UNIVERSITY OF GONDAR

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF STATISTICS

ENVIRONMENTAL DETERMINANTS OF CHILD MORTALITY IN ETHIOPIA: AN APPLICATION OF SURVIVAL ANALYSIS

M.Sc. THESIS

BY

YOHANNES TADESSE ASNAKEW

ADVISOR: SHIBRU TEMESGEN (Ph.D)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE, UNIVERSITY OF GONDAR IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

AUGUST, 2015

GONDAR, ETHIOPIA

APPROVAL SHEET-1

This is to certify that the thesis entitled **" ENVIRONMENTAL DETERMINANTS OF CHILD MORTALITY IN ETHIOPIA: AN APPLICATION OF SURVIVAL ANALYSIS"** submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics, Post Graduate Program, Gondar University, and is a record of original research carried out by **Yohannes Tadesse Asnakew Id.No. GUR/5625/06** under my supervision and no part of the thesis has been submitted for any other degree or diploma.

The assistance and the help received during the course of this investigation have been duly acknowledged. Therefore, we recommend that the thesis would be accepted as partial fulfillment of the requirement for Master of Science in Biostatistics.

Shibru Temesgen(PhD)	____________	______________
Main Advisor	**Signature**	**Date**
Kasim Mohammed (M.Sc.)	_____________	_____________
Co-Advisor	**Signature**	**Date**

APPROVAL SHEET-2

We, the undersigned, members of the board of examiners of the final open defense by **Yohannes Tadesse Asnakew** have read and evaluated his thesis entitled "**ENVIRONMENTAL DETERMINANTS OF CHILD MORTALITY IN ETHIOPIA: AN APPLICATION OF SURVIVAL ANALYSIS "** and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in statistics with specialization of Biostatistics.

Name of Chairperson	Signature	Date
Name of Main Advisor	Signature	Date
Name of Co-advisor	Signature	Date
Name of External Examiner	Signature	Date
Name of Internal Examiner	Signature	Date

Declarations

I declare that this thesis is my original work and that all sources of materials used for this thesis has been duly acknowledged. This thesis has been submitted in partial fulfillments of the requirements for MSc. Degree in Biostatistics at University of Gondar. I seriously declare that this thesis was not submitted to any other institution and anywhere for the award of any academic degree or diploma.

Name: **Yohannes Tadesse Asnakew**

Signature: __________

Date.21/08/2015

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES iv

ACRONYMS v

ACKNOWLEDGEMENTS vi

Abstract vii

Chapter One .... - 1 -

Introduction .... - 1 -

1.1 Background .... - 1 -

1.2 Statement of the Problem .... - 7 -

1.3 General Objective of the Study .... - 9 -

1.4 Hypotheses of the study .... - 9 -

1.5 Significance of the Study .... - 10 -

1.6 Limitations of the Study .... - 10 -

Chapter Two 11

Literature Review 11

2.1 Theoretical Literature 11

2.2 Empirical Literature 14

Chapter Three 20

Data and Methodology 20

3.1 Data Source 20

3.2 Definition of variables 20

3.2.1 The response (dependent) variable 20

3.2.2 Predictor (independent) variables 20

3.3 Methodology 22

3.3.1 Survival Analysis 22

3.3.2 Theoretical Model 22

3.3.3 Descriptive methods for survival data 23

3.3.4.1 Cox-Proportional Hazards Model 28

3.3.4.2 Fitting the Proportional Hazards Model 30

3.3.4.3 Variable Selection Procedures 34

3.4 Parametric Survival Regression Models 40

3.4.1 The Exponential Survival Regression Model 41

3.4.2 Fitting the Exponential Survival Regression Model 42
3.4.3 The Weibull Survival Regression Model 42
3.4.3 Fitting the Weibull Survival Regression Model 43
3.4.4 Model Selection in Parametric Survival Regression Models 44
CHAPTER FOUR 46
RESULTS AND DISCUSSION 46
4.1 Descriptive Analysis of the Survival Data 46
4.2 Results of Cox proportional hazard regression model 49
4.3 Model Adequacy 53
4.3.1 Overall Goodness of Fit 53
4.3.2 Testing the proportional hazards assumption 53
4.4 Interpretation and Discussion of the Results 55
4.5 Parametric Regression Modelling of Survival of Child 57
4.5.1 Model Selection for Survival of Child 57
4.5.2 Univariate unadjusted Weibull regression model 59
4.5.3 Multivariable Analysis Weibull Regression Model 60
4.5.4 Assessment of Adequacy of the Weibull Regression Model 61
4.5.5 Interpretation and Discussion of the Weibull Regression Model 61
4.5.6 Discussion of the Results 63
CHAPTER FIVE 65
CONCLUSIONS AND RECOMMENDATIONS 65
5.1 Conclusions 65
5.2 Recommendations 66
References 67
Appendixes 72

LIST OF TABLES

Table 3.1: Operational definition and categorization of the covariates, EDHS, 2014. 21

Table 4.3: Result of log-rank test of equality of survival distribution for the different categorical covariates..48

Table 4.5: The preliminary final model with Parameter estimates and hazard ratio of the covariates model...50

Table 4.6: Percentage changes in the coefficients of the variables included in Table 4.4, when the variables that were not significant in the univariable proportional hazards Cox regression models are added one at a time..52

Table 4.7: The Likelihood Ratio, Score and Wald tests for overall measures of goodness of fit of the preliminary final model in table 4.5..53

Table 4.8: Result of test of proportionality assumption for each covariate in the final model 54

Table 4.9: Statistical results for model comparison...58

Table 4.10:The result of un-adjusted univariate analysis using Weibull regression.........59

Table 4.11: Parameter estimates of the final multivariate weibull regression model........60

Table 4.12: The likelihood ratio and significance of the Weibull regression model.........61

Table 4.1 Category variables coding 72

Table 4.2: Summary of some important environmental characteristics of child Mortality in Ethiopia. 73

Table 4.4: Results of the univariable proportional hazards Cox regression model...........74

Table 4.13: Result of the Exponential regression model with corresponding AIC and BIC values...75

LIST OF FIGURES

Figure 2.1: Conceptual and theoretical framework for child mortality Mosely and Chen (1984)..12

Figure 4.1: Overall product limit estimate of survival function.......................................47

Figure 4.2 Nelson-Aalen Cumulative hazard estimate 47

Figure 4.3:The Cox Snell plot after fitting Exponential Regression model......................58

Figure 4.4:The Cox Snell plot after fitting Weibull Regression model............................58

Figure 4.5: Kaplan-Meier survival estimates of different covariates...............................76

Figure 4.6: The scaled Schoenfeld residuals and their lowess smooth plots of different covariates 77

Figure 4.7: -log(-log(survival probability)) plot..79

ACRONYMS

AIC	Akaike's information criterion
BIC	Bayesian Information Criterion
CI	Confidence Interval
CSA	Central Statistical Agency
DF	Degree of Freedom
EA	Enumeration Area
EAG	Empowered Action Group
EDHS	Ethiopia Demographic and Health Survey
GDP	Growth Domestic Product
HR	Hazard Ratio
KM	Kaplan-Meier
LL	Log Likelihood
LOWESS	Locally weighted scatter plot smoothing
LR	Likelihood Ratio
MDG	Millennium Development Goal
MPLE	Maximum partial likelihood estimator
NFHS	National Family and Health Survey
PH	Proportional Hazards
UNICEF	United Nations Children's Fund
UN-IGME	United Nations Inter-agency Group Child Mortality Estimation
USD	United States of America
USAID	United States Agency for International Development
WDI	World Development Indicator
WHO	World Health Organization

ACKNOWLEDGEMENTS

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study in the Department of Statistics, University of Gondar.

My special and heartfelt gratitude goes to **Dr. Shibru Temesgen**, my advisor and instructor, for his excellent scientific guidance, continual help, and tireless efforts to make this work a reality.

I would like to thank also my Co-Advisor **Kasim Mohammed** (MSc. in Biostatistics) who gave me all the necessary advices and valuable comments.

I am highly indebted to my wife Workie G/hiwot , she was the source of strength towards the successful completion of the study.

I acknowledge with appreciation the help rendered by all my dear friends who encouraged me to work hard.

Abstract

This paper focuses on the environmental determinants of child mortality in Ethiopia. The data for this study were obtained from the demographic and health survey conducted in 2014. It specifically examines how child mortality is related to the household's environmental characteristics, such as mother's education, source of drinking water, type of toilet used, type of cooking fuels, antenatal visit and place of delivery. A survival analysis was used to analyze the determinants of child mortality. As expected the Kaplan-Meier estimation show that most of the deaths occurred at first birth day of life. As the result of this we employed Cox proportional hazard and weibull regression models to select factors affecting child mortality in Ethiopia. According to the Cox proportional hazard and weibull regression models, mothers' education, source of drinking water, type of toilet used, antenatal visit, place of delivery and type of cooking fuel were found to have significant impact on child mortality in Ethiopia. Child's mother who had primary, secondary and above educational level were lower risk of mortality than mothers' who had no education level and children whose parents use non-improved source of drinking water have less survival chance than those who use improved source of drinking water. With regard to source of cooking fuel, children born in households using high polluting fuels (fire woods and charcoal) as their main source of cooking fuel have higher mortality rates as compared to those using low polluting fuels (electricity). Children born in household’s with either flush toilets or pit latrines have lower mortality rate than those born in households without any toilet facility. Policies aimed at achieving the goal of reduced child mortality should be directed on improving the household’s environmental status if this goal is to be realized.

Key Words: Child mortality , Kaplan-Meier estimator, Cox-PH model, weibull model.

Chapter One

Introduction

1.1 Background

Child mortality, commonly on the agenda of public health and international development agencies, has received renewed attention as a part of the United Nation's Millennium Development Goals. Approximately 6.3 million infants and children under five years of age die each year, with large variations in under-five mortality rates, across regions and countries (WHO, UNICEF, 2013).

Globally, the under-five mortality rates have declined from 85 per 1000 to 51 per 1000 (UNICEF 2012). However, it is estimated that more than 7 million children will die before attaining the age of five. Of these, India, Pakistan, Ethiopia, Nigeria and Democratic Republic of Congo will suffer half of all under-five children deaths (UNICEF 2008). India alone shares the burden of 24% of world's under-five mortality followed by Nigeria which shares 11% of this burden (UNICEF 2012). It is obvious that health policies in these five countries need to be reviewed and new imputes provided to bring down the high under-five mortality rate.

The world has made enormous progress in improving child survival since 1990, reducing the under-five mortality rate by nearly half from 90 to 46 deaths per 1,000 live births in 2013. Currently, the global under-five mortality rate is falling faster than at any other time over the past two decades. Yet, progress is insufficient to meet the Millennium Development Goal 4 (MDG 4) which calls for reducing the under-five mortality rate by two-thirds between 1990 and 2015. According to the report ,which examines trends in child mortality since 1990, analyses the main causes of under-five deaths, and highlights national and global efforts to save children's lives – the annual number of under-five deaths has fallen from 12.6 million in 1990 to 6.6 million in 2012.

Unger (2013) observed that areas of broad economic and social disadvantage (due to overcrowding, substandard housing, poor water and sanitation) tend to have higher under-five mortality compared to socially and economically advantaged areas. Becares et al., (2013) suggest that addressing neighborhood poverty and area deprivation is essential to improving health outcomes of individuals.

Environmental conditions are a major direct and indirect determinant of human health. In developing societies, modern forms of exposure to urban, industrial and agrochemical pollution add to the health burden caused by traditional household and community-based risks. The vicious cycle, intrinsically linking poverty, environmental degradation and ill health needs to be broken.

In most developing countries, especially in Sub- Saharan Africa (SSA), the basic child mortality causes of more than 80% of the diseases are inadequate and unsafe water supply, and improper disposal of waste(WHO 2010).

Child mortality varies among world regions but the highest prevalence is concentrated in Sub-Saharan Africa where mortality of children under five decreased from 177 in 1990 to 98 deaths per 1,000 live births in 2012 (UNICEF, 2013). Despite the overall decline in the prevalence of child mortality, it remains still at unacceptably high levels. About half of all deaths of children under five has been concentrated in Sub-Saharan Africa in 2012 (UNICEF, 2013). Hence, the need to reduce child mortality is one of the major challenges in improving child health, in particular in Sub-Saharan Africa.

Several single country studies based on micro data have shown the impact of individual's or household's endowments of resources (e.g. income, assets, land) as well as access to safe drinking water, food, energy, and improved sanitation on infant and child mortality (Kembo and Van Ginneken (2009) (Zimbabwe); Mesike and Mojekwu (2012) (Nigeria); Gemperli et al. (2004) (Mali); Nuwaha et al. (2011) (Uganda); Manda (1999) (Malawi); Kandala and Ghilagaber (2006) (Malawi); Adeyemi et al. (2008) (Nigeria); Adebayo and Fahrmeir (2012) (Nigeria); Ogunjuyigbe (2004) (Nigeria); Wang (2003) (Ethiopia). For example, using demographic and health survey (DHS) data from Nigeria, Fahrmeir (2012) shows strong positive impacts of socioeconomic and environmental factors on child survival. He also investigates the relative importance of socioeconomic endowments and environmental factors by the age of the child. While birth spacing and breastfeeding are found to be relatively more important for the survival probability during the period of infancy, socioeconomic variables and environmental factors such as access to safe water, improved sanitation, or indoor air pollution are relatively more important with increasing age of child. Similar results are found by Kyei (2012) for South Africa.

Similarly empirical research used aggregated macro data to study the determinants of child health outcomes. For example, Hmwe H. et al. (2013) show, based on a longitudinal study for 193 countries using annual data between 2000 and 2009 from the World Development Indicators, that GDP per capita, access to safe drinking water, improved sanitation, and public health expenditure per capita increases the probability of child survival.

The health effects of such environmental determinants were highlighted in the World Health Organization’s 2010 World Health Report (World Health Organization 2010), which showed that unsafe water, poor sanitation, and hygiene are the cause of 4%–8% of the overall burden of diseases in developing countries and nine-tenths of diarrheal diseases, which is a major contributor to infant mortality.

According to World Bank (2013), environmental health risks fall into two broad categories. The first are the traditional hazards related to poverty and lack of development, such as lack of safe water, inadequate sanitation and waste disposal, indoor air pollution, and vector-borne diseases. The second category is the modern hazards such as rural air pollution and exposure to agro industrial chemicals and wastes that are caused by development that lacks environmental safeguards.

Unsafe water and sanitation, indoor air pollution from household solid fuel use, and ambient urban particulate matter (PM) pollution are responsible for an estimated 3.4%, 2.7%, and 0.6% of the global burden of disease, respectively, with 90%, 71%, and 7% of the disease burden from these risk factors borne by infants and young children in low- and middle-income countries.

As the world enters into the 21st century, debate on childhood mortality remains a big issue for developing countries. Their commitment is reflected in their desire to reduce the level of child mortality by two- thirds of their 1990 levels by the year 2015, as expressed in the Millennium Development Goals. To achieve this goal, it is imperative to attempt and determine what factors contribute to the high levels of child mortality in developing countries and in particular, Ethiopia.

Although enormous literature exists on child mortality, evidence on why infant and child mortality rates remain high in many sub-Saharan African countries despite action plans and interventions made is still insufficient. Environmental risk factors account for about one-fifth of the total burden of disease in low income countries according to recent estimates (World Bank, 2010). WHO (2012) reports that among the 10 identified leading mortality risks in high-mortality developing countries, unsafe water, sanitation and hygiene ranked second, while indoor smoke from solid fuels ranked fourth. About 3% of these deaths (1.7 million) are attributable to environmental risk factors and child deaths account for about 90% of the total.

Worldwide, safe and adequate drinking water is still not accessible to 1.1 billion people, and 2.4 billion people lack adequate sanitation. The recent figures for Ethiopia (2010) indicate a water supply coverage of 38% (98% in urban areas and 26% in rural areas), and a sanitation coverage of 15% (58% in urban areas and 8% in rural areas). Unchecked urban growth has its price in terms of environmental health: disposal of municipal and hazardous waste, particularly health care waste, remains a problem in many regions. Up to 60% of the global burden of Acute Respiratory Infection (ARI) is associated with indoor air pollution and other environmental factors. Occupational diseases and injuries, grossly underreported are responsible for more than 1 million deaths annually all over the world; health care workers, miners and manufacturing workers are at highest risk.

Ethiopia has one of the lowest rates of coverage for improved water and sanitation in the world. Just over 54 per cent of households have access to an improved source of drinking water, with a higher proportion among urban households (75%) and among rural households (49%). According to Joint Monitoring Program (JMP) 2012 update, the proportion of the population having access to improved and unimproved sanitation facilities stands at 54 % (21% improved and 33 % unimproved).

- Nearly 39 million Ethiopians – most of them in rural areas don’t have access to safe water.

- Nearly 48 million lack access to basic sanitation.

Ethiopia is the second most populous country in Africa after Nigeria with a population of nearly 85 million in 2010 (World Bank, 2013). The population grows at a rate of 2.6 percent per annum which is slightly greater than the sub-Saharan African countries average growth of 2.5 percent and the majority of people (84%) reside in rural areas, with agriculture being the major source of livelihood. The age structures suggest nearly 45 percent of the populations are under age 15 and the percentages of the population above age 65 are only about 3.2 percent. High mortality, high fertility and low life expectancy characterize the demography, as in most sub-Saharan African countries (Ringheim *et al*,2009).

In Ethiopia, results from the 2011 EDHS data showed that a remarkable decline in all levels of childhood mortality. The same report showed that infant mortality has declined by 42 percent over the 15-year period preceding the survey from 101 deaths per 1,000 live births to 59 deaths per 1,000 live births. Furthermore, under-five mortality has declined by 47 percent over the same period from 166 deaths per 1,000 live births to 88 deaths per 1,000 live births. Even though not to the same extent, the neonatal mortality has also decreased over the 15-year period preceding the survey by 31 percent from 54 deaths per 1,000 live births to 37 deaths per 1,000 live births. This reduction in neonatal mortality, as in other parts of the world, was slower than for infant, and under-five mortality, which fell by 42 percent and 47 percent respectively over the 15 year period (EDHS Report 2011). In addition, the country is experiencing a high neonatal mortality rate at 37 per 1000 live births, comparable to the average rate of 35.9 per 1000 live births for the African region overall (Oestergaard *et.al*, 2011).

The Ethiopian situation is similar with that of the Sub Saharan Africa which is characterized by high infant mortality rate; it also ranks 6th in the world by total number of death of infants. Infant and child mortality in Ethiopia had shown a continuous decline since 1960 onwards with a more pronounced reduction in the recent decades. The trend of infant mortality rates has been about 200 per 1000 live births in 1960, 153 per 1000 live births in 1970, 110 per 1000 live births in 1984, 97 per 1000 live births in 2000 and 77 per 1000 live births in 2005, 59 per 1000 live birth in 2011. This means that infant mortality declined by 20.6% and 23% between 2000 to 2005 and 2006 to 2011 respectively . This decline may be attributed to expansion of Health Extension program, high coverage of immunization, community based intervention, and rapid expansion of health facilities. According to the 2011 Ethiopian Demographic Health Survey (EDHS) there were 59 deaths per 1000 live births. But it contributes to 67% of the under-five children mortality.

The statistics, contained in a 2013 progress report, Committing to Child Survival: A Promise Renewed, compiled by the UN children's fund UNICEF, the World Health Organization (WHO), and the World Bank Group, showed Ethiopia has reduced child deaths by more than two thirds over the past 20 years.

Government commitment and resources have contributed to Ethiopia's progress on the issue. "The government has set some very bold and extremely ambitious targets. It has then backed them up with real resources and real commitment sustained over the last 10 years," said Dr Peter Salama, UNICEF country representative for Ethiopia, pointing to the country's health extension program. "The program put on the government payroll more than 36,000 health workers and deployed them to more than 15,000 health posts across Ethiopia . That is the single most important reason why Ethiopia has reduced its under-five mortality rate."

There is limited research conducted on child mortality in Ethiopia. Most of the information for any program planning and implementation has been based on Ethiopian Demographic and Health Survey (EDHS) conducted every five years. EDHS describes only the rate of mortality and does not provide information on the causes of death distribution, and health interventions differ from older children. This study focused on the determinants and risk factors associated with child mortality in Ethiopia. particularly interested in how child survival is affected by environmental factors.

Poverty is one of the most important factors affecting the infant mortality rate in Africa. Ethiopia is one of the poorest African countries with, according to UNICEF (2013) report, with a Gross National Income per capita of about USD 530. The impact of poverty on the health of children is due to lack of access to a variety of material and nonmaterial resources, as well as environmental and psychological deprivation at cultural, social and health levels. Low socio-economic position has been found to be associated with low birth weight and increased neonatal mortality (Bradley and Corwyn 2012).

While medical interventions can in principle prevent most early child deaths, they cannot eliminate the underlying causes of poor health, which are linked directly to those severely deprived or 30 percent of the world's children living in absolute poor conditions (UNICEF 2010). Eliminating extreme poverty is the key to improving global child survival rates, particularly over the long term.

Environmental conditions , in particular, a safe source of drinking water , appear to be important determinants of infant mortality risks in both urban and rural locations. In the latter, the very few households with an electricity supply have a greatly reduced probability of infant death. In urban areas, the mortality risk is substantially higher among households living in premises with no finished floor. It seems likely that this characteristic identifies slum dwellings and the poor public health conditions found there. In rural areas, the majority of dwellings have no finished floor, and this is not significantly correlated with mortality risk. Surprisingly, having a toilet is not significantly correlated with mortality risk in either urban or rural areas. Children in households with fewer assets face a greater risk of death in urban but not in rural areas. This is consistent with a greater socio economic gradient in child health in urban areas that has been found in other studies (Kyei,2012).

1.2 Statement of the Problem

One of the targets of the Millennium Development Goals (MDGs) is to reduce the under-five mortality rate by two-thirds between 1990 and 2015. Since 1990 the under-five mortality rate has dropped 35 percent, with every developing region seeing at least a 30 percent reduction. However, at the global level progress is behind schedule, and the target is at risk of being missed by 2015. The global under-five mortality rate needs to be half from 57 deaths per 1,000 live births to 29 that imply an average rate of reduction of 13.5 percent a year, much higher than the 2.2 percent a year (UN-IGME, 2011).

The environment, which sustains human life, is also a profound source of ill health for many of the world's people. In the least developed countries, one in five children do not live to see their fifth birthday, mostly because of avoidable environmental threats to health. This translates into approximately 6.3 million avoidable childhood deaths each year (UNICEF 2013). Hundreds of millions of others, both children and adults, suffer ill health and disability that undermine their quality of life and hopes for the future. These environmental health threats, arguably the most serious environmental health threats facing the world's population today, stem mostly from traditional problems long since solved in the wealthier countries, such as a lack of clean water, sanitation, adequate housing, and improved toilet.

Poverty also influences health because it largely determines an individual's environmental risks, as well as access to resources to deal with those risks. Throughout the developing world, the greatest environmental health threats tend to be those closest to home. Many in these countries live in situations that expose their health through steady exposure to biological pathogens in the immediate environment. More than 1 billion people in developing countries live without adequate shelter or in unacceptable housing. Further about 1.4 billion lack access to safe water, while another 2.9 billion people have no access to adequate sanitation (WDI, 2010), all of which are essential for good hygiene. Unable to afford clean fuels, the poor largely rely on biomass fuels for cooking and heating. Inside the smoky dwellings of developing countries, air pollution is often higher than it is outdoors in the world's most crowded cities.

The study of child mortality becomes one of the most important researches in developing countries like Ethiopia, because there is high level of child mortality. There is little research on the patterns of environmental determinants of child mortality, by analyzing how child mortality is differently affected by environmental factors. This paper presents an analysis of the impact of environmental variables on child mortality. The data used in this study were obtained from the Demographic and Health Survey conducted in Ethiopia in 2014. The overall purpose of the paper was to determine the relative importance of various environmental factors on child mortality in Ethiopia. In particular, the study should focus on the relationship between child mortality and sex of child, mother's educational status, area of residence, place of delivery, source of drinking water , type of cooking fuel, antenatal visit, and toilet type used.

1.3 General Objective of the Study

The general aim of the study was to explore the environmental factors on child mortality in Ethiopia by using survival analysis using Non-parametric, parametric and semi-parametric methods.

The specific objectives are:

- To assess the relationship between the environment and child mortality in Ethiopia.
- To identify the environmental determinants of child mortality, controlling for other covariates.

1.4 Hypotheses of the study

In order to meet the above objectives, the following hypotheses were tested:

- Household's access to safe water has no effect on child mortality
- Children born in households with sanitation facilities are more likely to die than those in households without.
- The household's main source of cooking fuel has no effect on child mortality.

1.5 Significance of the Study

- It is hoped that the study would provide an in-depth use of Demographic and Health Survey data (EDHS 2014). It is expected to improve the understanding of the child mortality situation in Ethiopia.
- The results could be of interest to other studies related to child mortality environmental risks in Ethiopia.
- The result of this study could provide information to government and other concerned bodies in setting policies, strategies, and further investigation for reducing child mortality.

1.6 Limitations of the Study

- The estimates of child mortality are based on retrospective birth histories which are subject to possible reporting errors that may affect the quality of the data. A lack of accurate information on the age at death may distort the age pattern of mortality.
- Only surviving women age 15-49 were interviewed; therefore, no data were available for children of women who had died.
- Underreporting of events- respondents are likely to forget events that occurred in the past.

Chapter Two

Literature Review

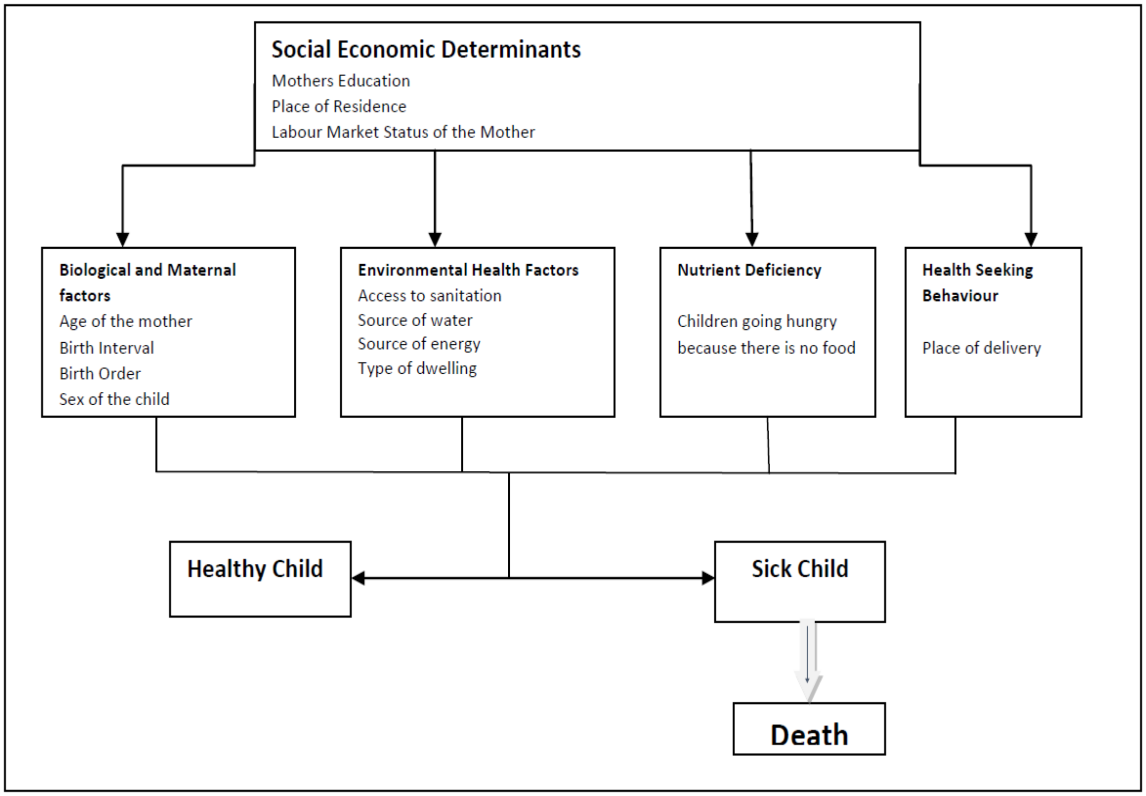
2.1 Theoretical Literature

Researchers used a number of different conceptual frameworks to analyze the impact of different factors on child survivals. Among these Mosley and Chen (1984) and Schultz (1984), classified the determinants of infant and child mortality as exogenous (socioeconomic or extrinsic) such as cultural, socioeconomic, community and regional determinants and endogenous (bio-medical or intrinsic) such as maternal, environmental, nutrition, injuries and personal illness. environmental factors affect indirectly infant and child mortality, they operate through the proximate factors while proximate determinates affect infant and child mortality directly (Mosley and Chen, 1984; Schultz, 1984).

Accordingly, Adebayo, and Samson, (2013) defined child mortality as the likelihood for a child born alive to die between its first and fifth birthday. Desta (2011) described infant mortality as the probability of dying between birth and the first birth day, while, Child mortality is the probability of dying between the first and the fifth birth day.

Mosley and Chen (1984) set the framework of child survival based on the assumption of all socioeconomic factors of child mortality necessarily operate through a common set of intermediate factors, they identify clearly the proximate and socioeconomic determinants of infant and child mortality and they categorized fourteen proximate determinants of infant and child mortality into five general groups based on some reasons; in an optimal setting, over 97 percent of children born can be expected to survive until the fifth birthday, proximate determinants through the socioeconomic factors operate to influence the infant.

Figure 2.1: Conceptual and theoretical framework for child mortality Mosely and Chen (1984)



Source: Based on Mosely and Chen (1984) and Desta (2011) theoretical framework and child mortality and socioeconomic, biological and environmental factors are the driving forces behind the reduction of infant and child mortality. Given these assumptions, we present the theoretical framework graphically above:

Several studies on infant and child mortality have been carried out using census and survey data. Most of these studies have estimated child mortality using indirect methods such as Trussel's technique and Preston method (Mojekwu & Ajijola, 2011; Jada,).

Antai et al. (2010) employed the multilevel logistic regression while Doctor (2011) uses multivariate logistic regression. All these studies find demographic, socio-economic and environmental factors (source of drinking water, sanitation facilities) to be significantly related to infant and child mortality.

A study on inequalities in child mortality in ten major African cities by Wilm Quentin et al., (2014), and being the first study to systematically investigate socio-economic inequalities in child mortality within and across African cities and their development over time, found out that in most cities, child mortality is considerably higher among the poor than among the rich, with differences between the poorest quintile and the richest quintile reaching as much as 108 deaths per 1000 live births in Abidjan in 2011 to 2012. And that around the year 2000, Dar-es-Salaam had the highest level of inequality while Abidjan and Cairo had rather low absolute (Cairo) and relative (Abidjan) inequality. However, the study did not investigate the underlying reasons for the identified differences in inequalities across the cities, a weakness that this research tries to unravel.

In another study on association of urban slum residency with infant mortality and child stunting in low and middle income countries by Hmwe et al., (2013), found out that living in a slum neighborhood was associated with infant mortality irrespective of individual and household characteristics and this association was consistent across countries. This study adds to concerns raised by Timaeus and Lush about the harmful effects of poor environmental conditions on child health. Nevertheless, this study failed to examine the association longitudinally to establish causal relations. It concluded that living in a slum neighborhood was associated with infant mortality irrespective of the socio-economic status and other characteristics of households and families and further showed that the risk of stunting in slum neighborhoods was greater for older children.

Espo (2002) in his study, used indirect methods to estimate levels and trends of mortality in Malawi. The results indicate that source of drinking water and sanitation facilities are strong predictors of child mortality. Also, Folasade (2010) in her study to determine the relative significance of environmental and maternal factors on childhood mortality in southwestern Nigeria find that child mortality rate continued to be a function of an environmental factor namely source of drinking water and a child care behavior factor, where the child was kept when mother was at work.

Similarly, Hmwe et al. (2013), in a comparative study of rural areas of Ghana, Egypt, Thailand and Brazil, discovered that children’s health is affected by environmental conditions and the economic status of the household. Nuwaha et al. (2011) utilized duration modeling to assess the impacts of water and sanitation on child mortality in Egypt. Though sanitation is found to have more pronounced impact than water, the results also show that access to municipal water reduces the risk of mortality.

A Bayesian geoadditive survival model was introduced by Adebayo et al. (2002) to analyze child mortality in Nigeria. The results showed that the existence of a district-specific geographical variation in the level of child mortality.

Klaauw (2003) developed a flexible parametric hazard rate framework for analyzing child mortality. Their model predicts significant correlation between child mortality and access to electricity, provision of sanitation facilities, improving maternal education and reducing indoor air pollution.

Jacoby (2003) in a related study, examined the linkages between child mortality, morbidity, and household quality and community environment in rural China using a competing risks approach. Their findings among others show that the use of clean cooking fuels, access to safe water and sanitation reduces the risks of child mortality.

2.2 Empirical Literature

Empirically, many studies have shown that child mortality is influenced by a number of socio economic and demographic factors such as sex of the child, mother‘s age at first birth, birth order, preceding birth interval among others (Adebayo, Samson, 2013). However, Adeyemi et al. (2010) gesticulates that the cause of disease and death over which not much controversies and uncertainties exist is the total environment of man. Malaria, acute respiratory infections, measles, and diarrhea which are today major causes of mortality for children under five are consequence of the built environment of man. In developing countries like Nigeria, one in eight children does not live to see their fifth birthday due to avoidable environmental threats, resulting into approximately 6.3 million avoidable childhood deaths yearly (World Bank, 2011). According to World Bank (2010), environmental risk factors were estimated to account for about one-fifth of the total burden of disease in low income countries. The WHO (2012) similarly, reported that among the ten identified leading mortality risks in high mortality developing countries, unsafe water, sanitation and hygiene ranked second while smoke from solid fuels ranked fourth.

Figures published by the United Nations Statistics Division (2012:1–4) indicated that child mortality and morbidity could not be reduced significantly in most sub-Saharan African countries as a consequence of severe economic crisis, lack of economic and political stability, and the inability of national governments to make the necessary resources and infrastructure available to the rural population. The authors argue that recent mortality rate trends of children less than five years could be substantially reduced if governments were to demonstrate political commitment to meet the basic needs of children and mothers. Sub-Saharan Africa is the region most affected by poverty, which leads to child mortality and accounts for more than one-third of all deaths of children younger than five years. Numerous studies have shown a close association between child mortality and poor environment status. Examples of socio-economic factors that adversely affect the survival of children in South Africa, amongst others, (United Nations Statistics Division 2012:2–5).

Kumar and File (2010) used data from the Ethiopia Demographic and Health Survey [EDHS] conducted in 2005 to investigate the predictors‘ of child [0-5 years] mortality in Ethiopia. The cross tabulation technique was used to estimate the predictors of child mortality. Results revealed that birth interval with previous child and mother‘s standard of living index were the vital factors associated with child mortality. Furthermore, Mother‘s education and birth order were found to have substantial impact on child mortality in Ethiopia. The study concluded that an increase in Mothers‘ education and improved health care services are significant in reducing child mortality in Ethiopia.

Thai.et al. (2010) employed was logistic regression to examine the effect of some environmental and economic factors that determine childhood mortality in Eritrea, using data from the 2005 Eritrea Demographic and Health Survey (EDHS). The results show that type of floor material, household economic status and place of residence are significant predictors of child mortality.

Mesike and Mojekwu (2012) in their study examined the environmental determinants of child mortality in Nigeria using principal component analysis and simultaneous multiple regression for child mortality modeling in Nigeria. Estimation from the stepwise regression model showed that household environmental characteristics do have significant impact on mortality as lower mortality rates were experienced in households that had access to immunization, sanitation facilities, good and proper refuse and solid waste disposal facilities, good healthy roofing and flooring materials as well as those using low polluting fuels as their main source of cooking.

Duration modeling is applied by Hala (2002) to assess water and sanitation's impacts on child mortality in Egypt. The results showed that access to municipal water decreases the risk and sanitation is found to have a more pronounced impact on mortality than water.

The meta-study by Blunch et al. (2010) suggests that, although there can be little doubt that household income is a crucial factor in determining child health, it appears that income is not a significant determinant of infant mortality in the majority of cases. This can partly be explained by the fact that as mortality falls, the bulk of under-five-mortality is rather those of infants than child death, and these deaths are more sensitive to health provision than socio-economic conditions.

Ikamari, L.D.E.(2013) find out that demographic factors are more important in explaining infant (under 12 months) mortality, socioeconomic, socio-cultural and hygienic factors are more important in explaining child (under five) mortality. Younger (2007), however, do not find significant effects of variables related to the quality of drinking water and sanitation on infant mortality.

Hmwe H. et al. (2013). evaluate empirically the Solow model with human capital, the model was estimated through a panel data analysis, which includes the growth rates of physical capital, labor, schooling and health indices. The heath index includes four determinants of health; lifestyles, environment, health services and socio-economic conditions. It was observed that variables were all significant showing the impact health has on economic growth. It was observed that among the determinants of health considered, health service result became the most significant. They concluded that a higher awareness of the health of the people is necessary if sustainable growth is pursued especially for the third world for policy implications.

Blunch,Niels-Hugo (2013) used data from 2003 and 2008, DHS surveys in Ghana to examine the determinants of infant and child mortality in three northern regions by using multivariate logistic regression model found that education of mothers, birth order of child and marital status of mothers are powerful significant determinants for infant mortality, while only mothers education have a significant impact for child mortality.

Similarly, Jinadu et al. (2010), in a study, found dirty feeding bottles and utensils, inadequate disposal of household refuse and poor storage of drinking water to be significantly related to the high incidence of diarrhea.

Twum et al. (2011) using the result of 2009 Burkina Faso DHS , indicated that children born to mothers with higher educational level associated with lower risk of infant and child mortality as compared to children born to mothers with primary education level or non-educated.

Kombo and Ginneken (2010) using the result of 2005-2006 Zimbabwean DHS investigate the maternal, socioeconomic and sanitation factors on infant and child mortality by using Cox regression model. They found an evidence of birth order (6+) with short preceding interval significantly associated with high risk of infant and child mortality. Multiple births tend to increase infant and child mortality. On the other hand the expected U shape relationship between birth order and infant and child mortality, and mothers age and infant and child mortality is not conformed in their analysis, that children who are first born and those born to mothers aged 40-49 years are found tend to decrease infant and child mortality. However socioeconomic determinants are rather small and insignificant effect on infant and child mortality. They suggest that the influence of birth order, preceding birth intervals, maternal age, type of birth and sanitation factors are more pronounced on infant mortality while weak effect on child mortality.

Kimani R.R and Kimani E R. (2012) in their study in Kenya, for children by using logistic regression models. They examined socioeconomic determinants of infant mortality rate both urban and rural setting. They found similar result like in the case of Ghana above that regional variation exists in infant and child mortality between the different provinces of Kenya. Most of the socioeconomic factors are not associated with the risk of infant and child mortality while children born in the richest household has lower probability of infant mortality relative to children born in the poorest households. However ethnicity and breast feeding in both urban and rural areas have a significant influence on infant mortality and sex of the child in urban areas and birth order and birth interval in rural areas are important determinants for the risk of infant mortality. Although they found that the incidences of HIV/AIDS in both urban rural areas increase the risk of dying at infancy period.

In addition many studies have documented role of individual or household socio-economic and demographic factors associated with rising levels of child mortality in Kenya. These factors include maternal education, income or well-being, place of residence, breastfeeding, water and sanitation, and access to and utilization of health (Ombok M et al. 2010; Ikamari, L.D.E. 2013;).

Wafula S.W., et al.(2012) used data from 2008 DHS in Kenya to investigate the impact of socioeconomic and environmental variables of infant and child mortality in urban areas of Kenya. The results show that the infant and child mortality were lower for those who were of birth order 2-3, birth interval more than 2 years, single births, living in wealthier households, had a access to drinking water and sanitation facilities, and users of low polluting fuels as their main source of cooking. However, maternal age, maternal education and gender of the child had no significant association with child mortality. Other study in Kenya by Hill (2011) found that mother‘s educational levels and economic status have a significant impact on infant and child mortality while urban areas are associate with high risk of infant and child mortality than rural areas, however, controlling for HIV prevalence child mortality are lower in urban areas.

The hazard rate framework is elegantly utilized by Klaauw (2003), in which a flexible parametric framework for analyzing infant and child mortality is developed. Their model predicts that a significant number of under 5 years deaths can be averted by providing access to electricity, improving the education of women, providing sanitation facilities and reducing indoor air pollution. In particular, reducing indoor air pollution and increasing the educational level of women might have substantial impacts on child mortality. In a related study, Jacoby (2003) examine the linkages between child mortality and morbidity, and the quality of the household and community environment in rural China using a competing risks approach. The key findings are that (1) the use of unclean cooking fuels (wood and coal) significantly reduces the neonatal survival probability in rural areas; (2) access to safe water or sanitation reduces child mortality risks by about 34% in rural areas; (3) a higher maternal education level reduces child mortality and that female education has strong health externalities (4) access to safe water/sanitation, and immunization reduce diarrhea incidence in rural areas, while access to modern sanitation facilities (flush toilets) reduces diarrhea prevalence in rural areas; (5) significant linkages between Acute Respiratory Infections (ARI) incidence and use of unclean cooking fuels are found using the city level data constructed from the survey.

UN (2011) Sub-Saharan Africa countries: The study has shown that under-five mortality is affected by the practice of breastfeeding, the mother's marital status, and the mother's level of education, ownership of flush toilet facilities, the residential area and place of delivery of the child. Based on results obtained from Pearson's Chi-square tests of association, children who died are characterized by a low immunization-coverage rate (P = 0.0000), poor nutrition (P = 0.0000), poor access to tap water (P = 0.0000), no access to flush toilets (P = 0.0108) and the children belonged to rural and unemployed mothers who did not attend antenatal and postnatal health care services (P = 0.0013). This indicates that rural mothers and children are relatively disadvantaged with regard to basic health services in comparison with urban mothers and children.

Kamal (2012) investigated the effect of maternal education on neonatal mortality in Bangladesh using data from the 2007 Bangladesh Demographic and Health Survey. Both bivariate and multivariate statistical analyses were used to assess the relationship between neonatal mortality and contextual factors focusing on maternal education. The results revealed that the sequential multivariate logistic regression analyses yielded a strong significant negative association between maternal education and neonatal mortality.

Chapter Three

Data and Methodology

3.1 Data Source

The source of the data used in this study was the 2014 Ethiopia Demographic and Health survey (EDHS, 2014) conducted in Ethiopia as part of the worldwide demographic and health survey project. The 2014 Ethiopia Demographic and Health Survey were conducted by the Central Statistical Agency (CSA) with the support of the Ministry of Health. This was the fourth Demographic and Health Survey (DHS) conducted in Ethiopia, under the worldwide MEASURE DHS project, a USAID-funded project providing support and technical assistance in the implementation of population and health surveys in countries worldwide.

The primary objectives of the 2014 EDHS were to provide up-to-date information for planning, policy formulation, monitoring, and evaluation of population and health programs in the country. The survey was intentionally planned to be responded at the beginning of the last term of the MDG reporting period to provide data for the assessment of the Millennium Development Goals (MDGs).

Information on child mortality was found from the birth history of women who were included in the survey. Since the interest of this study is about children from age one until age five.

3.2 Definition of variables

3.2.1 The response (dependent) variable

The variables used in the estimations are defined in this section. The time is dependent variable and is defined as the time that a child who has survived to the beginning of the respective interval (12 months-59 months) will fail (die) in that interval.

3.2.2 Predictor (independent) variables

The **explanatory variables** are classified into three groups: environmental, socioeconomic and demographic. The choice of these variables was guided by the determinants of child mortality literature. The main focus of this study was on the environmental variables only.

Operational Definition of Variables and Concepts

1. Child mortality – refers to deaths among children aged between exact age one and under five years (12-59 months).
2. Infant mortality – refers to deaths that occur to children who were born alive between the time of birth and die before they celebrate their first birth day.
3. Maternal education level- refers to the highest level of formal schooling attained by the mother and recognizes no education, primary education and secondary level and above.
4. Source of drinking water- refers to the main source of water for use in the household.
5. Type of toilet facility- refers to the type of facility used to dispose human waste.
6. Place of residence-This variable indicates where the household is located, either in urban areas (cities, towns) or in rural areas. In addition, we assume that the various effects might differ across regions.

Table 3.1: Operational definition and categorization of the covariate variables, EDHS, 2014.

Variables	Definition and Categorization
Sex of child	Sex of child(1=Male,2=Female)
Residence	Place of residence for women(1=Rural;2=Urban)
Women education	Women level of education (1= No education;2= Primary; 3= Secondary and Higher)
Antenatal visit	Antenatal visit during pregnancy(0=No,1=Yes)
House hold access water	Source of drinking water (1= improved source; 2= Non-Improved source ; 3=other source;).
House hold access toilet	House hold toilet type (1=Improved, not shared;2= shared facility ; 3= Non-Improved)
Cooking fuel	Type of Cooking fuel(1= fire hood ; 2= charcoal; 3=kerosene; 4=electricity)
Place of Delivery	Mothers delivery place(1=Home; 2=Health center;3=Others

3.3 Methodology

3.3.1 Survival Analysis

Survival Analysis involves the modeling and analysis of data that have a principal end point, the time until an event occurs (time-to-event data). Generally, survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. Survival data analysis involves a dependent variable, time-to-an event, which is always nonnegative and has a positively skewed distribution.

3.3.2 Theoretical Model

In survival analysis, we usually refer to the time variable as survival time, because it gives the time that an individual has 'survived' over some follow-up period.

We also use the term `failure' to define the occurrence of the event of interest (even though the event may actually be a `success' such as recovery from therapy) (Kleinbaum and Klein, 2005).

Survival analysis is different from the other statistical procedures due to following reasons:

1. In survival analysis, the response variable is time.
2. Staggered entries are more common in medical research. By staggered entries we mean that all individuals in the study do not have the same entrance time. This does not affect the survival analysis, as the analysis deals with the length of the observation time and not based on the same entrance.
3. The assumption of normality does not hold in survival analysis, as survival data are generally skewed.
4. The covariates can be time dependent.

One of the most important differences between the outcome variables modeled via linear and logistic regression analyses and the time variable in the survival data is the fact that we may only observe the survival time partially. The variable time actually records two different things. For those subjects who died, it is the outcome variable of interest, the actual survival time. However, for subjects who were alive at the end of the study, or for subjects who were lost to follow-up, time indicates the length of follow-up (which is a partial or incomplete observation of survival time). These incomplete observations are referred to as being censored. Censoring may occur when a person does not experience the event before the study ends, a person is lost to follow-up during the study period, and a person withdraws from the study.

There are three common forms of censoring:

a. **Right Censoring**: The most common form of incomplete data is right censoring. A survival time is said to be right censored if it is recorded from its beginning until a well defined time before its end time. It means a subject's follow-up time terminates before the outcome of interest is observed.
b. **Left Censoring**: A survival time is said to be left censored if an individual developed the event of interest prior to the beginning of the study. This situation is less common in survival studies and is often not a focus.
c. **Interval Censoring**: A survival time is categorized as interval censored if it is only known that the event of interest occurs within an interval of time without the knowledge of when exactly it occurs. Interval censoring can occur in clinical trials, industrial experiments, etc.

3.3.3 Descriptive methods for survival data

Descriptive analysis for survival data is used to present numerical or graphical summaries of the survival times in a particular group. In general, a statistical analysis should begin with a thoughtful and through univariate description of the data. The survivor function and hazard function are the two functions of central interest in summarizing survival data.

Survivor Function

Let T be a random variable associated with the survival times, t be the realization of the random variable T and $f(t)$ be the underlying probability density function of the survival time t. The cumulative distribution function $F(t)$, which represents the probability that a subject selected at random will have a survival time less than some stated value t, is then given by:

$$F(t) = pr(T \leq t) = \int_0^t f(u)du, \quad t \geq 0 \tag{3.1}$$

where T is the length of a completed spell and t is the elapsed time since entry to the state at time 0. The survivor function is obtained from the failure function and is given as:

$$S(t) = pr(T \geq t) = 1 - F(t) \quad t \geq 0 \tag{3.2}$$

The survivor function S(t) and the Failure function F(t) are each probabilities, and therefore inherit the properties of probabilities. The survivor function lies between zero and one, and is a strictly decreasing function of t. The survivor function is equal to one at the start of the spell (t = 0) and is zero at infinity. Since $S(t)$ is a probability, $S(0) = 1$ and as t approaches ∞, $S(t)$ approaches 0. From equations (3.1) and (3.2) the relationship between $f(t)$ and $S(t)$ can be given as:

$$f(t) = -\frac{dS(t)}{dt}, \quad t \geq 0 \tag{3.3}$$

Hazard Function

The hazard function is used to express the risk of death at some time t and is obtained from the probability that an individual dies at some time t, conditional on he or she having survived to that time. it is defined as:

$$h(t) = \lim_{\delta t \to 0} \frac{F(\delta t + t)F(t)}{S(t)\delta t} = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \tag{3.4}$$

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

$$S(t) = \exp^{[-H(t)]} \tag{3.5}$$

where $$H(t) = \int_0^t h(u)du = -\ln[S(t)] \geq 0 \tag{3.6}$$

is the integrated hazard function.

The important result is that, whatever functional form is chosen for (t), one can derive S(t) and F(t) from it (and also f(t) and H(t)), and vice versa.

Estimation of the Survivor Function

Among the other estimators of the survivor function the Kaplan-Meier estimator is the most common one. The Kaplan-Meier or product limit estimator is the limit of the life table estimator when intervals are taken so small that only at most one distinct observation occurs within an interval. Kaplan and Meier (1958) demonstrated that this estimator is a "maximum likelihood estimator". The estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps defined by the observed survival and censored times. This method is non-parametric or distribution-free, since it does not require specific assumptions to be made about the underlying distribution of the survival times (Hosmer and Lemeshow, 1999).

Let $d(x)$ denote the number of deaths at time x. Generally $d(x)$ is either zero or one, but we allow the possibility of tied survival times in which case $d(x)$ may be greater than one. Let $n(x)$ denote the number of individuals at risk just prior to time x; *i.e.*, number of individuals in the sample who neither died nor were censored prior to time x. Then the Kaplan-Meier estimator of the survival function at time t is obtained from the equation:

$$\hat{S}_{KM}(t) = \prod_{x \le t} \left(\frac{1 - d(x)}{n(x)} \right) \tag{3.7}$$

with the convention that $\hat{S}_{KM} = 1 \text{ if } t < t_1$

In the notation above, the product changes only at times x where $d(x) \ge 1$; *i.e.*, only at times where we observed deaths.

From equation (3.6) the KM estimator of the cumulative hazard function can be estimated

$$H_{KM}(t) = -\ln(\hat{S}_{KM}(t)) \tag{3.8}$$

The variance of the Kaplan-Meier estimators which is referred to as Greenwood's formula is given as:

$$\hat{V}ar(\hat{S}_{KM}(t)) = \left(\hat{S}_{KM}(t)\right)^2 \sum_{x \le t} \frac{d(x)}{n(x)[n(x) - d(x)]} \tag{3.9}$$

Another alternative estimator of the survival function and the corresponding commutative hazard function at time t due to Nelson and Aalen as stated in Collett (2003), which is based on the individual failure times is given by:

$$\tilde{H}_{NA}(t) = \sum_{x \le t} \frac{d(x)}{n(x)} \quad \text{and it implies} \quad \tilde{S}_{NA}(t) = \exp(-\tilde{H}(t)) = \prod_{x \le t} \exp\left(-\frac{d(x)}{n(x)} \right) \tag{3.10}$$

It is merely in the case of small samples that the Nelson-Aalen estimate of the survivor function prevails over the KM estimate (Hosmer and Lemeshow, 1999).

Comparison of Survivorship Function

After providing a description of the overall survival experience in the study, we turn our attention to a comparison of the survivorship experience in key subjects in the data. The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for these groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due merely to chance variation. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests.

The standard statistical procedures may be used when there are no censored observations. But modifications of these procedures are required when censored observations are present in the data. In comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. The figure in general shows if the pattern of one survivorship function lies above another, meaning that the group defined by the upper curve lived longer, or had a more favorable survival experience, than the group defined by the lower curve. In other words, at any point in time the proportion of subjects estimated to be alive is greater for one group (represented by the upper curve) than the other (represented by the lower curve). Now the statistical question is whether the observed difference seen in the figure is significant. A number of statistical tests have been proposed to answer this question such as Log-rank, Gehan's generalization of Wilcoxon test, and Peto-Peto-Prentice's test and so on.

The calculation of each test is based on a contingency table of groups by status at each observed survival time. The general form of these test statistics for the comparison of survival functions between two groups can be defined as follows:

$$Q=\frac{\left\{\sum_{x} w(x)\left[dN_1(x)-\frac{dN(x)*Y_1(x)}{Y(x)}\right]\right\}^2}{\sum_{x} w^2(x)\left[\frac{Y_1(x)Y_0(x)dN(x)\left[Y(x)-dN(x)\right]}{Y^2(x)\left[Y(x)-1\right]}\right]} \tag{3.11}$$

where:

- ✓ $Y_0(x)$ is the number of individuals at risk at time x from group 0
- ✓ $Y_1(x)$ is the number of individuals at risk at time x from group 1
- ✓ $Y(x)$ is the total number of individuals at risk at time x from both groups
- ✓ $dN_0(x)$ is the number of observed deaths from group 0 at time x
- ✓ $dN_1\, x$ is the number of observed deaths from group 1 at time x
- ✓ $dN(x)$ is the total number of deaths observed at time x
- ✓ $w(x)$ is the weight for censor adjustment at failure time x

The test statistic Q has chi-square distribution with 1 degree of freedom under the null hypothesis that the two survivorship functions are the same when the total number of observed events and sum of expected number of events are large and assuming that the censoring experience is independent of group. The statistic Q can be extended for comparing more than two groups of survival experience (Collett, 2003). The weight function $w(x)$ can be used to emphasize differences in the hazard rates over time according to their relative values. The most commonly used test is the log-rank test where $w(x) = 1$ for all x. The log-rank test is a non-parametric test for comparing two or more survival curves. Since it is a non-parametric test, no assumptions about the distributional form of the data need to be made. This test is however most powerful when used for non-overlapping survival curves. This test can be generalized to accommodate other tests that are equally used sometime in practice such as Generalized Wilcoxon test, and Peto-Peto-Prentice test. Each of these tests uses different weights to adjust for censoring that is often encountered in survival data. For instance, the Wilcoxon test weights the j^{th} failure time by $Y(x)$ (the number still at risk), and the Peto-Peto-Prentice test weights the j^{th} failure time by the survival estimate $\tilde{S}(x)$ calculated over all groups combined. Since both $Y(x)$ and $\tilde{S}(x)$ are non-increasing functions of x, both tests emphasize the difference early in the survival curves (Kleinbaum and Klein, 2005).

3.3.4 Models for Survival Data

Our aim is to estimate the hazard ratio of the probability of a child dying within the next month after surviving for t months, as a result of environmental factors, among others. In the context of child mortality, the hazard rate is often referred to as the mortality rate. The mortality rate at age t can be interpreted as the intensity at which a child dies at this age, given that the child survived until age t. We focus on children who are born alive and model their mortality probabilities until the age of five. To check robustness, we implement two models, a parametric (Exponential and Weibull) and a semi-parametric model (Cox -PH).

A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches. Semi-parametric models are models that parametrically specify the functional relationship between the lifetime of an individual and his/her characteristics of environment. But leave the actual distribution of life times arbitrary. The most popular of the semi-parametric models is the proportional hazards model. It has the property that the ratio of the hazards depends on the values of their explanatory variables but does not depend on time t. A hazard model is a regression model in which the "risk" of experiencing an event (death in our case) at a certain time point is predicted with a set of covariates.

3.3.4.1 Cox-Proportional Hazards Model

This model was proposed by Cox (1972) and has also come to be known as the Cox regression model. Cox introduced the model to cater for covariate effects for single event failures. This model is valid under the assumption of proportional hazards (PH) , no particular form of the probability distribution is assumed for survival time. Cox observed that if proportional hazards assumption holds (or is assumed to be hold), then it is possible to estimate the effect of parameter(s) without any consideration of the hazard function.

Suppose the set of values of the explanatory variables in the PH model will be represented by a vector X. Let $h_0(t)$ be the hazard function for an individual for whom the values of all explanatory variables that make up the vector X are zero. The function $h_0(t)$ is called the baseline hazard function. The hazard function for the individual can then be written as:

$$h(t, x, \beta) = ho(t)\exp(\beta' x) \qquad (3.12)$$

where **β** is a p x l vector of unknown regression parameters that are assumed to be the same for all individuals in the study and measure the influence of the covariate on the survival experience with β_i representing increase in the log hazards as x_i increases one unit relative to the baseline hazard function. **X** is a $p\times 1$ vector of covariates such as treatment indicators, prognostic factors, and etc. The baseline hazard function $h_o\,(t)$ can take any shape as a function of *t*. The only requirement is that $h_o\,(t) > 0$. This is the nonparametric part of the model and $\beta' x$ is the parametric part of the model. That is why Cox's proportional hazards model is a semi parametric model.

A key reason for the popularity of the Cox model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and survival curves can be obtained for a wide variety of data situations. Another way of saying this is that the Cox PH model is a “robust” model, so that the results from using the Cox model will closely approximate the results for the correct parametric model (Kleinbaum and Klein, 2005).

An important feature of the Cox proportional hazards model, which concerns the proportional hazards assumption, is that the baseline hazard is a function of *t*, but does not involve the **X**’s. In contrast, the exponential expression, involves the **X**’s, but does not involve *t*. The **X**’s, here, are assumed to be time-independent. The other assumption of the proportional hazards refers to the fact that the effects of covariates are the same for all values of *t*. Putting it in other words, the Cox proportional hazards model assumes that changes in the hazard of any subject over time will always be proportional to changes in the hazard of any other subject and to changes in the underlying hazard over time (Kleinbaum and Klein, 2005).

From equation (3.12) one can notice a couple of features. First, if the vector of covariate is a zero vector, then the hazard function for the i^{th} individual is the baseline hazard function. It is the hazard function in the absence of covariates or when all of the coefficients of the covariates are assumed to be zero. Second, if we divide both sides by $h_o(t)$, we get equation (3.13) below that indicates where the term proportional comes from. Since for each individual, $\exp(\beta' x)$ is constant across time, equation (3.14) below shows that at every value of t, the i^{th} individual's log hazard ratio is constant. This implies that each individual's hazard function is parallel to the $h_o(t)$.

$$\frac{h(t, x_i, \beta)}{h_o(t, x_i = 0, \beta)} = \frac{h_o(t)\exp(\beta' x_i)}{h_o(t)} = \exp(\beta' x_i) \tag{3.13}$$

The logarithm of the hazard ratio for two individuals having two distinct covariate values x_j and x_i can be expressed as

$$\ln\left(\frac{h(t, x_j, \beta)}{h_o(t, x_i = 0, \beta)}\right) = \ln\left(\frac{h_o(t)\exp(\beta' x_j)}{h_o(t)\exp(\beta' x_i)}\right) = \beta'\left(x_j - x_i\right) \tag{3.14}$$

Clearly the above ratio is independent of time which means that the log hazard ratio is constant at any given time. Moreover, the hazard ratio does not depend on the value of the covariate; rather it depends on the difference between the covariate values. The Cox proportional hazards model can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$\ln\left(\frac{h(t, X, \beta)}{h_o(t)}\right) = \beta' X \tag{3.15}$$

The cumulative hazard functions at time t for a subject with covariate x is given by:

$$H(t, X, \beta) = H_o(t)\exp(\beta' X) \tag{3.16}$$

Consequently, from the proportional hazard function, we obtain the survivor function given by:

$$S(t, X, \beta) = \left[S_o(t)\right]^{\exp(\beta' x)} \tag{3.17}$$

where S_0(t) is the baseline survival function (Hosmer and Lemeshow, 1999).

3.3.4.2 Fitting the Proportional Hazards Model

Fitting the proportional hazards model to observed survival data entails estimating the unknown regression coefficients. Since the baseline hazard $h_o(t)$ is left completely unspecified, ordinary likelihood methods can't be used to estimate β. Cox conceived of the idea of a partial likelihood to remove the nuisance parameter $h_o(t)$ from the proposed equation.

Suppose we have a random sample of individuals of size n from a specific population whose true survival times are $Z_1, Z_2, \ldots, Z_n$. Denote by C the censoring process and by $C_1, C_2, \ldots, C_n$ the (potential) censoring times. The observed data are the minimum of the survival time and censoring time for each subject in the sample and the indication whether or not the subject is censored. Statistically, we have observed triplet data (t_i, δ_i, X_i) where $t_i = \min(Z_i , C_i)$, δ_i is the event indicator $\delta_i = 1$ if the event has occurred and $\delta_i = 0$ if it is censored, and X_i is the vector of covariates or the risk factors for the i^{th} individual. Under the assumption of independent observations, the full likelihood function is obtained by multiplying the respective contributions of the observed triplets, a value of $f(t, X, \beta)$ for uncensored observation and a value of $S(t, X, \beta)$ for censored observations. Thus, the contribution of each triplet to the likelihood is the expression

$$[f(t,X,\beta)]^{\delta_i} \times [S(t,X,\beta)]^{1-\delta_i} \tag{3.18}$$

Since the observations are assumed to be independent, the likelihood function is the product of the expression in (3.18) over the entire sample and is formulated as:

$$l(\beta) = \prod_{i=1}^{n} \left\{ [f(t_i, X_i, \beta)]^{\delta_i} \times [S(t_i, X_i, \beta)]^{1-\delta_i} \right\} \tag{3.19}$$

It can be further simplified as:

$$l(\beta) = \prod_{i=1}^{n} \left\{ [h(t_i, X_i, \beta)]^{\delta_i} \times [s(t_i, X_i, \beta)] \right\} \tag{3.20}$$

Cox showed that the relevant likelihood function which considers the baseline hazard rate as a nuisance parameter; he called it a partial likelihood function, for the proportional hazards model assuming no tied survival times is given by (Hosmer and Lemeshow, 1999)

$$l_p(\beta) = \prod_{i=1}^{n} \left(\frac{e^{x_{(i)}\beta}}{\sum_{j \in R(t_{(i)})} e^{x_{(j)}\beta}} \right)^{\delta_i} \tag{3.21}$$

where, $R(t_{(i)})$ represents the risk set just prior to time $t_{(i)}$ The corresponding log-partial likelihood function is given by

$$L_p(\beta) = \sum_{i=1}^{n} \delta_i \left\{ X_{(i)}\beta - \ln\left[\sum_{j \in R(t_i))} \exp(X_j \beta) \right] \right\} \tag{3.22}$$

We obtain the maximum partial likelihood estimator (MPLE) by differentiating the right hand side of (3.21) with respect to β, setting the derivatives equal to zero and solving for the unknown parameters. This is using iterative numerical analysis techniques such as Newton-Raphson which make use of the efficient scores and the observed information matrix. Let $U(\beta)$ be the $p\times1$ vectors of first derivatives of the log-likelihood function with respect to the β -parameters. This quantity is known as the vector of efficient scores. The negative of the second derivative of the log-partial likelihood is known as the observed information matrix (Hessian matrix) and denoted by

$$I(\beta)=\frac{\partial^2 L_p(\beta)}{\partial\beta\partial\beta} \tag{3.23}$$

According to the Newton-Raphson procedure an estimate of β at the $(k+1)^{th}$ cycle of the iterative procedure , $\hat{\beta}_{k+1}$, is $\hat{\beta}_{k+1}=\hat{\beta}_k+I^{-1}(\hat{\beta}_k)\bigcup(\hat{\beta}_k)$, $\mathrm{k}=1,2,...$. The process can be started by taking $\hat{\beta}_o=(0,0,...,0)'$ and continue until the change in the likelihood function is sufficiently low. The estimator of the covariance matrix of the MPLE can be approximated by the inverse of the observed information matrix, evaluated at $\hat{\beta}$, that is

$$\hat{V}ar(\hat{\beta})=I(\hat{\beta})^{-1} \tag{3.24}$$

The partial likelihood function methods described above are based on the assumption that there were no tied values among observed survival times. Hence to incorporate tied survival times in analyses there are two approaches. These are the Breslow and the Efron approximations. The MPLE for β in the presence of ties is obtained in the same manner as in the non-tied data case, with exception that derivates are taken with respect to the unknown parameters in the log of either the Breslow or Efron approximation to the partial likelihood. In many applied settings there will be little or no practical difference between the estimators obtained from the two approximations. Because of this, and since the Breslow approximation is more commonly available in many software packages, unless stated otherwise, analysis presented in this study will be based on it (Hosmer and Lemeshow, 1999).

After estimation of the regression coefficients, we go for assessing the significance of the coefficients and the construction of the confidence interval as well. The three different tests used to assess the significance of the coefficients are explained below.

a) The partial likelihood ratio test

It is used for testing the significance of a subset of q explanatory variables from p explanatory variables, and fit both the unrestricted and the restricted models. Then we obtain the value of the log-partial likelihood function $L_P(\hat{\beta}_{P-q})$ in the unrestricted model and $L_p(\hat{\beta}_p)$ when the model imposes the restrictions under H_o. The partial likelihood ratio test statistic is given by:

$$Q_{LR} = 2(L_p(\hat{\beta}_p) - L_p(\hat{\beta}_{p-q})) \tag{3.25}$$

Under the null hypothesis H_o for large sample size the statistic Q_{LR} is asymptotically distributed as chi-squared with q degrees of freedom.

b) The Wald test

To test $H_0 = (0,0,...,0)$, we use the multivariable Wald statistic

$$Q_w = \hat{\beta}'_q \left[I_q(\hat{\beta})\right]^{-1} \hat{\beta}'_q \tag{3.26}$$

where $\hat{\beta}_p$ and $I_q(\hat{\beta})$ are the corresponding estimates of β_q and sub matrix of the inverse of the observed information matrix from the full model. Under Ho and for large sample size the statistic $Q_w \sim \chi^2_{(q)}$ at α level of significance. The Wald test can also be used to test the significance of individual variables. The Wald test statistic is

$$Z = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \tag{3.27}$$

Under the null hypothesis H_o: $\beta_j=0$ the statistic $Z \sim N(0,1)$. Consequently, the 100(1-α)% Wald statistic-based confidence interval for β_j is $\hat{\beta}_j \pm Z_{\alpha/2} Se(\hat{\beta}_j)$ where, $Z_{\alpha/2}$ is the upper α/2 percentile of the standard normal distribution.

c) The Score test

The score test statistic, to test $H_0 : \beta_q\ (0,0,...,0)$ is defined as:

$$Q_s = U'(\beta_q, \hat{\beta}_{p-q}) I'(\beta_q, \hat{\beta}_{p-q}) U(\beta_q, \beta_{p-q}) \tag{3.28}$$

where $U(\beta_q, \hat{\beta}_{p-q})$ and $I'(\beta_q, \hat{\beta}_{p-q})$ are the score vectors and inverse of the observed information matrix evaluated at the hypothesized value of β_q and the restricted partial maximum likelihood estimator of β_{p-q} .Under the null hypothesis and for large sample, $Q \sim \chi^2(q)$. When there is a disagreement among the three tests of the significance of the coefficient, the partial likelihood ratio test will prevail.

3.3.4.3 Variable Selection Procedures

The variable selection procedures in proportional hazards regression analysis requires critical decisions in selecting subsets of covariates. The methods available to select a subset of the covariates to include in a proportional hazards regression model are essentially the same as those used in the other regression models, like purposeful selection, stepwise (forward selection and backward elimination). When the number of variables is relatively large, it can be computationally expensive to fit all possible models. In this situation, automatic routines for variable selection that are available in many software packages might seem an attractive prospect. But they lead to the identification of one particular subset, rather than a set of equally good ones. The subsets found by these routines often depend on the variable selection process that has been used, that is, whether it is forward selection, backward elimination or the stepwise procedure, and generally tend not to take any account of the hierarchic principle. They also depend on the stopping rule that is used to determine whether a term should be included in or excluded from a model.

Thus, instead of using automatic variable selection procedures, the following general strategy for model selection is recommended by Collet (2003).

1. The first step was to fit models that contain each of the variables one at a time. The values of $-2\hat{L}_p$ for these models were then compared with that for the null model to determine which variable on their own significantly reduce the value of this statistic. A significance level from 20% to 25% is recommended in Hosmer and Lemeshow (1999).

2. The variables which appear to be important from Step 1 were then fitted together. In the presence of certain variables others may cease to be important. As a result, backward elimination was used to omit non-significant variables (i.e., those variables that do not significantly increase the value of $-2\hat{L}_p$ from the model). Only those that lead to a significant increase in the value of $-2\hat{L}_p$ were retained in the model.
3. Variables that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added the model from step 2 with forward selection method (i.e., any that reduce $-2\hat{L}_p$ significantly were retained in the model).
4. A final check was made to ensure that no term in the model could be omitted without significantly increasing the value of $-2\hat{L}_p$, and that no term not included significantly reduces the value of $-2\hat{L}_p$.

3.3.4.4 Assessment of Model Adequacy

Following a model has been fitted, the adequacy of the fitted model needs to be assessed. Model based inferences depend completely on the fitted statistical model. For these inferences to be valid in any sense of the word, the fitted model must provide an adequate summary of the data upon which it is based. Many model checking procedures are based on residuals. A residual is the difference between the observed value of the outcome variable and that value predicted by the model. The two key assumptions in the definition of a residual are the value of the outcome is known and the fitted model provides an estimate of the mean of the dependent variable or systematic component of the model. However, the two assumptions are not valid when using partial likelihood to fit the proportional hazards model to censored survival data. The absence of an obvious residual has lead to the development of several different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazard model. These include the Cox-Snell, martingale and Schoenfeld residuals.

a. Cox-Snell residuals (rc_i): The Cox-Snell residual for the i^{th} individual with observed survival time t_i is given by:

$$rc_i = \hat{H}_i(t_i) = -\hat{S}_i(t_i) \tag{3.29}$$

where $\hat{H}_i(t_i)$ and $\hat{S}_i(t_i)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} subject at time t_i respectively. In general, Cox-Snell residuals are useful in assessing an overall model fit (Cox and Snell, 1968).

b. Martingale residuals (rM_i): are also called modified Cox-Snell residuals and, expressed as:

$$rM_i = \delta_i - \hat{H}_i(t) = \delta_i - rc_i \tag{3.30}$$

where $\delta_i = 1$ for uncensored observations and zero otherwise, and rc_i are Cox-Snell residuals. The martingale residuals take values between negative infinity and unity. They have a skewed distribution with mean zero. In large samples, the martingale residuals are uncorrelated with one another and have an expected value of zero. However, the martingale residuals are not symmetrically distributed about zero (Barlow and Prentice,1988).

c. Schoenfeld residuals (rs_{ik}): All the above residuals are residuals for each individual. We will describe covariate-wise residuals: Schoenfeld residuals. These residuals are calculated for each individual and for each covariate (Schoenfeld, 1982). Thus, the Schoenfeld residual for the i^{th} individual on the k^{th} covariate is given by:

$$rs_{ik} = \delta_i - (x_{ik} - \bar{\hat{x}}_{w_i k}) \tag{3.31}$$

where $\bar{\hat{x}}_{w_i k} = \dfrac{\sum_{j \in R(t_{(i)})} x_{jk} \exp(x'_j \hat{\beta})}{\sum_{j \in R(t_{(i)})} \exp(x'_j \hat{\beta})}$ is a weighted mean of covariate value for those in the risk set at the given event time.

The sum of these residuals is zero and they have a large sample property that, their expected value is zero and they are uncorrelated with one another. The vector of these residuals for the i^{th} observation can be written as $rs_i = (rs_{i1}, rs_{i2}, ..., rs_{ip})'$ and the convention is that rs_{ik} is set to be missing for censored observations. Scaling a vector of Schoenfeld residuals by an estimator of its variance is more effective in detecting departures from the assumed model. The vector of the scaled Schoenfeld residuals is then given by:

$$rs_i * = \left[\operatorname{var}(rs_i)\right]' rs_i \approx m \operatorname{var}(\hat{\beta}) rs_i \tag{3.32}$$

where, m is the number of events (deaths) (Grambsch and Therneau,1994).

Each of these residuals provides a useful tool for examining one or more aspects of model adequacy.

1. Testing for the form (linearity) of covariates

After identification of a particular set of explanatory variables on which the hazard function depends, it is important to check that the correct functional form has been adopted for the continuous covariates. Linearity assumption can be checked by using the plot of martingale residuals. The plot of martingale residuals obtained from fitting the model, excluding the covariate whose functional form needs to be determined, against the excluded covariate display the functional form required for the covariate. In such a way that, LOESS smoothed curve can be superimposed on the scatter plots to give interpretation. If the resulting plot is showing no systematic pattern and the smoothed plot is a horizontal straight line through zero. This indicates that the covariate is linear in the model.

2. Subject-wise diagnostic measures

In the assessment of model adequacy, it is important to determine whether there are any subjects have an unusual configuration of covariates, exert an undue influence on the estimates of the parameters or have an undue influence on the fit of the model. Such observations may be termed as influential observations and the data from such individuals will need to be the subject of further analysis. Conclusions from survival analyses are often framed in terms of estimates of the relative hazard, which depends on the estimated values of the coefficients in the Cox regression model. For that reason, it has particular importance to examine the influence of each observation on these estimates (Hosmer and Lemeshow, 1999). It may happen that the structure of the fitted model is particularly sensitive to one or more observations in the data set. Such observations can be analyzed through diagnostics that are designed to highlight observations that influence the complete set of parameter estimates in the linear predictor. This could be done by fitting the model to all n observations in the data set, and then fitting the same model to the sets of n-1 observations obtained by omitting each of the n observations in turn.

Suppose that $\hat{\beta}_k$ denotes the partial likelihood estimator of the coefficient computed using the entire sample of size n and $\hat{\beta}_{k(-1)}$ denotes the value of the estimator if the i^{th} subject is removed. Thus, the DFBETA statistic, which can be used as a measure of how the j^{th} parameter estimate would change if the i^{th} observation was deleted from the data set, is defined as:

$$\Delta\hat{\beta}_{ki} \approx \hat{\beta}_k - \hat{\beta}_{k(-1)} \tag{3.33}$$

Observations that influence a particular parameter estimate have a large absolute value of DFBETA than for other observations in the data set. However, this procedure involves a significant amount of computation if the sample size is large. We would like to use an alternative approximate value that does not involve an iterative refitting of the model. To check the influence of observations on a parameter estimate, an approximate estimator of (33) is the k^{th} element of the vector of coefficient changes

$$\Delta\hat{\beta}_i = (\hat{\beta} - \beta_{(-i)}) = \hat{\mathrm{var}}(\hat{\beta})\hat{L}_i \tag{3.34}$$

where $\hat{L}_i$ is the vector of score residuals which are modifications of Schoenfeld residuals and are defined for all the observations, and $V\hat{a}r(\hat{\beta})$ is the estimator of the covariance matrix of the estimated coefficients. These are commonly referred to as the scaled Schoenfeld residuals.

3. Methods for Assessing the Proportional Hazards Assumption

The main assumption of the Cox hazards model is the proportionality of hazard. The assumption is vital to the interpretation and use of a fitted proportional hazards model. If hazards are not proportional, this means that the linear component of the fitted model varies with time in some manner. As a result, we need to plot the logarithm of the Kaplan-Meier cumulative hazards function based on different factors so that it helps in assessing the proportional hazards assumption before fitting a Cox model. If this assumption is met, then the plots will be more or less parallel. However, looking at the plot is not enough to be certain of proportionality since they are univariate analysis and do not shows whether hazards will still be proportional when a model includes many other predictors. But they support our argument for proportionality (Hosmer and Lemeshow, 1999).

The other method, which could be used after the fit of the model, is extending the proportional hazards model by defining several product terms involving each time independent variable with some function of time. That is, if the j^{th} time-independent variable is denoted as x_j , then we can define the j^{th} product term as $x_j \times g_j(t)$ where $g_j(t)$ is some function of time for the j^{th} variable. Usually the function $g_j(t)$ is chosen to be the natural logarithm of survival time i.e. $g_j(t) = \ln(t)$ Likewise, Grambsch and Therneau (1994) also considered a specific form of time-varying coefficient as:

$$\beta_j(t) = \beta_j + \theta_j x_j g_j(t) \tag{3.35}$$

where θ_j is a coefficient of the product term. Thus, the extended Cox model that simultaneously considers all time-independent variables of interest can be formulated as:

$$h(t, x, \beta) = h_o(t)\exp\left(\sum_{j=1}^{p}\beta_j x_j + \sum_{j=1}^{p}\theta_j x_j g_j(t)\right) \tag{3.36}$$

To check the proportional hazards assumption, we consider the null hypothesis that all the θ_j terms are equal to zero so that the model reduces to the proportional hazards model. The hypothesis all θ_j's are zero $(H_o : \theta_j = 0)$ is tested via the partial likelihood ratio test, score test or Wald test. If the time-dependent covariate is insignificant then the assumption of proportionality is satisfied for that particular covariate. Moreover, the other statistical test of the proportional hazards assumption is based on the scaled Schoenfeld residual. If the PH assumption holds for a particular covariate then the scaled Schoenfeld residual for that covariate will not be related to survival time. So this test is accomplished by finding the correlation between the scaled Schoenfeld residuals for a particular covariate and the ranking of individual survival times. The null hypothesis is that the correlation between the scaled Schoenfeld residuals and the ranked survival time is zero. Rejection of null hypothesis concludes that PH assumption is violated.

4. Overall Goodness of Fit

Residual plots can be used in the graphical assessment of the adequacy of a fitted model. For instance, if the fitted model is adequate, the Cox-Snell residuals will behave as *n* observations from a unit exponential distribution. Thus, the plot of the estimated hazard rate of the Cox-Snell residuals $\hat{H}_i(t)$, versus rc_i will give a straight line with unit slope and zero intercept if the fitted model is correct. However, the drawback is that they do not indicate the particular departure from the model fitted, if there is any. The other method of checking goodness of fit of the model is to use R^2. In proportional hazards regression model as in all regression analyses there is no single, simple method of calculating and interpreting R^2, because in Cox proportional hazards model, R^2 depends on the proportion of the censored observations in the data. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to high percent of censored data (Hosmer and Lemeshow, 1998). The measure of goodness of fit R^2_p based on partial likelihood is given by:-

$$R_P^2 = 1 - \exp\left[\frac{2}{n}\left(L_0 - L_p\right)\right] \tag{3.37}$$

where,

- ❖ L_o is the log partial likelihood for empty/null model, the model with no covariates.
- ❖ L_p is log of partial likelihood for the fitted model with *p* covariates, and n is the total number of observations in the model.

3.4 Parametric Survival Regression Models

In previous topics it was focused entirely on the use of semi-parametric model, proportional hazards Cox regression model, in the analysis and prediction of the survival time of child mortality. The basis of this method is to avoid having to specify the hazard function completely. However, there may be setting in which the distribution of the survival time is in specific parametric distribution that justifies the use of a fully parametric model to better address the goal of the analysis. A parametric survival model assumes that the survival time follows a known distribution. The popularity of this approach is due to the fact that plausible models may be easily fit, evaluated and interpreted.

3.4.1 The Exponential Survival Regression Model

The simplest model for the hazard function is to assume that it is constant over time. The hazard of death at any time after the time origin the study is then the same, irrespective of the time elapsed (Collett, 2003). Under this model, the hazard function is written as:

$$h_o(t) = \lambda \tag{3.38}$$

From the constant baseline hazard function, the corresponding survivor function is:

$$S_o(t) = \left\{ -\int_0^t \lambda du \right\} = \exp(-\lambda t) \tag{3.39}$$

And so the implied probability density function of the survival times is

$$f_o(t) = \lambda \exp(-\lambda t) \tag{4.40}$$

This is the probability density function of a random variable T that has an exponential distribution with a mean of λ^{-1}.The parameter λ with $\lambda > 0$,is often called the intensity. The median event time can be obtained by solving the equation $S_o(t_{0.5}) = 0.5$ which leads to $t_{0.5} = \log 2/\lambda$. More generally, the p^{th} quantile can be obtained by solving the equation $S(t_p) = 1 - p$ and thus $t_p = \dfrac{-\log(1-p)}{\lambda}$ (3.41)

The main feature of the exponential distribution is thus that the instantaneous hazard does not vary over time. Another important property is the lack of memory property. Consider a random variable $T \sim Exp(\lambda)$. We now study the survival function of a subject conditional on having survived up to time t_0, the excess survival time is described by the same exponential distribution with constant hazard rate λ. An empirical check for this distribution for a set of survival data is provided by plotting the log of the survival function estimate versus t. Such a plot should resemble a straight line through the origin, as $\log S_o(t) = -\lambda t$ if the exponential distribution assumption holds.

3.4.2 Fitting the Exponential Survival Regression Model

In the parametric setting, estimates of the parameters are obtained by maximizing the likelihood function. The survival likelihood for survival data with event times and right censored data is generally given by:

$$L = \prod_{i=1}^{n} (f(x_i))^{\delta i} (S(x_i))^{1-\delta i} \tag{3.42}$$

which leads for exponentially distributed event times to:

$$\begin{aligned} L &= \prod_{i=1}^{n} (\lambda \exp(-\lambda x_i))^{\delta i} (\exp(-\lambda x_i))^{1-\delta i} \\ &= \prod_{i=1}^{\mathrm{n}} \lambda^{\delta i} \exp(\lambda x_i) \end{aligned} \tag{3.43}$$

By differentiating the log likelihood function with respect to λ and equating this expression to zero leads to the maximum likelihood estimator

$$\lambda = \frac{d}{\sum_{i=1}^{n} x_i} \tag{3.44}$$

3.4.3 The Weibull Survival Regression Model

The Weibull distribution is a generalization of the exponential distribution. However, unlike the exponential distribution, it does not assume a constant hazard rate and therefore has broader application. The distribution was proposed by Weibull (1939) and its applicability to various failure situations discussed again by Weibull (1951). The baseline hazard function for Weibull distributed event times is given by:

$$h_o(t) = \lambda \gamma t^{\gamma-1} \tag{3.45}$$

It follows that the survival function for the Weibull distribution is given by:

$$S_o(t) = \exp(-\lambda t^{\gamma}) \tag{3.46}$$

and the density function is

$$f_o(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^{\gamma}) \tag{3.47}$$

with λ, $\lambda > 0$, the scale parameter and $\gamma, \gamma > 0$, the shape parameter.

The median event time can be obtained by solving the equation $S_o(t_{0.5}) = 0.5$ which leads to $t_{0.5} = \left(\frac{\log 2}{\lambda}\right)^{1/\gamma}$. More the p^{th} quantile can be obtained by solving the equation $S(t_p) = 1 - p$ and thus

$$t_p = \left(\frac{-\log(1-p)}{\lambda}\right)^{1/\gamma} \tag{3.48}$$

The shape of the hazard function critically depends up on the values of γ.

If $\gamma < 1$: hazard decreases monotonically with time

If $\gamma > 1$: hazard increases monotonically with time

If $\gamma = 1$: constant hazard (equivalent to exponential distribution)

The Weibull hazard model can be generally presented as

$$h_i(t) = h_o(t)\exp(\beta' x_i) \tag{3.49}$$

$$S_i(t) = (\exp - \lambda \exp(\beta' x_i) t^{\gamma}) \tag{3.50}$$

$$f_i(t) = \lambda \gamma t^{\gamma-1} \exp(\beta' x_i)(\exp - \lambda \exp(\beta' x_i) t^{\gamma}) \tag{3.51}$$

with $h_o(t) = \lambda \gamma t^{\gamma-1}$ and β a p × 1 vector containing the parameters. The event time of the i^{th} subject is then characterized by the Weibull distribution with scale parameter $\lambda \exp(\beta' x_i)$ and shape parameter γ. Thus, all subjects share the shape parameter but differ with respect to their scale parameter. The model assumes that individual i and j with covariates X_i and X_j have proportional hazard function of the form:

$$\frac{h(t; x_i)}{h(t; x_j)} = \frac{\exp(\beta' x_i)}{\exp(\beta' x_j)} = \exp\left(\beta'(x_i - x_j)\right) \tag{3.52}$$

The quantities $\exp(\beta)$ can be interpreted as hazard ratios.

3.4.3 Fitting the Weibull Survival Regression Model

The survival likelihood for Weibull distributed survival data with event times and right censored data is generally given by

$$L = \prod_{i=1}^{n} \left\{ (t_i \lambda \gamma x_i^{\gamma-1} \exp(-\lambda x_i^{\gamma})^{\delta i} (\exp(-\lambda x_i^{\gamma}))^{1-\delta_i} \right\} \tag{3.53}$$

resulting in the log likelihood function

$$l = d + \log(\lambda\gamma) + (\gamma - 1)\sum_{i=1}^{n}\delta_i \log x_i - \lambda\sum_{i=1}^{n} x_i^{\gamma} \tag{3.54}$$

with d the total number of events. Maximum likelihood estimators can be obtained by equating the first derivatives of l with respect to λ and γ to zero and we get.

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^{n} x_i^{\hat{\gamma}}} \qquad \text{and}$$

$$\frac{d}{\hat{\gamma}} + \sum_{i=1}^{n}\delta_i \log x_i - \frac{d}{\sum_{i=1}^{n} x_i^{\hat{\gamma}}}\sum_{i=1}^{n} x_i^{\gamma} \log x_i = 0 \tag{3.55}$$

which is nonlinear in $\hat{\gamma}$ and can only be solved by a numerical procedure such as the Newton Raphson algorithm.

3.4.4 Model Selection in Parametric Survival Regression Models

To be select the model that can predict the survival of child, we have two methods. The first is graphical approach. For this method the cox-Snell plot is the common one. It is a graph of the minus ln of Kaplan-Meier plotted against the cox-Snell residual values. It is used to determine how well a specific distribution fits to the observed data. This plot would be approximately linear if the specified theoretical distribution is the correct model. Easy fit displays the reference diagonal line along which the graph points should fall along with the goodness of fit tests; the distribution plots can be helpful to determine the best fitting model. The fundamental difference of this approach is that it is quite subjective to come on conclusion while the goodness of fit tests are “exact” in the sense that the results do not depend on the researcher (provided that the tests are performed correctly), using plot is a more empirical way to use in model selection. Akaikie (1974) proposed an informative criterion (AIC) statistic to compare different models and/or models with different numbers. For each model the value is computed as:

$$AIC = -2\log likelihood + 2\,(p+1+\text{k}) \tag{3.56}$$

Where, p denotes the number of covariates in the model without including the constant term and k is the number of parameters minus one *i.e.* s=0 for the Exponential regression and k=1 for Weibull regression models. According to the criterion, a model with small AIC value will be considered as it fits for the data.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Descriptive Analysis of the Survival Data

The total number of live births considered for this study were 2120 with 181 of children death aged from 12 up to 59 months. Of the total live birth, 47.5% and 52.5% of Child death have occurred for male and female, respectively. Regarding Place of residence, from the total of 2120 children included in the study, 1732 (58.6%) were born in rural and 388 (41.4%) born in Urban part of Ethiopia. When we see mothers' education, the child mortality was 43.4%, 33.2% and 23.2% for children whose mothers' have no education ,Primary and Secondary & above educational status, respectively.

On the other hand, type of toilet facility children death were 90.6%,1.1% and 8.3% used in Non-improved, shared facility and improved, respectively. Among the total number of mothers' (2120), 1759 (83%) were not pre-birth follow up during pregnancy period and 361(17%) mothers' dose pre-birth follow up during pregnancy period. In addition, 372 mothers' delivered in health center, 1741 mothers' delivered at home and the rest 7 delivered on other places like road. Finally, 144 (79.6%) children death were due to Non-improved source of drinking water and 37(20.4%) of death of children were in the case of improved water source. All the results have been summarized in Table 4.2 (annex A).

The different survival estimates are displayed in Estimates in annex A (Table 4.1), together with the numbers at risk and the number failing. The median survival time of a child was 36 months with a standard error of 0.2039869 for the follow-up period of time. In addition, the plot of overall Kaplan-Meier estimate indicate that for child mortality monotonically decreases as follow up time increases (see figure 4.1).

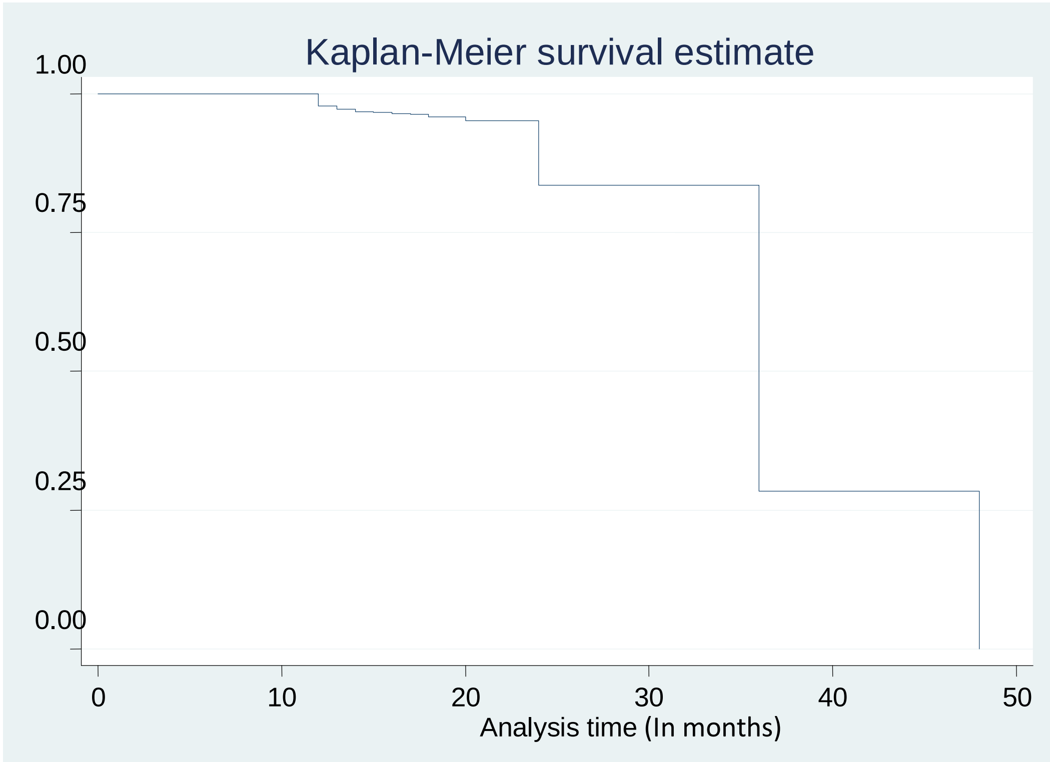


Figure 4.1: Overall product limit estimate of survival function

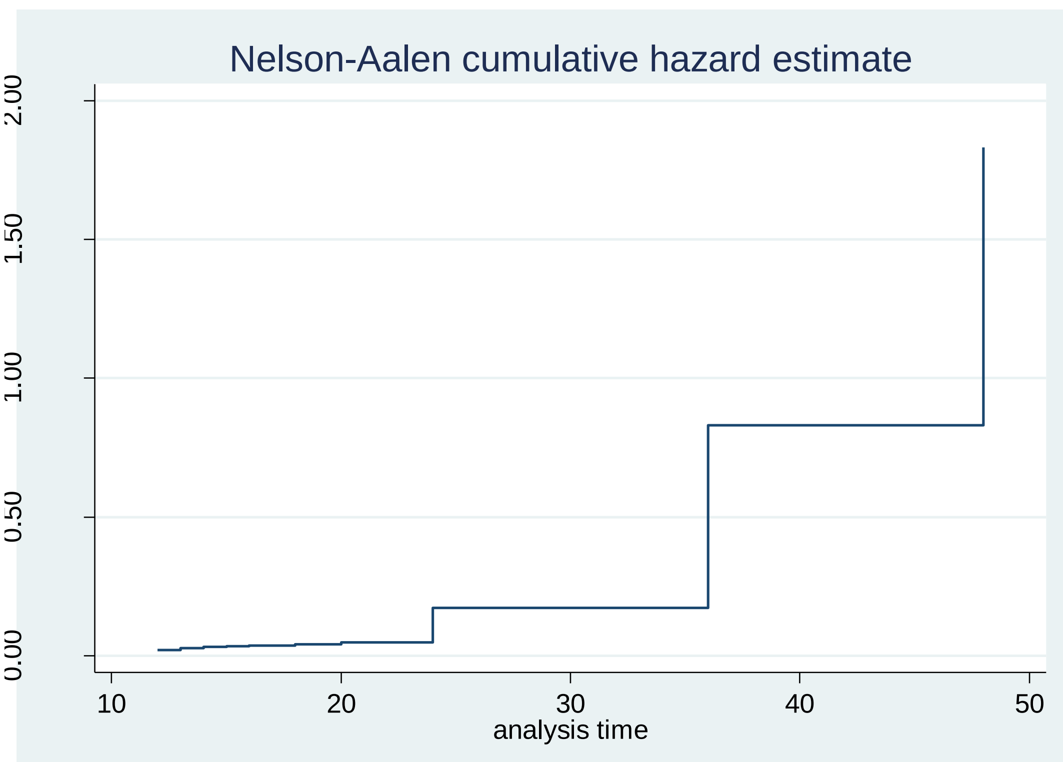


Figure 4.2 Nelson-Aalen Cumulative hazard estimate

The log-rank test, with the null hypothesis that the survival distributions being compared are equal at all follow-up times and the alternative hypothesis that the two survival curves differ at one or more points in time, was performed to see if there is a significant difference among survival experience of two or more groups of the covariates.

We would like to point out that comparing the differences among survival curves utilizing graphical method is more or less subjective and we need formal statistical tests to assess the observed difference is the real difference between groups. Hence, we employed log-rank statistical test to check for significance differences among different categories of factors that had been demonstrated by using the Kaplan-Meier estimates of the survivor functions. As we can see from Table 4.3, the results of log-rank test shows that there was no significant difference in survival experience in covariates, sex of child and place of delivery. However, the p-values of the log-rank test showed that the survival experience of children in the various categories antenatal visit, mothers educational status, toilet type used , type of cooking fuel, source of drinking water and place of residence were differ significantly (i.e. all of these covariates have P-value less than 0.05).

Table 4.3: Result of log-rank test of equality of survival distribution for the different categorical covariates

Covariates	Chi-Square	Df	p-value
Sex of child	2.785	1	0.095
Place of Residence	52.99	1	0.000
Mothers Educational Status	34.748	2	0.000
Toilet Type Used	36.253	2	0.001
Type of Cooking Fuel	26.697	3	0.004
Antenatal visit during pregnancy	59.058	1	0.000
Place of Delivery	8.670	2	0.067
Source of drinking Water	41.378	1	0.000

4.2 Results of Cox proportional hazard regression model

The Cox proportional hazard model is the most widely used procedure for modeling the relationship of covariates to a survival time by incorporating censored outcome in the analysis. It can be employed for estimating the regression coefficients, conducting statistical tests, constructing confidence intervals and making interpretation based on the hazard function. Checking the adequacy of model and its development precede interpretation of results obtained from the fitted model.

In model development procedures, fitting all possible models is computationally expensive when the number of covariates considered in study is relatively large. For this reason, the variable selection procedures given by Collet (2003).

The first step is to select covariates which are important in a study at some relaxed level of significance. Results from univariable proportional hazards Cox regression model are presented in Table 4.4 (Appendix). From the table, variables which are significant in relation to the time to child mortality at the 10 percent level of significance were included in multivariable analysis. The univariable analysis showed that not all explanatory variables are statistically important to be included in the multivariable analysis stage. Thus, the most appropriate subset of these covariates to be included in the multivariable model will be selected based on their contribution to the maximized log- partial likelihood of the model ($-2L\hat{\beta}$) . The value of $-2L(\hat{\beta})$ for the null or empty model was 2165.072. Therefore, inclusion of covariates (explanatory variables) was based on the amount of reduction of this value. Based on Table 4.4, the highest reduction in $-2L(\hat{\beta})$ was observed for type of cooking fuel. It reduces the value from 2165.072 to 2023.691. This difference is 141.381 and it is statistically significant (P-value <0.0001) when compared with percentage points of the χ^2 distribution on 3 degree of freedom. The next highest change obtained for place of residence where the difference equal to 131.194 and statistically significant. The third highest change was obtained for household type of toilet used were the difference equal to 126.893.

All potential variables that were supposed to have statistically significant impact (at P-value < 0.1) at univariable analysis were included in the initial multivariable proportional hazards model which leaded to a value of $-2L(\hat{\beta})$ to 2019.760. Thus, removal of variables from the model will be based on the increasing $-2L(\hat{\beta})$ and P-value.

Results from Table 4.4 indicate that the least important covariate in the model were place of delivery and sex of child, since the removal of these covariates led to insignificant increment (P-value 0.516 and 0.120) in the value of $-2L(\hat{\beta})$. Continuing the fitting processes by eliminating the variable place of residence and sex of child , the model consisted of the remaining six variables were fitted and the effect of eliminating variables from the model was assessed.

Table 4.5: shows the increase in $-2L(\hat{\beta})$ and P-values after eliminating the variables place of delivery and sex of child from the model. All of the covariates included in this table were significant at 5% level of significance. Hence, we obtained a multivariable model that included six covariates, namely antenatal visit, source of drinking water, mothers educational status, toilet type used, type of cooking fuel and place of residence.

Table 4.5: The Preliminary Final Model with parameter estimates and hazard ratios of the covariates

	B	SE	Wald	Df	Sig.	HR	95.0% CI for Exp(B)	
							Lower	Upper
Place of Residence Urban(1)	-.648	.272	5.674	1	.017	.523	.307	.891
Rural(2)(Ref)								
Mothers' Education Status			12.293	2	.002			
No Education(1)	1.469	.748	3.857	1	.050	4.343	1.003	18.811
Primary(2)	-.634	.270	5.504	1	.019	.531	.312	.901
Secondary&above 3(Ref)								
Toilet Type Used			8.204	2	.017			
Improved(1)	-1.301	.461	7.968	1	.005	.272	.110	.672
Shared facility(2)	.014	.798	.000	1	.986	1.014	.212	4.842
Non-improved 3(Ref)								

Type of Cooking Fuel			8.471	3	.003			
Fire wood (1)	1.252	.458	7.461	1	.006	3.497	1.424	8.588
Charcoal (2)	.989	.514	3.703	1	.045	2.688	1.982	8.359
Kerosene (3)	.844	.656	1.655	1	.198	2.326	.643	8.421
Electricity 4(Ref)								
Antenatal Visit No	1.085	.281	14.868	1	.000	2.959	1.705	5.135
Yes(1) (Ref)								
Source of Drinking Water Improved (1)	-1.050	.193	29.590	1	.000	.350	.240	.511
Non-improved 2(Ref)								

Remark: Reference category is marked by parenthesis (Ref).

Table 4.5**:** shows the Cox regression analysis by stepwise method . In the table, estimated coefficients (β's) of covariates, standard error of β estimates (SE), Wald's test statistic values, p-values of Wald's test, relative risks of covariates on child survival (e^{β}) and 95% confidence interval of relative risks are shown.

The final Cox -PH model as shown in Table 4.5 looks like this

$$h(t,x,\beta) = ho(t)\exp(x\beta')$$

$$h_i(t) = h_0(t)\exp\begin{pmatrix} -0.648PR1_i + 1.469ME1_i - 0.634ME2_i - 1.301TT1_i + 0.014TT2_i + 1.252TCF1_i \\ +0.989TCF2_i + 0.844TCF3_i + 1.085AV1_i - 1.050SDW1_i \end{pmatrix}$$

where:

- PR1 is place of residence urban.
- ME1 is mothers educational status no education and ME2 mothers educational status primary level.
- TT1 is type of toilet used improved and TT2 type of toilet used shared facility.
- TCF1 is type of cooking fuel which is fire wood, TCF2 is type of cooking fuel charcoal and TCF3 type of cooking fuel kerosene.
- AV1 is no antenatal visit during pregnancy period.
- SDW1 is source of drinking water which is improved.

The next important step is to consider variables that were non-significant at univariable and multivariable analyses for possibility of confounders. This can be checked by considering the change in coefficients of variables remaining in the multivariable model when those insignificant variables were added one at the time. A value of 20% change is generally considered as an important change in a coefficient (Hosmer and Lemeshow, 1999). Thus, the variables sex of child and place of delivery were included one at a time; the change in the coefficients of the significant variables was depicted. The results show that the percentage changes in the coefficients of the variables were by far less than 20% revealing that none of them was significant confounder. Hence, variables that were neither significant at univariable analysis nor at multivariable analysis were not confounders of the main factors in the preliminary model of Table 4.5.

Table 4.6: Percentage changes in the coefficients of the variables included in Table 4.4, when the variables those were not significant in the univariable proportional hazards Cox regression models are added one at a time.

Covariates/factors	Sex of child	Place of delivery
Place of Residence	-0.07	0.09
Mothers' Education Status	0.11	0.19
Toilet Type Used	-0.13	0.16
Type of Cooking Fuel	-0.90	0.00
Antenatal Visit	-0.43	-0.20
Source of Drinking Water	-0.90	-0.12

The final step in model development strategy was consideration of interaction terms that may be useful in the improvement of the model fit. Thus, all possible interactions among covariates that were significant at multivariable analysis were formed and the significance of adding each of the interaction terms in the main effects model, one at a time, was checked. The SAS results from Table 4.8 indicate that none of` the interaction terms were significant at 5% level. Hence, the last model was the one which contains only the main effects . However, the interpretation based on this model should not be tested until the basic assumptions associated with the proportional hazards Cox regression model have been checked.

4.3 Model Adequacy

At this point we have a preliminary model and the next step is to assess its fit and adherence to key assumptions before we move to interpretation of the results obtained. We start here first by checking the overall goodness of fit using r-square and LR, Score and Wald tests. We then proceed to check the proportionality assumption for each covariate included in the final model.

4.3.1 Overall Goodness of Fit

The value of $R^2{}_p$ is calculated as:

$$R^2{}_p = 1 - \exp[\frac{2}{n}(L_0 - L_m)] = 1 - \exp[\frac{2}{2120}(-1082.536 + 1013.529)] = 0.063$$

Due to the presence of high censoring the value of $R^2{}_p$ is very low and indicates that the model is adequate.

Table 4.7: The Likelihood Ratio, Score and Wald tests for overall measures of goodness of fit of the preliminary final model in Table 4.5.

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	121.1679	6	<.0001
Score	108.1439	6	<.0001
Wald	102.8381	6	<.0001

Testing Global Null Hypothesis: BETA=0

As we can see from Table 4.7: the p-values associated with the likelihood ratio, Score and Wald test statistics are all less than 1% indicating goodness of the fitted model at 5% level of significance.

4.3.2 Testing the proportional hazards assumption

Two basic assumptions of the Cox-PH model are log-linearity and proportional hazards. Just as with other regression models, these assumptions need to be examined. Since all covariates used in the final model are categorical, there is no need of checking linearity assumption.

The validity of Cox regression analysis relies heavily on the assumption of proportionality of the hazard rates of individuals with distinct values of a covariate. If the proportionality assumption holds the LOWESS smoothing curve should be approximately horizontal line around zero and the distribution of residuals over time is random, with no particular trend with time. Alternatively, we can run a model with each covariate (individually) by introducing a time-dependent interaction term for that covariate. If the proportional hazards assumption is valid for the covariate, the time-dependent interaction term should not be significant. The following table display the SAS output of test of proportionality assumption.

Table 4.8: Result of test of proportionality assumption for each covariate in the final model

Analysis of Maximum Likelihood Estimates						
Variables	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>Chisq	Hazard Ratio
Place of Residence	1	0.65993	0.25796	6.5448	0.0105	1.935
Mothers Education status	1	-0.14136	0..21051	0.4509	0.0500	0.868
Toilet Type used	1	0.47784	0.23641	4.0854	0.0433	1.613
Type of Cooking fuel	1	-0.36084	0.13775	6.8616	0.0088	0.697
Antenatal Visit	1	-1.20911	0.27070	19.9503	<.0001	0.298
Source of Drinkin Water	1	1.01003	0.19156	27.7998	<.0001	2.746
PRes *(log(time))	1	0.14754	0.09193	2.5754	0.2085	1.159
MEdu*(log(time))	1	-0.21120	0.22840	0.8551	0.3551	0.810
Toilet*(log(time))	1	0.14534	0.16919	0.7380	0.3903	1.156
TCook*(log(time))	1	0.00725	0.14246	0.0026	0.9594	1.007
AVisit*(log(time))	1	-0.05220	0.13115	0.1585	0.6906	0.949
SoDrinWater*(log(time))	1	0.19142	0.42633	0.2016	0.6534	1.211

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
Test proportionality	8.1448	6	0.4195

From **Table 4.8:** above we can see that Wald chi-square values and the corresponding p-values for each covariate. Since the p-values for each interaction of covariate with logarithm of time are greater than 0.05, indicates that the proportionality assumption is satisfied. The global fit test also shows that the Wald chi-square test statistic is not significant which indicates that the proportional hazards assumption is not violated. Annex B (Figures 5-8) show the plots of the scaled Schoenfeld residuals for each covariate against time. All LOWESS smoothed curves seem to approximate a horizontal line through zero. The residuals look random showing no trend with time. Hence the proportionality assumption is satisfied.

4.4 Interpretation and Discussion of the Results

The study assessed child mortality and examined the environmental determinants of child mortality in Ethiopia. In survival analysis the measure of effect is the hazard ratio. It is interpreted in the same way as the odds ratio. The higher the hazard ratio the lower is the survival probability, and vice versa. For an exposed group the hazard ratio is high, the survival probability would be equivalently low. From the final model in Table 4.5 we obtained six significant main effects: mothers' educational status, toilet type used, type of cooking fuel, antenatal visit, place of residence, and household source of drinking water.

Place of residence has a negative significant association with child mortality. After adjusting other covariates, risk of dying for a child born in a family live in rural area is higher by 47.7% relative to those born in a family live in urban. The 95% confidence interval (0.307, 0.891) implies that the risk of death of children who born in a family lives in urban is 0.307 as low and 0.891 as high as those in the reference group.

After adjusting other covariates, the estimated coefficients of improved and shared toilet types are -1.301 and 0.014 respectively. The hazard ratio or relative risk of the covariate toilet type used improved is 1.014 and it is as little as 0.272. It means that the hazard rate of child reduce by 72.8 % in household with sanitary latrine as compared with the household without sanitary latrine (shared facility). In favor of this finding ,Klaauw and Wang(2004)**,** suggest that good public sanitation systems may constitute a more important preventive aspect of child survival. In the latter study of Kabir & Amin (2013) **,** in Bangladesh also highlights that the households with sanitary latrines have low risks of child mortality.

In urban Kenya, access to modern sanitation facilities (flush toilets) reduces diarrhea prevalence in urban areas and ultimately reduces the child mortality (Mutunga (2004)). In a study of Balk *et al*. (2005) , the principal component analysis is used to combine the correlated variables which influence on mortality. From this analysis it is found that the mortality is correlated positively with the complete lack of toilet facilities and negatively with access to flush toilets. It is also suggested by Vos *et al*. (2005) that the availability of better sanitation will decrease the probability of infant death since better sanitation and drinking water access by the household should positively improve hygienic and health conditions for all members.

For the mothers' education level, we have three categories (no education, primary level and secondary and above). Taking secondary and above as reference group and the risk of death for children, mothers' education who had no education and primary relative to mothers' education level secondary and above are 4.343 and 0.531, respectively. This result shows that children with mother's having no education level were 4.343 times more likely to die than those with secondary and above education level controlling for other variables in constant. These results clearly indicate that the child survival is increasing with increasing of parent's education and it is also found that parent's education has significant effect on child survival. This result may be due to fact that child survival is mainly affected by environmental factors and educated parents may be more conscious to the environment where child grow up.

The risk of dying for a child born in a family without access to improved (pipe) source of drinking water is higher by 35% relative to those born in a family with access to improved drinking water. The 95% confidence interval (0.240, 0.511) implies that the risk of death of children whose source of water is not improved water is 0.240 as low and 0.511 as high as those in the reference group. This result is in accordance with Unger (2013). But other researchers depicts that source of drinking water has no significant effect on child mortality.(Abdul Hamid Chowdhury 1, Mohammad Emdad Hossian 2, Md. Musa Khan 3,Mohammad Nazmul Hoq 4, Asian Journal of Social science and humanities , Bangladish Vol.2 No.2 May 2013).

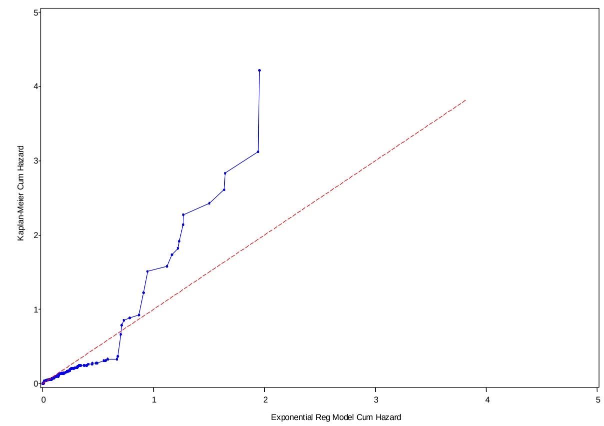
The estimated hazard ratio for children whose mother's attended antenatal visits during pregnancy when compared to those mothers who did not attend antenatal visit was 2.959 (95% CI:1.705-5.135) keeping effects of other covariates constant. That is, children whose mothers attended antenatal visits during pregnancy had 95.9% lower risk of child mortality than those who did not attend antenatal visit. In other words the risk of death for children, mother's who did not attend antenatal visit was 2.959 times relative to whose mother's attended antenatal visits during pregnancy.

With regard to households' source of cooking fuel, the risk of dying for children with fire hood, charcoal and kerosene were 3.497, 2.688, and 2.326, respectively. These figures shows that the risk of death children whose house hold cooking type of fuel, fire hood is 3.497 times relative to households type of cooking fuels is electricity. In other words, after adjusting other covariates, the hazard of death of children with households use fire hood cooking is 3.497 times higher than households use electricity (adjusted HR=3.479, 95% Cl:1.424-8.588). The hazard death of children for household use charcoal cooking is 2.688 times higher than households use electricity (adjusted HR=2.688, 95% Cl:1.982-7.36). The hazard death children for households use kerosene cooking fuel is 2.326 times higher than households use electricity(adjusted HR=2.326,95%Cl:1.643-8.421). All these findings are consistent with Hala (2002), Klaauw and Wang (2004) and Jacoby and Wang (2003).

4.5 Parametric Regression Modelling of Survival of Child

4.5.1 Model Selection for Survival of Child

For the child mortality the parametric regression models were fitted. We consider model comparison after adjusting for the effect of covariates. In this case the graphical displays are based on the Cox-Snell plots. That is, if the model is good fitted, the plot of Cox-Snell residuals versus Nelson-Aalen cumulative hazard estimates should lie along the 45 degree diagonal line that passes through the origin. Using all the covariates in the study, we fitted two parametric regression models which are Exponential and Weibull models with the corresponding AIC and BIC values. Here we present the Cox-Snell plots for model comparison in Figures 4.2 to 4.3.



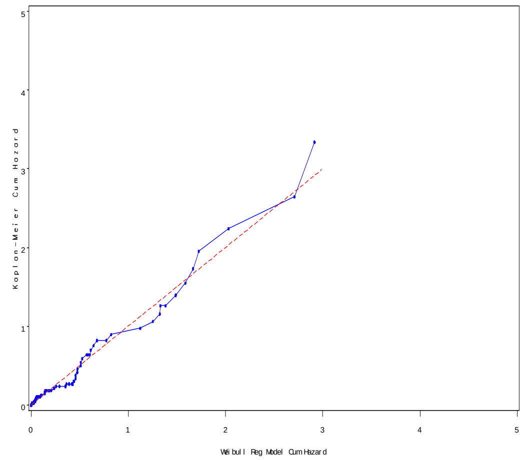


Figure 4.3: The Cox Snell plot after fitting Exponential regression mode

Figure 4.4: The Cox Snell plot after fitting Weibull regression model

As we can see from the above Cox-Snell plots that Weibull regression model seems the best fit among the two models. But graphical methods may not assure the result. The common applicable criterion to select the model is the Akaikie information criterion (AIC) statistic proposed by Akaikie (1974). So, In addition to the graphical comparison of the two parametric regression models, we used Akaikie information criterion (AIC) and Bayesian information criterion (BIC) to choose the best model out of the two possible models. The STATA output of the two parametric survival regression models are displayed in appendix B with the corresponding AIC and BIC values.

Table 4.9: Statistical results for model comparison

Model	**Observation**	**ll (null)**	**ll (model)**	**Df**	**AIC value**	**BIC value**
Exponential	2120	-600.9143	-542.6691	11	1107.338	1169.589
Weibull	2120	-395.6862	-324.8979	12	673.7959	741.7059

According to the results in Table 4.9 above, the Weibull regression model with the smallest value of AIC and BIC seems to be the best fit of the two models. Nevertheless,

the results of cox-snell were consistent with the results based on Akaikie's information criterion. Thus, the Weibull regression model was preferable to discuss the effect of covariates on the survival of Child mortality in Ethiopia.

4.5.2 Univariate unadjusted Weibull regression model

As Weibull regression is selected, according to the Weibull analysis of single covariate, the selected risk factors for further analysis and interpretation are made here below. To have an idea about the individual effects of the different explanatory variables on survival of children, we fitted Weibull regression model separately for each explanatory covariate.

Table 4.10: The result of un-adjusted univariate analysis using Weibull regression model

Covariate	**Hazard ratio**	**Std.error**	Z	P>|z|	**-2*LL**	**95% CI**	
						Lower	**Upper**
Sexochild	1.292267	.195206	1.70	0.090	845.996	.961108	1.737529
PResidence	1.731415	.4757156	2.00	0.046	686.865	.010483	.966698
MEducationStatus	1.424048	.2234352	2.25	0.024	652.547	1.047056	1.936776
ToiletType Used	1.634508	.3430492	2.34	0.019	652.214	1.083272	2.466247
TCookinkFuel	.7454722	.0899263	-2.44	0.015	652.214	.5885062	.944304
AntinatalVisit	.3071224	.0924874	-3.92	0.000	791.434	.1702076	.5541713
PoDelivery	.8110552	.2047389	-0.83	0.407	731.372	.4945124	2.330221
SoDrinkinWater	3.349726	.6283278	6.44	0.000	652.214	2.319233	4.838092
_cons	8.80e-11	1.25e-10	-16.35	0.000		5.48e-12	1.41e-09
/ln_p	1.683363	.0544476	30.92	0.000		1.576648	1.790079
P	.383632	.2931259				4.838709	5.989924
1/p	2.60666	.0101135				.166947	.2066667

As we can see from the above Table 4.10, it shows that covariates like mothers' educational status, households toilet type used, households type of cooking fuel, antenatal visit during pregnancy, place of residence, and household source of drinking water are statistically significant at 5% level of significance. But covariates like sex of children and place of delivery are not significant. The risk factors those were statistically significant included in the final Weibull regression model for the prediction of survival probability of child mortality.

4.5.3 Multivariable Analysis Weibull Regression Model

When there are a number of explanatory variables of possible relevance, the effect of each term cannot be studied independently one the others. The effect of any given term therefore depends on the other terms currently included in the model. However, in the univariate analysis technique the relations that are obtained for one factor do not take into account the other factors. So the multivariable analysis is used to know the most important factors associated with mortality of children in relation to the covariates included in the model. After fitting the univariate weibull survival regression analysis the next step is selecting the most important variables to fit the multivariate weibull regression model. In order to select the most important covariates in the final model, we used stepwise variable selection.

Table 4.11: Parameter estimates of the final multivariate weibull regression model

Covariate	Hazard ratio	Std.error	Z	P>|z|	95% CI	
					Lower	Upper
PResidence(1)	.5877572	.168221	1.86	0.043	.3354108	.929957
PResidence(Ref)						
Mothers' Educational status						
No education(1)	.5081227	.1555982	-2.21	0.027	.2788125	.9260298
Primary(2)	.4457565	.1128284	-3.19	0.001	.2714216	.7320673
Secondary&above(Ref)						
Toilet type used						
Improved(1)	.3312088	.1464452	-2.50	0.012	.1392322	.7878869
Shared facility(2)	1.142547	.8782884	0.17	0.862	.2532478	5.154685
Non-improved(Ref)						
Type of Cooking Fuel						
Fire wood (1)	3.323802	1.29274	3.09	0.002	1.550865	7.123548
Charcoal (2)	2.752342	1.269107	2.20	0.028	1.11484	6.79504
Kerosene (3)	2.149882	1.382608	1.19	0.328	.6095359	7.582803
Electricity(Ref)						
Antenatal Visit No (0)	3.174416	.9498236	3.86	0.000	1.765932	5.706287
Yes(1) (Ref)						
Source of Drinking						

Water Improved (1)	.2843618	.0539307	-6.63	0.000	.1960814	.4123881
Non-improved(Ref)						
_cons	2.00e-09	2.44e-09	-16.43	0.000	1.84e-10	2.19e-08
/ln_p	1.713488	.0558461	30.68	0.000	1.604032	1.822944
P	.54828	.3098497			4.973042	6.190057
1/p	1.8238	.0100655			.1615494	.2010842

Remark: The reference category is marked by parenthesis (Ref).

4.5.4 Assessment of Adequacy of the Weibull Regression Model

To assess the adequacy of weibull regression model, we used the likelihood ratio test presented in Table 4.12 below and it illustrated that the model was significantly fit the data of child mortality and in using the log likelihood values of the null model and the full model, it can be seen that the model has a significant improvement after the covariate is incorporated in the model.

Table 4.12: The likelihood ratio and significance of the Weibull regression model

Log likelihood (intercept only)	Log likelihood (Model)	LR chi-square	DF	Prob > chi2
-395.6868	-325.96576	139.44	6	0.0000

4.5.5 Interpretation and Discussion of the Weibull Regression Model

All variables significant in Cox- model were also significant in the Weibull models with the expected signs. Households with access to safe water have significantly lower mortality rates. Access to sanitation facilities is also significantly related to child mortality. Children born in household with either flush toilets or pit latrines have lower mortality rate than those born in households without any toilet facility.

The result of this study also showed that infants whose parents use unprotected (non-improved) source of drinking water have less survival chance than those who use improved source of drinking water. A study in China showed that access to safe water or sanitation reduces child mortality risks by about 34% in rural areas, which means access to safe water/sanitation, and immunization reduce diarrhea incidence in rural areas

(Jacoby and Wang (2003)). In Kenya, Mutunga (2004) found that child survival was found better for those who had access to safe drinking water and sanitation facilities.

Several single country studies based on micro data have shown the impact of individual's or household's endowments of resources , access to safe drinking water, and improved sanitation on infant and child mortality (Kembo and Van Ginneken (2009) (Zimbabwe); Mesike and Mojekwu (2012) (Nigeria); Gemperli et al. (2004) (Mali); Nuwaha et al. (2011) (Uganda); Manda (1999) (Malawi); Kandala and Ghilagaber (2006) (Malawi); Adeyemi et al. (2008) (Nigeria); Adebayo and Fahrmeir (2012) (Nigeria); Ogunjuyigbe (2004) (Nigeria); Wang (2003) (Ethiopia).

Mother's education is the most important determinant of child mortality among the mother's characteristics that are considered in this study. Children whose mothers' have no education are 50.81% likely to die as infants compared with children whose mothers have secondary and above education (HR = .5081; 95% CI=(.2788-.9260)). The risk of death of children whose mothers' have primary education level was 44.58% compared to the reference group secondary and above education level (HR=.4457; 95% CI=(0.2714-0.7320)) keeping effects of other covariates constant.

With regard to source of cooking fuel, children born in households using high polluting fuels (fire woods) as their main source of cooking fuel have higher mortality rates as compared to those using low polluting fuels (electricity). Higher incidence of respiratory infections which are responsible for child deaths is expected in households which use "dirty" fuels as opposed to those using clean cooking fuels. This finding is consistent with Mutunga, clive j. Kenya Institute for Public Policy Research and Analysis (KIPPRA) (2004).

Finally, from the Weibull model estimates, the shape parameter γ which is shown as ρ in STATA has a value of 0.548 which implies that the hazard rate decreases monotonically with time or in other words there is negative time dependence. This means that children face a higher hazard (mortality rate) in the initial birth day than in later periods.

4.5.6 Discussion of the Results

The main aim of this study was to identify factors of child mortality in Ethiopia using the nationally representative of 2014, EDHS data. Both univariate and multivariate statistical analyses were employed to examine factors affecting child mortality. The analyses revealed that environmental variables were statistically significant effect on child mortality in Ethiopia. The variables influencing child mortality are mothers' educational status, source of drinking water, place of residence, household type of toilet used, type of cooking fuel and antenatal visits. But covariates like sex of child and place of delivery were not statistically significant on child mortality in Ethiopia.

The findings of this study showed that children whose mothers attended antenatal visits during pregnancy had lower risk of child mortality than those who did not attend antenatal visits. A study in the Gaza Strip, occupied Palestinian territory, by Antai D. and Moradi T. (2010) found that newborn babies born to mothers who attended fewer than four antenatal sessions during pregnancy had a risk of dying that was almost twice that of those born to mothers who attended antenatal session four or more times. A study in Indonesia also revealed that the risk of children death was higher among women who did not attend antenatal care visits during pregnancy (Kamal S.M.M. (2012). A study in Ethiopia by Desta, M. (2011) showed that child mortality was associated with antenatal care follow-up: there was better survival with at least one antenatal care follow-up. Thus, antenatal care follow-up is a prominent predictor of survival time of children.

The result of this study also showed that children whose parents used unprotected drinking water have less survival chance than those who use piped drinking water or improved source of drinking water. A study in China showed that access to safe water or sanitation reduces child mortality risks by about 34% in rural areas, which means access to safe water/sanitation, and immunization reduce diarrhea incidence in rural areas (Jacoby and Wang (2003)). In Kenya, Mutunga (2004) found that child survival was found better for those who had access to safe drinking water and sanitation facilities. A study in Egypt by Hala (2002) showed that access to municipal water decreases sanitary risks. Access to municipal water and improved sanitation facilities had significant positive impact on children mortality (Unger (2013)). Therefore, higher mortality rates are experienced in households that have access to unprotected source of drinking water drinking water.

There was higher mortality in children whose mothers' were not educated or had primary education than children whose mothers were attending secondary and above education in this paper. The study in rural china by Jacoby and Wang (2003) showed that a higher maternal education level reduces child mortality and that female education has strong health externalities. In Nigeria a similar study showed that women's average educational level in their community exerts a great influence on child survival (Mesike, C.G., Mojekwu J. (2012)). A study In Ethiopia by Wang (2003) also showed that female education attainment has significant effect on reducing infant mortality. Therefore, improving the knowledge of mothers in the societies is important to reduce risk of child death. Twum et al. (2011) using the result of 2009 Burkina Faso DHS , indicated that children born to mothers with higher educational level associated with lower risk of infant and child mortality as compared to children born to mothers with primary education level or non-educated.

The probability of dying child for females compared to males found in this study was the same. It was not significant impact on environmental determinant of child mortality. but other study showed that there is a significant impact on child mortality. Likewise, more boys die before their first birth day than girls in Kenya (Hill et al., 2001).

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The study has empirically examined the environmental determinants of child mortality in Ethiopia using survival analysis method. It has utilized the national representative data from the Ethiopian Demographic and Health Survey (EDHS) - 2014. The study employed survival statistical analysis to determine risk factors associated with child mortality in Ethiopia. Both Cox Proportional hazard model and Weibull regression model analysis techniques have been applied to identify the important predictors of child survival.

The results from the Kaplan-Meier estimate showed that most of the deaths occurred during the first birth days of life. Results based on Proportional Hazards model and weibull model revealed that environmental factors had statistically significant effect on child mortality. Specifically, the study demonstrated that various factors such as mothers' education, household source of drinking water, antenatal visit, place of delivery, type of cooking fuel and type of toilet used had statistically significant impacts on the survival experience of children. But covariates like sex of child and place of residence were insignificant on survival of child.

The two parametric regression models: Exponential and Weibull regression models, for survival probability of children were compared. The Weibull regression model was found to better fit to the data. The findings further suggested the following: Mothers' educational and households source of drinking water had a significant effect on survival of child, that is, child mothers' who had primary, secondary and above educational level were lower risk of mortality than mothers' who had no education level. Children whose parents use non-improved has less survival chance than those who use improved source of drinking water. With regard to source of cooking fuel, children born in households using high polluting fuels (fire woods and charcoal) as their main source of cooking fuel have higher mortality rates as compared to those using low polluting fuels (electricity). Children born in household with either flush toilets or pit latrines have lower mortality rate than those born in households without any toilet facility.

5.2 Recommendations

Based on the study findings and keeping the limitations in mind, the study forwarded the following recommendations.

- ☞ Greater efforts need to be put in place to ensure provision of basic services like water for all. Availability of safe sources of drinking water will significantly reduce child mortality and therefore investments in this sector will be rewarding.
- ☞ Access to sanitation facilities like constructing toilets entail a private cost but do have significant social benefits. The government should work closely with both the private sector and civil society to ensure that households have universal access to sanitation facilities do great extend reduce the number of infant deaths. In addition, the proposed housing policy should make it mandatory for each housing unit to have a sanitation facility such that all households have access to sanitation facilities.
- ☞ The government policy should be focused towards promoting the use of low polluting fuels and in particular discouraging the use of firewood and charcoal. Through the use of economic instruments, incentives should be created for promotion of cleaner fuel sources. This will also create employment opportunities which will translate into increased earnings and reduced poverty.
- ☞ In general, the government policies should focus on improving child survival and health intervention policies should revise and implement to achieve the Millennium Development Goals (MDGs) of reducing child mortality.

References

1. Abdul Hamid Chowdhury 1, Mohammad Emdad Hossian 2, Md. Musa Khan3,Mohammad Nazmul Hoq 4, (2013): "Environmental determinants of Under-Five Mortality " Asian Journal of Social Science and Humanities, 2(2).
2. Adebayo, Samson B. and Ludwig Fahrmeir, (2012) “Analyzing Child Mortality in Nigeria with Geoadditive Discrete Time Survival Methods”, Statistics in Medicine **24** 709-728.
3. Adeyemi, S L, U A Raheem, and F B Olorunfemi, (2010) “Under-Five Mortality and the Environment of Health in the Third World: A Nigerian Example”, Journal of Human Ecology **24**, 117–124.
4. Akaike, H. (1974). A new look at the statistical model identification", IEEE Transactions on Automatic Control, **19** (6): 716–723, doi:10.1109/TAC.1974.1100705, MR 0423716.
5. Antai D. and Moradi T. (2010). Urban Area Disadvantage and Under-5 Mortality in Nigeria: The Effect of Rapid Urbanization. *Environ Health Perspect, **118**(6), 877–883.*
6. Becares, L. et al. (2013). Ethnic density and area deprivation: Neighbourhood effects on Maori health and racial discrimination in Aotearoa/New Zealand. Soc Sci Med, **88**, 76-82.
7. Blunch, Niels-Hugo, (2013) “Staying Alive: Adult Literacy Programs and Child Mortality in Rural Ghana”, World Development **42**, 114–126
8. Bradley RH, Corwyn RF,(2012) Socioeconomic status and child development. Empirical Comparisons of Calverton, Maryland, USA: Central Statistical Agency.
9. Collett, D. (2003). *Modelling survival data in medical research,* Second edition. Chapman and Hall/CRC, London.
10. Cox, D. R., and Snell, E. J.(1968). A general definition of residuals with discussion. *Journal of the Royal Statistical Society. Series B **30** (1968),248-275.*
11. Cox, D.R. (1972). “Regression models and life Tables (with Discussion).” *Journal of the Royal Statistical Society, **34**: 187-220.*

12. Desta, M., (2011) Infant & Child Mortality in Ethiopia: The role of Socioeconomic, Demographic and Biological factors in the previous five years period of 2000 and 2005.
13. Espo, M. (2002). Infant mortality and its underlying determinants in rural Malawi. dissertation, University of Tampere Medical School.
14. Ethiopian Demographic and Health Survey 2011. Addis Ababa, Ethiopia .
15. Grambsch P.M and Therneau T.M (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika; 81(3): 515-52.*
16. Hala A. (2002) "The effect of water and sanitation on child mortality in Egypt**".**
17. Hill, K., Bicego, J. and Mahy, M. (2011). Childhood mortality in Kenya: An examination of trends and determinants in the late 1990s to mid 2008s.
18. Hmwe H. et al. ,(2013). Association of Urban Slum Residency with Infant Mortality and Child Stunting in Low and Middle Income Countries.
19. Hosmer, D.W. and Lemeshow S. (1999). *Applied Survival Analysis.* John Wiley and Sons, Inc., New York.
20. Hosmer, D.W., Lemeshow, S. and May, S. (2008), Applied survival analysis regression modeling of time to event data. John Wiley and Sons, Inc., New York. http://www.worldbank.org/en/country/ethiopia.
21. International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Sweden.
22. Ikamari, L. D. E , (2013) Factors affecting Child Survival in Kenya. Unpublished PhD Thesis. Australian National University. Canberra.
23. Ikamari, L.D.E. (2013). "Regional variation in neonatal and post-neonatal mortality in Kenya. AfricanPopulation Studies, **27** (1): 1-15.
24. Jacoby, H. and L. Wang (2003) "Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach" World Bank,Washington D.C.
25. Kabir M, Amin R. (2013), Factors influencing child mortality in Bangladesh and their implications for the National Health Programme. Asia-Pacific Population Journal: 8(3): 31-46.
26. Kamal S.M.M. (2012). Maternal Education as a Determinant of Neonatal Mortality in Bangladesh. *Journal of Health Management **14**(3): 269–281.*

27. Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc 1958; **53**:457-81.*
28. Kimani R.R. E., Kimani E.R.R. (2012), "Determinants of under-five mortality in rural and urban Kenya". Rural and Remote Health **12**: 1812.
29. Klaauw, van der B. and L. Wang (2003) Child Mortality in Rural India', World Bank Working Paper, Washington DC: World Bank.
30. Kleinbaum, D.G. and Klein,W. (2005). *Survival Analysis a self learning text,* Second edition. Springer Science+Business Media, Inc., New York.
31. Kombo, J. & Ginneken, V. (2009). Determinants of infant and child mortality in Zimbabwe: Result of multivariate hazard analysis.
32. Kumar, PP, File G (2010). Infant and Child Mortality in Ethiopia: A Statistical Analysis Approach, *Ethiopian Journal of Education and Science*, **5**(2):51-7.
33. Kyei, Kwabena A., (2012), "Determinants of Childhood Mortality in South Africa: Using Categorical Data Modelling", Journal of Human Ecology **37** (47–56).
34. Mesike, C.G., Mojekwu J. N. (2012), Environmental Determinants of Child Mortality in Nigeria. *Journal of Sustainable Development,* **5**(1): 65-75.
35. Mutunga CJ. ,(2004) Environmental determinants of child mortality in Urban Kenya. Abdus Salam ICTP; Trieste.
36. Mosley, W. and Chen, L. (1984) "An Analytical Framework for the Study of Child Survival in Developing Countries", *Population and Development Review* **10**: 25 -45. http://dx.doi.org/10.2307/2807954
37. Nuwaha, Fred, Juliet Babirye, Olico Okui, and Natal Ayiga.(2011), "Understanding socio-economic determinants of childhood mortality: A retrospective analysis in Uganda", BMC Research Notes (**4)** 484.
38. Oestergaard MZ, Inoue M, Yoshida S, Mahanani WR, Gore FM, Cousens S,Lawn JE, Mathers CD: (2009) Neonatal mortality levels for 193 countries with trends since 1990: a systematic analysis of progress, projections, and priorities. PLoS Med 2011, **8**(8):e1001080.
39. Ombok, M., Adazu K., Odhiambo, F., Bayoh, N., Kiriinya, R., Slutsker, L., Hamel, M.J., Williamson, J., Hightower A., Laserson, K.F., Feikin, D.R. (2010),

"Geospatial distribution and determinants of child mortality in rural western Kenya ". Tropical Medicine and International Health **15**, (4): 423-433.

40. Ringheim, K. Teller, C. and Sines, E. (2009), Ethiopia at a crossroads: Demography,Gender, and Development. Washington, D.C.,Population Reference Bureau [PRB].
41. Schoenfeld, D. (1982), Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika; (7): 145-153.
42. Susman, A. S. (2012), Child Mortality Rate in Ethiopia. *Iranian J Publ Health, **41**(3): 9-19.*
43. Trussell, J. & Slough, H. C. (2003). A hazard model analysis of the covariance of infant and child mortality in Sri Lanka. *Demography, 20*(1), 1-26.
44. Unger, A. (2013), Children's health in slum settings. Arch Dis Child. Doi: **10**.1136/archdischild- 2011 -301621.
45. UNICEF (2012), Situation Analysis of children and women in Kenya. Unicef, Nairobi.
46. UNICEF (2012), Levels & Trends in Child mortality, estimate developed by the UN Inter_agency group for Child Mortality Estimation.
47. UNICEF (2013), "Levels & Trends in Child Mortality", United Nations Children's Fund
48. UN-IGME (2011), Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation Report 2011.
49. Union For African Population Studies (2009), 'Child Mortality and Socioeconomic Status in Sub-Saharan Africa', viewed 02 February 2011, from http://www.uaps.org/journal/VOL 19 N1/1.
50. United Nations Children's Fund (UNICF) ,(2011), *The state of world's children Maternal and newborn health*.
51. Vos R, Cuesta J, Leon M, Lucio R, Orsero J. (2005), Reaching the Millennium Development Goal for child mortality: improving equity and efficiency in Ecuador's health budget. The Hague and Quito **[28]** Claeson M, Bos ER

52. Wafula, S.W., Ilamari, D.E., and K'Oyugi, B.O., (2012). "In search for an explanation to the upsurge in infant mortality in Kenya during the 1988-2003 period. BMC Public Health **(12):** 441.
53. Weibull, W. (1939). A Statistical Theory of the Strength of Materials. Ingenioers vetenskaps akakemien Handlingar, **151**, 29 3-297.
54. Weibull, W. (1951). A Statistical Distribution of Wide Applicability. Journal of Applied Mathematics, **18**, 29 3-297.
55. Wilm Quentin, et al., (2014). Inequalities in child mortality in ten major Africa cities, BMC Medicine 2014, **12**:95 doi: 10.1186/1741-7015-12-95.
56. WHO (2006), in Global Burden of Disease and Risk Factors, Comparative quantification of mortality and burden of disease attributable to selected risk factors, eds Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL (The World Bank, Washington, DC), pp 241–396.
57. WHO. (2010). Countdown to 2015 decade report (2000-2010): Taking stock of maternal, newborn and child survival. Geneva, Switzerland.
58. World Data Bank (2013): World Development Indicators. The World Bank Group.
59. World Health Organization (2005). *World health report 2005*: Make every mother and child count. Geneva: WHO. **67**, Maternal Health and child Mortality.
60. Younger (2007), " Environmental Determinants of Childhood Mortality in South Africa: Using Categorical Data Modelling", Journal of Human Ecology .**(37)** 42-56.

Appendixes

Annex A.

Table 4.1: Category variables coding

		Frequency	(1)[h]	(2)	(3)
SexoChild[b]	1=Male	1134	1		
	2=Female	986	0		
PResidence[b]	1=Urban	388	1		
	2=Rural	1732	0		
EducationStatus[b]	1=No Education	1547	1	0	
	2=Primary	400	0	1	
	3=Secondary&above	173	0	0	
ToiletType[b]	1=Improved(not shared)	15	1	0	
	2=Shared facility	24	0	1	
	3=Non-Improved	2081	0	0	
TCookinkFuel[b]	1=Fire Wood	1783	1	0	0
	2=Char Coal	224	0	1	0
	3=Kerosene	14	0	0	1
	4=Electricity	99	0	0	0
AntinatalVisit[b]	0=No	1759	1		
	1=Yes	361	0		
PoDelivery[b]	1=Home	1741	1	0	
	2=Health Center	372	0	1	
	3=Other	7	0	0	
SoDrinkinWater[b]	1=Improved	1105	1		
	2=Non Improved Source	1015	0		

Table 4.2: Summary of some important environmental characteristics of child Mortality in Ethiopia.

Covariates/ Factor	Category	Death	Censored		Total
			Censored	Censored Percent	
Sex of child	1(Male)	86	1048	92.4%	1134
	2(Female)	95	891	90.4%	986
Place of residence	1(Urban)	75	313	80.7%	388
	2(Rural)	106	1626	93.9%	1732
Mothers Education Status	No Education(1)	79	1468	94.9%	1547
	Primary(2)	60	340	85.0%	400
	Secondary &above (3)	42	131	75.7%	173
Type of toilet facility	Improved(not shared)(1)	15	0	0.0%	15
	Shared facility(2)	2	22	91.7%	24
	Non-Improved(3)	164	1917	92.1%	2081
Type of cooking Fuel	Fire Wood(1)	142	1641	92.0%	1783
	Char Coal(2)	14	210	93.8%	224
	Kerosene(3)	3	11	78.6%	14
	Electricity(4)	22	77	77.8%	99
Antenatal Visit during pregnancy	No (0)	106	1653	94.0%	1759
	Yes(1)	75	286	79.2%	361
Place of Delivery	Home(1)	103	1638	94.1%	1741
	Health Center(2)	77	295	79.3%	372
	Other(3)	1	6	85.7%	7
Source of Drinking water	Improved(1)	37	1068	96.7%	1105
	Non Improved Source(2)	144	871	85.8%	1015

Table 4.4: Results of the univariable proportional hazards Cox regression model

	B	SE	Wald	df	Sig.	Exp(B)	-2LogL	LR(Sig)
SexoChild	-.243	.156	2.417	1	.120	.784	2142.832	0.670
PResidence	-.650	.272	5.715	1	.017	.522	2033.878	0.0025
EducationStatus			11.869	2	.003		2053.296	0.035
EducationStatus(1)	1.438	.748	3.689	1	.055	4.211		
EducationStatus(2)	-.625	.271	5.329	1	.021	.535		
ToiletType			9.794	2	.007		2038.179	0.000
ToiletType(1)	-1.456	.469	9.619	1	.002	.233		
ToiletType(2)	-.137	.807	.029	1	.865	.872		
TCookinkFuel			6.480	3	.090		2023.691	0.000
TCookinkFuel(1)	1.105	.468	5.577	1	.018	3.018		
TCookinkFuel(2)	.832	.525	2.515	1	.113	2.299		
TCookinkFuel(3)	.744	.658	1.275	1	.259	2.104		
AntinatalVisit	.961	.302	10.144	1	.001	2.614	2115.474	0.002
PoDelivery			1.324	2	.516		2138.481	0.422
PoDelivery(1)	-.822	1.008	.666	1	.415	.439		
PoDelivery(2)	-1.049	1.048	1.003	1	.317	.350		
SoDrinkinWater	-1.010	.195	26.928	1	.000	.364	2072.392	0.000
	-2 Log Likelihood(null model)=2165.072							

Remark: The value of -2L for the model containing all the covariates in this table is 2019.760

Table 4.13: Result of the Exponential regression model with corresponding AIC and BIC values

Exponential regression -- log relative-hazard form

No. of subjects = 2120 Number of obs= 2120

No. of failures = 181

Time at risk = 39020

LR chi2(6) = 115.22

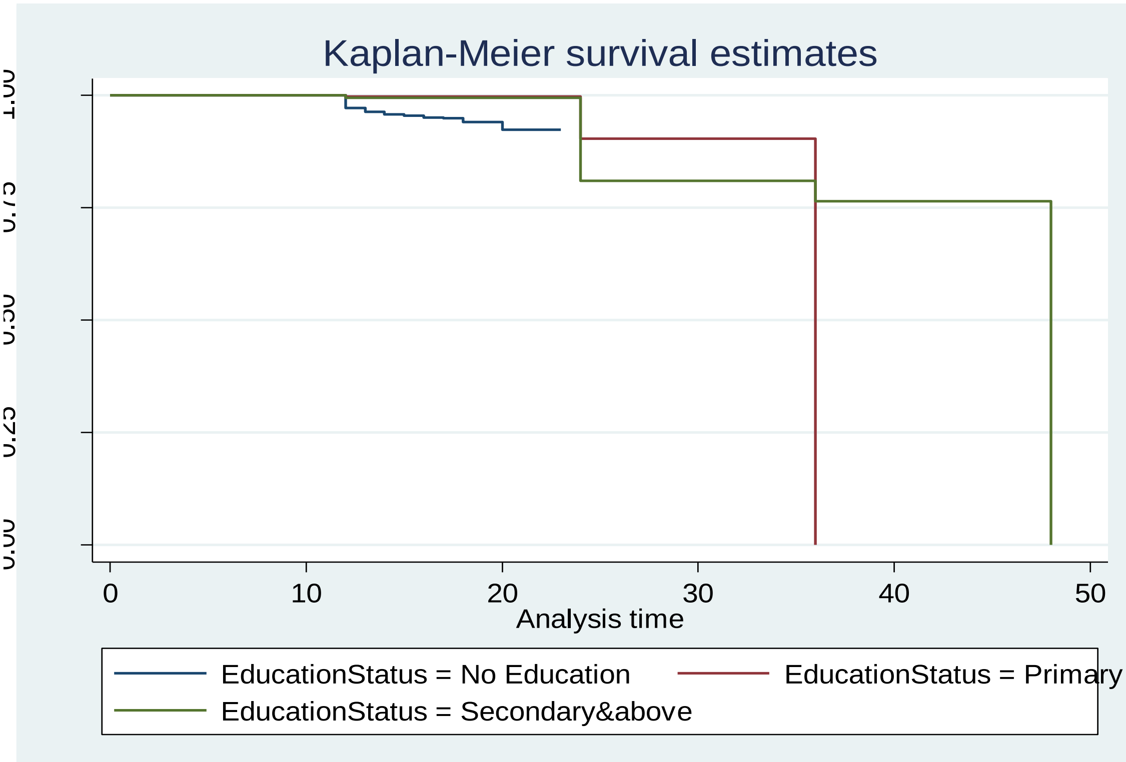
Log likelihood = -543.30261 Prob > chi2 = 0.0000

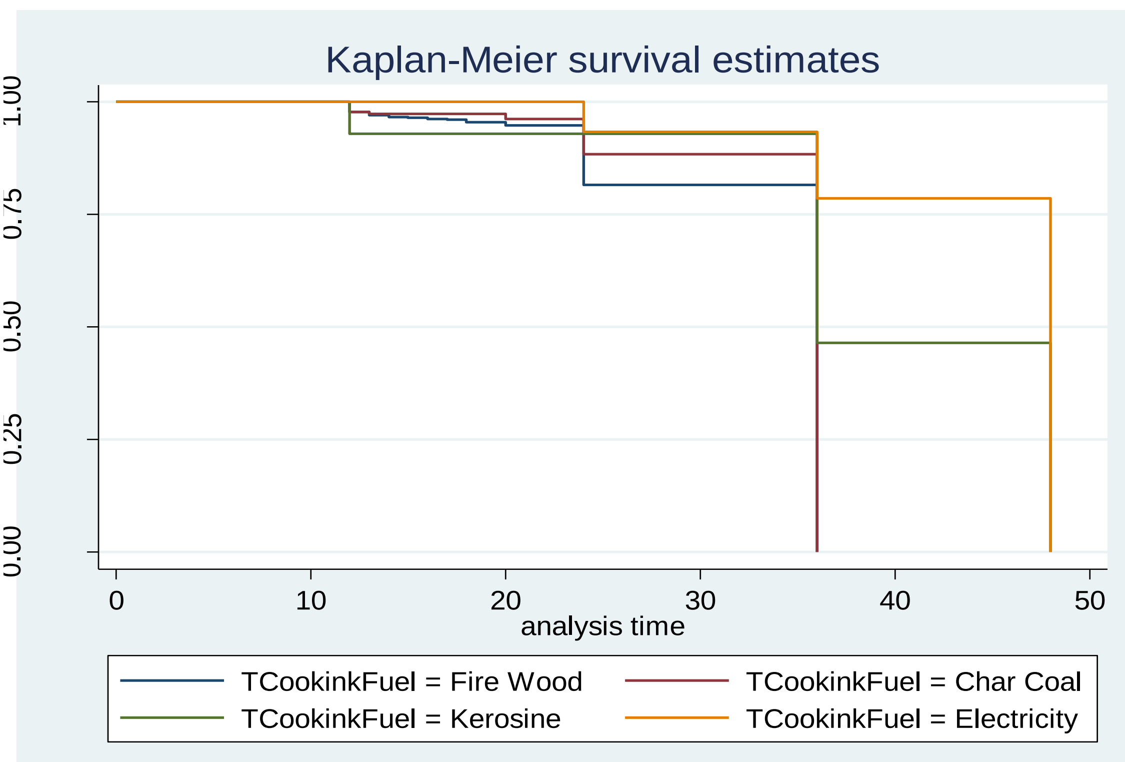
Covariates	Haz. Ratio	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
EducationStatus	1.677208	.2377105	3.65	0.000	1.270417	2.214255
ToiletType	.7981148	.1630596	-1.10	0.270	.5347611	1.191162
TCookinkFuel	.9931768	.1129584	-0.06	0.952	.7947228	1.241188
AntinatalVisit	.7244037	.2316033	-1.01	0.313	.3871134	1.355574
PoDelivery	1.401772	.4155004	1.14	0.255	.7841011	2.50601
SoDrinkinWater	4.195479	.7835058	7.68	0.000	2.909509	6.049834
_cons	.0002578	.0002284	-9.33	0.000	.0000454	.0014637

Model	Observation	ll (null)	Ll (model)	Df	AIC value	BIC value
Exponential	2120	-600.9143	-543.3026	7	1100.605	1140.219

Annex B

Figure 4.5: Kaplan-Meier survival estimates of different covariates





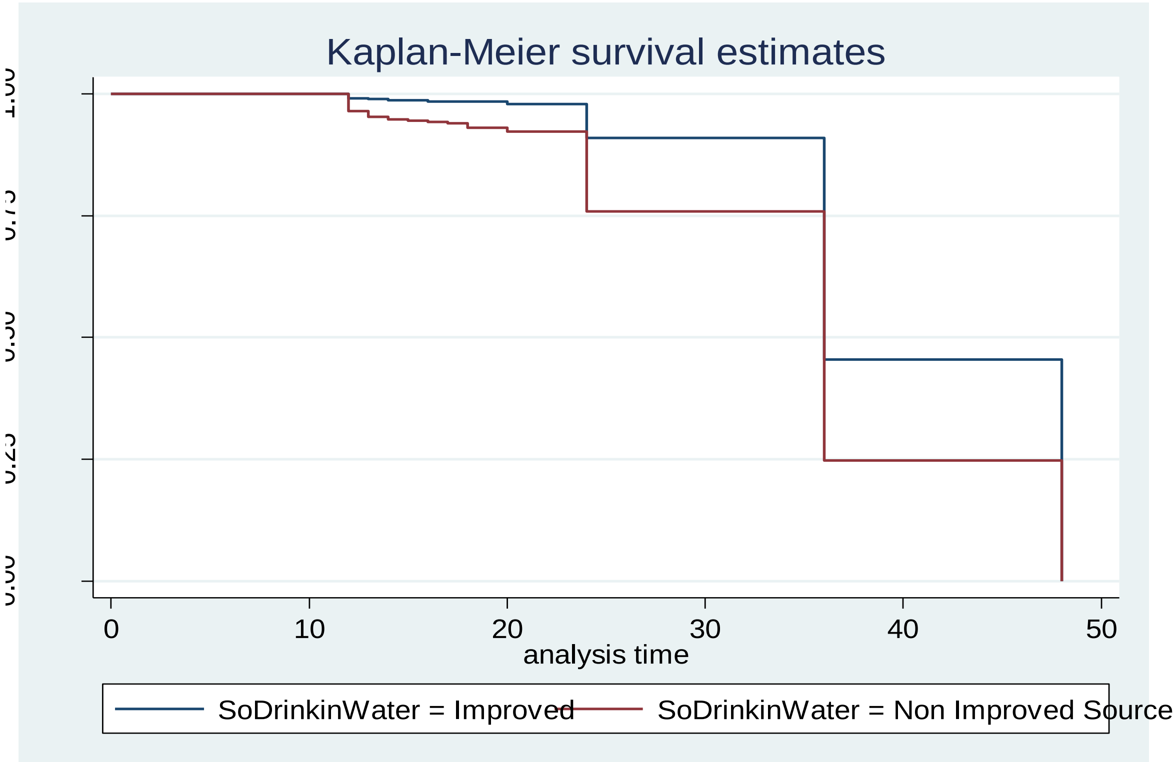


Figure 4.6: The scaled Schoenfeld residuals and their lowess smooth plots of different covariates

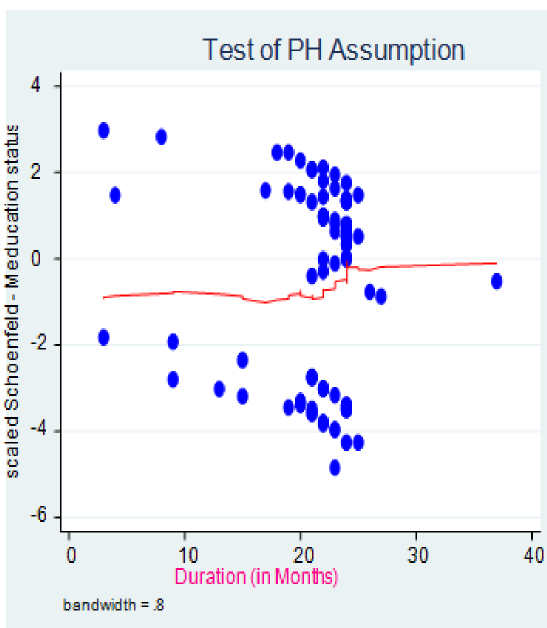


Figure 5: Plots of Scaled Schoenfeld residuals and their LOWESS smoothed obtained from the final model for the covariate Mothers educational status

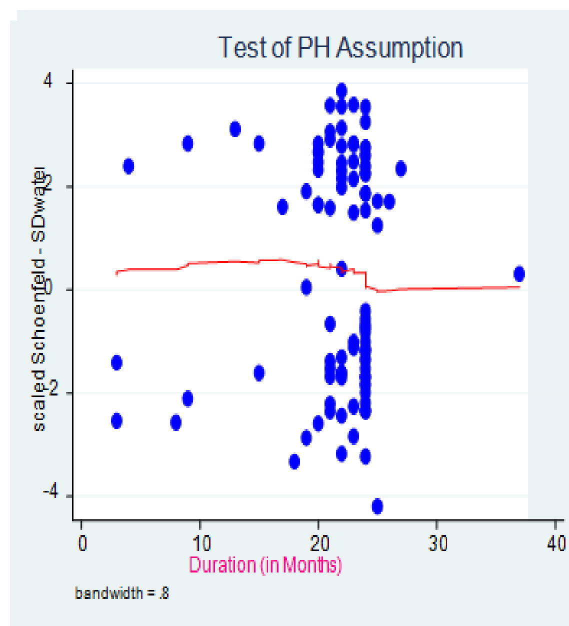


Figure 6: Plots of Scaled Schoenfeld residuals and their LOWESS smoothed obtained from the final model for the covariate source of drinking water

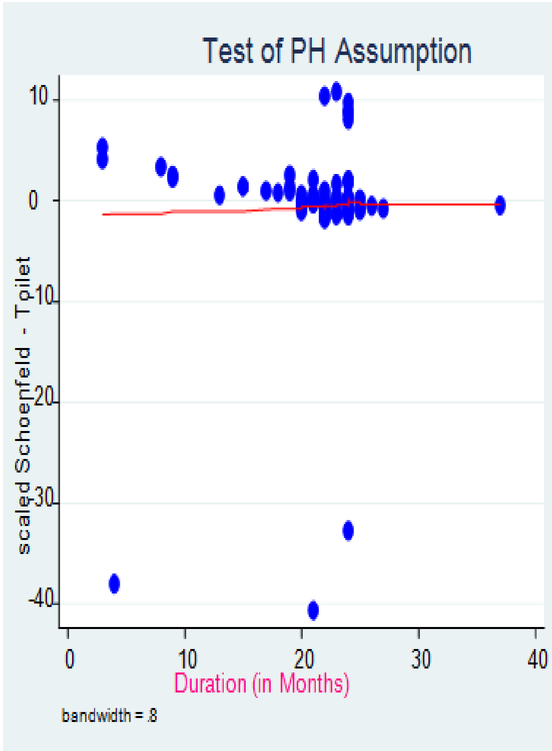


Figure 7: Plots of Scaled Schoenfeld sresiduals and their LOWESS smoothed obtained from the final model for the covariate toilet type used

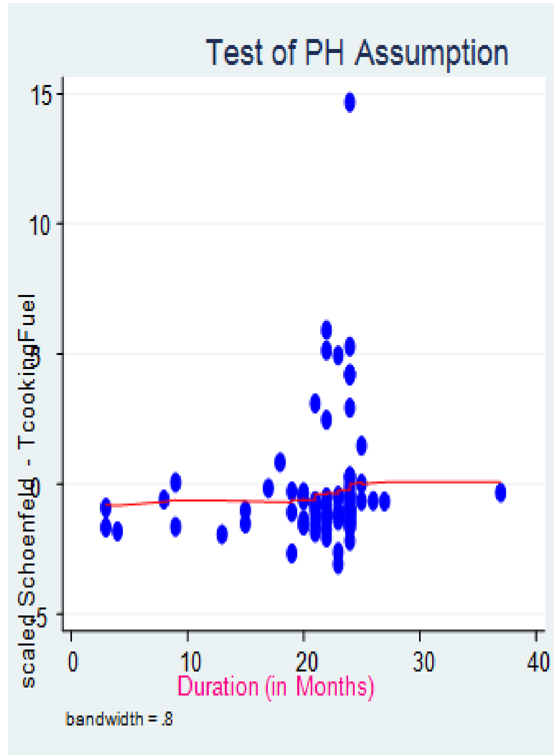


Figure 8: Plots of Scaled Schoenfeld residuals and their LOWESS smoothed obtained from the final model for the covariate type of cooking fuel

Figure 4.7: -log(-log(survival probability)) plot