MODELING TIME-TO-FIRST ANTENATAL CARE (ANC) VISIT IN ETHIOPIA: A COMPARISON OF ACCELERATION FAILURE TIME AND PARAMETRIC SHARED FRAILTY MODELS

By
ZUFAN ASAYE (BSc)

A thesis submitted to department of Statistics, College of Natural Sciences, Jimma University, in partial fulfilment of the requirement for Masters of Science (MSc) in Biostatistics.

Jimma, University

December, 2018

MODELING TIME-TO-FIRST ANTENATAL CARE (ANC) VISIT IN ETHIOPIA: *A COMPARISON OF ACCELERATION FAILURE TIME AND PARAMETRIC SHARED FRAILTY MODELS*

MSc Thesis

BY: ZUFAN ASAYE (BSc)

Advisor: Dr Girma Taye (PhD)

Co-Advisor: Fikadu Zewude (MSc.)

Jimma, Ethiopia

December, 2018

JIMMA UNIVERSITY

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared by **Zufan Asaye**, entitled: Modeling Time-to-First Antenatal Care (ANC) visit in Ethiopia: A Comparison Of Acceleration Failure Time And Parametric Shared Frailty Models and summited in partial fulfilment of the requirements for the Degree of Master of Science (Biostatistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Dr Girma Taye (PhD) ______________________ __________________________

Main advisor Signature Date

Fikadu Zewude (MSc.) ______________________ __________________________

Co-advisor Signature Date

As the members of the board of examiners of MSc. thesis open defence examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfils the requirements for the degree of Master of Science in Biostatistics.

________________________ ________________________ ____________________

Name of Chairman Signature Date

________________________ ________________________ ____________________

Name of Advisor Signature Date

________________________ ________________________ ____________________

Name of Co-Advisor Signature Date

________________________ ____________________ ____________________

Name of Internal Examiner Signature Date

________________________ ________________________ ____________________

Name of External Examiner Signature Date

Table of Contents

ACKNOWLEDGMENT VII

ABSTRACT VIII

ACRONYMS IX

List of Table X

List of Figure XI

CHAPTER ONE 1

1. INTRODUCTION 1

1.1. Background 1

1.2. Statements of the problem 4

1.3. Objective of the Study 6

1.3.1. General objective 6

1.3.2. Specific objectives 6

1.4. Significance of the study 6

CHAPTER TWO 7

2. LITERATURE RIVIEW 7

2.1. Overview of ANC coverage and Timing of first visits 7

2.2. Factors affecting time of first ANC visit 8

2.2.2. Pregnancy (obstetric) related factors. 9

2.2.3. Knowledge of ANC schedule and Women autonomy 10

2.3. Survival Models 10

2.3.1. Implication of Random Effect of Frailty 11

2.3.2. Common Regression Model in Survival Analysis 13

CHAPTER THREE 14

3. DATA AND METHODOLOGY 14

3.1. Data Source 14

3.2. Sample Design 14

3.3. Study population 14

3.4. Inclusion and Exclusion criteria 14

3.5. Variables in the study 15

3.5.1. The response (dependent) variable 15

3.5.2. Predictor (independent) variables 15

3.6. Survival Models 16

3.6.1. Non-Parametric Survival Model 16

3.6.1.1. Kaplan-Meier estimator 17
3.6.2. Non parametric Comparison of Survival Functions 18
3.6.2.1. Non-Parametric Survival Analysis 18
3.6.3. Cox PH Regression Model..................... 18
3.6.3.1. Checking for Proportional Hazard Assumption..................... 19
3.6.4. Accelerated Failure Time Model 19
3.6.4.1 Estimation of AFT model 21
3.6.4.2. Weibull AFT model 21
3.6.4.3. Log-logistic AFT model 22
3.6.4.4. Log-normal AFT model..................... 23
3.6.5. Modeling Frailty 24
3.6.5.1. Multivariable frailty models 24
3.6.5.3. Gamma Frailty Distribution..................... 25
3.6.7. Prediction of Frailties..................... 28
3.6.8. Comparison of Models..................... 28
3.6.8.1. Model Diagnosis 28
3.6.8.1.1. Evaluation of the Parametric Baselines 28
3.6.8.1.2. The Cox Snell Residuals..................... 29
3.6.8.1.3. The Quantile - Quantile plot 30
CHAPTER FOUR..................... 31
4. RESULT AND DISCUSION 31
4.1. Descriptive Statistics..................... 31
4.2. Non-parametric Survival Analysis..................... 33
4.2.1. The Kaplan-Meier Estimate of Time-to-First ANC visit..................... 33
4.2.2. Comparison of place of residence in terms of survival time to first ANC visit.......... 34
4.2.3. Comparison of ‘survival’ time to first ANC visit by women education..................... 34
4.2.4. Comparison of survival time to first ANC visit by wealth index 35
4.2.5. Comparison of survival time to first ANC visit by Distance to get health care 36
4.3. Comparison of ‘Survival’ Experience by log-rank test 36
4.4. Testing the Validity of Proportional Hazard Assumption 37
4.5. Accelerated Failure Time Model Result..................... 39
4.5.1. Univariate Analysis..................... 39
4.5.2. Multivariable AFT Analysis 39
4.6. Parametric Shared Frailty Model Results 42

4.6.1. Log-logistic Gamma Shared Frailty Model Result 42
4.7. Comparison of Log-normal AFT and Log-Logistic-Gamma Frailty Model 45
4.8. Model Diagnostics 46
4.8.1. Diagnostic Plots of the Parametric Baselines 46
4.8.2. The Cox-Snell Residual Plot... 47
4.8.3. Quantile-Quantile plot 48
4.9. DISCUSSION 50
CHAPTER FIVE 54
5. CONCLUSION AND RECOMMENDATIONS 54
5.1. Conclusion 54
5.2. Recommendation 55
5.3. Limitation of the study.. 55
REFERENCES 56
APPENDEX... 62

ACKNOWLEDGMENT

First, I would like to thank the almighty God for his grace, wisdom, favour, faithfulness and protection.

Secondly, I am grateful for the excellent guidance and for his meticulous feedback and great encouragement Dr. Girma Taye. I would also like to thank my co-advisor Mr. Fikadu Zewude for his valuable suggestions and comments to the successful realization of this thesis.

I would like to thank the Jimma University and my sponsor Mizan Tepi University for providing me to attend my training and Ethiopian Central Statistics Agency for providing me with all the relevant secondary data used in this study.

My sincere appreciation also goes to my husband, who support and encouragement allowed me to finish this journey and joined us when I was preparing for my thesis defense, for giving me unlimited happiness.

ABSTRACT

Background: *The survival of pregnant women is one of great interest of the world and especially to a developing country like Ethiopia which had the highest maternal mortality ratios in the world due to low utilization of maternal health services including Antenatal care (ANC).Survival analysis is a statistical method for data analysis where the outcome variable of interest is time to occurrence of an event. AFT and frailty model is an extension of Cox's PH model in which the hazard function depends upon time and an unobservable random quantity called frailty. Regional states of the women were used as a clustering effect in all frailty models.*

Methodology: *The study aimed to model the determinants of time-to-first antenatal care visits to Ethiopia. The data for the study were taken from the 2016 EDHS and data of 7161 women in the age group of 15-49 years, who got pregnancy during five years survey whom survival information available were included in the analysis. The AFT and gamma shared frailty models with weibull, log-normal and log-logistic baseline distribution were employed to identify the best model fit for the timing of first ANC visit using health-related risk factors, socio-economic and demographic factors. All the fitted models were compared by AIC.*

Results: *The median of time of first ANC visit was 5 months. The log-logistic with Gamma shared frailty model is an appropriate model when compared with other models for a time at first ANC visit dataset based on AIC and graphical evidence. The clustering effect was significant for modelling the determinants of time-to-first ANC visit dataset. The final model showed that place of residence, perceived problem to get medical care due to distance, wanted pregnancy; women and husband education level, religions, wealth index, and parity were found to be significant determinants of time at first ANC visit at 5% level of significance. The estimated acceleration factor for the group of women's who had secondary and higher educational level was highly earlier time at first ANC visit by the factor of $\phi=0.89$ and $\phi=0.86$ respectively.*

Conclusion: *The log-logistic with gamma shared frailty model described time at first ANC visit data set better than other models and there was heterogeneity between the regions on time-to-first ANC initiation. Specific efforts are needed to target women of lower socioeconomic status, access to informal education for woman and husband, accessing health facilities due to distance and give awareness about having few numbers of children was an important avenue for rising women's time at first ANC visit.*

Keywords*: acceleration failure time, frailty, gamma shared frailty, ANC, Visit*

ACRONYMS

ANC	Antenatal Care
AFT	Accelerated Failure Time
AIC	Akaike's Information Criterion
AOR	Adjusted Odd Ration
BIC	Bayesian Information Criterion
CSA	Central Statistical Agency
EDHS	Ethiopia Demographic and Health Survey
HIV	Human Immunodeficiency Virus
HSDP	Health Sector Development Program
PH	Proportional Hazard
WHO	World Health Organization

List of Table

Table 4. 1. Descriptive statistics of time to first ANC visit among pregnant women Ethiopia (EDHS2016). 32
Table 4. 2. Log-rank test for factors affecting time to first ANC visit in Ethiopia...................... 36
Table 4. 3. Test of proportional hazard assumptions………………………………………… 39
Table 4. 4. Comparison of Accelerated Failure Time Models using AIC, BIC and -2loglik in for the assessment of time to first ANC visit in Ethiopia. 39
Table 4. 5. Summary result of the final Log-Normal AFT model in the assessment of time to first ANC visit in Ethiopia.. 40
Table 4. 6. Comparison of shared gamma frailty model with different baseline distributions..... 42
Table 4. 7. Log-logistic Gamma Frailty Model Result in assessments of time to first ANC visit in Ethiopia. 43
Table 4. 8. Comparison of Log-normal AFT and Log-logistic Gamma Frailty Model in the assessment of time to first ANC visit in Ethiopia. 45

List of Figure

Figure4.1. The K-M plots of Survival and hazard functions of time to first ANC visit among pregnant mothers in Ethiopia. 33
Figure 4. 2. Survival time to first ANC visit among pregnant women by place of residence 34
Figure 4. 3. Comparison time to first ANC visit among pregnant women in Ethiopia by women educational status 35
Figure 4. 4. Comparison of survival time to first ANC visit among pregnant women in Ethiopia by wealth index 35
Figure 4. 5. Comparison of waiting time to first ANC visit among pregnant women in Ethiopia by perceived problem to get medical care due to distance 36
Figure 4. 6. Graphs of Weibull, Log-logistic and Log-normal baseline distribution for time-to-first ANC visit 47
Figure 4. 7. Estimated cumulative hazard plot of the Cox-Snell residuals for Weibull, lognormal and log-logistic distribution 48
Figure 4. 8. Quantile-Quantile plot to check the adequacy of the accelerated failure time model 49

CHAPTER ONE

1. INTRODUCTION

1.1. Background

According to the World Health Organization (WHO) definition, antenatal care is a pregnancy-related essential health care, which could be given either in a health facility or at home and it is an integral component of maternal and child health (Wang et al., 2011). It is considered as one of the most important factors for the health of the mother and optimal development of the fetus as well as for preventing or minimizing the complication of pregnancy (World Bank Group, 2016). Currently, a minimum of four ANC visits is recommended by WHO for women whose pregnancies are progressing normally with personalized visit interval, the first visit in the first trimester should be before 16 weeks; 2nd visit 24-26 weeks; 3rd visit 28-32 weeks and 4th visit; 36-38 weeks (Kibaru, 2007).

Complications of pregnancy and childbirth are the leading causes of disability and death among women in the reproductive age especially in developing countries (Ononokpono and Odimegwu, 2014; World Bank Group, 2016). In 2015, an estimated 303,000 women died as a result of pregnancy and childbirth-related complications and about 99% of these deaths were occurred in low and middle-income countries, with sub-Saharan Africa alone accounting for roughly 66% (201,000) of these deaths. In addition, maternal mortality ratio (MMR) in Sub-Sahara Africa, in 2015 was 546 maternal deaths per 100,000 live births; in which it was 17 maternal deaths per 100,000 live birth for high-income countries (WHO, 2016). Ethiopia is also among countries with the highest maternal mortality ratios in the world. According to 2016 Ethiopia demographic and health survey (2016 EDHS), pregnancy-related maternal mortality ratio was 412 maternal deaths per 100,000 live births (Central Statistical Agency (CSA) [Ethiopia] and ICF, 2017).

One of the key strategy in reducing maternal mortality is access to and utilization of antenatal care services (Were et al., 2013; WHO et al., 2010; World Health Organization, 2010). Timely and appropriate ANC provides an opportunity for early detection of diseases and timely treatment (Halim et al., 2010). In addition, it provides opportunities for preventive health care services such as immunization against neonatal tetanus, prophylactic treatment of malaria and HIV counselling and testing, monitoring of chronic conditions such as anaemia. Furthermore, antenatal care exposes pregnant women to counselling and education about their own health and the care of their children (Babalola and Fatusi, 2009). Promoting the use of ANC could be instrumental in encouraging women to seek skilled assistance at birth (Gage and Calixte, 2006; Kabir et al., 2005).

Under normal circumstances, WHO recommends pregnant woman should visit ANC service at least four times at specified intervals during pregnancy starting the first visit at 16 weeks of gestational age (CSA Ethiopia and ICF, 2016; World Health Organization, 2010). However, in developing countries, the majority of pregnant women start ANC visit at a later time than recommended by WHO hence a larger number of pregnant women had fewer four ANC visits. For instance, globally, only 64% of women receive ANC four or more times throughout their pregnancy. Reports showed slower progress in Sub-Saharan Africa than in other regions (Lincetto et al., 2013). Although substantial progress has been made over the past two decades in Ethiopia, many pregnant women did not receive ANC or receive less than the recommended number of ANC visit and fewer women start ANC at an appropriate time. EDHS 2011 and 2016 showed, and 34% and 62% of women who had a live birth in the 5 years before the survey received ANC from a skilled provider at least once for their last birth respectively. In addition, only 19% and 32% of women had four or more ANC visits during their last pregnancy respectively. Furthermore, 11% and 20% of women made their first ANC visit before the fourth month of pregnancy during these two surveys (CSA Ethiopia and ICF, 2016).

The time of the first ANC visit, as well as the total number of ANC visit also affect the quality of ANC that a pregnant woman receives, and under-attending the recommended ANC service may lead to adverse pregnancy outcomes. This is primarily because different services and interventions are available for different gestational ages. Mothers who attend ANC late miss the opportunity to receive health information and interventions such as early detection of HIV, malaria, and anaemia prophylaxis, and prevention or management of complications (Belayneh, Adefris, and Andargie, 2014).

Time at ANC visit has affected by multiple factors; therefore, solutions are not come through a single detection but rather from an array of innovations addressing multiple biological, clinical, and social factors. Therefore, survival analysis is used in several fields of data analysis existing between the event occurrence and situation changing time. The two classes of regression models for survival data are Cox PH models as semi-parametric models (Cox, 1972) and Accelerated Failure Time (AFT) models and frailty model as parametric models (Collett, 2003). Cox PH models relate the hazard function to covariates, but no assumptions are made nature or shape of hazard function, while the AFT models specify a direct relationship between the failure time and covariates.

Frailty model accounts for unobserved heterogeneity that occurs because some observations are more disposed to failure. Therefore, it introduces an additional parameter to the hazard function have been developed and lead to hidden heterogeneity or frailty models that account for random frailties. Vaupel, (1979) introduced a random effects model in order to account for unobserved heterogeneity due to unobserved susceptibility to the event. There may be an association between the times to events of some subgroups of the population since that individual share the common trait (family, litter, study centre) that cannot be measured. In this study, time-to-first ANC visit was clustered by the region. Hence, the effect of the region considered as the frailty term in the survival model. The study uses non-parametric, semi-parametric and parametric with (weibull, log-logistic and log-normal) gamma shared frailty model in determining the factors which affect the time-to-first ANC visit and AFT to compare and get the best model which fits the time-to-first ANC visit data appropriately by using AIC and BIC.

1.2. Statements of the problem

The survival of pregnant women is one of great interest of the world and especially to a developing country like Ethiopia. Ethiopia is one of the countries with the highest maternal mortality ratios in the world. Approximately 80% of maternal deaths globally occur due to haemorrhage, sepsis, unsafe induced abortion, a hypertensive disorder of pregnancy and obstructed labour. These deaths were unjust and preventable with improved key interventions like antenatal care, safe delivery and postpartum care (Chimankar and Sahoo, 2011). Likewise, studies revealed that this highest rate of maternal death in developing countries has been attributed to the inadequate use of maternal health care service including ANC (Berhe et al., 2014; World Health Organization, 2010).

Studies have been conducted to identify factors affecting the timing of first ANC visit but almost all of them used binary logistic regression by categorizing time of first ANC visit as timely and delayed visit (Belayneh, Adefris, and Andargie, 2014; Berhem *et al.*, 2014; Gupta *et al.*, 2015). However, binary logistic regression does not account for censored observations hence survival analysis is more appropriate for analysing data where the outcome variable is the time until the occurrence of an event of interest. Also in such applications, it is assumed that all heterogeneity is captured by theoretically relevant covariates (Trussell and Richards, 1985). In many situations, however, there are ample reasons to suspect omitted or unmeasured factors. That is, while some individuals are more at risk of experiencing the event, it is unlikely that the underlying reason for this variability is fully captured by the observed covariates. If there is unmeasured frailty, the hazard would not only be a function of the covariates but also the frailty introduced independently by (Clayton, 1978; Vaupel et al., 1979).

To assess the true effects of the observed covariates under this circumstance, some have stressed the need to explicitly account for unobserved heterogeneity (Gutierrez, 2002; Lancaster, 1979; Vaupel et al., 1979). Indeed, results from several empirical and simulation studies have shown that accounting for significantly improves overall model fitness (Sastry, 2017). Therefore, in this study, we argued that clustering (frailty) has an effect on modelling the determinants of time to first ANC, which might be due to heterogeneity in regions of study.

This research aimed to explore factors that affect time-to-first ANC visit after pregnancy by comparison of parametric accelerated failure time and gamma shared frailty model. Frailty term is added to account for the correlation which comes from the cluster, accounts for unobservable random effect, while AFT models specify a direct relationship between the failure time and covariates, which may be appropriate when a covariate acts to speed up or slow down the expected time to failure by contracting or expanding time scale (Lambert et al., 2004). In such applications, it is assumed that all heterogeneity is captured by theoretically relevant covariates (Trussell and Rodriguez, 1990). In this study, demographic, reproductive, and socioeconomic characteristics of pregnant mothers was assessed by using AFT and frailty models for investigating model that best fit time to first ANC visit data set and predictors of time to first ANC visit.

In general, the motivation behind this study is to address the following major research questions:

- What are the key socio-economic, socio-demographic and reproductive health-related factors are predicting time-to-first ANC visit among pregnant women in Ethiopia?
- Which baseline distributional assumption among the weibull, log-logistic and lognormal, as well as frailty distributions the gamma and accelerated failure time, describes well time-to-first ANC visit?
- Does the time-to-first ANC visit of pregnant women varies across the regional states of Ethiopia?

1.3. Objective of the Study

1.3.1. General objective

To model the determinants of time-to-first ANC receipt among pregnant women using various parametric model approaches.

1.3.2. Specific objectives

The specific objectives of this study are to: -

- Identify factors associated with time-to-first ANC receipt among pregnant women in Ethiopia.
- Determine parametric baseline hazard to help model the determinants at first ANC visits.
- To test whether there is a clustering effect on modeling the time of first ANC visit, which might be due to the heterogeneity in regions of study pregnancy woman.
- To compare the performance of AFT and parametric frailty model in modelling time-to-first ANC dataset.

1.4. Significance of the study

The result of this study will provide information on time-to-first ANC visit by analysing the effect of regional difference on survival time to ANC visit.

Specifically;

- The result evaluates variation of survival time of pregnant women and variation of regions based on the selected model which is best fitting time to first ANC visits data and hence it helps to popularize these models to be used by medical and public health researchers.
- The findings from this study may have a contribution to the improvement of the health status of the mother in the country in general by providing new information regarding specific barriers to timely ANC initiation for pregnant women.
- This, in turn, play a significant role in reducing the maternal mortality rate and quicken the country's footsteps in its journey towards sustainable development goals and health sector transformation plan in which maternal and new-born health are priorities for both international and Ethiopian governments.

CHAPTER TWO

2. LITERATURE RIVIEW

2.1. Overview of ANC coverage and Timing of first visits

The slow progress in the reduction of maternal and new-born mortalities in developing countries can be partially attributed to the low utilization of continuum care before, during, and after pregnancy. Existing evidence from developing countries including Ethiopia indicates that few women seek antenatal care at the early stage of their pregnancy. Globally, only 64% of women receive ANC four or more times throughout their pregnancy. In addition, 71% of women worldwide receive at least one visit ANC but in industrialized countries, over 95% of pregnant women have access to ANC. In Sub-Saharan Africa, 69% of pregnant women have at least one ANC visit, more than in South Asia which was 54% (Lincetto et al., 2010). Studies in Ethiopia showed that larger proportion of pregnant women did not attend ANC during their pregnancy (Berhe et al., 2014; Francis, 2017; Gedefaw et al., 2014; Gurmesa, 2014; Shewa et al., 2014).

Late attendant of ANC (not well-timed) is also another challenge to the effort to reduce maternal death, especially in developing countries. Among women who had ANC visit, a greater number of them start ANC visits later than the recommended time. For instance, a study conducted in Zambia showed that 81% of ANC attendant started their first visit between $6^{th}$and 9^{th}months gestation (Nyambe et al., 2016). Analysis of demographic health survey from three Africa countries; Botswana, Nigeria, and South Africa showed about 58% pregnant mother accessed ANC in the second trimester (Fagbamigbe et al., 2017). Other study conducted in Tanzania also reported only 29% of women enrolled in antenatal care within the first trimester (Gross et al., 2012).

Similarly finding from a study conducted Nigeria revealed that less than one-third (32.3%) of the women accessed ANC within the first 3 months of pregnancy (Francis, 2017). Likewise according to EDHS 2016 report only 20% of women made their first ANC visit before the fourth month of pregnancy(CSA Ethiopia and ICF, 2016). Furthermore, studies conducted in Addis Ababa reported that about 42.0% (95% CI of 38.9%, 45.1%) of ANC attendant start their first visit after 16 weeks of gestation (Hanna and Berhane, 2017).

2.2. Factors affecting time of first ANC visit

2.2.1. Socio-Demographic and Socio- Economic Factors

The body of research identified that different socio-demographic and socio-economic factors are associated with time to first ANC initiation. A study conducted in three Africa countries; Botswana, Nigeria, and South Africa revealed that older age pregnant mothers were more likely to be enrolled into ANC services earlier than women who were younger age (Fagbamigbe et al., 2017). In contrast to this, a study conducted in Gondar, northern Ethiopia showed that mothers who are aged 25 years and below were about 2 times more likely to start ANC within the recommended time than those whose age was above 25 years (Temesgen et al., 2014). Opposite to these finding from other study showed that adolescent pregnant women started antenatal care no later than adult pregnant (Gross et al., 2012).

Analysis of Nepal demographic and health survey showed that women with higher-level education were more likely (AOR: 11.40, 95% CI: 5.05–25.73 to initiate ANC early when compared to women who had never attended school (Gilles, 2017). Similarly, a study conducted among pregnant women in Nigeria and Uganda find out that women with higher educational level were more likely to start their first ANC visits at earlier gestational age relative uneducated (Francis, 2017; Turyasima et al., 2015). Furthermore, a cross section survey conducted in Ambo, central Ethiopia, reported that pregnant mother who had attended Grade 12 and above had a higher likelihood of initiated ANC visits in the first trimester than illiterate mother (AOR: 2.10, 95% CI =1.13, 3.82) (Damme et al., 2015).

In addition, a meta-analysis showed that maternal education was significantly associated with time of ANC initiation in which women who have attended primary or above level of education were less likely to delay their first ANC visit as compared to women without formal education (Gezahegn et al., 2017). Furthermore, the association between husband's education and time of ANC initiation was identified by different. Accordingly, meta-analysis done on delayed initiation of antenatal care and associated factors in Ethiopia showed that women having a husband who attended formal education were less likely to delay their first antenatal care visit as compared to those women whose husband had never attended formal education (OR, 0.44; 95% CI: 0.23, 0.85) (Gezahegn et al., 2017).

The importance of place of residence in determining the time of first ANC initiation acknowledged by different studies. For instance, cross-sectional study conducted in central Ethiopia revealed urban residents were 2.86 times (AOR: 2.86, 95% CI: 1.11, 4.38) more likely to be booked for ANC within appropriate time than the rural resident (Damme et al., 2015). Similarly, Yaya et al., (2017) also reported rural resident was associated with higher odds of delayed initiation of ANC visits. Other studies also showed there were differences in time of first ANC visit and place of residence (Francis, 2017; Gezahegn et al., 2017).

Furthermore, studies also showed that there was a difference in timing of the first ANC visit along different religious groups. It was revealed that Christianity women were more likely to start ANC earlier (Francis, 2017) than Muslims. In addition, reports indicated that the odds for a Muslim woman to achieve the recommended number of ANC visits and deliver in the modern health facility was very low compared to their Christian counterparts residing in the same (Umar, 2017).

In addition to socio-demographic factors discussed above, socio-economic factors such as household income (wealth index) have been found to be equally important in explaining the timing of the first ANC visit. Gilles, (2017) suggested that women from richest wealth quintile were significantly more likely to initiate ANC early (AOR: 3.74, 95% CI: 2.31–6.05) compared to the poorest. Other studies conducted in Ethiopia showed that pregnant mother with low monthly income was more likely present late for ANC visit (Damme et al., 2015; Gebremeske, 2014; Girum, 2016).

2.2.2. Pregnancy (obstetric) related factors.

Different studies concluded that number of factors related to either current or past pregnancy affect the time of first ANC receipt. Gross et al., (2012) reported that prim-parity and previous experience of a miscarriage or stillbirth were associated with an earlier antenatal care attendance. Similarly, cross-sectional study conducted in Zimbabwe reveals that lower parity was significantly associated with early ANC initiation (Erica, 2012). In addition, different studies revealed that women whose pregnancy was unwanted were significantly less likely to attend first ANC at recommended time in comparison to women whose pregnancy was wanted (Gebremeske, 2014; Gilles, 2017; Girum, 2016). Furthermore, women who have ever give birth were more likely to book timely (within 16 weeks of pregnancy) for the first visit of antenatal care (Gidey et al., 2017). In addition, meta-analysis study showed nulliparous were less likely to have delayed their ANC initiation as compared to women who were prime para and above (Gezahegn et al., 2017).

2.2.3. Knowledge of ANC schedule and Women autonomy

In addition, the factors mentioned above, knowledge of ANC schedule and women's decision making power in the household also play an important role in determining the time of ANC initiation during pregnancy. A systematic review different researches showed that first antenatal care visit was later for women who did not know the antenatal care initiation schedule correctly compared to women who knew the schedule correctly (Hanna and Berhane, 2017). Similar different studies conducted in Ethiopia concluded that women who received advise on when to start ANC visits were more likely to ANC on recommended time than those who did not advise (Damme et al., 2015; Gebremeske, 2014; Gidey et al., 2017). Furthermore, the study's revealed that having decision power to use antenatal care was significantly associated with the time of ANC initiation (Temesgen et al., 2014).

2.3. Survival Models

The source of survival analysis goes back to the time when life tables were specified. Life tables are one of the ancient statistical techniques and are extensively used by medical statisticians and biological by actuaries. Cox, (1972) was concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and the incorporation of regression like influences into life table analysis. Survival analysis is generally defined as a set of approaches for analysing data where the outcome variable is the time until an event occurred. The event can be death, the occurrence of a disease, marriage, divorce or time to data (Despa, 2017). Also the term, failure is used to define the occurrence of the event of interest that may actually be a 'success' such as recovery from therapy (Kleinbaum and Klein, 2005). Multilevel clustered time to data arises when the clustering of data occurs at more than one level. So a flexible survival model with two nested random effects at regions and district level seems appropriate (Sastry, 2017), with an application to the study of child survival in northeast Brazil by nested frailty model for survival data.

2.3.1. Implication of Random Effect of Frailty

The concept of frailty is a statistical modeling concept which accounts for unobserved heterogeneity, association and convenient way to introduce random effects caused by unmeasured covariates (Wienke, 2010) investigated a bivariate frailty model with a cure fraction for modeling familial correlations in diseases. This random effect describes the dependence in the frailty models indicates groups simply a stratified model with some additional structure enacted upon the strata (Lancaster, 1979) suggested this model duration of unemployment. In statistical terms, a frailty model is a random effect model for time to event data, where the random effect has a multiplicative effect on the baseline hazard function (Samia A. and Amani A., 2016) studies on survival analysis using frailty models under the Bayesian mechanism and its further scope.

A frailty model is a multiplicative hazard model consisting of three components: a frailty or unobserved heterogeneity called random effect, a baseline hazard function (parametric or non-parametric) and observed covariates (fixed effects). Vaupel, (1979), introduced with gamma distribution how unobserved heterogeneity of individual frailty has an influence on the dynamic of mortality of population level and explained the frailty models that are used to take the individual hidden differences into survival analysis to account for unobserved heterogeneity or missing covariates in the study population level. Banbeta et al., (2015) discussed the applications of the parametric model to survival data in his study on time to cure severe acute malnutrition. In frailty models, the variability of survival times can be divided into two parts. One part is observed risk factors, known as covariates, and the other part is unobserved risk factors, known as frailty. The univariate frailty model presents the population as a mixture in which baseline hazard is common to all individuals but each individual has his/her own frailty.

The proportional hazards model assumes that conditional on the frailty, the hazard function for an individual at a time greater than zero. In the case of univariate frailty models, independent lifetime is used to describe the influence of unobserved covariates in a proportional hazard model heterogeneity. The variability of survival data is split into a part that depends on risk factor and is therefore theoretically predictable and a part that is initially unpredictable, even when all relevant information is known. A separation of these two sources of variability has the advantage that heterogeneity can explain some unexpected results or give an alternative interpretation of some

results. For example, crossing-over effects or convergence of hazard functions of two different treatment arms (Vaupel et al., 1979) or leveling-off effects that means the decline in the increase of mortality rates, which could result in a hazard function old ages parallel to the x-axis.
There are various frailty models that have been developed in frailty models because of censoring and truncation. The one-parameter gamma distribution is the most widely used frailty distribution proposed by (Clayton, 1978) which showed without the notion of frailty on epidemiological studies of familial tendency in chronic disease incidence. Hougaard, (1984) suggested the gamma, degenerate and inverse Gaussian distribution on positively skewed distribution for the frailty model for heterogeneous populations derived from stable distributions.

Although frailty distribution for the mathematical reasons emphasized that there are no biological reasons for choosing the frailty distribution. The inverse Gaussian or inverse normal distribution is introduced as a frailty distribution alternative to the gamma distribution by (Hougaard, 1984) and (Vaupel, Manton and Stallard, 1979; Klein, 1992) study the impact of heterogeneity in populations derived from stable distribution and individual frailty on the dynamics of mortality uses similar to the gamma frailty model, simple closed-form expressions exist for the unconditional survival and hazard functions, this makes the model attractive.

The particular interest in the multivariate case is the association between related event times. Indeed, different dependence structures result from different frailty distributions. Generally, gamma frailties typically create very strong dependence at late times and the Inverse Gaussian frailties at mid times. The optimal of a family of frailty distributions should hence be accompanied by an assessment of fit. It is natural to consider the mean of the frailty variable conditionally on the observed categorization, which should fluctuate around (Bressen and Torben, 2004). This is due to the simplicity of the derivative the Laplace transform, meaning that traditional maximum likelihood procedures can be used for parameter estimation (Hougaard, 1984; Locatelli et al., 2003). Its flexible shape is another reason given for the selection of the gamma distribution as the frailty distribution (Sastry, 2017).

2.3.2. Common Regression Model in Survival Analysis

In survival analysis, one of the most common assumptions is that the event times are independent of from one observation to another given survival to a specific time and observed covariates. When there are dependencies among observed event times, model based on these assumptions are not appropriate. Dependence also exist when observations are clustered (Lambert et al., 2004), in the multi-center study of the kidney of patients from the same transplant center were associated with the transplants might be carried out by the same surgical group.

In survival analysis, deviations from PH may be explained by unaccounted random heterogeneity or frailty (Weipan, 2001). This work also omitted covariates in survival analysis and shows unbiased or unstable frailty models might behave when asked to account for unobserved heterogeneity in survival analysis with no replications per heterogeneity unit. It would be valuable to upgrade the AFT approach alongside the hazard modeling approach to survival analysis. Lambert et al., (2004) discuss AFT models with shared frailty to determine prognostic factors for the survival time of a kidney graft patient. More recently, AFT models applied to shared gamma frailty model to the analysis of the time time-to-recovery from obstetric fistula patients (Million, 2018). AFT models identify a direct linear relationship between the log of the failure time and covariate, which may be accelerated or decelerated. In this thesis, we will study random effects into the AFT model to allow for correlation and propose an estimation procedure for AFT models with random effects.

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1. Data Source

Data for this analysis were extracted from Ethiopia Demographic and Health Survey (EDHS) 2016. The survey was conducted by Central Statistical Agency (CSA) under the auspices of the Ministry of Health from January 18, 2016, to June 27, 2016, based on a nationally representative sample that provides estimates at the national and regional levels and for urban and rural areas. The primary purpose of the EDHS is to furnish policymakers and planners with detailed information on fertility, sexual activity, family planning, breast feeding practices, nutrition, child hood, maternal mortality, maternal and child health, nutrition and knowledge of HIV/AIDS and other sexually transmitted infections.

3.2. Sample Design

The 2016 EDHS sample was selected using a stratified, two-stage cluster design and enumeration areas (EAs) were the sampling units for the first stage. During the 2007 census each kebele was subdivided into census enumeration areas (EAs), which were convenient for the implementation of the census. The sample included 645 EAs, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling. All women aged 15-49 were eligible for interview. In the interviewed households 16,583 eligible women were identified for individual interview; complete interviews were conducted for 15,683. In all, a total of 7161 women from nine regions and two city administration were included in the study.

3.3. Study population

The study was conducted on pregnant women ages of 15-49 years form all regions and two city administrations in Ethiopia by survey done obtained from EDHS 2016.

3.4. Inclusion and Exclusion criteria

Pregnant women of age 15-49 years and whose their gestational age (duration of pregnancy) was known at first ANC visit was included the study (event). In addition, women who did not accessed ANC throughout pregnancy and the duration of pregnancy were recorded at delivery or termination of pregnancy recorded were also included as censored observation. However, women who had ANC visit but their gestational at ANC visits was unknown (unrecorded) were excluded from analysis. For women who became pregnant for more than one times during the 5 years preceding the survey only the more recent pregnancy were included in the analysis.

3.5. Variables in the study

3.5.1. The response (dependent) variable

The response variable is the time-to-first ANC receipt among pregnant woman in Ethiopia which measured in months. For this study, the survival time was the duration of pregnancy (in months) measured from time of conception to the first ANC visit (event) and others who did not attend ANC throughout of pregnancy period regardless the outcome of pregnancy were considered as (censored).

3.5.2. Predictor (independent) variables

Several predictors were considered in this study to investigate the determinant factors for the time to first ANC visit. Regional state of the women was considered as a clustering effect in all frailty models.

Table3.1. List of predictor variables for the assessment of time to first ANC visit in Ethiopia

Variables	Descriptions	Categories/codes
Region	Region of residence in which the household resides. It is classified as: Tigray, Affar, Amhara, Oromia, Somali, SNNP, Benishangul-Gumuz, Gambela, Harari, Addis Ababa or Dire Dawa.	0=Tigray 1=Affar 2=Amhara 3=Oromia 4=Somalia 5=Benishangul-Gumuz 6=SNNPR 7=Gambela 8=Harari 9=Addis Ababa 10=Dire Dawa
Mother's age	It is the age of the mother at the time of birth coded as: 15-19 ,20-24 ,25-34 , 35-49	0=15-19 1= 20-24 2=25-34 3=35-49
Place of residence	Type of place of residence where the household resides as either urban or rural	0=Urban 1=Rural
Mother's Education	The highest level of education women attained with categories: No education, Primary, Secondary or higher	0=No education 1=Primary 2=Secondary 3=Higher
Wealth index	This is the measure that indicates inequalities in household characteristics, in the use of health and other services. It was categorized as Poor, Middle and Rich	0= Poorest 1=Middle 2=Rich

household head	Sex of person who headed the of house hold categorized as male or female	0=Male 1=Female
Husband Education level	The highest level of education husband attained with categories: No education, Primary, Secondary or higher	0=No education 1=Secondary 2=Primary 3=Higher
Decision maker on respondent's health care	Person in the family who decided on the respondent's health care coded as :Respondent alone , Together husband and respondent), Husband or partner, Other	0= Together (husband and respondent) 1=Respondent alone 2= Husband or partner alone 3=Other
Wanted pregnancy	At the time the respondent became pregnant with the current pregnancy, whether the current pregnancy was wanted then, later or not at all (no more)	0=Then 2=Later 3=No more
Religion	Is religion a respondent which categorised as Muslim, Catholic, Protestant, Orthodox or Others	0=Muslim 1=Catholic 2=Orthodox 3=Protestant 4=Others
perceived problem to get medical care due to distance	Perceived problem to get medical care due to distance from health care which as Not big problem or Big problem	0= Not big problem 1=Big problem
Parity	It the number children ever born including the current pregnancy categorized as: 1, 2 -3 ,4 -5 or >=6	0= 1 1= 2-3 2=4-5 4= >=6

3.6. Survival Models

3.6.1. Non-Parametric Survival Model

Non-parametric survival analyses are more widely used in situations where there is uncertainty about the exact form of distribution or distribution free. In survival analysis, the data are conveniently summarizing through estimates of the survival function and hazard function. 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 the most widely used to estimate the survival and hazard functions (Collet, 1994).

3.6.1.1. Kaplan-Meier estimator

The Kaplan-Meier estimator (Kaplan and Meier, 1958) is also called the Product-Limit estimator. The Kaplan-Meier estimator of the survivorship function (survival probability consider at time$t$). Suppose $t_1, t_2, \dots, t_n$be the survival times of n independent observations and $t_1 \le t_2 \le \cdots t_m, m \le n$ be the m distinct ordered ANC times. The Kaplan-Meier estimator of the survivorship function at time$t$,$\hat{S}(t) = P(T \ge t)$ is defined as:

$$\hat{S}(t) = \prod_{ts<t}\left(\frac{r_j - d_j}{r_j}\right) = \prod_{ts<t}\left(1 - \frac{d_j}{r_j}\right)$$

Where: d_j is the number of women who had ANC visit in the j^{th}interval and $r_j$is the number of women pregnancy at beforej^{th} time.

The cumulative hazard function and variance of the KM estimator can be estimated as:

$$\hat{H}(t) = -\ln(\hat{S}_{(t)})\, V\big(\hat{S}_{(t)}\big) = \big(\hat{S}_{(t)}\big)^2 \sum_{j:y_{j\le t}} \frac{d_j}{R\big(y_{(j)}\big)[R\big(y_{(j)}\big) - d_j]}$$

The variance of the product-limit estimator is estimated by Greenwood's formula (Greenwood and Yule, 1920) and given by;

$$Var(\hat{S}_{(t)})^2 = [\hat{S}_{(t)}]^2 \sum_{j:t_{j\le t}} \frac{d_j}{r_j(r_j - d_j)}; j = 1,2, \dots, r$$

Since, the distribution of survival time tends to be positively skewed, the median is preferred for a summary measure. Median survival time is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value $t_{(50)}$which is such that $\hat{S}\big(t_{(50)}\big) = 0.5$. The estimated median survival time, is defined to be the smallest observed survival time for which the value of the estimated survival function is less than 0.5.

$$\hat{t}(50) = \min\{t_{(i)}/\hat{S}\big(t_{(j)}\big) < 0.5\}$$

Where t_i is the observed survival time for the i^{th}individual,$i = 1,2, \dots, n$ and t_i is the j^{th} ordered ANC visit time,$j = 1,2, \dots, n$,

$$\hat{t}_p = \min\big\{t_i/\hat{S}(t_i)\big) < 1 - \frac{p}{100}\}$$

A confidence interval for the percentiles can be obtained delta method, the variance of the estimator of the p^{th}percentile is

$$V\left(\hat{S}(t_p)\right) = \left(\frac{d\hat{S}(t_{(p)})}{dt_{(p)}}\right)^2 V(t_p) = \left(f(t_p)\right)^2 V(t_p)$$

The standard error of $\hat{t}$(p) is therefore given by

$$SE(\hat{t}_p) = \frac{1}{f(\hat{t}_p)} SE(\hat{S}(\hat{t}_p))$$

The standard error of $\hat{S}(\hat{t}_p)$ can be obtained by using Greenwoods formula

$$\hat{f}(\hat{t}_p) = \frac{\hat{S}(\hat{u}_p) - \hat{S}(\hat{l}_p)}{\hat{l}_p - \hat{u}_p}$$

$$\hat{u}_p = \max\{t_j / \hat{S}(t_j) \geq 1 - \frac{p}{100} + \varepsilon\}$$

$$\hat{l}_p = \min\{t_j / \hat{S}(t_j) \leq 1 - \frac{p}{100} - \varepsilon$$

Median survival time, $\hat{u}_{50}$is the largest observed survival time from the K-M curve for which$\hat{S}_t \geq 0.55$, and $\hat{l}_{50}$ is the smallest observed survival time from the K-M for which$\hat{S}(t_p) \leq 0.45$. The 95% confidence interval for the p^{th}percentile $\hat{t}_p$ has limits of $\hat{t}_p \pm 1.96SE(\hat{t}_p)$

3.6.2. Non parametric Comparison of Survival Functions

3.6.2.1. Non-Parametric Survival Analysis

Between the various non-parametric tests one can find in the statistical literature, is the Mantel - Haenzel test, currently called the "log-rank" is the one commonly used non-parametric tests for comparison of two or more survival distributions. The log rank test statistic for comparing two groups is given by (Cox, 1984)

$$Q = \frac{[\sum_{i=1}^{m} w_i \, (d_{1i} - \hat{e}_{1i}]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \sim X_{k-1}^2$$

Where $\hat{e} = \frac{n_{1i} d_i}{n_i}$ and $\hat{v} = \frac{n_{1i} n_{oi} d_i (n_{1i} - d_i)}{n_i^2 (n_i - 1)}$, n_{oi} and n_{1i} are the number at risk at observed survival t_i in group 0 and 1. n_i is the total number of individuals or risk before t_i, d_{1i} number of observed event in group 1 and d_i is the total number of event at $t_i$and $w_i$weight.

3.6.3. Cox PH Regression Model

Cox, (1972), introduced Cox proportional hazards model is known as semi-parametric model which used to quantify the effect of one or more explanatory variables on failure time which describes the relationship between the event incidences. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time.

$h_i(t/x) = h_o(t)\exp(xi^T\beta)$, $h_o(t)$ is the baseline hazard function; xi is a vector of covariate and β is a vector of parameters for fixed effects.

The corresponding survival function for Cox-PH model is given by:

$S(t,x) = [S_o(t)]^{\exp\{\sum_{i=1}^{p}\beta_i x_i\}}$, $S_o(t)$is the baseline survival function

In this model, no distributional assumption is made for the survival time; the only assumption is that the hazards ratio does not change over time. Even though the baseline hazard is not specified and depends on time not covariate. We can still get a good estimate for regression coefficientsβ, hazard ratio and adjusted hazard curves.

The hazard ratio of two individuals with different covariates x and x^* is given by:

$$\widehat{HR} = \frac{h_o(t)\exp(\widehat{\beta'}_x)}{h_o(t)\exp(\hat{\beta}'x^*)} = \exp\{\sum \widehat{\beta'}\,(x - x^*)\}$$

This hazard ratio is time-independent, which is why this is called the proportional hazards model.

3.6.3.1. Checking for Proportional Hazard Assumption

Let $\left(\widehat{H}_0(t)\right)$is the cumulative baseline hazard function, and K is the number of disjoint categories. To check the proportionality assumption we could plot $\ln(\hat{H}_{10}(t))\ldots\ln\hat{H}_{k0}(t))$ versus t. If the assumption holds, then these should be approximately parallel and the constant vertical separation between $ln(\hat{H}\,g_o(t))$ and $ln(\hat{H}\,bo(t))$ should give a crude estimate of the factor needed to obtain $\widehat{H}_{bo}(t)$ from$\widehat{H}_{go}(t)$. An alternative approach is to plot $ln\,(Hg0\,(t)) - ln\,(H10\,(t))$ versus t for g = 2… k. If the proportional hazards model holds, each curve should be roughly constant (Klein, J. 1992).

3.6.4. Accelerated Failure Time Model

AFT is an alternative to the PH model for the analysis of survival time data. Under AFT models we measured the direct effect of the explanatory variables on the survival time instead of hazard. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. The model works to measure the effect of covariate to "accelerate" or to "decelerate" survival time. When PH assumptions are not satisfied, the parametric AFT model can be used instead of Cox model.

The AFT model states survival function of an individual with covariate x at time t is the same as the survival function of an individual with a baseline survival function at a time ($t * \exp(\alpha' x)$), where $\alpha' x = (\alpha_1, \alpha_2, \cdots, \alpha_p)$ is a vector of regression coefficients. In other words, the accelerated failure-time model is defined by the relationship (Klein & Moeschberger, 2003).

$S(t/x) = S_o\{t \,/\, \eta(x)\}$ for all x , where $\eta(x) = \exp(\alpha' x)$

Under the AFT model, the covariate effects are assumed to be constant and multiplicative on the time scale that is the covariate impacts on survival by a constant acceleration factor. According to the relationship of survival function and hazard function, the hazard function for an individual with covariate $x_1, x_2, \cdots x_p$ is given by:

$$h(t/x) = h_o(t/\eta(x))(1/\eta(x))$$

Here by we can consider on a log-scale of the AFT model with respect to time is given analogous to the classical linear regression approach. In this approach, the natural logarithm of the survival time $y = \log(T)$ is will be modelled.

When we denote by S_o the survival function when $X = 0$ then we find that

$$\begin{aligned} P(T > t/x) &= P(y > \log(t/x) \\ &= P\{\mu + \sigma\varepsilon > \log t - \alpha' x/x\} \\ &= P\{\exp(\mu + \sigma\varepsilon) > t^{*}\exp(-\alpha' x)/x\} \\ &= S_0\{t^{*}\exp(-\alpha' x)\} \end{aligned}$$

The effect of the covariates on the survival function is the time scale is changed by a factor $\exp(-\alpha' x)$, and we call this an acceleration factor.

We note that, $\exp(-\alpha' x) > 1,\ when\ \alpha < 1$ →the survival process accelerates

$\exp(-\alpha' x) < 1,\ when\ \alpha > 1$ → the survival process decelerates.

The survival function of T_i can be expressed by ε_i (Klein & Moeschberger, 2003). For each distribution ofε_i, there is a corresponding distribution for$T$. The members of the AFT model class include the Weibull AFT model, log logistic AFT model and log-normal AFT model.

$$S_i(t) = S_{\varepsilon i}\left(\frac{\log t - (\mu + \alpha' x}{\sigma}\right)$$

The effect size for the AFT model is the time ratio. The time ratio comparing two levels of covariate$xi = 1\ vs\ xi = 0$, after controlling all the other covariates is$\exp(\alpha_i)$, which is interpreted as the estimated ratio of the expected survival times for two groups. A time ratio above 1 for the covariate implies that this covariate prolongs the time to event, while a time ratio below 1 indicates that an earlier event is more likely. Therefore, the AFT models can be interpreted in terms of the speed of progression of a disease. The effect of the covariates in an accelerated failure time model is to change the scale, and not the location of a baseline distribution of survival times.

3.6.4.1 Estimation of AFT model

AFT models are fitted using the maximum likelihood method. The likelihood of the n observed survival times, $t_1, t_2, \cdots, t_n$, is given by:

$$L(\alpha, \mu, \sigma) = \prod_{i=1}^{n} \{f(t_i)\}^{\delta i} \{S_i(t_i)\}^{1-\delta i}$$

Where $f(t_i)$ and $S_i(t_i)$ are the density and survival functions for the i^{th} individual at t_i and δ_i is the event indicator for the i^{th} observation. The logarithm of the above equation yields;

$$\log L(\alpha, \mu, \sigma) = \sum_{j=1}^{n} \{-\delta_i \log(\sigma t_i + \delta_i \log f_i\,(x_i) + (1-\delta_i) \log S_i(w_i))\}$$

Where, $W_j = \left\{\log t_i - \frac{\mu + \alpha_1 x_1 + \cdots + \alpha_{pi} x_{pi}}{\sigma}\right\}$, Z=$\{z_{ji}\}$ is vector of covariates for the j^{th} subject. The maximum likelihood parameters estimates are found by using Newton-Raphson procedure.

3.6.4.2. Weibull AFT model

Weibull distribution including exponential distribution as a special case can be parameterized as AFT model and they have only family of distribution to have this property. The results of fitting a weibull model can therefore be interpreted in either framework (Klein & Moeschberger, 2003). Then the weibull distribution is very flexible model for time-to-event data. It has a hazard rate which is monotone increasing, decreasing, or constant. The survival and hazard function of weibull model with scale parameter and shape parameter is given by: If

$$f(t, \mu, \alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left(\frac{-t}{\mu}\right)^{\alpha}, where\ \mu > 0\ and\ \alpha > 0$$

$$S_{\varepsilon i}(t) = exp\left(-exp\left(\frac{\log t - (\mu + \tau\alpha' x))}{\sigma}\right)\right) = exp\left(-exp\left(\frac{-(\mu + \alpha' x)}{\sigma} t^{\frac{1}{\sigma}}\right)\right)$$

From the above equation, the PH representation of the survival function of the Weibull model is given by:

$$S_i(t) = exp\{-exp(\beta_1 x_{1i} + \cdots + \beta_p x_{pi})\lambda t^{\gamma}\}$$

Comparing the above two formulas the parameter λ, γ, β_j in the PH model can be expressed by the parameters μ, σ, α_jin the AFT model:$\lambda = \exp(-\mu/\sigma)\,, \gamma = 1/\sigma, \beta_j = \alpha_j/\sigma$

The AFT representation of hazard function of the Weibull model is given by:

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} exp\left(\frac{-\mu - \alpha' x1i}{\sigma}\right)$$

3.6.4.3. Log-logistic AFT model

The log-logistic distribution has a fairly flexible functional form; it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump shaped that is it initially increases and then decreases. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions. The log-logistic model has two parameters λ and ρ, where, $\lambda$is the scale parameter and ρ is the shape parameter. Its pdf is given by (Bennett, 1983) and (Cox, 1972).

The cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring (Bennett, 1983). The log-logistic distribution is very similar in shape to the log-normal distribution, but it more suitable for use in the analysis of survival data. The log-logistic model has two parameters λ andρ, where λ the scale parameter is and ρ is the shape parameter. Its pdf is

$$f(t) = \frac{\lambda \rho t^{\rho-1}}{(1+\lambda t^{\rho})^2}$$

The corresponding survival and hazard functions are given by;

$$S(t) = \frac{1}{1+\lambda t^{\rho}}$$

$$h(t) = \frac{\lambda \rho t^{\rho-1}}{1+\lambda t^{\rho}}$$

Where; $\lambda \epsilon R, \rho > 0$.When $K \leq 1$,the hazard rate decreases monotonically and when $k > 1$, it increases from zero to a maximum and then decreases to zero. Under the AFT model the hazard function for the i^{th} individual is

$$h_{i(t/x)} = h_0\big(texp(-\alpha' x_i)\big) \exp(-\alpha' x_i) = \frac{\rho \exp((\lambda)texp(-\alpha' x_i)}{1+\exp(\lambda)\{texp(-\alpha' x_i\}^{\rho}}$$

The log-logistic AFT model with a covariate x may be written as;

$Y = logT = \mu + \alpha' x_i + \delta\varepsilon,$ where; $\alpha' = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_p)$; ε has standard logistic distribution.

The survival with covariate x is given as follows:

$$S_T(t/x) = \frac{1}{1+\lambda \exp(\alpha' x) t^{\rho}} = \frac{1}{1+\exp(log\lambda + \beta' x)}$$

$$h_T(t/x) = \frac{\rho t^{\rho-1}\lambda \exp(\alpha' x)}{1+\lambda \exp(\alpha' x) t^{\rho}} = \frac{\rho t^{\rho-1}\lambda \exp(\alpha' x)}{1+\exp(log\lambda+\alpha' x)}$$

To interpret the factor $\exp(\beta' x)$ for log-logistic model, one can notice that the odds of survival beyond time t for log-logistic model is given by $\frac{S_T(t)}{1-S_T(t)}$.

We can see that the log-logistic distribution has the proportional odds (PO) property. So this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as;

$$\frac{S_T(t)}{1-S_T(t)} = \exp(-\alpha' x)\frac{S_0(t)}{1-S_0(t)}$$

The factor $\exp(-\alpha' x)$ is an estimate of how much the baseline odds of survival at any time changes when individual has covariate x. And $\exp(-\alpha' x)$ is the relative odds of experiencing the event for an individual with covariate x relative to an individual with the baseline characteristics. As this representation of log-logistic regression is as accelerated failure time model with a log logistic baseline survival function, then the log logistic model is the only parametric model with both a proportional odds and an accelerated failure-time representation. If $T_i$has a log-logistic distribution, then ε_i has a logistic distribution. The survival function of logistic distribution is given by (Collett, 2003)

$$S_{\varepsilon i}(\varepsilon) = \frac{1}{1+\exp(\varepsilon)}$$

Then, the AFT representation of log-logistic survival function is given by

$$S_t(t) = \left[1 + t^{\frac{-1}{\sigma}} exp\left(\frac{-\mu - \alpha' x}{\sigma}\right)\right]^{-1}$$

And the associated hazard function for the individual is given by

$$h_t(t) = \frac{1}{\sigma t}\left[1 + t^{\frac{-1}{\sigma}} exp\left(\frac{-\mu - \alpha' x}{\sigma}\right)\right]^{-1}$$

If the plot of $log\left[\frac{S_T(t)}{1-S_T(t)}\right]$ against $log(t)$ is linear, the log-logistic distribution is appropriate for the given data set.

3.6.4.4. Log-normal AFT model

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by (Collett, 2003): Simply assumes that $\epsilon \sim N(0,1)$.

$$S_o(t) = 1 - \phi\left(\frac{\log t - \mu}{\sigma}\right)$$

$$h_o(t) = \frac{\phi\left(\frac{\log t}{\sigma}\right)}{[1-\phi\left(\frac{\log t}{\sigma}\right)]\sigma t\prime}$$

Where μ and σ are parameters, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ is the probability density function and $\phi(x) = \int_{-\infty}^{\chi} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{u}} du$ is the cumulative distribution function of the standard normal distribution. The survival function for the i^{th}individual is

$$S_i(t) = S_o(t/\eta i) = S_o(t * expexp(\mu + \alpha' x_i)$$
$$= 1 - \phi\left(\frac{\log t - \alpha'^{x_i} - \mu}{\sigma}\right)$$

Where $\eta i = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p)$,the log survival time for the i^{th}individual has normal $\mu + \alpha' x_i, \sigma$.then the log normal distribution of AFT property to be

$\phi^{-1}[1 - S_{(t)}] = \frac{1}{\phi}(\log t - \alpha'^{x_i} - \mu)$, where x_i is the value of a categorical variable which takes the value 0 in one group and 1 in the other group. This implies that the plot $\phi^{-1}[1 - S_{(t)}]$ against $\log(t)$ will be linear if the log-normal is appropriate for the given data set.

3.6.5. Modeling Frailty

Frailty models have been used frequently for modeling dependence in multivariate time-to-event data. (Wienke et al., 2003) suggested the variability of lifetimes is formulated as arising from two different sources. The first one is natural variability, which is included in the baseline hazard function, while the second one is explained by the frailty. Lifetimes are conditionally independent given the frailty (as individual random effect), and the frailty term represents unobserved covariates. It is assumed that, given the unobserved frailty, the hazard for each survival time follows a proportional hazards model with the frailty variable and the covariate effect acting multiplicatively on the baseline hazard.

3.6.5.1. Multivariable frailty models

The shared frailty model is used with multivariable survival data where the unobserved frailty is shared among groups of individuals, and thus a shared frailty model may be thought of as a random effects model for survival data. Frailty models for univariate data have long been used to account for heterogeneous times-to-failure. The term frailty was first suggested by (Vaupel et al., 1979) in the context of mortality studies and (Lancaster, 1979) incorporated the frailty concept into a study of duration of unemployment.

For a random time-to-data, T, we define the probability density function of T as $f(t)$ and the cumulative distribution function as $F(t) = P(T \leq t)$.

Two other functions that prove useful in this context are the survivor function $S(t) = P(T > t) = 1 - F(t)$, and the hazard function$h(t) = f(t)/S(t)$, which can be interpreted as the instantaneous rate of data given survival up until time$t$. Consider a parametric survival model characterized by its hazard function, $h(t)$.

Implicit in the definitions of all these functions are the effects of any covariates, whether we parameterize the model as having proportional hazards (PH) with respect to changes in covariate values, or accelerated failure time (AFT) due to the covariates. Suppose we have k observations and i subgroups. Each subgroup consists of $n_i$observation and$\sum_{i=1}^{G} ni = n$, where n is the total sample size. The hazard rate for thej^{th}individual in the i^{th} subgroup is given by:

$$h_{ij}(t) = h_o(t)u^i \exp({z_{ij}}^t \beta), i = 1, \cdots, ni$$

Where u_i frailty terms for subgroups and their distribution are is again assumed to be independent with a mean of 0 and a variance of 1. If the number of subjects n_i is 1 for all groups, the univariate frailty model is obtained (Wienke, 2017); otherwise the model is called the shared frailty model (Duchateau et al., 2002) and (Klein, 1992) because all subjects in the same cluster share the same frailty value. In general, we use the notation$h_i(t) = h(t/x_i)$. The shape parameter p and regression coefficients β are estimated from the data. A frailty model in the univariate case introduces an unobservable multiplicative effect α on the hazard, so that conditional on the frailty,$h(t/\alpha) = \alpha h(t)$, α is some positive quantity assumed mean one and variance θ. Where Klein, (1992) discusses the ramifications of the assumed distribution of the frailty, whether gamma or inverse Gaussian.

3.6.5.2. Frailty Distribution

3.6.5.3. Gamma Frailty Distribution

The gamma distribution is very-well known and has simple densities. Even though gamma models have closed form expressions for survival and hazard functions, from a computational view, it fits well to frailty data and it is easy to derive the closed form expressions for unconditional survival and hazard functions.

Gamma frailty model belongs to the power variance function (Hougaard, 1986)and can be expressed in terms of its Laplace transform from which properties such as mean and variance are easily derived (Duchateau et al., 2002). It is widely used due to mathematical tractability

(Wienke, 2010). Assuming a two-parameter gamma density with $\delta > 0$ and $\gamma > 0$ as shape and scale parameters respectively, the density function is given by: -

$$f_z(z) = \frac{\gamma^{\delta} zi^{\delta-1} \exp(-\gamma zi)}{\Gamma(\delta)}$$

With $\delta > 0$ and $\gamma > 0$ where $\Gamma(.)$ is the gamma function, it corresponds to a gamma distribution $Gam(\mu, \theta)$ with μ fixed to 1 for identify ability and its variance is θ. The corresponding Laplace transformation is: -

$L(s) = \gamma^{\delta}(s + \gamma^{-\delta})$

In gamma frailty models, restriction $\delta = \gamma$ is used, which results in expectation of 1.

The variance of the frailty variable is then 1. Assuming that the frailty term z_i is a gamma with $E(z) = 1$ and$Var(z) = \theta,$ then $\delta = \gamma = \frac{1}{\Theta}$. The distribution function of the frailty term z_i is therefore a one-parameter gamma distribution given by:-

$$f_z(z_i) = \frac{zi^{\frac{1}{\Theta}-1} exp^{\frac{-zi}{\Theta}}}{\Gamma(\frac{1}{\Theta})\Theta^{\frac{1}{\Theta}}}$$

Where $\theta > 0$ and $z_i > 0$ indicates that individuals in group i are frail, wherea $z_i < 0$ indicates that individuals are strong and have lower risk.

The corresponding Laplace transform is given by;

$$L_{(s)} = (1 + \frac{1}{\theta})^{-\theta}$$

Note that if $\theta > 0$, there is heterogeneity. So the large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups.

The conditional survival function of the gamma frailty distribution is given by:(Gutierrez, 2002).

$$S_{\theta}(t) = [1 - \Theta \ln\{S(t)\}]^{\frac{-1}{\Theta}} \; , \theta > 0$$

The conditional hazard function of the gamma frailty distribution is given by:(Gutierrez, 2002).

$$h_{\theta}(t) = h(t)[1 - \Theta \ln\{S(t)\}]^{-1}$$

Where $S(t)$ and $h(t)$ are the survival and the hazard functions of the baseline distributions.

The variance θ of the frailty term represents the heterogeneity among clusters while the mean is constrained to 1 in order to make the average hazard identifiable (Duchateau et al., 2002); larger variance indicates a stronger association within groups.

For the Gamma distribution, the Kendall's Tau Klein, (1992), measures the association between any two event times from the same cluster in the multivariate case and given by:

$$\tau = \frac{\Theta}{\Theta + 2}, where\ \tau \epsilon (0,1)$$

3.6.6.2. Parameter Estimation

For right-censored clustered survival data, the observation for subject $j\epsilon j_i = \{1, \cdots, n\}$ from cluster $i\epsilon I = \{1, \cdots, s\}$ is the couple(y_{ij}, δ_{ij}), where $y_{ij} = \min(t_{ij}, c_{ij})$ is the minimum between survival time t_{ij} and the censoring time c_{ij} and where $\delta_{ij} = I(t_{ij} \leq c_{ij})$is the event indicator. When covariate information are been collected the observation will be$(y_{ij}, \delta_{ij}, x_{ij})$, where x_{ij} denote the vector of covariates for the ij^{th}observation. In the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution.

Under assumptions of non-informative right-censoring and independence between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data can be written as.

$l_{marg}(\psi, \beta, \theta, z, x)$

$$= \prod_{i=1}^{s} [\left(\prod_{j=1}^{ni} (h_o(y_{ij}) \exp\left(X^T{}_{ij} X_{ij}{}^T \beta\right)^{\delta ij}\right) \int_0^\infty z_i{}^{di} \exp(-z_i \sum_{j=1}^{ni} h_o(y_{ij} \exp(X^T{}_{ij}\beta))) f(z_i) dz_i]$$

$$= \prod_{i=1}^{s} [(\prod_{j=1}^{ni} \left(h_o(y_{ij}) \exp(X_{ij}{}^T \beta)\right)^{\delta ij}))(-1)^{di} L^{di}(\sum_{j=1}^{ni} H_o\,(y_{ij}) exp(X_{ij}{}^T)]$$

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

$$L(S) = E(\exp(-Zs) \int_0^\infty \exp(-Z\,is) f(Zi) dzi, s \geq 0,$$

Where ψ represents a vector of parameters of the baseline hazard function, β the vector of regression coefficients and θ *the* variance of the random effect. Estimates of ψ, β, θ are obtained by maximizing the marginal log-likelihood above. This can be done if one is able to compute higher order derivatives $L^q(.)$ of the Laplace transform up to $q = \max\{d_1, \cdots, d_s\}$.

3.6.7. Prediction of Frailties

Besides parameter estimates, prediction of frailties is sometimes desirable. The frailty term z_i can be predicted by $\hat{z}_i = E(Z/zi; \hat{\psi}, \hat{\beta}, \theta)$with zi the data of ith cluster. This conditional expectation can be achieved as: -

$$\hat{Z}_i = E(Z/zi, \theta) = -\frac{L^{(di+1)}(\sum_{j=1}^{ni} H_o(y_{ij})\exp(x^T{}_{ij}\beta)}{L^{(di)}(\sum_{j=1}^{ni} H_o(y_{ij})\exp(x^T{}_{ij}\beta))}$$

3.6.8. Comparison of Models

In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Munda, 2012). This thesis used the AIC and BIC criteria to compare various candidates of parametric frailty models. In addition to these criteria, we used likelihood ratio test in order to compare models that are nested, particularly the effect of the random effect.

Manipulation of the comparison will be use R and Stata software.

3.6.8.1. Model Diagnosis

3.6.8.1.1. Evaluation of the Parametric Baselines

The graphical methods can be used to check if a parametric distribution fits the observed data. Model with the Weibull baseline has a property that the $\log(-\log(S(t)))$ is linear with the log of time, where$S(t) = \exp(-\lambda t^p)$. Hence, $\log(-\log(S(t))) = \log(\lambda) + \rho \log(t)$. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting $\log(-\log(\hat{S}(t)))$ versus $\log(t)$ where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler and Stucki, 2011). The log-failure odd versus log time of the log-logistic model is linear. Where the failure odds of log-logistic survival model can be computed as:

$$\frac{1 - s(t)}{s(t)} = \frac{\frac{\lambda t^p}{1+\lambda t^p}}{\frac{1}{1+\lambda t^p}} = \lambda t^\rho$$

Therefore, the log-failure odds can be written as:

$$\log(1 - s(t)/s(t)) = \log(\lambda t^\rho) = \log(\lambda) + \rho \log(t)$$

Therefore, the appropriateness of model with the log logistic baseline can graphically be evaluated by plotting $\log((1 - \hat{S}(t)/\hat{S}(t))$ versus log time where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler and Stucki, 2011). If the plot is straight line, log-logistic distribution fitted the given dataset well. If the plot $\phi^{-1}[1 - s(t)]$ against log(t) is linear, the log-normal distribution is appropriate for the given data set.

3.6.8.1.2. The Cox Snell Residuals

For the parametric regression problem, analogy of the semi parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Kleinbaum and Klein, 2005). The first such residual is the Cox–Snell residual that provides a check of the overall fit of the model. The Cox–Snell residual for the ith individual with observed ti, r_{ci}, is defined as:

$$r_{ci} = \hat{H}(ti/xi) = -\log[\hat{S}(ti/xi)]$$

Where ti is the observed survival time for individual i, $\hat{H}$ is the cumulative hazard function of the fitted model, xi is the covariate values for individual i, and $\hat{S}(ti)$ is the estimated survival function on the fitted model. The estimated survival function of the ith individual is given by

$$\widehat{S_\iota}(t) = s_{\varepsilon i}\left(\frac{logt - \hat{\mu} - \hat{\alpha}xi}{\hat{\sigma}}\right)$$

Where μ, $\hat{\alpha}$ and $\hat{\sigma}$ are the maximum likelihood estimator of μ, α and σ respectively, $S_{\varepsilon i}(\varepsilon)$ is the survival function of ε_i in the AFT model, and $r_{si} = \frac{logt - \hat{\mu} - \hat{\mu}xi}{\hat{\sigma}}$ is known as the standard residual. The Cox-Snell residual can be applied to any parametric model. Under the Weibull AFT model since $S_{\varepsilon i}(\varepsilon)$ = exp($-e^{\varepsilon}$), the Cox-Snell residual for Weibull is then given as

$$r_{ci} = -log\{\hat{S}(t)\} = -logS_{\varepsilon i}(r_{si}) = \exp(r_{si})$$

Similarly, with the log-logistic AFT model, since $S_{\varepsilon i}(\varepsilon) = (1 - e^{\varepsilon})^{-1}$, the Cox-Snell residual for the for the log-logistic is then given as $r_{ci} = \log[1 + \exp(r_{si})]$

Also under the lognormal AFT model, $S_{\varepsilon i}(\varepsilon) = 1 - \phi(\varepsilon)$ hence the Cox-Snell residual for the lognormal becomes

$$r_{ci} = \log[1 - \phi(r_{si})]$$

If the fitted model is appropriate, the plot of $\log(-logS(r_{ci}))$ versus $logr_{ci}$ is a straight line with the unit slope through the origin.

For the three baseline hazard functions considered in this thesis, the Cox–Snell residuals are:

For Log-normal,$r_j = \log[1 - \phi(rsi)]$

Weibull, $r_j = \hat{\lambda}exp(\widehat{\beta}'_{xj})t_j{}^p$

Log-logistic, $r_j = \ln(\frac{\hat{\iota}}{1+\hat{\lambda}exp(\widehat{\beta'}_{xj})t_j{}^p})$

3.6.8.1.3. The Quantile - Quantile plot

A quantile-quantile or q-q plot is made to check if the accelerated failure time model provides an adequate fit to the data. The plot is based on the fact that, for the accelerated failure-time model, $S_1(t) = S_0(\phi y)$, Where S_0 and S_1 are the survival functions in the two groups and ϕ is the acceleration factor. Let $t_o p$ and $t_1 p$ be the p^{th} percentiles of groups 0 and 1, respectively, that is

$$t^{kp} = Sk^{-1}(1 - p), k = 0,1$$

Using the relation $S_1(t) = S_o(\phi t)$ we must have $S_o(t_0 p) = 1 - p = S_o(\phi t 1_p)$ for all t. If the accelerated failure time model holds, $t_o p = \phi t_1 p$. To check this assumption we compute the Kaplan–Meier estimators of the two groups and estimate the percentiles $t_1 p, t_o p$ for various values of$p$. If we plot the estimated percentile in group 0 versus the estimated percentile in group 1 (i.e., plot the points $t_1 p, t_o p$ for various values of p), the graph should be a straight line through the origin, if the accelerated failure time model holds. If the curve is linear, a crude estimate of the acceleration factor ϕ is given by the slope of the line (Klein, 1992).

CHAPTER FOUR

4. RESULT AND DISCUSION

4.1. Descriptive Statistics

The descriptive summary of covariates is given in Table 4.1. A total of 7161 of women who got pregnancy during five years' survey were included in this study from nine regions and two administrative cities of whom, 4680 (65.4%) received first ANC visit (events) and 2481(34.6%) did not receive first ANC visit (right censored). Among pregnant women included in study the highest number were from Oromia 1028 (14.4%) while the lowest numbers were from Addis Ababa, 371(5.2%) followed by Dire Dawa 384(5.4%) and Harari regional state 410 (5.7%). In contrast majority of women from Addis Ababa, 359 (96.8%) had first ANC visit and majority, 455 (56.6%) of pregnant mothers from Somali regional state had no ANC visit. On the other hand, 4339 (60.6%) women had no education and among these 2019 (46.5%) were censored (had no ANC visit). Regarding educational status of husband 3121(47.1%) had no education and only 567 (8.6%) had attended higher education. Of 5656 (79%) pregnant women from rural setting, 3292(58.2%) had ANC visit. Married women accounted for 91.4% of total pregnant women included in this study and 2281(34.8%) of them were censored. Majority were Muslim, 1335(46.3%) followed by orthodox Christian, 2349 (32.8%) and of these 41.9% and 21.8% had no ANC visit (censored) respectively.

Furthermore, among pregnant mothers included in the analysis, 3780 (52.8%) perceived distance as being a major problem to seek medical care; while the remaining 3381(47.2%) reported distance was not a major problem to seek medical care. But 45.2% of those perceived distance was a big problem to seek medical care had no ANC visit, whereas only 22.8% of those who perceived distance was not a major problem had no ANC visit. About 3810(53.2%) pregnant mothers were from poorest wealth index households. Of these mothers, 45.7% did not have ANC visit records. From 2132 women in the richest wealth index, only 25.5% did not visit facilities for ANC. Regarding decisions on respondent's health care, it was found that 1176 (17.7%) decided by themselves, 4157(62.7%) decided together with their husband/partner while 1282(19.3%) decided by husband/Partner alone and out of these, 16.7%, 32.3% and 32.3 had no ANC visit respectively.

Table 4.1. Descriptive statistics of time to first ANC visit among pregnant women Ethiopia (EDHS2016).

Variables	***Categories***	***Frequency***	***Status variable***	
			Event N°- (%)	***Censored N°- (%)***
Region	Tigray	761(10.6%)	682(89.6%)	79(10.4%)
	Afar	646(9%)	286(44.3%)	360(55.7%)
	Amhara	758(10.6%)	500(66%)	258(34%)
	Oromia	1028(14.4%)	525(51.1%)	503(48.9%)
	Somalia	804(11.2%)	349(43.4%)	455(56.6%)
	Benishangul	575(8%)	389(67.7%)	186(32.3%)
	SNNPR	890(12.4%)	623(70%)	267(30%)
	Gambela	534(7.5%)	319(59.7%)	215(40.3%)
	Harari	410(5.7%)	317(77.3%)	93(22.7%)
	Addis Ababa	371(5.2%)	359(96.8%)	12(3.2%)
	Dire Dawa	384(5.4%)	331(86.2%)	53(13.8%)
Education of mother	No education	4339(60.6%)	2320(53.5%)	2019(46.5%)
	Primary	1933(27%)	1534(79.4%)	399(20.6%)
	Secondary	574(8%)	518(90.2%)	59(9.8%)
	Higher	315(4.4%)	308(97.8%)	7(2.2%)
Type of place of residence	Urban	1505(21%)	1388(92.2%)	117(7.8%)
	Rural	5656(79%)	3292(58.2%)	2364(41.8%)
Age group of respondent	15-19	355(5%)	247(69.6%)	108(30.4%)
	20-24	1487(20.8%)	1022(68.7%)	465(31.3%)
	25-34	3535(49.4%)	2391(67.6%)	1144(32.4%)
	35-49	1784(24.9%)	1020 (57.2%)	764(42.8%)
Husband/Partner's education level	No education	3121(47.1%)	1599(51.2%)	1522(48.8%)
	Primary	2150(32.4%)	1560(72.6%)	590(27.4%)
	Secondary	741(11.2%)	632(85.3%)	109(14.7%)
	Higher	567(8.6%)	496(87.5%)	71(12.5%)
	Don't know	52(0.8%)	37(71.2%)	15(28.8%)
Wanted pregnancy when became pregnant	Then	5721(79.9%)	3745(65.5%)	1976(34.5%)
	Later	983(13.7%)	687(69.9%)	296(30.1%)
	No more	457(6.4%)	248(54.3%)	209(45.7%)
Religion	Orthodox	2349(32.8%)	1836(78.2%)	513(21.8%)
	Catholic	49(0.7%)	28(57.1%)	21(42.9%)
	Protestant	1335(18.6%)	852(63.8%)	483(36.2%)
	Muslim	3315(46.3%)	1927(58.1%)	1388(41.9%)
	Traditional/Other	113(1.6%)	37(32.7)	76(67.3%)
household head	Male	5576(77.9%)	3670(65.8%)	1906(34.2%)
	Female	1585(22.1%)	1010(63.7%)	575(36.3%)
perceived problem to get medical care due to distance	Big problem	3780(52.8%)	2070(54.8%)	1710(45.2%)
	Not a big problem	3381(47.2%)	2610(77.2%)	771(22.8%)

Wealth index	Poorest	3810(53.2%)	2068(54.3%)	1742(45.7%)
	Middle	1219 (17%)	908(74.5%)	311(25.5%)
	Richest	2132(29.8%)	1704(83.3%)	428(16.7%)
Person who usually decides on respondent's health care	Respondent alone	1176(17.7%)	796(67.7%)	380(32.3%)
	Respondent and Husband/partner	4157(62.7%)	2815(67.7%)	1342(32.3%)
	Husband/Partner alone	1282(19.3%)	705(55%)	577(45%)
	Other	16(0.2%)	8(50%)	8(50%)
Parity	1	1459(20.4%)	1140(78.1%)	319(21.9%)
	2-3	2210(30.9%)	1578(71.4%)	632(28.6%)
	4-5	1634(22.8%)	999(61.1%)	635(38.9%)
	>=6	1858(25.9%)	963(51.8%)	895(48.2%)

4.2. Non-parametric Survival Analysis

4.2.1. The Kaplan-Meier Estimate of Time-to-First ANC visit

Non-parametric survival analysis (K-M) is used to visualize the survival time-to-first ANC visit of pregnant women in Ethiopia under different covariates. It also provides information on the shape of the survival and hazard function of ANC data set. The survival plot in Figure 4.1 sharply decreased first and slowly decline at later times. This implies probabilities of not starting ANC visit is higher at early gestational age and tends to sharply decrease later as gestational age increased. Furthermore, the median time of first ANC visit for pregnant women in Ethiopia was at the 5th month of gestation with 95% CI: 4.88th 5.12th month.

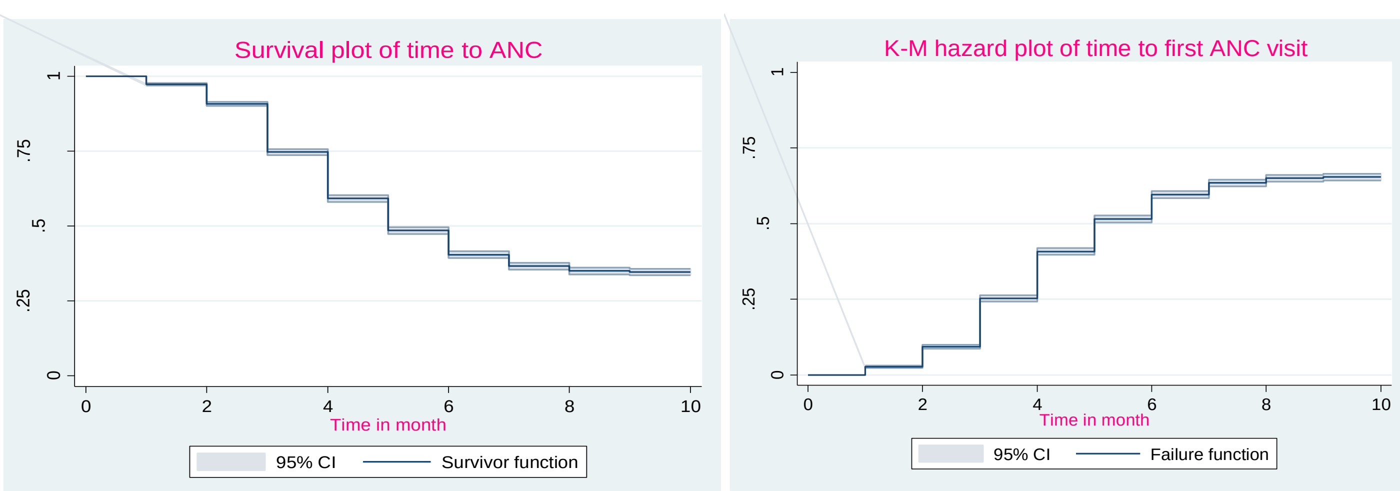


Figure4.1. The K-M plots of Survival and hazard functions of time to first ANC visit among pregnant mothers in Ethiopia

4.2.2. Comparison of place of residence in terms of survival time to first ANC visit

Kaplan Meier graphs are used to depict the waiting time to first ANC visit of pregnant women for different covariates (mother's characteristics). Figure 4.2, shows that pregnant mother from rural area started first ANC visit late compared to those from an urban area or the probability of not starting ANC visit were higher through gestational age for pregnant women from rural compared to urban residence. In addition, the median time of first ANC visit was 3rd and 6th month for urban and rural residents. In addition, the log-rank test (Table 4.2) shows there is a statistically significant difference between them in terms of waiting time to first ANC visit (p-value<0.001).

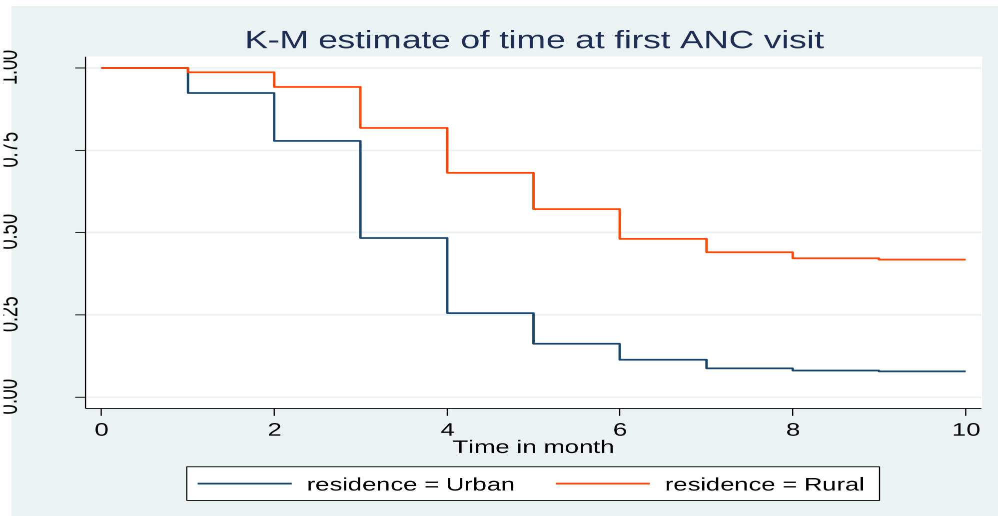


Figure 4. 2. Survival time to first ANC visit among pregnant women by place of residence

4.2.3. Comparison of 'survival' time to first ANC visit by women education

Figure 4.3 illustrates waiting time-to-first ANC visit. The graph shows there are differences in waiting time at first ANC visit for different categories of educational status. Women who attended primary, secondary or higher education started ANC visit earlier than women who did not attend education. The median survival time of women who had no education attended primary, secondary and higher education were 7, 4, 4 and 3 months respectively. Furthermore, the result of log-rank from Table 4.2 shows there is a statistically significant difference in survival time of first ANC visit among the educational levels (p-value <0.001).

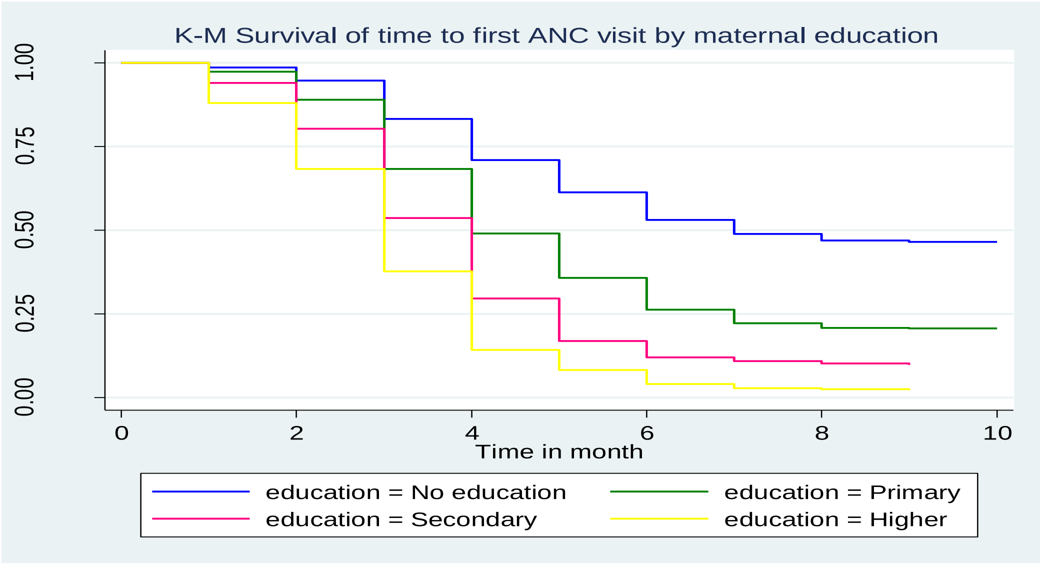
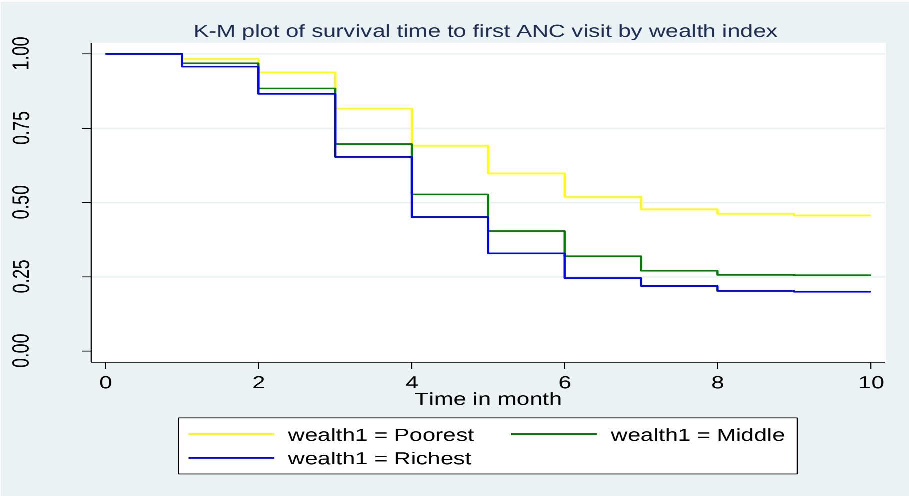


Figure 4. 3. Comparison time to first ANC visit among pregnant women in Ethiopia by women educational status

4.2.4. Comparison of survival time to first ANC visit by wealth index

K-M plot used to compare the survival time to first ANC visit for pregnant women in different wealth index (Figure 4.4). Thus, it was found women in the poorest wealth index household received first ANC later than both middle and richest wealth categories, and middle wealth index started ANC later than richest index categories. Furthermore, the median survival times of first ANC visit for poorest, middle and richest wealth were 7th, 5th and 3rd month respectively.



-

Figure 4. 4. Comparison of survival time to first ANC visit among pregnant women in Ethiopia by wealth index

4.2.5. Comparison of survival time to first ANC visit by Distance to get health care

The probability of not starting first ANC visit is higher for pregnant women who perceived problem to get medical care due to distance was a big problem to seek health care when compared to those who distance was not a big problem to get health care (Figure 4.5). The log-rank test also shows a statistically significant difference in survival time to first ANC visit between these groups (Table 4.2).

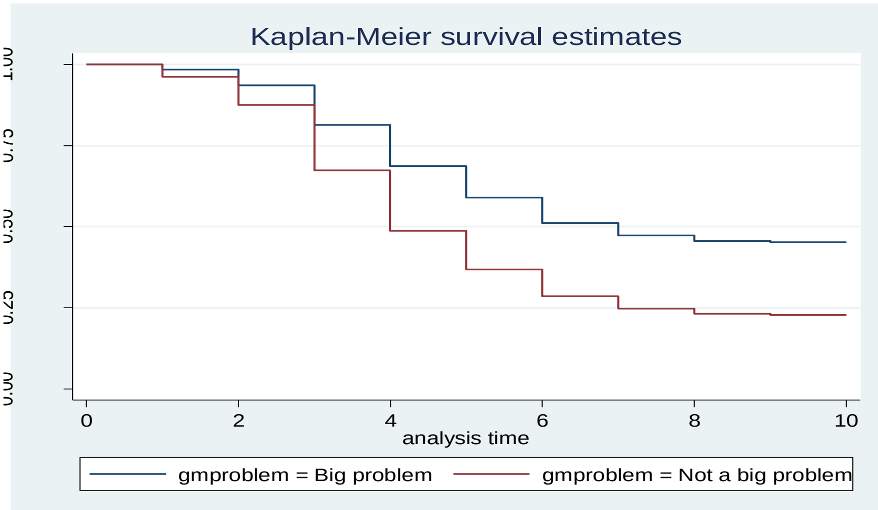


Figure 4. 5. Comparison of waiting time to first ANC visit among pregnant women in Ethiopia by perceived problem to get medical care due to distance

4.3. Comparison of 'Survival' Experience by log-rank test

A formal test was carried out using the log-rank to compare the difference between each categorical variable. The general hypothesis states that there is no difference against there are differences among the groups in survival experience time to first ANC visit.

The log-rank test result (Table 4.2) shows that there is no significant difference in the 'survival' experience of different categories for household head (p= 0.744). Similarly, the log-rank test performed for different covariates indicate there is statistically significant difference in survival experience among age group (p<0.001), region (p <0.001), residence (p <0.001), education of woman (p <0.001), wealth index (<0.001), education of husband (p <0.001), perceived problem to get medical care due distance (p <0.001), parity (p <0.001), religion (p<0.001) and wanted pregnancy (p=<0.001).

Table 4.2. Log-rank test for factors affecting time to first ANC visit in Ethiopia

	Degree Freedom	**Log-rank test**	
		Chi-square	P-value
Age	3	79.2	<0.001
Region	10	1401	<0.001
Residence	1	1301	<0.001
Education	3	1310	<0.001
Wealth index	2	550	<0.001
perceived problem to get medical care due to distance	1	464	<0.001
Wanted pregnancy	2	37.2	8.32e-09
head of house hold	1	0.1	0.744
Person who usually decides health care	3	85.1	<0.001
Parity	3	417	<0.001
Husband education	4	873	<0.001
Religion	4	321	<0.001

4.4. Testing the Validity of Proportional Hazard Assumption

Testing for proportional assumption is vital for interpretation and use of fitted proportional hazard model to look for other models that best fit to the data if the assumption of proportional hazard does not hold. Therefore, a test based on the interaction of the covariates with time and the plot of the scaled Schoenfeld residuals are used to see if the assumption of proportionality is violated or not. The test suggested that the PH assumption is violated for place of residence, getting medical care, total children ever born and religion, hence it was observed that each covariate has P-Value of <0.05 and all of the covariates simultaneously (GLOBAL test for Cox proportional hazard) has P-Value<0.05, so do not meet the proportional hazard assumption (Table 4.3).This departure from proportional hazards occurs when regression coefficients are dependent on time that is when time interacts with covariates. Thus, we doubt the accuracy of the PH assumption and consider the AFT model for this data set.

Table 4.3. Test of proportional hazard assumptions.

Covariate	Categories	rho	Chis-q	P-value
Place of residence	Rural	Ref		
	Urban	-0.039	7.24	0.0071
Woman education	No education	Ref		
	Primary	0.017	1.22	0.27
	Secondary	0.006	0.19	0.66
	Higher	0.017	1.25	0.26
Distance to get medical care	big problem	Ref		
	not big problem	0.04	7.47	0.006
Wanted pregnancy	Then	Ref		
	Later	0.005	0.12	0.73
	No more	0.018	1.46	0.226
Sex of house hold	Male			
	Female	-0.017	1.3	0.25
Birth order	1	Ref		
	2-3	0.02	1.77	0.183
	4-5	0.018	1.33	0.249
	>=6	0.042	7.67	0.005
Husband education	No education	Ref		
	Primary	0.017	1.33	0.248
	Secondary	-0.00016	0.00	0.991
	Higher	-0.026	3.08	0.079
	Don't know	-0.011	0.51	0.476
	Muslim	Ref		
Religion	Orthodox	0.074	24.32	0.00
	Catholic	-0.004	0.07	0.797
	Protestant	0.062	16.60	0.00
	Traditional/Other	0.0008	0.00	0.956
Person decided on respondent's health care	Respondent and Husband/partner	Ref		
	Respondent alone	0.0019	0.02	0.898
	Husband/Partner alone	0.0092	0.37	0.542
	Other	-0.0058	0.15	0.700
	15-19	Ref		
Age group	20-24	.-0.004	0.07	0.792
	25-34	-0.0023	0.02	0.877
	35-49	-0.0118	0.60	0.438
Wealth	Poorest	Ref		
	middle	-0.0019	0.02	0.896
	Richest	-0.0125	0.71	0.399
GLOBAL		NA	76.89	0.00

4.5. Accelerated Failure Time Model Result

4.5.1. Univariate Analysis

This study used univariate analysis in order to see the effect of each covariate on time-to-first ANC visit before proceeding to the multivariable analysis. The univariate analysis was fitted for every covariates by AFT models using different baseline distributions i.e. weibull, log-logistic, and log-normal. In all univariate analysis of AFT models, place of residence, woman education, perceived problem to get medical care due to distance, wanted pregnancy, husband education, parity, religion, person who decided on respondent's health care and wealth of index associated with waiting time to first ANC visit at 5% level of significance. The summary of univariate analysis is given in (Appendix I Table 1). Hence, based on univariate analysis except for the household head, all explanatory variables are candidates for further analysis.

4.5.2. Multivariable AFT Analysis

For time-to-first ANC visit data, multivariate AFT models of weibull, log-logistic and log-normal distribution were fitted by including all the covariates significant in the univariate analysis at 5% level of significance. To compare the efficiency of different models, AIC and BIC were used, which is most common applicable standard to select the model. A model having the smallest AIC & BIC values is considered a better fit model. Accordingly, Lognormal AFT model has smallest AIC = 12575.89 & BIC=12752.68 from Table 4.4 and selected for the time-to-first ANC visit during gestational age data set from the given alternatives.

Covariates which become non-significant in the multivariable analysis were removed from the model by using a backward elimination technique. Accordingly, this model age group of women was excluded. And finally, the effect of interactions terms was also tested and found to be statistically non-significant in multivariable log-normal AFT model at 5% level of significance. The final model kept the main effect of the covariate except age of women at first ANC visit. All AFT models and the corresponding AIC values are displayed in Table 4.3.

Table 4. 4 Comparison of Accelerated Failure Time Models using AIC, BIC and -2loglik in for the assessment of time to first ANC visit in Ethiopia

Distribution	AIC	BIC	-2Loglik
Weibull	12984.22	13161	23852.47
Log-normal	12575.89	12752.68	23518.95

Log-logistic	12641.32	12818.1	23521

Table 4.4 shows the result of log-normal multivariable AFT model and covariates such as place of residence, woman education, a perceived problem to get medical care due to distance, wanted pregnancy, husband education, parity, and religion and wealth index are statistically significant at 5% level of significance except age. This implies that the age has no effect on the survival time-to-first ANC visit among pregnant mothers in Ethiopia. Based on the lognormal AFT regression model, given that the effect of all other factors keep fixed, the estimated acceleration factor for place of residence is 0.696. This implies that women residing in urban start first ANC visit earlier than rural residents. The acceleration factors for women attending primary, secondary and above education are estimated at 0.832, 0.82 and 0.779 respectively. This implied that women who attended primary, secondary and higher education started their first ANC visit at an early gestational age compared to non-educated women. The acceleration factor for women who are getting medical care was a big problem due to a distance from health facility is 0.88. This means pregnant women who perceived distance as being a big problem to get medical care had prolonged waiting time to first ANC visit when compared to those who perceive distance is not a big problem.

Table 4. 5. Summary result of the final Log-Normal AFT model in the assessment of time to first ANC visit in Ethiopia

Covariate	**Categories**	**Estimate($\hat{\beta}$)**	ϕ	**95%CI SE(ϕ)**	**P-value**
Place of residence	Rural	Ref			
	Urban	-0.362	0.696	(0.660,0.733)	<0.001
Woman education	No education	Ref			
	Primary	-0.184	0.832	(0.794, 0.870)	<0.001
	Secondary	-0.195	0.822	(0.760, 0.890)	<0.001
	Higher	-0.249	0.779	(0.700, 0.870)	<0.001
Distance to get	big problem	Ref			
medical help	not big problem	-0.13	0.878	(0.846, 0.110)	<0.001
Wanted pregnancy	Then	Ref			
	Later	.023	1.023	(0.972, 1.073)	0.382
	No more	.173	1.188	(1.100, 1.284)	<0.001
Parity	1	Ref			
	2-3	.075	1.081	(1.020, 1.140)	0.008
	4-5	.083	1.086	(1.015, 1.163)	0.016
	>=6	.182	1.21	(1.112, 1.295)	<0.001
Husband education	No education	Ref			
	Primary	-.170	0.843	(0.806, 0.880)	<0.001

	Secondary	-.191	0.826	(0.773, 0.883)	<0.001
	Higher	-.208	0.810	(0.749, 0.880)	<0.001
	Don't know	-.114	0.892	(0.736, 0.925)	0.245
Religion	Muslim	Ref			
	Orthodox	-.097	0.907	(0.871, 0.951)	<0.001
	Catholic	.175	1.190	(0.963, 1.473)	0.107
	Protestant	.102	1.107	(1.054, 1.163)	<0.001
	Traditional/Other	.370	1.445	(1.240, 1.690)	<0.001
Person decided on respondent's health care	Respondent and Husband/partner	Ref			
	Respondent alone	-.00016	0.999	(0.955, 1.041)	0.995
	Husband/Partner alone	.067	1.067	(1.033, 1.123)	0.004
	Other	.083	1.086	(0.761, 1.612)	0.654
Wealth	Poorest	Ref			
	Middle	-.167	0.846	(0.806, 0.887)	<0.001
	Richest	-.230	0.794	(0.762, 0.830)	<0.001

*Scale σ=0.67 and shape α=1.484 ,constant=2.07,ϕ; Acceleration factor;** significant at 5% level; 95%CI: 95% confidence interval for acceleration factor; SE(β) standard error for β; Ref. Reference*

When the effect of other factors kept constant, the estimated acceleration factor of 'wanted pregnancy' later and no more were 1.023 and 1.2 respectively using then or the child was wanted at the time of pregnancy as a reference category. The acceleration factors for birth order women 2^{nd}-3^{rd}, 4^{th}-5^{th}, and >=6^{th} were 1.08, 1.086 and 1.2 respectively by using first birth order as a reference category. This implied that first birth order women have less survival of time-to-first ANC visit second, third and above. Regarding the education level of husband, the acceleration factor of the husband who attend primary, secondary, higher and don't know were 0.843, 0.826, 0.81, and 0.892 respectively. Since acceleration factor of all categories is less than 1, it indicates husband who attended primary, secondary and higher had shorter time-to-first ANC visit (started first ANC earlier) respectively than the reference category (no education). In addition, women whose decision on their health care was made by their husband/partner alone started first ANC visit later than those who decide by herself and together with husband (ϕ=1.07). On the other hand, women whose decision on their health care was made by the women alone or other categories are not statistically significant at 5% level and their ϕ confidence interval includes1.

Regarding religion, women who are followers of Protestant and other religions had prolonged time-to-first ANC visit by a factor of 1.107 and 1.445 when compared to their corresponding reference categories. Moreover, orthodox religion followers were started ANC visit earlier by acceleration factor of 0.907 compared to the reference category (Muslim) religion. Similarly, middle wealth index ($\phi = 0.846$) and richest women ($\phi = 0.794$) had less prolonged time-to-first ANC visit than reference category (poorest women). The value of the shape parameter in the log-normal model is $\alpha = 1.484$. Since, this value is greater than unity the hazard function is unimodal.

4.6. Parametric Shared Frailty Model Results

In the previous section, AFT models with different distribution were fitted and compared to analyse the survival time-to-first ANC visit to identify baseline distribution and factors affecting time to first ANC visit. To model the heterogeneity (random component) using the region as frailty term and investigate exposure factors associated with the survival of time-to-first ANC visit using gamma shared frailty with log-normal, weibull and log-logistic baseline distribution were used. The effect of random component (frailty) is significant for three baselines of gamma shared frailty models, and log-logistic gamma shared frailty model have minimum Akaike's information criteria (AIC =11954.01). This indicates Log-logistic gamma shared frailty model is a more efficient model to describe time-to-first ANC visit during gestational age of women.

Table 4. 6. Comparison of shared gamma frailty model with different baseline distributions

Baseline Distribution	**Frailty Distribution**	**AIC**	**BIC**
Weibull	Gamma	12503.23	12219.96
Log-logistic	Gamma	11954.01	12151.2
Log-normal	Gamma	12022.77	12700.41

4.6.1. Log-logistic Gamma Shared Frailty Model Result

This model is the same as the log-normal AFT model discussed, except that a frailty component has been included. The frailty term in this model assumed to follow a gamma distribution with mean one and variance equal to theta (θ). The estimated value of theta (θ) is 0.568. A likelihood ratio test for the hypothesis θ= 0 (Table 4.7) have chi-square value of 559.12 with one degree of freedom and highly significant (P <0.001). Moreover, the associated Kendall's tau (τ), which measures dependence within clusters (region), is estimated to be 0.221. The estimated value of the shape parameter in this selected model is (γ=3.53). This value is greater than unity that indicates the shape of hazard function is unimodal, that is, it increases up to some time and then decreases. Analysis based on log-logistic Gamma frailty model shows that place of residence, education level of women, a perceived problem to get medical care due to distance, wanted pregnancy, parity, husband education, religion and wealth of index women are significant predictors of time to first ANC visit. However, according to this model the 'age group' and 'person who decided on health care of woman' had no significant effect on time of first ANC visit during gestational age.

Table 4. 7. Log-logistic Gamma Frailty Model Result in assessments of time to first ANC visit in Ethiopia.

Covariate	Categories	Estimate($\widehat{\beta}$)	ϕ	95%CI SE(ϕ)	P-value
Place of residence	Rural	Ref			
	Urban	-0. 253	0.776	(0.74, 0.817)	<0.001
Woman education	No education	Ref			
	Primary	-0. 109	0.897	(0.86, 0.936)	<0.001
	Secondary	-0. 115	0.89	(0.83, 0.96)	0.002
	Higher	-0. 153	0.86	(0.78, 0.945)	0.002
Distance to get medical care	big problem	Ref			
	not big problem	-. 084	0.92	(0.886, 0.965)	<0.001
Wanted pregnancy	Then	Ref			
	Later	.023	1.023	(0.976, 1.073)	0.335
	No more	.17	1.185	(1.1, 1.275)	<0.001
Parity	1	Ref			
	2-3	.078	1.08	(1.026, 1.14)	0.003
	4-5	.069	1.07	(1.005, 1.143)	0.035
	>=6	. 16	1.17	(1.09, 1.26)	<0.001
Husband education	No education	Ref			
	Primary	-. 096	0.843	(0.806, 0.88)	<0.001
	Secondary	-.14	0.826	(0.773, 0.883)	<0.001
	Higher	-.193	0.81	(0.749, 0.88)	<0.001
	Don't know	-.117	0.892	(0.736, 0.925)	0.218
	Muslim	Ref			
Religion	Orthodox	-.14	0.907	(0.871, 0.95)	<0.001
	Catholic	.166	1.19	(0.963, 1.473)	0.113
	Protestant	.17	1.107	(1.054, 1.163)	<0.001
	Traditional/Other	.5	1.445	(1.24, 1.69)	<0.001
	Respondent alone	-.005	0.99	(0.955, 1.04)	0.818
	Husband/Partner alone	.041	1.07	(1.01, 1.12)	0.073
	Other	.035	1.086	(0.76, 1.6)	0.852
	20-24	.056	1.058	(0.96, 1.146)	0.221
	25-34	.032	0.968	(0.9, 1.085)	0.499
	35-49	.078	1.08	(0.92, 1.135)	0.145

Wealth	Poorest	Ref			
	Middle	-.11	0.895	(0.854, 0.94)	<0.001
	Richest	-.17	0.842	(0.808, 0.88)	<0.001

*τ=0.221, $\boldsymbol{\theta = 0.568(0.2301)}$,$\boldsymbol{\gamma = 3.53}$, $\boldsymbol{\lambda = 0.00782}$ log(scale)=-1.2616 *p < 0.05 was statistically significant. $\boldsymbol{\phi}$ =Acceleration factor, $\boldsymbol{\theta}$ =Variance of the random effect, $\boldsymbol{\tau}$= Kendall's tau, , CI=confidence interval, S.E=standard error, Ref=Reference, Likelihood-ratio test of theta=0: chibar2(01) =559.12 Prob>= chibar2=0.00*

Table 4.7 also shows the acceleration factor of 0.776 for urban women with 95% CI (0.74, 0.817), 22.4% less than those rural women. This indicates rural women have prolonged time-to start first ANC visit than urban women. Moreover, the acceleration factor for women who attended primary, secondary and higher education are 0.897, 0.89, 0.89 and 95% CI (0.86, 0.936; 0.83, 0.96; 0.78, 0.945) respectively does not include 1. An acceleration factor less than 1 indicates that decreasing of time of first ANC visit and hence, shorter expected waiting of time-to-first ANC visit than the reference group at 5% level of significance.

The acceleration factor of women wanted pregnancy no more is 1.185 with 95% CI (1.1, 1.275) this implies they had higher expected survival time than their corresponding reference categories. However, women wanted pregnancy later were not significantly different from the baseline women wanted pregnancy then at 5% level. The acceleration factor and 95% CI of parity 2-3 were 1.08 (1.026, 1.14), for parity 4-5 was 1.07 (1.005, 1.143) and for parity >=6 was 1.17 (1.09, 1.26), respectively, hence they had prolonged time-to-first ANC visit when compared with women in the reference group.

Husband education is also another significant covariate with acceleration factors greater than 1 for all categories. As a result, married women whose husband attended don't know, primary, secondary and higher education had prolonged time-to-first ANC visit by a factor of 0.892, 0.843, 0.826 and 0.81 respectively when compared with the reference category no education. Categories of significant covariates having acceleration factor less than 1 imply that women characterized by those categories of the same covariate started first ANC visit earlier relative to the reference category of the same covariates. The confidence interval of the acceleration factor for orthodox, Protestant, and other/ traditional religion are 0.907 (95% CI: 0.871, 0.95), 1.107(95% CI: 1.054, 1.163) and 1.445(95% CI: 1.24, 1.69) respectively and did not include 1 between the interval, indicating that these categories are also significant determinant factor for time-to-first ANC visit. Thus, women who are followers of orthodox had shorter waiting time to first ANC visit while Protestant and other/traditional religion had prolonged time-to-first ANC visit when compared to the reference category (Muslim). But the Catholic religion followers are not significantly different from the baseline Muslim religion group at 5% level of significance.

The acceleration factor for wealth index of women is 0.895 and 0.842 for the group of middle and richest respectively using the poorest as a reference category. This indicates that for middle and richest groups started ANC earlier by a factor ϕ=0.895 and ϕ=0.842, respectively, than the reference group at 5% level of significance (see Table 4.7)

4.7. Comparison of Log-normal AFT and Log-Logistic-Gamma Frailty Model

From the Table 4.5 and 4.7, we can see that the log-logistic gamma shared frailty model has a smaller AIC =11954.01 than log-normal AFT =12575.89, this implies log-logistic gamma frailty model fitted the survival of time-to-first ANC visit data better than the log-normal AFT model which did not take into account the clustering effect in the region. In AFT models exponentiation the coefficients it is possible to obtain time ratios. These time ratios used to calculate the factor change or percentage change in the expected survival time associated with a one unit increases in a covariate. The 95% CI of ϕ of all covariate does not include one except age.

Table 4.8. Comparison of Log-normal AFT and Log-logistic Gamma Frailty Model in the assessment of time to first ANC visit in Ethiopia

Covariate	Log-normal AFT			Log-logistic Gamma Frailty		
	$\hat{\beta}$	ϕ	**95%CI**	$\hat{\beta}$	ϕ	**95%CI**
Place of residence						
Rural	Ref			Ref		
Urban	-0.362	0.696	(0.66,0.733)	-0. 253	0.776	(0.74, 0.817)
Woman education						
No education	Ref			Ref		
Primary	-0.184	0.832	(0.794, 0.87)	-0. 109	0.897	(0.86, 0.936)
Secondary	-0.195	0.82	(0.76, 0.89)	-0. 115	0.89	(0.83, 0.96)
Higher	-0.249	0.779	(0.7, 0.87)	-0. 153	0.86	(0.78, 0.945)
Getting medical help						
Distance is big problem	Ref			Ref		
Distance is not big problem	-.13	0.88	(0.846, 0.911)	-. 084	0.92	(0.886, 0.965)
Wanted pregnancy						
Then	Ref			Ref		

Later	.023	1.023	(0.972, 1.073)	.023	1.023	(0.976, 1.073)
No more	.173	1.2	(1.1, 1.284)	.17	1.185	(1.1, 1.275)
Parity						
1	Ref			Ref		
2-3	.075	1.08	(1.02, 1.14)	.078	1.08	(1.026, 1.14)
4-5	.083	1.086	(1.015, 1.163)	.069	1.07	(1.005, 1.143)
>=6	.182	1.2	(1.112, 1.295)	. 16	1.17	(1.09, 1.26)
Husband education						
No education	Ref			Ref		
Primary	-.17	0.843	(0.806, 0.88)	-. 096	0.843	(0.806, 0.88)
Secondary	-.191	0.826	(0.773, 0.883)	-.14	0.826	(0.773, 0.883)
Higher	-.208	0.81	(0.749, 0.88)	-.193	0.81	(0.749, 0.88)
Don't know	-.114	0.892	(0.736, 0.925)	-.117	0.892	(0.736, 0.925)
Religion						
Muslim	Ref			Ref		
Orthodox	-.097	0.907	(0.871, 0.95)	.14	0.907	(0.871, 0.95)
Catholic	.175	1.19	(0.963, 1.473)	.166	1.19	(0.963, 1.473)
Protestant	.102	1.107	(1.054, 1.163)	.17	1.107	(1.054, 1.163)
Traditional/Other	.37	1.445	(1.24, 1.69)	.5	1.445	(1.24, 1.69)
Person decided on respondent's health care						
Respondent and husband/partner	Ref					
Respondent alone	-.00016	0.99	(0.955, 1.04)			
Husband/Partner alone	.067	1.07	((1.01, 1.12)			
Other	.083	1.086	(0.76, 1.6)			
Wealth						
Poorest	Ref			Ref		
middle	-.167	0.846	(0.806,0.887)	-.11	0.895	(0.854, 0.94)
Richest	-.23	0.794	(0.762, 0.83)	-.17	0.842	(0.808, 0.88)
AIC=12575.89				AIC=11954.01		
BIC=12752.68				BIC=12151.2		

Estimate ($\hat{\beta}$) =estimated value of$\hat{\beta}$; ϕ=acceleration factor; 95%CI for acceleration factor; Ref. = Reference; AIC=Akakie's information criteria; BIC= Bayesian information criteria

4.8. Model Diagnostics

4.8.1. Diagnostic Plots of the Parametric Baselines

To check the adequacy of baseline, hazard of the Weibull is plotted by the logarithm cumulative hazard function with the logarithm of time-to-first ANC visit in month; similarly, log-logistic is plotted by the logarithm failure odds with the logarithm of time of the study and the log-normal is plotted by the qnorm of log failure odds with the logarithm of time of the study. If the plot is linear the given baseline distribution is said to be appropriate for the given dataset. Accordingly, their respective plots are given in figure 4.6 and the plot for the log-logistic baseline distribution is closer to a straight line than that of Weibull and log-normal baseline distributions. Also, the log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions. This evidence supports the decision made based on the AIC value that log-logistic baseline distribution is appropriate for the given dataset.

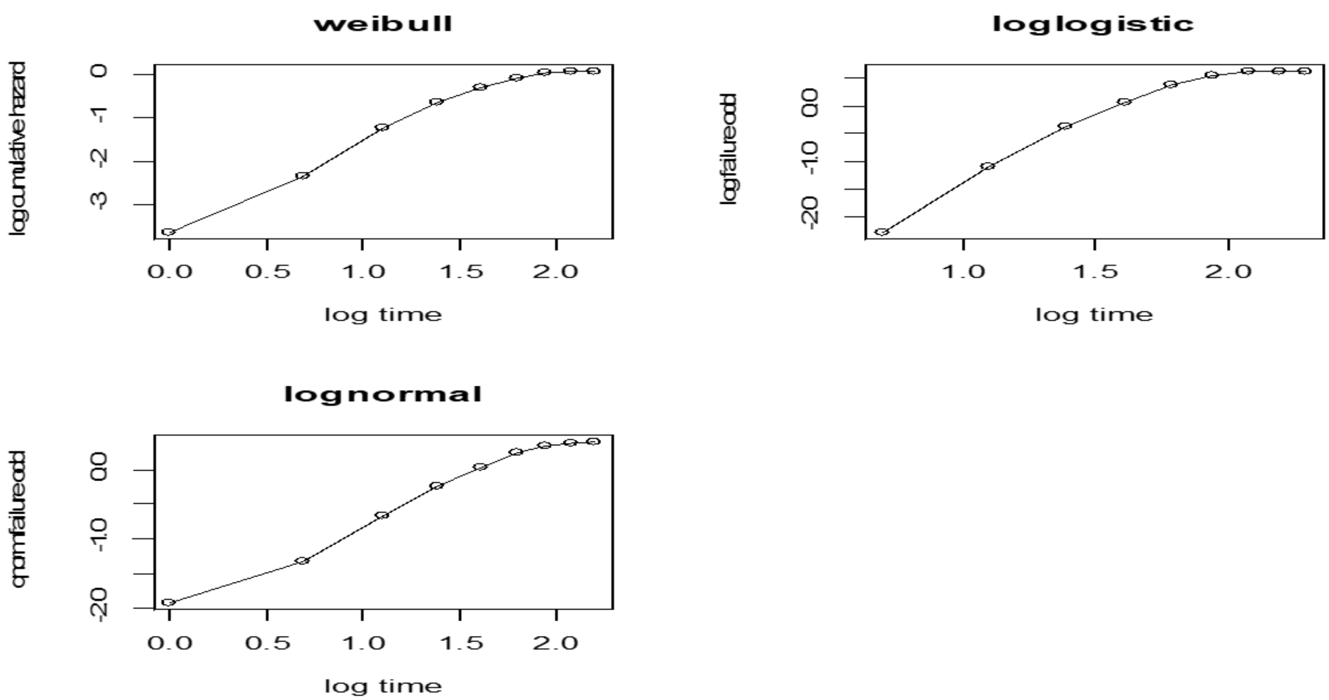


Figure 4. 6. Graphs of Weibull, Log-logistic and Log-normal baseline distribution for time-to-first ANC visit

4.8.2. The Cox-Snell Residual Plot

The Cox-Snell residual is one way to investigate how well the model fits the data. The plot for the fitted model of residuals for Weibull, log-logistic and log-normal distribution data via maximum likelihood estimation with their cumulative hazard function is given in figure 4.7. The plots indicate that the graph of the log-logistic plot gives us evidence or more appropriate linear than the log-normal and Weibull graph. These results are consistent with our previous results (in Table 4.8) based on Akaike's information criterion and Bayesian information criterion.

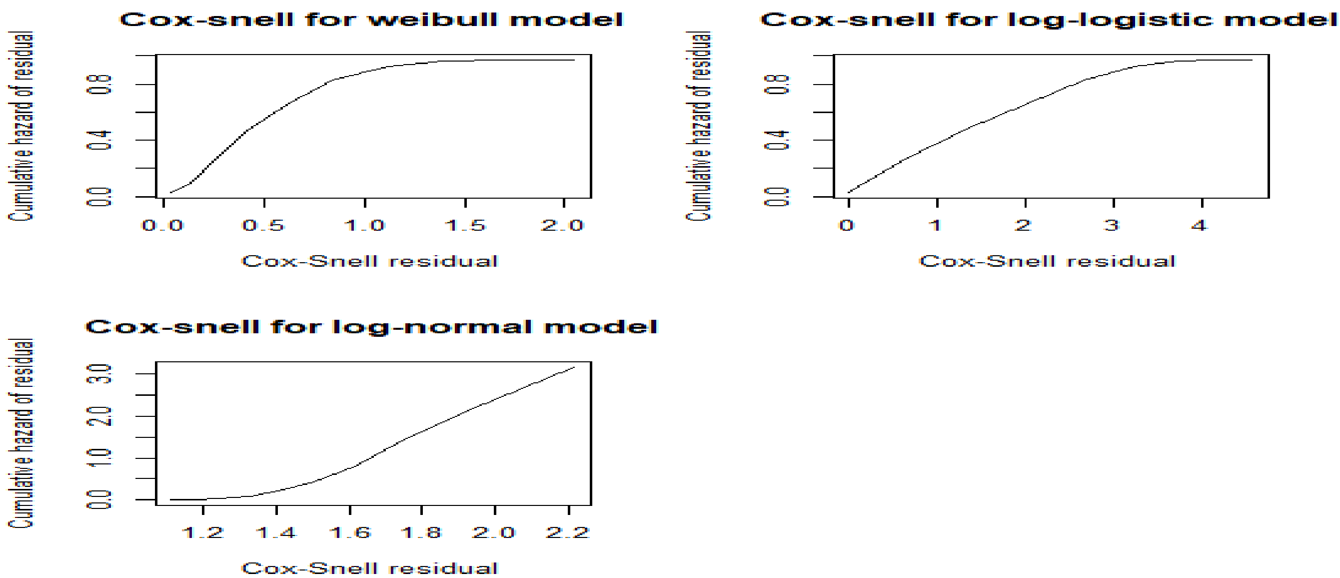


Figure 4. 7. Estimated cumulative hazard plot of the Cox-Snell residuals for Weibull, lognormal and log-logistic distribution

4.8.3. Quantile-Quantile plot

A quantile-quantile (q-q plot) was made to check if the accelerated failure time provided an adequate fit to the data using two different groups of the population. We shall graphically check the adequacy of the accelerated failure-time model by comparing the significantly different pregnancy women perceived problem to get medical care due to distance is a big problem and not big problem, pregnant women who were middle and poor wealth index. From figure 4.8, the quantile-quantile plot approximately linear for all pregnant women's. Therefore, a log-logistic gamma shared frailty model is the best fit for both covariates (perceived problem to get medical care due to distance and wealth index) with slopes equivalent to the acceleration factors 0.88 and 0.846 respectively. Therefore, for data time to first ANC log-logistic as a baseline was accelerated failure time model.

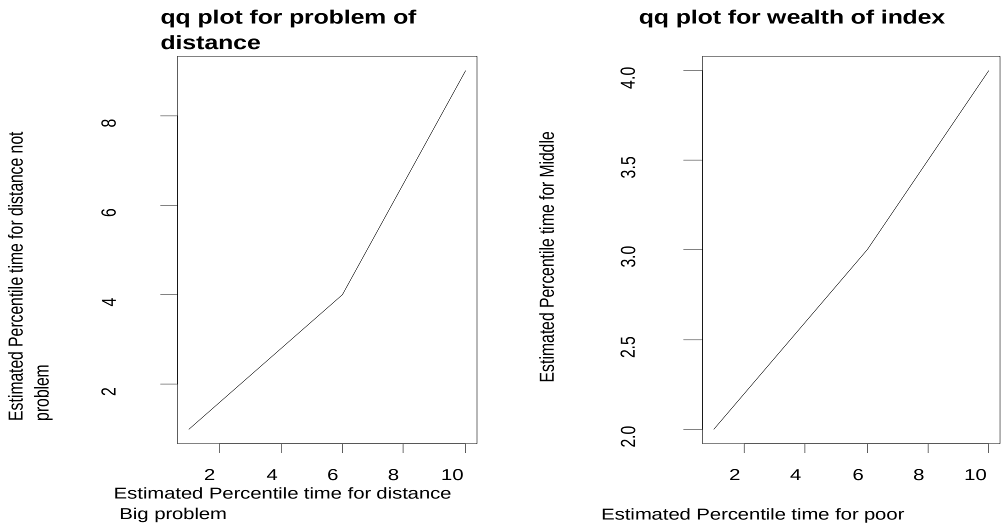


Figure 4. 8. Quantile-Quantile plot to check the adequacy of the accelerated failure time model

4.9. DISCUSSION

The main aim of the study was modelling determinants of time-to-first ANC receipt using AFT and gamma shared frailty models by considering three baseline distributions Weibull, log-normal and log-logistic to investigate model that is better in predicting time to first ANC visit in Ethiopia. The comparison of distributions of the models was done using AIC criteria, where a model minimum AIC was accepted (Munda, 2012).

Univariate analysis (AppendixI, Table 1) revealed that household head was not significantly related to time to first antenatal care visits but place of residence, woman's educational level, wealth index, problem to get medical care due distance, husband's educational level, religion, person who decided on health care, parity, wanted pregnancy and age group of women at pregnancy were found to be statistically significant. All significant variables in univariate analyses were included in all multivariable analysis of the AFT model and the best model was selected using AIC criteria. Log-normal AFT model is best over weibull and log-logistic AFT based on AIC value from (Table 4:4).The result comparable with the earlier study (Million, 2018). Place of residence, mother educational level, a perceived problem to get medical care due to distance, wanted pregnancy, husband education, parity, religion, persons who decided respondent health care and wealth of index are significantly associated with the timing of first ANC visit.

Moreover, after analysing the given data set by using log-normal AFT, parametric shared frailty models were fitted by considering weibull, log-logistic and log-normal baseline distributions by assuming gamma distribution for the frailty term. Consequently, log-logistic gamma shared frailty model was selected over weibull and log-normal gamma shared frailty models. This study is consistency the baseline with other study (Tessema et al., 2015). Gamma distribution is selected for the frailty term due to its mathematical tractability and flexibility of hazard function (Clayton, 1978; Vaupel, 1979). This study also exposed there is heterogeneity (frailty effect) between women's categorized as region and correlation within the same region of ANC visit women's. The clustering effect was significant (p-value <0.001) in log-logistic gamma shared frailty model. This showed that there was heterogeneity between the regions on the timing of first antenatal care visit during the gestational age. This heterogeneity could be due to populations in the same region relatively have some shared factors such as accessibility of health facilities, socio-cultural factors and others in determining the timing of the first ANC visit in Ethiopia.

Finally, log-normal AFT and log-logistic gamma shared frailty were compared and the results from the AFT and frailty models are somewhat similar to each other. In both models, place residence, woman education level, a perceived problem to get medical care due to distance, wanted pregnancy, husband education, parity, religion, a person decided on respondent on health care and wealth index are significant predictors of time to first ANC 5 % level of significance. However, some improvement was observed on the parameter estimates in which confidence interval of accelerated factors for log-logistic gamma shared frailty is a little bit wider than log-normal AFT due to the inclusion of frailty term. Moreover, the study revealed that AIC for Log-logistic gamma frailty model was smaller than lognormal AFT that indicates parametric frailty model fitted for time-to-first ANC visit data better than the AFT.

Based on the given dataset place of residence of the woman was the factors that affect survival time of first antenatal care during the gestational age. As it was indicated in both log-normal accelerated failure time model and log-logistic gamma shared frailty models the acceleration factor for women who lived in the urban area is much lower 0.696 and 0.776 respectively. That is the rural woman had more prolonged time-to-first antenatal care visit than urban women. The similar finding reported by a study conducted in Nigeria; where in earlier initiation of ANC was found to be more common in urban women than in the rural (Francis, 2017). Furthermore, other studies also publicize comparable finding (Damme et al., 2015). This later initiation of ANC among rural women could be due to better access to health facilities in urban areas than in rural areas. In addition, distances to health facilities are generally shorter in the urban area hence distance may not a big problem to get medical care for urban. This finding supported by that of study (Tsegay et al.,2013).

A problem to get medical care due to a distance from health facilities is another factor that significantly predicts the timing of first ANC visit of women during the gestational age. The result of both lognormal AFT and log-logistic gamma shared frailty models show, women who perceive distance is a big problem to get medical were started first ANC visit later than their counterpart whose distance was not a big problem to get medical care. This finding is consistent with a conclusion from another study (Francis, 2017). This finding implies the need for making health facilities more accessible to enhance proper ANC utilization.

The findings of this study also uncovered that increasing husband and women's education level significantly shortens the time-to-first ANC visit. The finding is comparable with the report from another study (Bahilu et al., 2012). In addition, similar findings were documented by studies conducted in Nigeria and Uganda which publicize that educational attainments of women had a significant effect on first visit of antenatal care and educated women had shortened time to first ANC visit (Francis, 2017; Turyasima et al., 2015). This could be due to educated women and partner are more likely to get better jobs which yield higher income, and better access to health information which could help them to start ANC at an appropriate time.

Moreover, the study also revealed that women from richest and middle wealth index households were started first ANC at earlier than women from poor households. This is similar to reports by Gilles, (2017) suggested that women from richest wealth index were more likely to initiate ANC at an earlier gestational age. A study conducted among Nigeria women also showed women from poorest households were started first ANC visit later than (after prolonged time) than women from middle or richest households (Francis, 2017). Furthermore, there were differences in timing of the first ANC visit among women who wanted pregnancy then, later and not at all (no more). Pregnant mothers who wanted pregnancy no more were initiated ANC later than those who wanted then but there is no difference in time of ANC initiation among mother who wanted pregnancy then and later.

In both lognormal AFT and Log-Logistic-Gamma Frailty Models, parity (the number of children ever born) also significantly associated with the timing of ANC initiation. Accordingly, women of para 2-3 were initiated first ANC visit later (after prolonged time) than para one women. Furthermore, para one women started first ANC visit earlier than para 4-5 and para >=6 categories. This finding is in line with previously conducted studies (Erica, 2012; Gross et al., 2012). In this study age of women has not statistically significant effect on the timing of ANC visit. But this study is not consistent with other studies with regard to the age of women. For instance, the result other studies (Fagbamigbe et al., 2017; Gross et al., 2012) suggest that age of women have a significant effect on first ANC visit and older age pregnant woman started ANC service earlier than younger age pregnant women. In addition, it was found that age of pregnant women was significantly determine timing of ANC visit and woman who had age 25 years and below more likely to start ANC visit timely than women who had above 25 years (Temesgen et al., 2014).

Furthermore, the finding from models suggested there are differences in timing of the first ANC visit along religion group. The study shows Orthodox follower started ANC earlier than Muslim while Protestant and other or traditional followers are categories started later than Muslim followers. The study conducted in Nigeria also documented there was a difference in timing of the first ANC visit between different religion group (Francis, 2017).

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

To model the determinants of time-to-first ANC visit, different parametric shared frailty, and AFT models by using different baseline distributions were applied. Among this using AIC, log-logistic gamma shared frailty model is better fitted to time-to-first ANC visit dataset than other parametric shared frailty and AFT models. There was a frailty (clustering) effect on the time-to-first ANC visit that arises due to differences in the distribution of timing of first ANC receipt among regions of Ethiopia. This indicates the presence of heterogeneity and necessitates the frailty models. This heterogeneity could be arising due to environment, socio-cultural differences in utilization of health care services and variation in accessing health services across the regions of Ethiopia.

In this study the major factors identified were, wealth index, educational levels of mother and husband, place of residence, number of children, pregnancy intention (wanted pregnancy), religion and problem due to a distance to seek health care are statistically significant. Likewise, being from high wealth index household, women and partner with better education, residing in urban areas, accessing health facilities due to distance, having a fewer total number of children born (few Para) and wanted pregnancy then were associated with earlier initiation of first ANC visit among pregnant women in Ethiopia.

In addition, the median time of first ANC visit was at 5th month which was later than the optimum time recommended by WHO that is every pregnant woman should start first visit at least at 4th month. The findings of the study suggest that specific efforts are needed to target women of lower socioeconomic status, especially those who are from the poor wealth index, access to informal education for woman and husband, especially in religious places.

5.2. Recommendation

Based on the findings of this study the following recommendations are forwarded for policy makers', stakeholders, and researches.

- To ensure that all pregnant women start ANC visit early, wealth status of households (economic status of family), mothers' and husband educational level, place of residence, number of birth (parity), religion, fertility preference (wanted pregnancy), distance to health facilities, and decision makers on women health care (women autonomy) need to be considered when planning and developing policies against antenatal care utilization.
- Furthermore, maternal health care national policy makers, planners and ANC providers should take into consideration the regional heterogeneity in time of initiation of ANC and proximity rate of health facility especially in rural area.
- The log-logistic gamma shared frailty model give better predictions to the timing of first ANC visit since the findings from the study suggest that there are some unobserved characteristics which were accounted for heterogeneity across regions. So, future studies should have to use parametric shared frailty models.

5.3. Limitation of the study

The study used the ANC visit history data of the respondents (mothers) from EDHS 2016 and the data are reported retrospectively. Retrospective ANC histories are subject to possible reporting errors that may adversely affect the quality of the data due to the lack of memory of the respondent (mothers). It may be affected by the completeness with which ANC visit, as well as the accuracy of information on current ANC visit times and the type of trimester visited. If the type of trimester visited is misreported and the net effect of this ANC visit misreporting results in transference from one visit bracket to another, it will bias the estimates. On the other hand, this study also includes only respondents' age of 15-49 years old.

REFERENCES

Babalola, S., & Fatusi, A. (2009). Determinants of use of maternal health services in Nigeria-looking beyond individual and household factors. *BMC Pregnancy and Children*, *9*, 45. https://doi.org/10.1186/1471-2393-9-43

Bahilu, T, Tewodros, Bahilu, Mariam, Abebe G, Dibaba, Y. (2012). Factors Affecting Antenatal Care Utilization in Yem Special Woreda , Southwestern Ethiopia. *Ethiop J Health Sci.*, *9*(1), 25–27. Retrieved from https://www.ajol.info/index.php/ejhs/article/viewFile/145941/135458

Banbeta, A., Seyoum, D., Belachew, T., Birlie, B., & Getachew, Y. (2015). Modeling time-to-cure from severe acute malnutrition application of various parametric frailty models. *Archives of Public Health*, *73*(6), Doi: 10.1186/2049-3258-73-6.

Belayneh, T., Adefris, M., & Andargie, G. (2014). Previous early antenatal service utilization improves timely booking: cross-sectional study at university of Gondar hospital, northwest. *Ethiopia. J Pregnancy, 132494.* https://doi.org/10.1155/ 2014/132494

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, *2*(2), 273–277.

Berhem, K., Welearegayn, H., Gerezgihe, A., & Hailemariam. (2014). Assessment of Antenatal Care Utilization and its Associated Factors Among 15 to 49 Years of Age Women in Ayder Kebelle , Mekelle City 2012 / 2013 ; A Cross Sectional Study. *American J of Advanced Drug Delivery*, *2*(1), 065–075.

Bressen, C. (2004). An Estimating Equation for Parametric Shared Frailty Modelswithmarginal Additive Hazards. *Journal of the Royal Statistical Society*, 66: 207-220.

Chimankar, D. A., & Sahoo, H. (2011). Factors influencing the Utilization of Maternal Health Care Services in Uttarakhand. *Ethno Med*, *5*(3), 209–216.

Clayton, R. (1978). Erectile Dysfunction (impotence) Erectile Dysfunction (impotence).

Collet, D. (1994). *Modelling Survival Data in Medical Research* (2nd editio). London, UK: Chapman & Hall/CRC.

Collett, D. (2003). Modeling survival data in medical research. In *CHAPMAN & HALL/CRC*.

Cox, D. R. (1972). Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society*, *34*(2), 187–220.

Cox, D. R. (1984). Kaplan EL and Meier P. (1958). *Royal Statistical Society*, *34*(2).

CSA Ethiopia and ICF. (2012). Ethiopia Demographic and Health Survey 2011. Addis Ababa, Ethiopia and Calverton, Maryland, USA: CSA and ICF International.

CSA Ethiopia and ICF. (2016). *Ethiopia Demographic and Health Survey. Addis Ababa, Ethiopia, and Rockville, Maryland, USA.*

Central Statistical Agency (CSA) [Ethiopia]and ICF (2017) *Ethiopia Demographic and Health Survey 2016. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF*.

Damme, Tolera, Gudissa, Gmariam, & Abebe. (2015). Time of Antenatal Care Booking and Associated Factors Among Pregnant Women Attending Ambo Town Health Facilities , Central Ethiopia. *Journal of Gynecology and Obstetrics*, *3*(5), 103–106. https://doi.org/10.11648/j.jgo.20150305.12

Datwyler, C., & Stucki, T. (2011). *Parametric Survival Model, Handout*.

Despa, S. (2017). *What is Survival Analysis?Cornell University*. Retrieved from http://www.cscu.cornell.edu/news/statnews/stnews78.pdf

Duchateau, Janssen, L., P Lindsey, P., Legrand, C., Nguti, R., & Sylvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics & Data Analysis*, 40:603-630.

Erica, K. (2012). The timing of first antenatal care visit and factors associated with access to care among antenatal care attendees at chitungwiza municipal clinics,Zimbabwe. *Health, Public Health, Public Supervisor, Western Cape*, (May).

Fagbamigbe, L., Francis, A., Abel, C., Lepetu, L., & Mashabe, B. (2017). Application of parametric survival analysis to timing of first antenatal care visit in three African countries. *Proceedings of Research for a 7 Th International Conference, Abstact.*

Francis, A. (2017). Are the timings and risk factors changing ? Survival analysis of timing of first antenatal care visit among pregnant women in Nigeria (2003 – 2013). *International Journal of Women's Health*, *7*, 807–819.

Gage, A., & Calixte, M. (2006). Effects of physical accessibility of maternal health services on their use in rural Haiti. *Population Studies*, *60*, 271–88.

Gebremeske, F. (2014). Timing of first antenatal care booking and associated factors among pregnant women attending public health facilities in Arbaminch town and Arbaminch Zuria woreda, Gamo-Gofa zone, Southern Ethiopia.

Gedefaw, M., Muche, B., & Aychiluhem, M. (2014). Current Status of Antenatal Care

Utilization in the Context of Data Conflict : The Case of Dembecha District , Northwest Ethiopia. *Open Journal of Epidemiology*, *4*(November), 208–216.

Gezahegn, T., Deborah, L., Catherine, C., Agumasie, S., & Roger, S. (2017). Delayed initiation of antenatal care and associated factors in Ethiopia: a systematic review and meta-ana lysis. *BMC Reprod Health*, *14*, 150.

Gidey, Gebreamlak, H., Birhane, N., Kidane, H., Tesfay, & Woldegebriel, G. G. H. (2017). Timing of first focused antenatal care booking and associated factors among pregnant mothers who attend antenatal care in Central Zone , Tigray , Ethiopia. *BMC Research Notes*, *10*, 608. https://doi.org/10.1186/s13104-017-2938-5

Gilles, M. T. (2017). Timing of First Antenatal Care (ANC) and Inequalities in Early Initiation of ANC in Nepal. *Front. Public Health 5:242.*, *5*, 242. https://doi.org/10.3389/fpubh.2017.00242

Girum, T. (2016). Medicine Assessment of Timing of First Antenatal Care Visit and Associated Factors Among Pregnant Women Attending Antenatal Care in Dilla Town Governmental, *5*(3), 1–5. https://doi.org/10.4172/2327-5162.1000220

Greenwood, M., & Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 255–279.

Gross, karin Alba and Sandra, G., Schellenberg, T. R., Armstrong, J., & Obrist, B. (2012). Timing of antenatal care for adolescent and adult pregnant women in south-eastern Tanzania. *BMC Pregnancy and Children*, *12*, 1–12.

Gupta, S., Yamada, G., Mpembeni, R., Frumence, G., & Callaghan-koru, J. A. (2015). Factors Associated with Four or More Antenatal Care Visits and Its Decline among Pregnant Women in Tanzania between 1999 and 2010. *Journal of Gynecology and Obstetrics*, *9*(7). https://doi.org/10.1371/journal.pone.0101893

Gurmesa, T. (2014). Antenatal Care Service Utilization and Associated Factors in Metekel Zone , Northwest Ethiopia ASSOCIATED FACTORS IN METEKEL ZONE ,. *Ethiop J Health Sci*, *19*(2), 111–119.

Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. *Stata Journal*, *2*(1), 22–24.

Halim, N., Bohara, A., & Xiaomin, R. (2010). Healthy mothers, healthy children: does maternal demand for antenatal care matter for child health in Nepal? *Health Policy and Planning*, *26*, 242–56.

Hanna, G., & Berhane, Y. (2017). Timing of First Antenatal Care Visit and its Associated Factors among Pregnant Women Attending Public Health Facilities in Addis. *Ethiop J Health Sci*, *27*(1), 136. https://doi.org/doi: http://dx.doi.org/10.4314/ejhs.v27i2.6

Hougaard, P. (1984). Survival models for heterogeneous populations derived from stable distribution. *Biometrika*, *72*(2), 387–396.

Kabir, M., Iliyasu, Z., Abubakar, I., & Asani, A. (2005). Determinants of utilization of antenatal care services in Kumbotso Village, Northern Nigeria. *Tropical Doctor*, *35*.

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *American Statistical Association*, 53:457-81.

Kibaru, J. (2007). *Focused Antenatal Care*.

Klein & Moeschberger. (2003). Survival Analysis Techniques for Censored and Truncated Data. In *Google Books.htm*.

Klein, J. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics, 48,* pp.795-806.

Kleinbaum, D. ., & Klein, W. (2005). *Survival Analysis a self -learning tex* (2nd editio). New York: Springer Science.

Lambert, P., Collett, D., Kimber, A., & Johnson, R. (2004). "Parametric accelerated failure time models with random effects and anapplication to kidney transplant survival." *Statistics in Medicine*, *23*, 3177–3192.

Lancaster, T. (1979). *Econometric methods for the duration of unemployment. Econometrica 47, 939–95*.

Lincetto, O. and M., Gomez, S., & Munjanja, P. (2010). *Antenatal Care: Opportunities for Africa's Newborns*.

Lincetto, O., Mothebesoane-anoh, S., Gomez, P., & Munjanja, S. (2013). Antenatal Care: Opprotunities for Africa's Newborns. *International Journal of Scientific & Technology Research*, *2*(2), 51–62.

Locatelli, I., Lichtenstein, P., & Yashin, A. (2003). A Bayesian correlated frailty model applied to Swedish breast cancer data in MPIDR. *Institute for Demographic Research, Biometrica,* 25:3-21.

Million, W. D. (2018). Determinants and Heterogeneity of Time -to- Recovery from Obstetric Fistula Patients ; Comparison of Acceleration Failure Time and Parametric Shared Frailty Models. *Canadian Center Of Science and Education*, *7*(1), 99–114. https://doi.org/10.5539/ijsp.v7n1p99

Munda, M. (2012). Parametric Frailty Models in R. *American Statistical Association*, 55:1-21.

Nyambe, S., Lungowe, S., Choolwe, J., Patrick, M., & Michelo, C. (2016). Factors associated with late antenatal care booking: population based observations from the 2007 Zambia demographic and health survey. *Pan Afr Med J*, *8688*, 1–11. https://doi.org/10.11604/pamj.2016.25.109.6873

Ononokpono, D. N., & Odimegwu, C. O. (2014). Determinants of Maternal Health Care Utilization in Nigeria : a multilevel approach. *Pan Afr Med J*, *17*(Supp 1), 1–6. https://doi.org/10.11694/pamj.supp.2014.17.1.3596

Samia A., A., & Amani A., A. (2016). Gamma and inverse Gaussian frailty models: A comparative study. *International Journal of Mathematics and Statistics Invention (IJMSI)*, *4*(4), 4101-05. https://doi.org/10.9790/5728-1204051521

Sastry, N. (1997). A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survivalin Northeast Brazil. *Journal of the American Statistical Association.*, *92*, 426–435.

Shewa, W., Gela, B. D., Bedada, N. D., Jaleta, F. T., & Sinkie, S. O. (2014). Antenatal care utilization and associated factors from rural health extension workers in Abuna Gindeberet. *American J of Health Research*, *2*(4), 113–117. https://doi.org/10.11648/j.ajhr.20140204.12

Temesgen, W., Solomon, M., & Abdella, A. (2014). Timing and factors associated with first antenatal care booking among pregnant mothers in Gondar. *BMC Pregnancy and Childbirth*, *14*, 287.

Tessema, B., Ayalew, S., & Mohammed, K. (2015). Modeling the Determinants of Time-to-age at First Marriage in Ethiopian Women : A Comparison of Various Parametric Shared Frailty Models, *3*(5), 707–718. https://doi.org/10.11648/j.sjph.20150305.27

Trussell, J., & Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. *Sociological Methodology*, 15, 242–276.

Trussell, J., & Rodriguez, G. (1990). “Heterogeneity in demographic research." In Convergent issues in genetics and demography. *Oxford University Press, USA*.

Turyasima, M., Tugume, R., Openy, A., Ahairwomugisha, E., Opio, R., Ntunguka, M., … Odongo, P. (2015). Determinants of First Antenatal Care Visit by Pregnant Women at Community Based Education, Research and Service Sites in Northern Uganda. *East Afr Med J*, *91*(9), 317–322.

Umar, A. S. (2017). The Use of Maternal Health Services in Nigeria : Does Ethnicity and Religious Beliefs Matter ? *MOJ Public Health*, *6*(6). https://doi.org/10.15406/mojph.2017.06.00190

Vaupel. (1979). *The impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality*. *Demography 16*.

Vaupel, J., Manton, K., & Stallard, E. (1979). *The Impact ofHeterogeneity in Individual Frailty on the Dynamics of Mortality*.

Wang, W., Alva, S., Wang, S., & Fort, A. (2011). *Of Maternal Health Service In Devloping Countries DHS Comparative Reports*.

Weipan. (2001). Using frailties in the accelerated failure time. *Model. Life Time Data Analysis*, *7*, 55–64.

Were, F., Afrah, N. A., Chatio, S., Manda-, L., Pell, C., Men, A., … Pool, R. (2013). Factors Affecting Antenatal Care Attendance : Results from Qualitative Studies in Ghana , Kenya and Malawi, *8*(1). https://doi.org/10.1371/journal.pone.0053747

WHO. (2016). *Trends in Maternal Mortality: 1990 to 2015:Estimates by WHO, UNICEF, UNFPA, World Bank, and the United Nations Population Division*. Geneva.

WHO, UNICEF, UNFPA, & World Bank. (2010). *Trends in maternal mortality 1990–2008*.

Wienke, A. (2010). Frailty models in survival analysis. In *Chapman & Hall/CRC* (Vol. 37).

Wienke, A., Arbeev, K., Locatelli, I., & Yashin, A. . (2003). A Simulation Study of different correlated frailty models. *MPIDR WORKING PAPER*, *49*(0), 1–17.

World Bank Group. (2016). *Disease control priorities:Reproductive, Maternal,Newborn and Child Health* (3rd ed.). (B. Robert, E, L. Ramanan, T. Marleen, & W. Neff, Eds.).

World Health Organization. (2010). *Integrated Management of Pregnancy and Childbirth WHO Recommended Interventions for Improving Maternal and Newborn Health*.

Yaya, S., Bishwajit, G., Ekholuenetale, M., Shah, V., Kadio, B., & Udenigwe, O. (2017). Timing and adequate attendance of antenatal care visits among women in Ethiopia. *PLoS ONE*, *12*(9), e0184934.

APPENDEX

Appendex I

Table 1. Accelerated Failure Time Model for the assessment of time to first ANC visit in Ethiopia.

Variables	***Baseline Distribution***		
	Weibull	***Log-logistic***	***Log-normal***
	$\hat{\boldsymbol{\beta}}(\mathbf{95\%}\ \boldsymbol{CI}\ \boldsymbol{\phi})$	$\hat{\boldsymbol{\beta}}(\mathbf{95\%}\ \boldsymbol{CI}\ \boldsymbol{\phi})$	$\hat{\boldsymbol{\beta}}(\mathbf{95\%}\ \boldsymbol{CI}\ \boldsymbol{\phi})$
Place of residence			
Rural	Ref	Ref	Ref
Urban	-0.73(0.47, 0.5)	-0.699(0.48, 0.52)	-.7263(0.47, 0.503)
Woman Education			
No education	Ref	Ref	Ref
Primary	-0.46(0.61, 0.654)	-0.46(0.608, 0.66)	-0.434(0.622, 0.674)
Secondary	-0.75(0.48, 0.5)	-0.716(0.46, 0.52)	-0.711(0.463, 0.524)
Higher	-1.024(0.335, 0.385)	-0.933(0.364, 0.425)	-0.96(0.353, 0.42)
Distance to health care			
Not big Problem	Ref	Ref	Ref
Big problem	-0.42(0.63, 0.68)	-0.416(0.64, 0.684)	-0.4(0.644, 0.696)
Sex of household			
Male	Ref	Ref	Ref
Female	0.0157(0.97, 1.06)	-0.0106(0.944, 1.0367)	-0.0064(0.95, 1.04)
Wanted pregnancy			
Then	Ref	Ref	Ref
Later	-0.068(0.89, 0.984)	-0.049(0.9, 1.006)	-0.05(0.9, 1.005)
No more	0.226(1.155, 1.34)	0.26(1.19, 1.405)	0.26(0.835, 1.404)
Parity			
1	Ref	Ref	Ref
2-3	0.143(0.0957, 0.19)	0.15(0.1, 0.202)	0.155(0.103, 0.206)
4-5	0.34(0.29, 0.395)	0.36(0.304, 0.415)	0.36(0.305, 0.416)
>=6	0.39(1.40,1.55)	0.42(1.45,1.60)	0.403(1.424,1.573)
Husband education			
No education	Ref	Ref	Ref
Primary	-0.662(-0.72, -0.606)	-0.37(-0.415, -0.33)	-0.352(-0.395, -0.309)
Secondary	-77(-0.832, -0.71)	-0.66(-0.71, -0.6)	-0.635(-0.695, -0.574)
Higher	-0.39(-0.584, -0.1955)	-0.79(-0.56, -0.73)	-0.78(-0.85, -0.712)
Don't Know	-0.377(-.42, -0.33)	-0.464(-0.68, -0.25)	-0.455(-0.66, -0.25)
Religion			
Muslim	Ref	Ref	Ref
Orthodox	-0.36(-0.4, -0.32)	-0.332(-0.375, -0.291)	-0.321(-0.364, -0.28)
Catholic	-0.0065(-0.24, 0.225)	-0.036(-0.28, 0.204)	-0.027(-0.26, 0.204)
Protestant	-0.083(-0.133, -0.033)	-0.051(-0.103, 0.0006)	-0.0384(-0.09, 0.013)
Traditional/Other	0.53(0.33, 0.73)	0.553(0.372, 0.733)	0.497(0.326, 0.67)
Person who decided on respondent health care			

Respondent and Husband/Partner	Ref	Ref	Ref
Respondent alone	-0.009(-0.0585, 0.04)	0.0194(-0.072, 0.033)	-0.0198(-0.0724, 0.033)
Husband/Partner alone	0.246(0.194, 0.298)	0.261(0.208, 0.314)	0.243(0.191, 0.296)
Other	0.305(-0.132, 0.74)	0.282(-0.152, 0.72)	0.24(-0.18, 0.652)
Age			
15-19	Ref	Ref	Ref
20-24	0.00086(-0.087, 0.089)	-0.0176(-0.112, 0.0764)	-0.019(-0.113, 0.0742)
25-34	0.027(-0.056, 0.11)	0.0133(-0.0753, 0.102)	0.013(-0.075, 0.1)
35-49	0.23(0.138, 0.31)	0.211(0.12, 0.304)	0.22(0.13, 0.313)
Wealth			
Poorest	Ref	Ref	Ref
Middle	-0.37(-0.42, -0.322)	-0.37(-0.42, -0.32)	-0.019(-0.113, 0.0742)
Rich	-0.49(-0.53, -0.45)	-0.48(-0.524, -0.44)	0.013(-0.075, 0.1)

Table 2. Result of multivariate log-normal Gamma frailty model.

Covariate	Categories	Estimate($\widehat{\beta}$)	ϕ	95%CI SE(ϕ)	P-value
Place of residence	Rural	Ref			
	Urban	-0. 314	0.730	(0.692, 0.770)	<0.001
Woman education	No education	Ref			
	Primary	-0. 135	0.874	(0.835, 0.913)	<0.001
	Secondary	-0. 141	0.868	(0.805, 0.935)	<0.001
	Higher	-0. 197	0.821	(0.742, 0.907)	<0.001
Distance to get medical care	big problem	Ref			
	not big problem	-.111	0.895	(0.863, 0.910)	<0.001
Wanted pregnancy	Then	Ref			
	Later	.028	1.028	(0.813, 1.079)	0.258
	No more	.180	1.197	(1.111, 1.290)	<0.001
Parity	1	Ref			
	2-3	.081	1.018	(1.028, 1.142)	0.003
	4-5	.082	1.085	(1.016, 1.159)	0.015
	>=6	.171	1.186	(1.101, 1.277)	<0.001
Husband education	No education	Ref			
	Primary	-.126	0.881	(0.845, 0.919)	<0.001
	Secondary	-.171	0.843	(0.791, 0.898)	<0.001
	Higher	-.213	0.808	(0.747, 0.874)	<0.001
	Don't know	-.112	0.894	(0.743, 1.075)	0.234
	Muslim	Ref			
Religion	Orthodox	.123	1.131	(1.075, 1.189)	<0.001
	Catholic	.204	1.226	(0.990, 1.518)	0.062
	Protestant	.179	1.196	(1.126, 1.271)	<0.001
	Traditional/Other	.477	1.611	(1.384, 1.877)	<0.001
Person decided on respondent's health care	Respondent and Husband/partner	Ref			
	Respondent alone	-.009	0.991	(0.946, 1.036)	0.681
	Husband/Partner alone	.039	1.039	(0.994, 1.088)	0.092

	Other	.064	1.066	(0.746, 1.525)	0.725
	15-19	Ref			
Age group	20-24	.055	1.056	(0.966, 1.156)	0.223
	25-34	.019	1.019	(0.928, 1.119)	0.688
	35-49	.066	1.068	(0.961, 1.187)	0.216
Wealth	Poorest	Ref			
	Middle	-.122	0.885	(0.844, 0.928)	<0.001
	Richest	-.201	0.818	(0.783, 0.853)	<0.001

τ=0.072 $\theta = 0.156$ constant=1.793 *p < 0.05 was statistically significant. ϕ =Acceleration factor, θ =Variance of the random effect, τ= Kendall's tau, , CI=confidence interval, S. E=standard error, Ref=Reference, Likelihood-ratio test of theta=0: chibar2(01) =442.97 Prob>= chibar2=0.00

Table 3. Result of Multivariate Weibull gamma shared frailty model

Covariate	Categories	Estimate($\hat{\beta}$)	ϕ	95%CI SE(ϕ)	P-value
Place of residence	Rural	Ref			
	Urban	-0. 332	0.717	(0.682, 0.754)	<0.001
Woman education	No education	Ref			
	Primary	-0. 166	0.847	(0.813, 0.882)	<0.001
	Secondary	-0. 164	0.849	(0.792, 0.908)	0.002
	Higher	-0. 262	0.769	(0.704, 0.840)	0.002
Distance to get medical care	big problem	Ref			
	not big problem	-.127	0.881	(0.850, 0.913)	<0.001
Wanted pregnancy	Then	Ref			
	Later	.048	1.049	(1.002, 1.098)	0.039
	No more	.163	1.177	(1.092, 1.268)	<0.001
Parity	1	Ref			
	2-3	.089	1.093	(1.043, 1.147)	0.003
	4-5	.104	1.109	(1.046, 1.177)	0.001
	>=6	.197	1.218	(1.137, 1.304)	<0.001
Husband education	No education	Ref			
	Primary	-.154	0.8577	(0.823, 0.892)	<0.001
	Secondary	-.189	0.828	(0.780, 0.877)	<0.001
	Higher	-.201	0.818	(0.762, 0.877)	<0.001
	Don't know	-.123	0.884	(0.746, 1.047)	0.155
	Muslim	Ref			
Religion	Orthodox	.0608	1.062	(0.988, 1.116)	0.015
	Catholic	.162	1.175	(0.959, 1.441)	0.119
	Protestant	.161	1.175	(1.109, 1.245)	<0.001
	Traditional/Other	.497	1.643	(1.378, 1.963)	<0.001
Person decided on respondent's health care	Respondent and Husband/partner	Ref			
	Respondent alone	-.005	0.995	(0.953, 1.037)	0.73
	Husband/Partner alone	.033	1.033	(0.989, 1.081)	0.135
	Other	.156	1.169	(0.815, 1.676)	0.395
	15-19	Ref			

Age group	20-24	.016	1.016	(0.939, 1.101)	0.681
	25-34	-.037	0.964	(0.886, 1.047)	0.381
	35-49	.003	1.003	(0.912, 1.104)	0.944
Wealth	Poorest	Ref			
	Middle	-.128	0.879	(0.840, 0.920)	<0.001
	Richest	-.225	0.798	(0.766, 0.832)	<0.001

τ=0.057 θ = 0.121 constant=2.193 *p < 0.05 was statistically significant. ϕ =Acceleration factor, θ =Variance of the random effect, τ= Kendall's tau, , CI=confidence interval, S. E=standard error, Ref=Reference, Likelihood-ratio test of theta=0: chibar2(01) =351.37 Prob>= chibar2=0.00

Appendix II

Figure.1 KM plot for time-to-first ANC visit for parity and religion

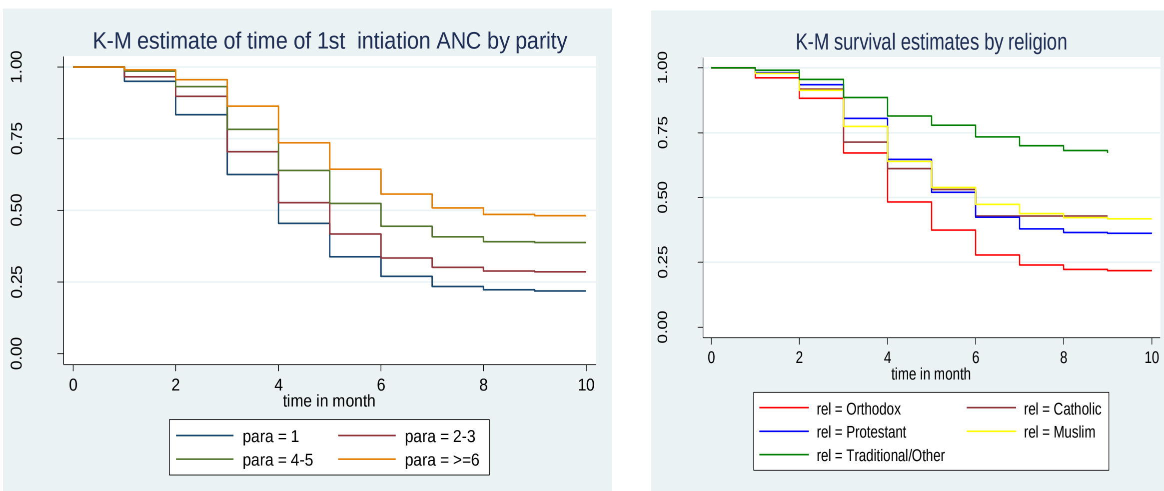


Figure.2 KM plot for time-to-first ANC visit for husband education and wanted pregnancy