statistically significant. All significant covariates in univariable analysis were included in multivariable analysis. By using purposeful selection, the final model has also included Employment status.

The Study revealed that the median time to first birth after marriage for couples is 2 years and a large portion of women gave birth within two years of marriage. From the hazard distribution shown in Table 4.2 the likelihood of a woman having her first birth is very high within the second and third years of marriage but in a fluctuating manner declines as her years in marriage increases until about 20 years. The married woman's possibility of giving birth is in the second year after her marriage where it is almost 46% (probability of 0.4577). This is closely followed by the third year in marriage where it is almost 40% (probability of 0.4057). Thereafter, their chances generally keep dropping until about 20 years in marriage. About 75% of married women had their first baby within the first four years of marriage (probability of 0.7537). This is probably due to the pressure of family and society members, and or explains the reason for the pressure mounted on women to prove their fertility early in the marriage.

Age at first marriage is found to be statistically significant positive influence on the first birth interval. The current study found that Age at first marriage is highly significant. $H\hat{R}$ =1.089 [95% CI 1.037, 1.145, p=0.001]

As per the study conducted by Ida (2013) in Ghana, women with higher educational attainment had shorter first birth interval. The relationship between education and length of first birth interval seems to be indirect. This study found that wives who had primary education had significantly different waiting times. Women who are primary educated have 95 percent lower likelihood of being a mother compared to illiterate wives. $H\hat{R}$ =0.256 [95% CI 0.117, 0.371, p=<0.001]. Moreover, education increases marital stability through secured financial resources (Ikamari, 2005). It is also believed to shorten first birth interval as at the time of entry to marital life they are emotionally prepared, biologically mature, and financially secured to have a child.

The study by Eshetu and Dula (2014) reveals that the likelihood of having first birth immediately after first marriage was significantly higher in Oromia and SNNP regions compared to Amhara region. For women in Oromia region, the likelihood of first birth soon after marriage had increased by 65% (p<0.001) while it was 63% higher for women in SNNP region (p<0.001). In this study region of respondent's is found to have a highly significant positive influence on the first birth interval for the women of Amhara and Oromia regions. As compared to Tigray region, Amhara region has 2.462 times higher likelihood of having first birth while Oromia has 58% higher likelihood of having first birth. $H\hat{R}$ =2.462 [95% CI 1.734, 3.498, p=0.000], $H\hat{R}$ =1.572 [95% CI 1.107, 2.231, p=0.011].

The study by Laurie (1986) suggested that the place of residence is a useful indication of change from traditional or rural behavior to modern behavior and that the first birth distribution for urban communities were more dispersed than those for rural distributions. These differences are associated with differences in marriage and fertility timing between rural and urban communities and are partly an impact of educational differentials between rural and urban areas on age at first marriage. This study also showed that place of residence significantly affect time to first birth distribution $H\hat{R}$=0.561 [95%, CI 0.451, 0.698, p=0.000].

A study in rural areas of Bangladesh by Bates et al. (2007) found that husband educational attainment leads to higher waiting time of first birth because husband roles give priorities to their children to reinforce their educational goals. Also similar findings are obtained by Choe et al. (2005) in Nepal and Zahangir et al. (2008) in rural Bangladesh. However, as per the findings of this study, husbands' educational level is partially significant on first birth interval. That is only women who are married with primary educated men have 32 percent lower likelihood of being a mother as compared to illiterate husbands of respondent's. $H\hat{R}$=0.648 [95% CI (0.465, 0.902), p=0.010]

Employment status is found to have statistically significant positive influence on the first birth interval. From the analysis, the respondents that are employed have 20 percent higher risk of being a mother than those who are not employed $H\hat{R}$=1.199 [95% CI (1.024, 1.404), p=0.024]

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

5.1 Conclusions

This study was based on a dataset of the waiting time to first birth after marriage obtained from the central statistical agency of Ethiopia with an aim of modeling the determinants of the waiting time to first birth by using Cox proportional hazard model. Out of the total 7594, about 89% were experienced an event (gave the first birth) and 11% were not experienced an event.

The Study revealed that the median time to first birth after marriage for couples is 2 years and a large portion of women gave birth within two years of marriage. The overall mean of first birth interval among Ethiopian women is 3.46 years.

The result of Cox proportional hazard model showed that the factors that determine the waiting time to first birth after marriage are age at first marriage, women educational level, region, place of residence, educational level of husband and employment status of the respondents. Where-as religion and wealth index were not statistically significant.

From the result of the study, urban women give birth as soon as they got married as compared to rural women. The rural women have lower likelihood of giving birth after marriage.

As the time of first birth is one of the prime determinants of fertility and is also determined largely by the demographical characteristics of the women. Therefore, it can be utilized for estimating various demographical determinants of human reproductions, the knowledge of which may be helpful in the assessment of impact of contraception along with the impact of socio-economic factors on fertility.

5.2 Recommendations

These recommendations are based on the findings of the analysis of the Ethiopian Demographic and Health Surveys (EDHS) dataset. Based on the result of the study different factors are identified for the waiting time of first birth after marriage of Ethiopian women.

These are the suggestions for future researchers and policy makers.

- This study has implications for government and concerned bodies that seek to increase age at first birth. It is crucial to continue improving girls and young women access to education, as this is an important avenue for rising the women's age at first marriage and for empowering women.
- Educate both spouses to change their reproductive behavior so that they delay their first child after marriage with respect to their age.
- Further investigations into the unexpected effects of modernization factors on first birth interval length by collecting relevant data on different aspects of these factors are needed.
- Delay in first birth interval should be suggested keeping in view the impact of socioeconomic and demographic characteristics. There is need to change the trend of people towards birth spacing irrespective of their education and status.

REFERENCES

Aalen O. (2008). Survival and Event History Analysis. Springer-Verlag, New York.

Abreham K. (2009). Survival of Tuberculosis Patients Treated under Directly Observed Treatment, Short Course in Addis Ababa, Ethiopia,

Allison, P.D. (1984). Event history analysis: Regression for longitudinal event data (Sage University paper series on quantitative applications in the social sciences, Number 07-046), Beverly Hills, CA: Sage.

Allison, P.D. and Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A coment on Gottman. Psychological Bulletin, 91, 393-403.

Assifa K. and Muhammad K.,(2013). Determinants of Marriage to First Birth Interval in Pakistan. Journal of Statistics Volume 20, 2013. pp. 44-68.

Ayalu, A. and Lindstrom, D. 2014. "Recent Trends in the Timing of First Sex and Marriage among Young Women in Ethiopia" African Population Studies 28, (2 supplement): 1157–1170.

Bates L. M., J. Maselko and S. R. Schuler (2007). Women's Education and the Timing of Marriage and Childbearing in the Next Generation: Evidence from Rural Bangladesh, Studies in Family Planning, 38(2): 101-112.

Bloessfeld, H.P., Hamerle, A. and Mayer, K.U. (1989). Event history analysis: Statistical theory and application in the social sciences. Hillsdale, NJ: Erlbaum.

Blossfeld H-P., (1995): Changes in the process of family formation and women's growing economic independence: A comparison of nine countries.

Casterline, J.B. and Trussell J., (1980): Age at First Birth: WFS Comparative Studies, No 15.Voorburg.

Choe M.K, S.Thapa and Mishra V. (2005). Early Marriage and Early motherhood in Nepal. Journal of Biosocial science 37(2):143-62.

Collett D. Modelling Survival Data for Medical Research (2003, 2nd Edition). Chapman-Hall, London.

Cox, D.R.(1972). Regression models and life tables. Journal of the Royal Statistical Society. Series B (Methodological), 34, No.2, 187-220.

Cox, D.R. and Oakes, D. (1984). Analysis of survival data. London: Chapman and Hall.

Davis, K. and J. Blake. 1956. Social structure and fertility: an analytic framework. *Economic and Cultural Change* 4(2):211-235.

Dress, K.A. (1986). The effect of gender identify on conversation. Social Psychology Quarterly, 49, 294-301.

Ermisch, J. and Pevalin, D.J. 2005. "Early Motherhood and Later Partnerships" Journal of Population Economics 18, (3): 469-489.

Eshetu G. and Dula E.,(2014). Age at First Marriage and First Birth Interval in Ethiopia: Analysis of the Roles of Social and Demographic Factor. African Population Studies Vol. 28, No. 3

Ezra, M. and Gurmu, E. 2002. "Correlates of Marriage and Family Patterns in Southern Ethiopia" Ethiopian Journal of Development Research 24,(1):58-91.

Feng W. and Quanhe, Y. 1996. "Age at Marriage and the First Birth Interval: The Emerging Change in Sexual Behavior among Young Couples in China" Population and Development Review 22, (2): 299-320.

Fergusson, D.M., Horwood, L.J. and Shannon, F.T. (1984). Proportional hazards model of a family background. Journal of Marriage and the Family, 46, 539-549.

Fichman, M. (1988). Motivational consequences of absence and attendance Proportional hazards estimation of a dynamic motivation model. Journal of Applied Psychology, 73, 119-134.

Finnas and Hoem, 1980 "Starting age and subsequent birth intervals in cohabitational unions in current danish cohorts", Volume 17, 275-295

Ford K., (1984). Timing and Spacing of Births: WFS Comparative Studies, No. 38. Voorburg.

Gardner, W. and Griffin, W.A. (1989). Methods for the analysis of parallel streams of continuously reported social behaviors. Psychological Bulletin, 446-455.

Gillespie R., (2001). Contextualizing Voluntary Childlessness Within A Postmodern Model of Reproduction: Implications for Health and Social Needs', Critical Social Policy, Vol. 21 (2).

Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnosis based on weighted residuals, Biometrika 81, 515-526.

Greenhouse, J.B., Stangle, D. and Bromberg, J. (1989). An introduction to the survival analysis method for the analysis of clinical trial data. Journal of Consulting and Clinical Psycology, 57, 536-544.

Heckman, J. and Singer, B. (1985). Longitudinal analysis of labor market data. New York: Cambridge University press.

Hoem B., (2000). Entry into motherhood in Sweden: The influence of economic factors on th rise and fall in fertility, 1986-1997. Demographic Research 2, Article 4.

Hong, Y. 2006. "Marital Decision-making and the Timing of First Birth in Rural China before the 1990s" Population Studies 60, (3): 329-341.

Hosmer, D.W. and Lemeshow S. (1999).Applied Survival Analysis. John Wiley and Sons, Inc., New York.

Ida Anuwoje L., (2013). Survival Analysis of Time to First Birth after Marriage. Research on Humanities and Social Sciences Vol.3, No.12.

Ikamari Lawrence D. E. (2005). The effect of education on the timing of marriage in Kenya.

Islam, S. 2009. "Differential Determinants of Birth Spacing Since Marriage to First Live Birth in Rural Bangladesh" Pertanika Journal of Social Science and Humanities 17, (1): 1-6.

Johnson, D. (1988). Panel analysis in family studies. Journal of Marriage and Family, 50, 949-955.

Joshi H., (2002). Production, Reproduction and Education: Women, Children and Work in Contemporary Britain. Population and Development Review 28 (3): 445-474.

Kaplan EL and Meier P. (1958). Nonparametric estimation from incomplete observations. Journals of American Statistical Association 1958; 53:457-81.

Khan, H.T.A. and R. Reside, 1998. The determinants of first and subsequent births in urban and rural areas of Bangladesh. Asia-Pacific Population Journal, 13(2): 39-72.

Kinfu, Y. 2000. "Below Replacement Fertility in Tropical Africa? Some Evidence from Addis Ababa" Journal of Population Research 17, (1): 63-82

Kleinbaum D.G. and Klein W. (2005). Survival Analysis a self-learning text, Second edition. Springer Science, Inc., New York.

Kohler H.P., Billary F.C. and Ortega J.A., (20002). The emergence of lowest-low fertility in Europe during the 1990s: population and Development Review28 (4): 641-680.

Kumar, G. and Danabalan, M. 2006. "Determinants of Delayed First Birth" Indian Journal of Community Medicine 31 (4).

Laurie, S. (1986): Age of Physical Maturation and First Intercourse in Black Teenage Males and Females. Demography 23(4):595-598.

MacInnes J., (2005). A Promising Relationship: Sociology and Demography.

Majumder, A.K., 1991. Breast feeding, birth interval and child mortality in Bangladesh. Journal of Biosocial Science, 23(3): 297-312.

Martin S. P., (2000). Diverging fertility among U.S. women who delay childbearing past age 30: Demography 37 (4): 523-533

Martin, W., D. Morley and M. Woodland, 1964. Interval between births in a Nigerian village. Journal of Tropical Pediatrics, 10(3): 82-85.

Miller, R.G. (1981). Survival analysis. John Wiley and Sons, New York.

Millman, S. and Hendershott, P. H. (1980). Early fertility and lifetime fertility. Family Planning Perspectives, 12, 139-149.

Molla, M., Berhane, Y. and Lindtjorn, B. 2008. "Traditional Values of Virginity and Sexual Behaviour in Rural Ethiopian Youth: Results from a Crosssectional Study" BMC Public Health 8(9).

Morita, J.G., Lee, T.W. and Mowday, R.T. (1989). Introducing survival analysis to organizational researchers: A selective application to turnover research. Journal of Applied Psychology, 74, 280-292.

Mosammat Z., Mohammad S. and S.M. Shafiqul Islam (2013). Age at first marriage and its relation to fertility in Bangladesh, Chinese Journal of Population Resources and Environment, 11:3, 227-235

Mukhlesur R., Mohitul A.,Mohammad A.(2013). Analysis of the determinant's of marriage to first birth interval in Bangladesh. International Journal of Management and Sustainability, 2(12):208-219

Murphy M., (1992). Economic models of fertilityin post-war Britain: A conceptual and statistical reinterpretation, Population Studies 46 (2): 235-258.

Murphy M., (1993). The contraceptive pill and women's employment as factors of fertility change in Britain 1963-1980: A challenge to conventional view, Population Studies, 47 (2): 221-243.

Ngalinda I. (2000). Age at first birth, Fertility and Contraception in Tazania

Nortman et al, 1976. The effect of breastfeeding and pace of chilbearing on mortality at early ages, 23:1. 1986.

Rao K.V. and Balakrishnan T.R., (1988). Age at First Birth in Canada: A Hazards Model Analysis.

Rindruss, R.R., J.A. Palmore and L. Bumpass, L., (1987). Analyzing birth intervals: Implications for demographic theory and data collection. Sociological Forum, 2(4): 811-828.

Rodriguez, G. and J. Hobcraft, (1980). Illustrative life table analysis of birth intervals in Colombia. World Fertility Survey Scientific Report No. 16. The Hague, International Statistical Institute, Voorburg, Netherlands. pp: 72

Salway, S., N.C. Roy, M.A. Koenig and J. Cleland, (1993). Levels and trends in postpartum amenorrhoea, breastfeeding and birth intervals in Matlab, Bangladesh: 1987-1989. Asia-Pacific Population Journal, 8(2): 3-21.

Shapiro, D.R., Quitkin, F.M. and Fliess, J.L. (1989). Response to maintenance therapy in bipolar illness: Effects of index episode. Archives of General Psychiatry, 46, 401-405

Shrestha, D.P. 1998. "Socio-economic Changes, Women's Autonomy, and Timing of First Birth in a Semi-urban Community in Nepal" Nepalese Studies 25, (Special Issue): 129-143.

Sobotka T., (2004). Postponement of Childbearing and Low Fertility in Europe: PhD thesis, University of Groningen, Dutch University Press, Amsterdam, 298 pp.

Stevens, V.J. and Hollis, J.F. (1989). Preventing smoking relapse using an individually tailed skills training technique. Journal of Consulting and Clinical Psychology, 57, 420-424.

Teachman, J.D. (1982). Methodological issue in the analysis of family formulation and dissolution. Journal of Marriage and the Family, 44, 433-445.

Trussell, J. and Menken, J. (1978). Early childbearing and subsequent fertility. Family Planning Perspectives, 10(4), 209-218.

Tuma, N.B. and Hannan, M.T. (1984). Social dynamics: Models and Methods. San Diego, C.A. Academic Press.

Yamaguchi, K. and Ferguson, L. (1995). The stopping and spacing of childbirths and their birth-history predictors: rational-choice theory and event history analysis. American Sociological Review, 60, 272-298.

Yerushalmy et al, 1956. The effect of breastfeeding and pace of chilbearing on mortality at early ages, 23: 1. 1986.

Zahangir M. S., M. A. Karim, M. R. Zaman, M. I. Hussain and M. S. Hossain, (2008). Determinants of age at first marriage of rural women in Bangladesh: A cohort analysis Trends Applied Science. Res. 4(3):335-343, Academic journal, Department of Statistics.

APPENDIX A

Results of the multivariable proportional hazards Cox regression model

Table 1: Results of the multivariable proportional hazards Cox regression model containing the variables significant at 20-25% level in the univariable proportional hazards Cox regression model

Analysis of Maximum Likelihood Estimates						
Covariates	DF	$\hat{\beta}$	s.e($\hat{\beta}$)	Chi-Square	Pr>ChiSq	Hazard Ratio
Age at first marriage	2			187.513	0.000	
		-1.751	0.129	183.227	0.000	0.174
		0.570	0.093	37.899	0.000	0.566
Women Education level	3			140.555	0.000	
Primary	1	-1.288	0.187	47.541	0.000	0.276
Secondary	1	-0.246	0.170	2.082	0.149	0.782
Higher	1	-0.125	0.175	0.509	0.476	0.883
Region	10			98.989	0.000	
Affar	1	0.298	0.189	2.481	0.115	1.348
Amhara	1	0.933	0.180	26.902	0.000	2.543
Oromia	1	0.446	0.179	6.211	0.013	1.563
Somalia	1	-0.132	0.189	0.486	0.485	0.876
Benishangul-Gumuz	1	0.257	0.211	1.476	0.224	1.293
SNNP	1	0.265	0.193	1.894	0.169	1.304
Gambela	1	-0.425	0.203	4.404	0.036	0.654
Harari	1	-0.060	0.198	0.093	0.76	0.941
Addis Ababa	1	-0.052	0.209	0.063	0.802	0.949
Dire Dawa	1	0.075	0.193	0.151	0.698	1.078
Residence	1	-0.569	0.122	21.876	0.000	0.566
Wealth	2			1.598	0.450	
Medium	1	0.031	0.101	0.097	0.756	1.032
Rich	1	0.148	0.123	1.452	0.228	1.159
Husband Education	3			19.391	0.000	
Primary	1	-0.443	0.17	6.809	0.009	0.642
Secondary	1	-0.087	0.152	0.323	0.570	0.917
Higher	1	0.105	0.149	0.491	0.483	1.110

Table 2: Results of the multivariable proportional hazards Cox regression model after elimination of the variable Wealth from the multivariable proportional hazards Cox regression model in Table 1.

Covariates	DF	Chi-Square	Pr>ChiSq
Age at first marriage	2	188.725	<0.001
Women Education level	3	141.068	<0.001
Region	10	100.05	<0.001
Residence	1	29.215	<0.001
Husband Education	3	19.373	<0.001

APPENDIX B

In the result of multivariable Cox hazard model, those variables not significant in both the univariable and multivariable analysis were fitted by including in the model containing variables significant in multivariate analysis one at a time**.**

I. Result of multivariate Cox proportional hazard model containing variables in Appendix A, Table 1 including those not significant in univariable analysis one at a time.

Table 1: When Religion is included

Covariates	DF	Chi-Square	Pr>ChiSq
Age at first marriage	2	189.251	<0.001
Women Education level	3	142.065	<0.001
Region	10	73.123	<0.001
Residence	1	29.301	<0.001
Husband Education	3	20.708	<0.001
Religion	3	4.496	0.213

Table 2: When Employment status is included and Religion is excluded

Covariates	DF	Chi-Square	Pr>ChiSq
Age at first marriage	2	187.872	<0.001
Women Education level	3	144.105	<0.001
Region	1	92.677	<0.001
Residence	1	26.762	<0.001
Husband Education	3	19.518	<0.001
Employment status	3	5.104	0.024

II. Factor not significant in multivariate model is included in the multivariate model of covariate in Table 1 (Appendix A) one at a time

Table 1: When Wealth index is included

Covariates	DF	Chi-Square	Pr>ChiSq
Age at first marriage	2	186.887	<0.001
Women Education level	3	143.486	<0.001
Region	1	92.064	<0.001
Residence	1	20.329	<0.001
Husband Education	3	19.327	<0.001
Employment status	3	5.156	0.023
Wealth index	2	1.402	0.437

APPENDIX C

Interaction of two covariates

Table 1: Wald statistics and corresponding p-values of possible interaction terms

Interaction between covariates/Factors		DF	Wald X^2	P-value
Age At First Marriage	Women Education	6	12.111	0.06
	Region	20	16.121	0.709
	Residence	2	3.401	0.183
	Husband Education	6	4.304	0.636
	Employment Status	2	7.931	0.019
Women Education	Region	30		
	Residence	3	8.544	0.036
	Husband Education	9	3.584	0.937
	Employment Status	3	7.539	0.057
Region	Residence	9	16.585	0.056
	Husband Education	30	41.764	0.075
	Employment Status	10	15.433	0.117
Residence	Educational level of Husband	3	3.493	0.322
	Employment Status	1	0.126	0.722
Husband Education	Employment Status	3	3.136	0.371

Table 2: Wald statistics and corresponding p-values of significant interaction terms including all at a time and eliminate the large significant interaction variable one at a time.

Interaction between covariates/Factors		DF	Wald X^2	P-value
Age At First Marriage	Employment Status	2	8.754	0.013

Women Education	Region	30	136.994	0.000
	Residence	3	3.486	0.323

Table 3: When the interaction of Women Education and Residence is excluded

Interaction between covariates/Factors		DF	Wald X^2	P-value
Age At First Marriage	Employment Status	2	8.868	0.012
Women Education	Region	30	141.246	0.000

APPENDIX D

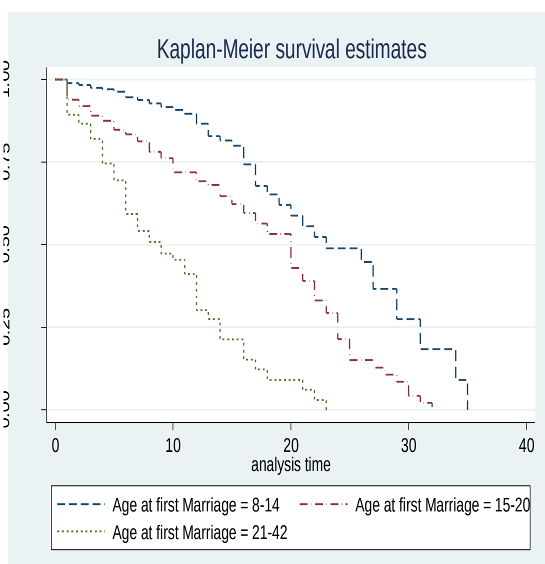
Table 1: Stata result of the assumption of proportionality test

Covariates	Parameter Estimate	Std. Err.	Wald X^2	P-value
Age at first Marriage				
15-20	1.317	0.139	9.55	0.000
21-42	1.973	0.182	10.84	0.000
Women Education				
Primary	1.093	0.103	10.61	0.000
Secondary	1.321	0.192	6.90	0.000
Higher	1.590	0.243	6.54	0.000
Region				
Affar	0.599	0.149	4.01	0.000
Amhara	0.188	0.143	1.32	0.187
Oromia	-0.378	0.157	-2.40	0.016
Somalia	-0.065	0.188	-0.35	0.730
Benishangul-Gumuz	0.009	0.164	0.06	0.955
SNNP	-0.684	0.175	-3.91	0.000
Gambela	-0.319	0.181	-1.76	0.079
Harari	-0.331	0.204	-1.63	0.104
Addis Ababa	-0.226	0.205	-1.11	0.269
Dire Dawa	-0.306	0.216	-1.41	0.157

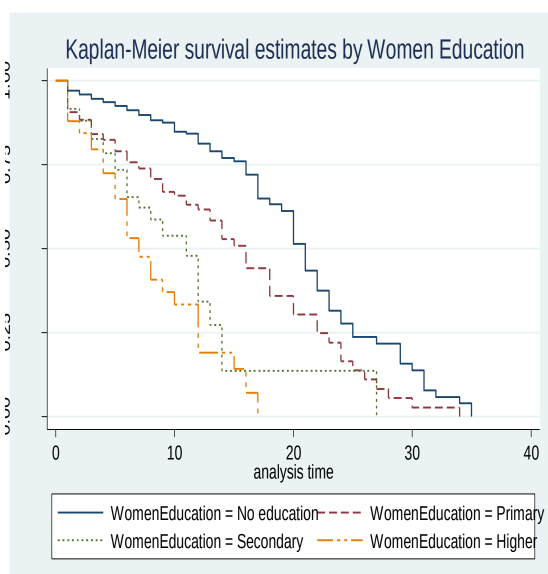
Residence	0.734	0.143	5.12	0.000
Husband Education				
Primary	0.282	0.102	2.77	0.006
Secondary	0.406	0.165	2.46	0.014
Higher	0.218	0.214	1.01	0.311
Employment status	-0.350	0.109	-3.22	0.001
Age at first Marriage	-0.107	0.066	-1.61	0.108
Women Education	-0.071	0.061	-1.16	0.246
Region	-0.000	0.009	-0.03	0.977
Residence	-0.158	0.102	-1.55	0.122
Husband Education	0.081	0.055	1.47	0.142

APPENDIX E

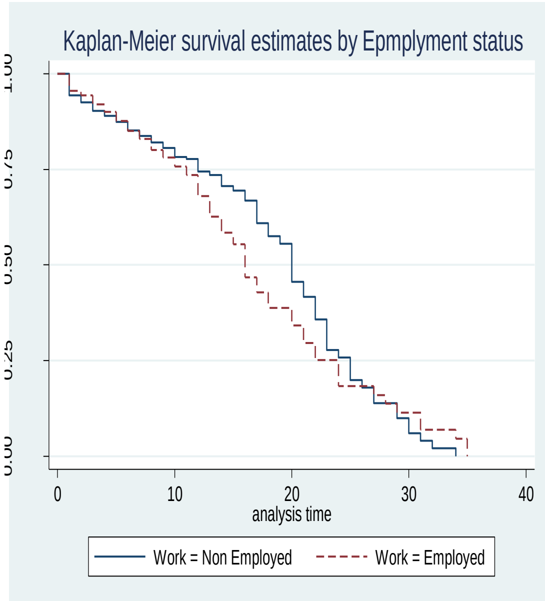
Figure 1 (1-8) Plots of Kaplan-Meier survivor functions based on different factors, waiting time of a woman to first birth after marriage in Ethiopia, EDHS, 2011

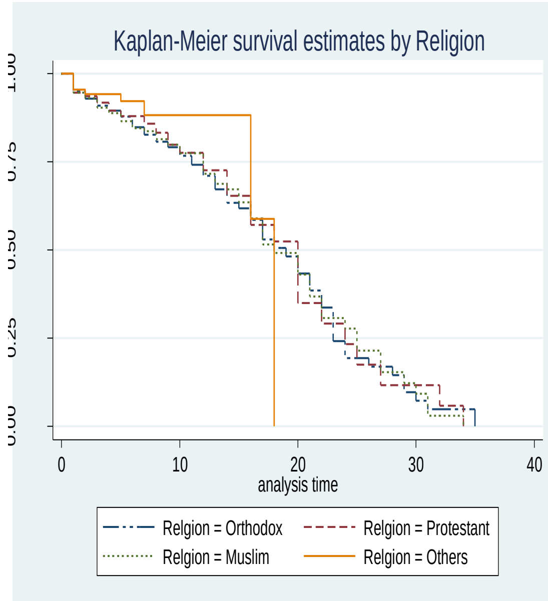


(1)



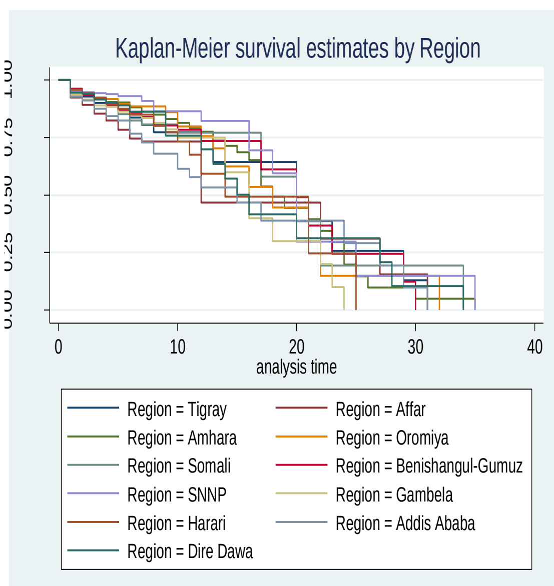
(2)

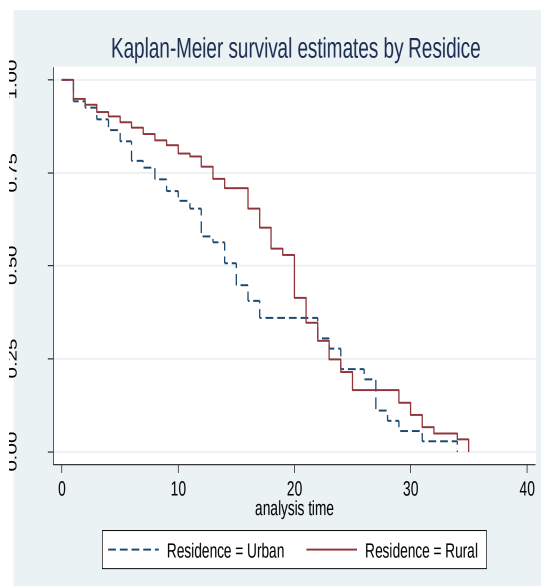

Kaplan-Meier survival estimates by Epmplyment status
0
10
20
30
40
analysis time
Work = Non Employed
Work = Employed


Kaplan-Meier survival estimates by Religion
0
10
20
30
40
analysis time
Relgion = Orthodox
Relgion = Protestant
Relgion = Muslim
Relgion = Others

(3)

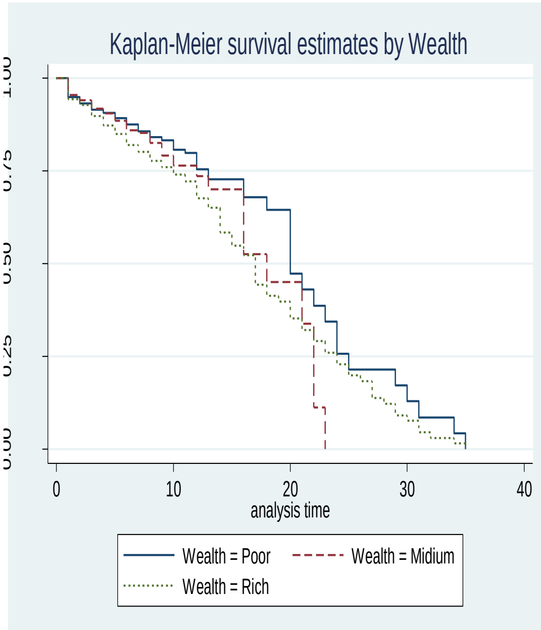
(4)


Kaplan-Meier survival estimates by Region
0
10
20
30
40
analysis time
Region = Tigray
Region = Affar
Region = Amhara
Region = Oromiya
Region = Somali
Region = Benishangul-Gumuz
Region = SNNP
Region = Gambela
Region = Harari
Region = Addis Ababa
Region = Dire Dawa

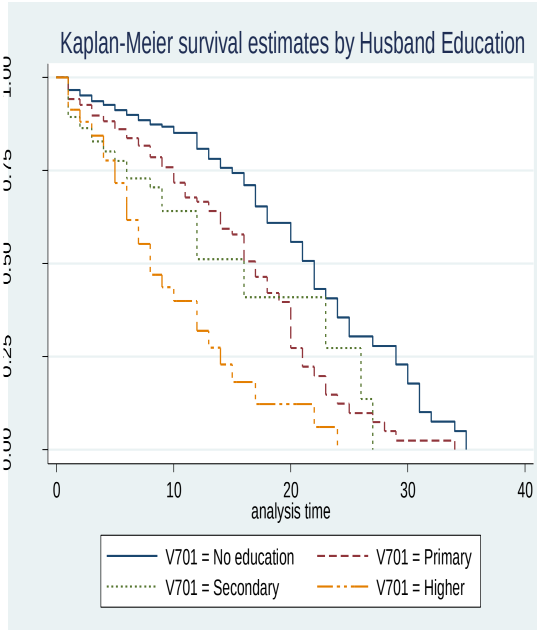

Kaplan-Meier survival estimates by Residice
0
10
20
30
40
analysis time
Residence = Urban
Residence = Rural

(5)

(6)


Kaplan-Meier survival estimates by Wealth
0
10
20
30
40
analysis time
Wealth = Poor
Wealth = Midium
Wealth = Rich

(7)


Kaplan-Meier survival estimates by Husband Education
0
10
20
30
40
analysis time
V701 = No education
V701 = Primary
V701 = Secondary
V701 = Higher

(8)