ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

Determinants of Mortality among One to Five Years Old Children in Ethiopia

Senayit Seyoum

A Thesis Submitted to the Department of Statistics

Presented in Partial Fulfillment of the Requirement for the Degree of Master of Science in Statistics

Addis Ababa University

Addis Ababa, Ethiopia

June, 2012

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Senayit Seyoum, entitled: *Determinants of Mortality among One to Five Years old Children in Ethiopia* and submitted in fulfillment of the requirement for the Degree of Master of Science (statistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner _____Signature _____Date _____

Examiner _____Signature _____Date _____

Advisor _____Signature _____Date _____

_____

Chair of Department or Graduate Program Coordinate

Table of contents

List of Tables

ACRONYMS

ARI Acute Respiratory Infection

CSA Central Statistics Agency

DHS Demographic and Health Survey

EDHS Ethiopian Demographic and Health Survey

MDG Millennium Development Goals

MoFED Ministry of Finance and Economic Development

UN United Nations

UNICEF United Nations Children's Fund

UNMC United Nations Millennium Campaign

Abstract

Determinants of Under-Five Child Mortality in Ethiopia

Senayit Seyoum

Addis Ababa University, 2012

Child mortality is a factor that is strongly associated with the well-being of a population and it is taken as an indicator of health development and socioeconomic status. Child mortality rates are falling, but not quickly enough**.** According to the 2011 UN report during the last 10 years, the death rate for children under five has decreased by 35% worldwide. UNICEF in 2008 reported that Ethiopia has reduced under-five mortality by 40 percent over the past 15 years. From the EDHS 2011 report child mortality rate in Ethiopia was reduced from 50/1000 deaths in 2005 to 31/1000 deaths in 2011. The Ethiopian Demographic and Health Survey data are used for the study. The study uses information on 12,710 children included in the survey for the entire country. The main objective of the study was to identify determinants of under-five child mortality. For this purpose survival analysis was employed. The results show that mother's education, mother's age, marital status, birth order and place of residence are the significant factors that contribute to under-five child mortality. However, sex of a child, family size, wealth index, water source and toilet facility are not found to be significant. In order to reduce the rate of under-five child mortality attention should be paid to improve mother's education and discourage early marriage. Equal distribution of infrastructures is also necessary.

Acknowledgement

First of all I would like to thank the Almighty God for helping me with everything. Next my great gratitude goes to my advisor Professor Eshetu Wencheko for helping me from the beginning until the end of this thesis. Finally I would like to thank my family and my friends for their support.

Chapter One

Introduction

1.1. Background

Child mortality is a factor that is strongly associated with the well-being of a population and it is taken as an indicator of health development and socioeconomic status. That is why reduction of infant and child mortality is a worldwide target and one of the most important key indicators of the Millennium Development Goals (MDGs). Hence its indication is very important for evaluation and public health strategy. Thus, it is an area that many researchers focus on and that has attracted the attention of policy-makers and program implementers worldwide.

Child mortality rates are falling, but not quickly enough**.** For the last 10 years, the death rate for children under five has decreased by 35%. That means that, worldwide, 12,000 fewer under-fives die each day. Many countries have shown considerable progress in tackling child mortality. Almost one third of the 49 least developed countries have managed to reduce their under-five mortality rates by 40 percent or more over the past twenty years. The number of countries with under-five mortality rates of 100 deaths per 1,000 live births or higher has been halved from 52 in 1990 to 26 in 2010. In addition, no country had an under-five mortality rate above 200 deaths per 1,000 live births in 2010, compared with 13 countries in 1990. The rate of decline has accelerated from 1.9 percent a year over 1990–2000 to 2.5 percent a year over 2000–2010. Moreover, in Sub-Saharan Africa, the region with the greatest burden of under-five deaths, the rate of decline

doubled. But these rates are still insufficient to achieve Millennium Development Goal Number 4 (MDG 4) by 2015 [UN, 2011].

Only three years remain to achieve MDG 4, which calls for reducing the under-five mortality rate by two-thirds between 1990 and 2015. However, at the global level progress is behind schedule. The global under-five mortality rate needs to be halved from 57 deaths per 1,000 live births to 29, which implies an average rate of reduction of 13.5 percent a year [UNICEF, 2010]. Large absolute reductions in under-five mortality have occurred in regions with the highest initial mortality rates, such as sub-Saharan Africa and South Asia. Performance in Western and Eastern Africa has been particularly impressive, with average annual reductions of 2.64 and 2.16 per 1000 live births. Sub-Saharan Africa has achieved only around a 30 percent reduction in under-five mortality, less than half that required to reach MDG 4. Significant relative reductions in child mortality can be found in regions with relatively lower initial mortality rates, including North Africa, South-East Asia, and Latin America [UNMC, 2010]. The reductions were achieved through interventions targeted at communicable diseases. It was also noticed that disease-oriented vertical programs were not effective alone. Maternal, environmental, and socio-economic factors were recognized as additional important determinants of Child survival.

Africa accounts for only 22 percent of births globally but half of the 10 million child deaths annually occur on the continent. Africa is the only continent that has seen rising numbers of deaths among children under-five since the 1970s [UNICEF, 2008]. By contrast, many countries with unacceptably high rates of child mortality, most notably in

sub-Saharan Africa, have made little or no progress in recent years. Sub-Saharan Africa has achieved only around a 30 percent reduction in under-five mortality, less than half that is required to reach MDG 4. A major reason for the limited progress in reducing child mortality at the global level, despite more than half the regions having already achieved reductions of more than 50 percent, is the large and growing share of under-five deaths that occur in Sub-Saharan Africa and South Asia [UN, 2010]. All 36 countries with child mortality rates above 100 per 1000 births are in sub-Saharan Africa, with the exception of Afghanistan and Myanmar [UNMC, 2010]. Many of these children die of preventable and curable diseases. According to UNICEF, malaria is the cause of 18 percent of under-five deaths in Africa. Diarrheal diseases and pneumonia - both illnesses that thrive in poor communities where sanitation is severely compromised, and where residents are often undernourished and exposed to pollution - account for a further 40 percent of child deaths. Another major killer is AIDS.

Ethiopia has reduced child mortality by 40 percent over the past 15 years between 1990 and 2006 [UNCEF, 2008]. Despite Ethiopia being one of the top populous nations with relatively high population growth rate, still remains to be an outstanding African nation taking active steps in assuring primary health services and in curbing child mortality rates. From the EDHS report child mortality rate in Ethiopia was reduced from 50/1000 deaths in 2005 to 31/1000 deaths in 2011. The major causes of child mortality include preventable or treatable diseases such as measles, malaria, diarrhea, pneumonia and respiratory infections. It was noticed that the decline of infant and child mortality had been achieved through the intervention of disease oriented programs. In recent decades the awareness of maternal, environmental, behavioral and socioeconomic factors

increased and recognized as additional important factors of child mortality. Understanding the current determinants of child mortality is essential to inform policies and strategies to accelerate the reduction of child mortality. Child mortality is often associated with poverty, maternal education, maternal fertility characteristics, maternal under-nutrition, intervals between births, access to adequate safe water and basic curative health services [MoFED, 2010].This study will consider most of the variables corresponding to the categories of these determinants and will identify the major factors that contribute to the death of children.

The data used in this study are obtained from Ethiopia Demographic and Health Survey conducted in 2011 (EDHS, 2011). The subjects of the study are children of age between one and five years. This study uses survival analysis to estimate mortality and identify socioeconomic and demographic determinants of mortality among one to five years old children in Ethiopia.

1.2. Statement of the problem

Child mortality rate is a basic indicator of a country‟s socioeconomic status and quality of life. The current levels of child mortality in Ethiopia are higher than the target of the minimum Millennium Development Goals (MDGs), which is 67 deaths per 1000 live births, internationally adopted in the 1990 world summit for children. It is not well understandable why infant and child mortality rates remain high and far from desired in Ethiopia, despite the intervention made. Previous studies suggest that like other sub-

Saharan regions, the impact of HIV/AIDS epidemic, poverty, economic crises, political unrest and civil war in the past contributed to the worsened of the levels of mortality particularly for infant and child mortality. It was noticed that the decline of infant and child mortality was achieved through the intervention of disease oriented programs. In recent decades the awareness of maternal, environmental, behavioral and socioeconomic factors increased and these are recognized as additional important factors of infant and child mortality.

As mentioned above child mortality rate in Ethiopia is still very high. Despite numerous interventions and action plans, very little evidence exists on why the infant and the child mortality rates in Ethiopia have not declined as desired. If Ethiopia is committed to achieving the MDG on child mortality, it is necessary to understand clearly the factors that are contributing to the high levels of mortality. This study therefore explores the maternal, environmental and socio-economic characteristics and their effect on child mortality in Ethiopia.

1.3. Objectives of the study

The general objective of the study is to identify determinants of child mortality in Ethiopia.

The specific objectives are to:

- investigate the causes of child mortality.
- asses the relationship between the risk factors and child mortality.

- suggest valuable strategies to reduce child mortality in Ethiopia.

1.4. Significance of the study

The United Nations declared the Millennium Development Goals, one of which focuses on child mortality. Global reports showed that the number of childhood deaths in developing countries particularly in Sub-Saharan Africa is still unacceptably high. Evidence-based interventions are needed to reduce child deaths. Decline of child mortality could be achieved through the intervention of disease-oriented programs, in recent decades the awareness of maternal, environmental, behavioral and socioeconomic factors were increased and recognized as additional important factors of infant and child mortality. This study, therefore, will investigate the causes of child mortality in Ethiopia and identify its determinants which could be used as knowledge input by health professionals and administrators.

Chapter Two

Literature review

Ezra and Gurum (2002) employed a logistic regression model to investigate the impact of birth interval on infant and child mortality in the context of communities characterized by high reproduction, prolonged breast feeding practice and poor living conditions in Ethiopia. They found that a short birth interval (<18 months) is significantly associated with infant and child mortality as compared to longer birth intervals (>24 months), implying the influence of short birth interval are more pronounced on infant mortality but weaker impact on child mortality. They observed that mother‟s in the age groups 15-19 and 35-49 have a significant effect on infant and child mortality as compared with children born to mothers in the age category 25-34. Education is also a significant determinant of infant and child mortality.

Desta (2011) using data from 2000 and 2005 EDHS employed logistic regression analysis to examine the socioeconomic, demographic and biological factors of infant and child mortality in Ethiopia. The study shows that marital status, birth order, type of births and preceding birth intervals are a significant proximate determinants of infant and child mortality. Breast feeding has an important significant effect on infant mortality but not on child mortality. Children born to women not currently married, first born children, multiple birth, children born within 18 months of the previous birth and children who were breastfed less than 6 months were exposed to the high risk of infant and child mortality. Among socioeconomic variables household size and to some extent, fathers

and mothers educational levels and sex of head of the household are the most significant determinants of infant and child mortality. Children born in small household size, children born in male headed household, children born to mothers and fathers with no education and to some extent children born to mothers and fathers with primary education are exposed to the high risk of infant and child mortality.

Kumar and Gemechis (2010) used data from Ethiopia DHS survey (2005) and employed a cross tabulation technique to examine selected socioeconomic, bio-demographic and maternal health care factors that determine child mortality in Ethiopia. Socioeconomic variables such that birth interval with preceding birth and mother"s education were found to have a significant impact in lowering the risk of child mortality. Their result confirmed that child mortality risk associated with children for whom birth interval is shorter than two years with previous child was highest (15 percent) and lowest (4. 2 percent) for the children whose birth interval was four years and longer. On the other hand, they reported that the mortality rate of children born to illiterate mothers and fathers is significantly correlated to the highest risk of child mortality. Birth order and place of residence are also an important determinates of child mortality. The result indicates that the first births (10.6%) and births of order 7 and higher (9.9%) also experience higher rates of mortality than births of orders 2 and 3 (7.3%) and 4 to 6 (8.3%). Among the total deaths, 9.67% deaths were found in urban areas and 90.33% deaths occurred in rural areas.

Amare et al (2007) used primary data to study the determinants of under-five mortality in Gilgel Gibe Field research center, Southwest Ethiopia. The study shows that there was no

significant difference in mortality between males and females. Higher under-five mortality was observed among children whose mother"s educational level was elementary school and below as compared to children whose mother had above elementary school level of education. Under-five mortality was highest for unvaccinated children compared to those who were vaccinated at least once. Higher risks of death of under-five children were observed while the next birth interval is shorter than 24 months. Other intermediate variables such as parity, birth order, age of mother at delivery, family planning utilization and housing conditions were not significantly associated with under-five mortality. After controlling the effect of potential confounding variables, birth interval was again significantly associated with under-five mortality. Among the behavioral variables, high risk of under-five mortality were observed when the mothers had no good practice, negative perception about severity of illnesses, negative perception on benefits of some modern treatment. There was a statistically significant association between practice and perceived benefits with under-five mortality after controlling the effect of other variables.

Wang (2003), using the results from the 2000 Ethiopia DHS, examines environmental determinants of child mortality. The results show that children born in rural area face higher mortality risk compared with those born in urban areas. Poor environmental conditions are related to high risk of infant and child mortality.

Mutunga (2004) used the 2003 DHS data to investigate the impact of socioeconomic and environmental variables of infant and child mortality in urban areas of Kenya. The study show that 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, having 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. Another study in Kenya by Hill (2001) shows that mother‟s educational level and economic status have a significant impact on infant and child mortality while urban areas are associated with high risk of infant and child mortality than rural areas. However, controlling for HIV prevalence child mortality is lower in urban areas. Infant and child mortality in urban areas was found to be lower than in rural areas.

Joshua and Jeroen (2009) using the result of 2005-06 Zimbabwean DHS investigate the maternal, socioeconomic and sanitation factors on infant and child mortality using Cox regression model. They found evidence of birth order six and more 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 confirmed in their analysis, that children who are first born and those born to mothers aged 40-49 years are associated with higher infant and child mortality. However, socioeconomic determinants have an 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 have a pronounced effect on infant mortality but weak effect on child mortality.

Goro (2007) used data from 1993, 1998, and 2003 DHS surveys in Ghana to examine determinants of infant and child mortality in three northern regions using multivariate logistic regression model and found that education of mothers, birth order of child and marital status of mothers are significant determinants of infant mortality. Mother‟s education has a significant impact on child mortality.

Sahn and Stifle (2003) using data from DHS for 24 African countries found that infant and child mortality in urban areas are lower relative to rural areas. Various factors contribute to urban-rural variation such as better education and better public health facilities in urban than in rural areas.

Hala (2002) used DHS data on Egypt to study child mortality using survival analysis. The results show that access to municipal water decreases the risk of mortality. Moreover, gender discrimination is found to be of an important effect beyond the neonatal period.

Hussain et al (1999) used primary data to study determinants of mortality among children in the urban slums of Dhaka city, Bangladesh. The results showed that gender difference in mortality may have been influenced by the patterns of treatment received during sickness and the choice of treatment was determined by the financial ability of the households. Household income, children‟s vaccinations, TT immunization of mothers and personal cleanliness seemed to be significantly associated with child mortality. Despite the relatively high vaccination coverage for this population, child mortality

remained alarmingly high, indicating that socioeconomic and environmental conditions must be improved to substantially reduce morbidity and mortality in this population.

Kaharuza et al (1999) used primary data to study determinants of child mortality in rural eastern Uganda. The result shows that child mortality is associated with low parental education, being born to adolescent mothers or mothers aged 35 or more.

Uddin (2009) used data from Bangladesh DHS. Cross-tabulation and multiple logistic regression techniques were used to estimate the predictors of child mortality. The cross-tabulation analysis shows that parents‟ education is the vital factor associated with child mortality risk but in logistic regression analysis only the father‟s education was found significant to reducing child mortality. The occupation of father was found a significant characteristic in both analyses. Breast feeding status and birth order have substantial impact on child mortality.

Kaldewei and Pitterle (2011) used data from Jordan DHS 2007 to reassess the main determinants of child mortality. The study shows that behavioral factors were important compared to the household and community factors that were found to be important in earlier studies. The results suggest that a short birth interval, being born to a mother who smokes, and avoidance of breast feeding all substantially increase a child‟s risk of dying before its first birthday. For the mother‟s age at birth, they found the expected U-shaped relationship, with higher risks of infant mortality of children born to mothers under 20 years and mothers over 30 years old.

Jacoby and Wang (2003) examine the linkages between child mortality and morbidity, and the quality of the household and community environment in rural and urban China using a competing risks approach. The key findings include (1) use of unclean cooking fuels (wood and coal) significantly reduce the neonatal survival probability in rural areas - an outcome that is also confirmed in two other studies (India and Guatemala); (2) access to safe water or sanitation reduces child mortality risks by about 34% in rural areas; (3) higher maternal education reduces child mortality, and female education has strong health externalities (i.e. controlling for other factors, a child living in a neighborhood with more educated mothers has about 50% lower mortality risk); (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 urban 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.

Baker (1999) used data from three administrative region of Malawi: the north, center and south to examine the pattern of regional variation of child mortality and selected maternal, socioeconomic and environmental factors. A significant variation of child mortality was observed between north and center, between north and south but not between south and center. Educational variations between those regions contribute for this regional variation of infant and child mortality. However, the level of education is associated with high child mortality variation if health service not readily available. On

the other hand sanitation and wealth index do not contribute to the regional variation of child mortality.

Manda (1999) investigated the effect of birth interval, maternal age at birth, birth order and breast feeding by considering other relevant determinants on infant and child mortality in Malawi. The study shows that birth interval and maternal age significantly affect infant mortality, however, the impact is much weaker on child mortality. Breast feeding status of mothers does not change the influence of preceding birth interval length on infant and child mortality risk while it has an effect on the influence of succeeding birth interval on infant and child mortality. Source of drinking water and sanitation facilities are highly correlated with the reduction of infant and child mortality.

Chowdhury et al (2010) using primary data in Bangladesh found that mother‟s occupation has a significant effect on infant mortality. Among the occupational groups, the category professional has a positive effect on infant mortality. Availability of electricity has a significant effect on infant mortality. Significant predictors of child mortality were found to be mother‟s education, father‟s education, mother‟s occupation, types of latrine and electricity.

Klaauw and Wang (2003) in India predict that a significant number of deaths of children under-five years can be averted by providing access to electricity, improving the education of women, providing sanitation facilities and reducing indoor air pollution.

Mondal et al (2009) used primary data to examine the differential patterns of infant and child mortality in Rajshahi District, Bangladesh. A multivariate technique is employed to investigate the effects of socioeconomic and demographic variables on infant and child mortality. The study indicates that the most significant predictors of neonatal, post-neonatal, and child mortality levels are immunization, ever breastfeeding, mother‟s age at birth and birth interval. Parents‟ education, toilet facilities and treatment places are significant predictors during neonatal and childhood period but father‟s occupation is significant during the post-neonatal period. The risk of child mortality decreased with increased female education and wider access to safe treatment places.

Mahfouz et al (2009) used primary data to estimate the levels of infant and under-five mortality and to determine the socioeconomic, demographic and environmental factors contributing to infants and child mortality in Malakal town, southern Sudan. It was found that child interval, child immunization, family size, family income, and mother's education, have significant influence on infant and under-five mortality.

Balk et al (2003) used data from DHS for 12 countries in West Africa. It was shown that the determinants that influence child survival are mother‟s age at birth and the birth order of the child. These maternal factors have differential impacts on infants and children: infant deaths among mothers under age 20 typically occur in early infancy; young motherhood has less impact for children age 1–4 years. Birth order is closely related to mother‟s age at birth. First births are less likely to survive infancy than higher order births. The impact of birth order on survival is greatly reduced for children age 1–4 years.

Multiple births are associated with much higher risk of death, especially during infancy. Maternal education has been observed to have a strong impact on child survival. Infants and children of mothers with no education both have only an 89 percent chance of survival at 12 months and at 59 months. Infants and children born to mothers with secondary or higher education have greatly improved chances of surviving, 95 percent and 97 percent, respectively. Infants and children residing in urban areas have, on average, better survival chances than those in residing in rural areas.

Chapter Three

Data and Methodology

3.1. Data

The source of the data used in this study is the 2011 Ethiopia Demographic and Health survey (EDHS, 2011) conducted in Ethiopia as part of the worldwide demographic and health survey project. The 2011 Ethiopia Demographic and Health Survey was conducted by the Central Statistical Agency (CSA) with the support of the Ministry of Health. This is the third 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 2011 EDHS are 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).

The survey interviewed a nationally representative population in about 18,500 households, and all women of age 15-49 and all men of age 15-59 in these households. Indicators relating to family planning, fertility levels and determinants, fertility preferences, infant, child, adult and maternal mortality, maternal and child health,

nutrition, women‟s empowerment, and knowledge of HIV/AIDS are provided for the nine regional states and two city administrations. In addition, data by urban and rural residence at the country level are provided.

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, a data set consisting of 12,710 children was used. This age group is selected because the risk is expected to be higher in this age group and the risk for children above age five is less since they are older.

3.2. Variables

3.2.1. Response variable

The response or outcome variable for this study is the survival time of a child measured in months from year one until death/censor (before age five).

3.2.2. Explanatory variables

The explanatory variables are classified into three groups: maternal, socioeconomic and environmental.

Maternal (bio-demographic) factors

- Birth order
- Mothers age at first birth
- Child‟s sex
- Marital status

Socio-economic factors

- Mothers education
- Residence
- Family size
- Wealth index (Three categories: poorest and poorer=poor, richest and richer=rich and middle)

Environmental factors

- Source of drinking water
- Toilet facility

3.3. Methodology

Survival analysis

Survival analysis is the phrase used to describe the analysis of data that corresponds to the time from a well-defined time origin until the occurrence of some particular event or end-point. Survival data are not amenable to standard statistical procedures used in data analysis. One of the features of survival data that renders standard methods inappropriate is that survival times are frequently censored. The survival time of an individual is said to be censored when the end-point of interest has not been observed for that individual (Collet, 2003).

3.3.1.Descriptive method of data analysis

The survivor function and hazard function are the two functions of central interest in summarizing survival data. The actual survival time, *t*, of an individual is the value of a random variable time T, which can take any non-negative value. When the random variable T has a probability distribution with underlying probability density function f(*t*), the distribution function of T is then given by

$$F(t) = P(T \leq t)$$

and represents the probability that the survival time is less than some value t.

The survivor function, S(*t*), is defined to be the probability that the survival time is greater than or equal to t, and so

$$S(t) = P(T > t) = 1 - F(t).$$

The survivor function can, therefore, be used to represent the probability that an individual dies at time t, conditional on having survived to that time. That is, the function represents the instantaneous death rate for an individual surviving to time *t*. Thus, the hazard function, h(*t*), is defined as

$$h(t) = \lim_{t \to \infty} \left\{ p \left(\frac{t \leq T < t + \delta t / T \geq t}{\delta t} \right) \right\}$$

From this definition the relationship between the survivor and hazard function, can be expressed as

$$h(t) = f(t) / S(t) = -d / dt \{\ln S(t)\}$$

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

$$H(t) = \int_0^t h(u)du = -\ln S(t)$$

Thus $S(t) = \exp(- H(t))$ consequently $f(t) = h(t) \exp(-H(t))$

Estimation of the survivor function

The survival and hazard functions are estimated using the Kaplan-Meier method as a preliminary analysis. 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.

To apply the Kaplan-Meier method suppose that there are n independent individuals in a random sample with observed survival times $t_1, t_{2,\ldots,}t_n$. The distinct ordered failure times observed among the n individuals are $t_{(1)}, t_{(2)}, \ldots, t_{(m)}, m<n$ as there are more than one individual with the same observed survival time and some of the observations may be right-censored, i.e., the survival status of the individual might not be known at the time of the analysis.

The probability of survival at time $t_{(j)}$, $P(t_{(j)})$ is then estimated by

$$P(t_{(j)}) = (n_j - d_j)/n_j$$

where n_j is the number of individuals who are alive just before time $t_{(j)}$ and d_j is the number of those who die during this time.

Consequently the estimated probability of surviving beyond t(j), S(t) is

$$\hat{S}(t) = \prod_{j/t_{(j)} \leq t} \frac{n_j - d_j}{n_j}$$

with the approximated standard error given by

$$s.e\{\hat{S}(t)\} = \hat{S}(t)\left\{\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}\right\}^{1/2}$$

Comparison of survivor functions

When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. In survival analysis, the Kaplan-Meier estimators of the survivor function for each group are plotted. The statistical question is whether the observed difference in the graph is significant. A number of statistical tests have been proposed to answer this question such as Log-rank, Generalized Wilcoxon, Tarone-Ware test and so on (Hosmer et al, 2008).

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:

$$Q=\frac{[\sum_{i=1}^{m} w_i(d_{1i}-\widehat{e_{1i}}]^2}{\sum_{i=1}^{m} w_i{}^2\widehat{v_{1i}}}$$

where:

m is the number of rank-ordered failure (death) times.

n_{1i} is the number of individuals at risk in group 1 just prior to failure time t_i

n_{2i} is the number of individuals at risk in group 2 just prior to failure time t_i

n_i is the number of individuals at risk in both groups 1 and 2 just prior to failure time t_i

$d1i$ is the observed number of failure (death) in group 1 at failure time t_i

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

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

w_i is the weight for censor adjustment at failure time t_i .

The most common method used to compare survival curves is then a statistical hypothesis test called log-rank test, this test is based on weights equal to one, i.e. $w_i = 1$. The null hypothesis for the log-rank test is that there is no difference between the survivals of two or more populations that are being compared (i.e. the probability of the event of interest occurring at any time point is the same for each population). The significance of Q may be tested using the chi-square distribution with one degree of freedom. We can also use the above test to compare more than two groups.

3.3.2.Cox-proportional hazard model

The Kaplan-Meier and log-rank methods described are useful in the analysis of a single sample of survival data, or in the comparison of two or more groups of survival times. However, the relationship between the outcome variable and the explanatory variables is identified by fitting a regression model. The basic model to be considered here is the proportional hazards model.

The assumption of proportional hazards is that the hazard of death at any given time for an individual in one group is proportional to the hazard at that time for an individual in the other group. When there are covariates in the analysis which are time dependent, this assumption may not hold.

The set of values of the explanatory variables in the Cox-proportional hazards model will be represented by 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

$$h_i(t) = h_0(t) \exp(\beta x_i)$$

$$h(t)=h_0(t)\exp(\beta'\mathbf{x})$$

where β is a p x 1 vector of regression coefficients.

x_i is the value of **x** for the i^{th} individual, i=1,2,…,*n*

The hazard function in the Cox model is called semi-parametric function since it does not explicitly describe the baseline hazard function, $h_o(t)$. The survival function is given by:

$$S(t, x, \beta) = e^{-H(t,x,\beta)}$$

where, H (t, x, β) is the cumulative hazard function at time *t* for a subject with covariate x. Since we have assumed that survival time is absolutely continuous, the value of the cumulative hazard function is expressed as:

$$H(t, x, \beta) = H_0(t) \exp(\beta' x)$$

Consequently, from the proportional hazards function, we obtained the survivor function given by:

$$S(t, x, \beta) = [S_0(t)]^{\exp(\beta' x)}$$

where $S_0(t)$ is the baseline survival function.

Fitting the Proportional Hazards Model

Fitting the proportional hazards model to observed survival data entails estimating the unknown regression coefficients. Also, the baseline hazard function must be estimated. It turns out that these two components of the model can be estimated separately. The coefficients should be estimated first and the estimates are then used to construct an estimate of the baseline hazard function. The regression coefficients in the proportional hazards Cox model, which are the unknown parameters in the model, can be estimated using the method of maximum likelihood.

Suppose the survival data based on *n* independent observations are denoted by the triplet (t_i, x_i, δ_i) for i= 1,2, ..., n among whom there are *r* distinct death times and *n-r* right

censored survival times. There is one individual dies at each death time, so that there are no ties in the data. Therefore, the data consist of n observed survival times, denoted by t_1, t_2, …, t_n, and that δ_i is an event indicator, which is zero if the i^{th} survival time is right censored, and unity otherwise.

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

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{exp\, \beta' x_i}{\sum_{j \in R(t_i)} \exp(\beta' x_i)} \right]^{\delta_i}$$

where, $R(t_i)$ is the risk set at time t_i. The corresponding log-partial likelihood function is given by

$$logL(\beta) = \sum_{i=1}^{n} \delta_i \left\{ \beta' x_i - log \sum_{j \in R(t_i)} exp\, \beta' x_i \right\}$$

The maximum likelihood estimates of the β-parameters in the proportional hazards model can be found by maximizing this log-likelihood function using numerical methods. This maximization is generally accomplished using Newton-Raphson procedure (Collet, 2003)

Variable selection procedures

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) and best subsets selection. 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. These routines are based on forward selection, backward elimination or a combination of the two known as the stepwise procedure (Collet, 2003).

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

1. The first step is to fit models that contain each of the variables one at a time. The values of $-2\log\hat{L}$ for these models are then compared with that for the null model. The null model is a model to determine which variables on their own significantly reduce the value of this statistic.
2. The variables that appear to be important from step 1 are then fitted. In the presence of certain variables others may cease to be important. Consequently, those variables that do not significantly increase the value of $-2\log\hat{L}$ when they are omitted from the model can now be discarded. We therefore compute the change in the value of $-2\log\hat{L}$ when each variable on its own is omitted from the

set. Only those that lead to a significant increase in the value of $-2\log\hat{L}$ are retained in the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

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 to the model from step 2, one at a time, and any that reduce $-2\log\hat{L}$ significantly are retained in the model. This process may result in terms in the model determined at step 2 ceasing to be significant.
4. A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of $-2\log\hat{L}$, and that no term not included significantly reduces $-2\log\hat{L}$.

When using this selection procedure, rigid application of a particular significance level should be avoided. In order to guide decisions on whether to include or omit a term, the significance level should not be too small. A level of around 20% - 25% is recommended (Hosmer et al, 2008).

3.3.3. Model diagnostics

The fit of a regression model involves assessment of the regression coefficients and the formation of confidence intervals for the parameters and related quantities. Under the assumption of proportional hazards, there are three different tests for model assessment

(the significance of the coefficients): the partial likelihood ratio test, the Wald test and the score test. These tests are presented below as discussed in Hosmer et al (2008).

The Partial Likelihood Ratio Test (LR) is the best of the three tests for testing the significance of a subset of *q* explanatory variables from *p* explanatory variables, and fit both the unrestricted and the restricted models. We shall obtain the value of the log-partial likelihood function $L_p(\hat{\beta}_{p-q})$ in the restricted model and $L_p(\hat{\beta}_p)$ in the unrestricted model. The likelihood ratio test statistic in this case becomes

$$Q_{LR} = -2[L_P(\hat{\beta}_{p-q}) - L_p(\hat{\beta}_p)]$$

The test statistic for H_0 is based on the difference of the log-likelihood values. Under H_0, the statistic is asymptotically distributed as chi-squared with *q* degrees of freedom at a significance level of α.

The Wald test requires fitting the unrestricted model, and is based on the partial likelihood estimator $\hat{\beta}$. The test statistic is

$$Q_W = \hat{\beta}' I^{-1}(\hat{\beta})\,\hat{\beta} \sim X_{(q)}^2$$

The quadratic form of the above equation requires the inverse of the variance-covariance estimates corresponding to the *q* parameters in H_0: $\beta q = (0, 0, \ldots 0)'$ and $\mathbf{I}q \times q$ is the

information matrix. Under H_0 is asymptotically distributed as chi-squared with *q* degrees of freedom.

The Score Test: The score test statistic, to test H_0 : βq= (0, 0, 0….0)', is defined as

$$Q_S = U'(\beta_{q,} \hat{\beta}_{p-q}) I^{-1}(\beta_{q,} \hat{\beta}_{p-q}) U(\beta_{q,} \hat{\beta}_{p-q})$$

where $U(\beta_{q,} \hat{\beta}_{p-q})$ and $I^{-1}(\beta_{q,} \hat{\beta}_{p-q})$ are the score vectors and inverse of the observed information matrix evaluated at the hypothesized value of β_q (the coefficient sub-vector of length *q*) and β_{p-q} is the sub-vector containing the remaining *p-q* components. Under the null hypothesis and for large sample the statistic Q_s is asymptotically distributed as chi-squared with *q* degrees of freedom.

Residual analysis

Under the proportional hazards model, residuals play a central role in evaluating the model assessment and adequacy. Many model checking procedures are based on quantities known as residuals. Residuals are values that can be calculated for each observation and have the feature that their behavior is known, at least approximately, when the fitted model is satisfactory. The following residuals have been proposed for use.

Cox-Snell residuals: The Cox-Snell residual for the i^{th} individual is given by

$$rc_i = \exp(\hat{\beta}' x_i) \hat{H}_0(t_i) = \hat{H}_i(t_i) = -\log \hat{S}_i(t_i)$$

where $\tilde{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i, the observed survival time of that individual, $\tilde{H}_i(t_i)$ and $\hat{S}_i(t_i)$ are the estimated values of the cumulative hazard and survivor functions of the ith individual at t_i.

Martingale residuals are modified Cox-Snell residuals and, defined as

$$r_{Mi} = c_i - r_i$$

where c_i is censoring indicator and r_i is the Cox-Snell residual.

It can be shown that these residuals sum to zero and, in large sample, the martingale residual are uncorrelated with one another and have an expected value of zero. In this respect, they have properties similar to those possessed by residuals encountered in linear regression analysis.

Schoenfeld residuals: Schoenfeld residuals are useful to check the proportionality of the covariates over time, that is, to check the validity of the proportional hazards assumption. If the model fits well then the residuals are randomly distributed without any systematic pattern around the zero line, the reference line.

The i^{th} Schoenfeld residual for x_j, the j^{th} explanatory variable in the model, is given by

$$r_{pji} = c_i\{x_{ji} - \hat{a}_{ji}\},$$

where x_{ji} is the value of the j^{th} explanatory variable, j= 1, 2, ……, p, for the i^{th} individual

$\hat{a}_{ji}=\frac{\sum_{l\in R(t_i)} x_{jl}\exp(\hat{\beta}'x_l)}{\sum_{l\in R(t_i)}\exp(\hat{\beta}'x_l)}$ and R(t_i) is the set of all individuals at risk at time t_i.

Testing for the form of linearity of covariates

The assumption of linearity 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. If the resulting plot is random showing no systematic pattern this indicates that the covariate is linear in the model.

Identifying influential observations

It may happen that the structure of the fitted model is particularly sensitive to one or more observations in the data set. Such observations which are referred to as influential observations can be detected using diagnostics that are designed to highlight observations that influence the complete set of parameter estimates in the linear predictor. In other words, it may happen that the structure of the fitted model is particularly sensitive to one or more observations in the data set. Another important aspect of model evaluation is a thorough examination of regression diagnostic statistics to identify which subjects have an unusual configuration of covariates that exert an undue influence on the estimates of the parameters and on the fit of the model. In many occasions, the influence that each observation has on the estimated hazard function will be of interest, and it will then be

important to identify observations that influence the complete set of parameter estimates in the model.

Suppose that we wish to determine whether any particular observation has an *awkward* effect on $\hat{\beta}_j$, the j^{th} parameter estimate, j=1,2,…….,p, in a fitted Cox regression model. One way of doing this would be to fit the model with all *n* observations, $\hat{\beta}_j$ is the j^{th} parameter estimate, 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 the value of the j^{th} parameter estimate on omitting the i^{th} observation is denoted by $\hat{\beta}_{j(i)}$. Then, the statistic $\Delta_{j(i)} = \hat{\beta}_j - \hat{\beta}_{j(i)}$, known as DFBETA, 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. An index plot, or a plot of the likelihood displacements against the rank order of the survival times, provides information visual summary of the values of the diagnostic. Observations that have relatively large values of the diagnostic are influential.

Moreover, examining martingale residuals is helpful in identifying outliers. The effect of outliers on the regression model may be easily checked by dropping these points and refitting the regression equation.

Methods for testing the assumption of proportional hazards

The proportional hazards 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. To test the assumption of the proportionality of the hazards we plot the Schoenfeld residuals against time. 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 to see if the variable is time dependent. 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. The extended Cox model that simultaneously considers all time-independent variables of interest can be expressed as:

$$h(t,x,\beta) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_j + \sum_{j=1}^{p}\delta_j x_j g_j(t)\right)$$

To check the proportional hazards assumption using a statistical test, we consider the null hypothesis that all the δ terms, which are coefficients of the $x_j \times g_j(t)$ product terms in the model, are zero. Usually the function $g_j(t)$ is chosen to be the logarithm of survival time i.e. $g_j(t)=ln(t)$. Under the null hypothesis that all the δ terms are zero, the model reduces to the proportional hazards model **(**Hosmer et al, 2008**)**.

If the assumption of proportional hazards model is not satisfied, the survival analysis of the data can be done by using stratification variable (a variable that reveals the violation of proportionality). This means there will be separate baseline hazard functions for each stratum.

Chapter Four

Results and Discussion

4.1. **Descriptive statistics**

Causes of child mortality

From the children taken as a sample full information about the causes of disease was obtained for 8409. From these 4401(52.4%) of the total experienced one of the causes of mortality. Table 1 shows the share of the three major causes of child mortality in Ethiopia.

Table 1: Causes of child mortality

Causes	Number	Percentage (%)
Diarrhea	1199	14.3
Pneumonia(ARI)	1612	19.2
Fever (Malaria)	1590	18.9

Table 2: Summary results about the covariates included in this study

Covariates	Category	Censored	Events (%)	Total
Place of residence	Urban	1768	311(15%)	2079
	Rural	8466	2165(85%)	10631
Mother Education	No Education	7294	2103(22.4)	9397
	Primary and Higher	2940	373(11.3)	3313

Birth order	1 2-5 6+	2011 5600 2623	571(22.1) 1502(21.1) 403(13.3)	2582 7102 3026
Family size	1-4 5-8 9+	2439 6389 1406	648(21.0) 1552(19.5) 276(16.4)	3087 7941 1682
Wealth index	Poor Medium Rich	4951 1657 3626	1319(21.0) 439(20.9) 718(16.5)	6270 2096 4344
Source of drinking water	Pipe protected source unprotected source	2718 2638 4878	561(17.1) 697(20.9) 1218(20)	3279 3335 6096
Toilet facility	With toilet No facility	4631 5603	1035(18.3) 1441(20.5)	5666 7044
Mother‟s age at birth	15-19 20+	6387 3847	1821(22.2) 655(14.5)	8208 4502
Sex of a child	Female Male	5031 5203	1179(19.0) 1297(20.0)	6210 6500
Marital status	Currently married Currently not married	8885 1349	1993(18.3) 483(26.4)	10878 1832

Table 2 provides a summary of censored and uncensored data based on the 12,710 children in the sample.

The graph of the estimate of overall Kaplan-Meier survivor function is given in Figure 1(Appendix). Separate graphs of the estimates of the Kaplan-Meier survivor functions for the covariates mothers age, mothers education, marital status, wealth index, family size, place of residence, sex of a child, birth order, water source and toilet facility are also presented in Figures 2 up to 11 of the Appendix in order to see whether there is difference in survival experience between different categories of individuals. All of the graphs show differences between different categories except for sex of a child. The upper curve in each Figure indicates that the particular group experiences better survival than the one below.

Results of the log rank test

Table 3: Log-Rank Test for covariates

Covariates	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
Residence	-99.3965	18.4468	28.0334	<.0001
Sex of a child	27.3719	24.8661	1.2117	0.2710
Mothers age	220.1	23.7734	85.7499	<.0001
Marital status	-162.5	17.4688	52.4307	<.0001

Mothers education	280.1	21.9169	163.3	<.0001
Water source	-96.0946	41.6180	5.3313	0.0209
Toilet facility	-63.5978	24.7152	6.6215	0.0101
Birth order	254.8	32.8937	59.9973	<.0001
Wealth index	233.2	44.8756	27.0118	<.0001
Family size	59.9147	30.2079	3.9339	0.0473

The log rank test is performed to test if there are statistically significant differences among the survival experience of the different groups of the covariates at 5% level of significance. The results of the log rank test in Table 3 indicate that there are significant differences between the survivals of different groups of all the covariates except for sex of a child. Sex is not significant and it implies that there is no significant difference between the survival of males and females. To assess the relationship between the outcome and explanatory variables model development is necessary.

4.2. Results of Cox-proportional hazards model

Univariate Cox proportional hazards model

The first step in model development process is to select explanatory variables that have the potential to be included in the proportional hazards model. Table 4 provides results of the univariate Cox-proportional hazards model.

Table 4: Univariate Cox-proportional hazards model

Covariates	DF	$\hat{\beta}$	SE($\hat{\beta}$)	Wald X^2	p-value	-2 LOGL
Residence	1	0.29976	0.06089	24.2389	<.0001	44,509.198
Sex of a child	1	0.06283	0.04024	2.4376	0.1185	44,532.925
Water source	2			12.7209	0.0016	44,520.368
Piped water	1	-0.15532	0.05103	9.2661	0.0023	
Protected water	1	0.07183	0.04805	2.2349	0.1349	
Toilet facility	1	-0.12309	0.04075	9.1244	0.0025	44,526.188
Mothers age	1	0.41106	0.04556	81.3939	<.0001	44,408.902
Mothers education	1	0.65718	0.05618	136.8165	<.0001	44,376.775
Marital status	1	0.36285	0.05071	51.2092	<.0001	44,488.079
Family size	2			27.2342	<.0001	44,507.967
1-4	1	0.35792	0.07191	24.7776	<.0001	
5-8	1	0.18564	0.06532	8.0756	0.0045	
Wealth index	2			27.1323	<.0001	44,507.393
Poor	1	0.23355	0.04638	25.3598	<.0001	
Medium	1	0.22128	0.06059	13.3394	0.0003	
Birth order	2			84.3125	<.0001	44,441.460
1st	1	0.52134	0.06506	64.2106	<.0001	
2-5	1	0.49010	0.05610	76.3151	<.0001	

Discussion about the multivariate Cox-proportional hazards model

Covariates will be included stepwise in the multivariable model if they reduce the value of -2log L of the null model considerably on the basis of p-values; otherwise they are not candidates for inclusion. For the null model the value of -2log L is 44,535.369. Table 5 below shows that all the variables considered lead to reductions in the value of -2 log L. The highest reductions are obtained by the covariates mother‟s education (158.594), mothers age (126.467), birth order (93.907) and marital status (47.29). Therefore, all the covariates will be included in the multivariable study. The next step is to check the significance of the covariates in the multivariable model. The covariates which are not significant at each step are eliminated from the model and the final model is fitted using the remaining significant covariates. The final model is shown in Table 5.

Table 5: Multivariate Cox-proportional hazards model significance test result

Covariates	DF	Wald Chi-Square	Pr > Chi Sq
Mother‟s age	1	96.9988	<.0001
Marital status	1	65.3205	<.0001
Mothers education	1	155.4187	<.0001
Residence	1	5.9968	0.0143
Birth order	2	139.2011	<.0001

The final model contains the covariates mothers age, marital status, mother education, place of residence and birth order which are found to be significant at 5% level. The parameter estimates and the hazard ratios of the covariates for the final model are presented in Table 6.

Table 6: Results of the final Cox proportional hazards model

Covariates	DF	$\widehat{\beta}$	SE($\widehat{\beta}$)	Chi Square	P-value	$\widehat{HR}$	95% CI for $\widehat{HR}$	
							LCL	UCL
Mother‟s age (15-19) Ref (20 and above)	1	0.45269	0.04596	96.9988	<.0001	1.573	1.437	1.721
Marital status (Currently not married) Ref (currently married)	1	0.41398	0.05122	65.3205	<.0001	1.513	1.368	1.673
Mother‟s education (No education) Ref (primary and higher education)	1	0.73031	0.05858	155.4187	<.0001	2.076	1.851	2.328
Residence (Rural) Ref (urban)	1	0.15501	0.06330	5.9968	0.0143	1.168	1.031	1.322
Birth order 1 Ref (6+)	1	0.75128	0.06630	128.3992	<.0001	2.120	1.861	2.414
Birth order 2-5 Ref (6+)	1	0.57183	0.05643	102.7044	<.0001	1.772	1.586	1.979

Interpretation of the results in Table 6

The estimated hazard ratio ($\widehat{HR}$) of a child born in rural area is 1.168(95% CI: 1.031, 1.322) implying that the risk of dying for a child born in a rural area is 16.8% higher than a child born in an urban area (reference group) controlling for other covariates in the model. The confidence interval indicates that the risk of dying for a child born in rural area is as low as 1.031(3.1%) and as high as 1.322(32.2%) times the risk of dying for a child born in urban areas.

The estimated $\widehat{HR}$ for covariate mothers of age group 15 up to 19 is 1.573(95%CI: 1.437, 1.721). This implies that children born to mothers of age group 15 up to 19 have 57.3% higher risk of dying than children born to a mother at age group 20 and above (reference group) controlling for other covariates in the model. The confidence interval suggests that the hazard rate for a child born to mother of age group 15-19 is as low as 1.437(43.7%) and as high as 1.721(72.1%) times the risk of dying for a child born to a mother of age group 20 and above.

The hazard ratio $\widehat{HR}$ for a child born to currently not married mothers is 1.513(95% CI: 1.368, 1.673) showing that children born to a single mother has 51.3% higher risk of dying than children born to currently married mother (reference group) controlling for other covariates. The confidence interval indicates that the risk of dying for a child born to currently not married mothers is as low as 1.368(36.8%) and as high as 1.673(67.3%) times the risk of dying for a child born to currently married mothers.

The estimated $\widehat{HR}$ of a child born to a mother with no education is 2.076 (95% CI: 1.851, 2.328). The estimated risk of dying for a child born to a mother with no education is 2.13 times higher as compared to a child whose mother has primary and higher education (reference group) controlling for other covariates. The confidence interval indicates that the risk of dying for a child born to a mother with no education is as low as 1.851 times (85.1%) and as high as 2.328 times (133%) the risk of dying for a child born to a mother with primary and higher education.

The estimated $\widehat{HR}$ of a first born child is 2.120 (95% CI: 1.861, 2.414) and that of a child with birth order 2 up to 5 is 1.772(95% CI: 1.586, 1.979). A first born child has 2.12 times higher risk of dying than a child whose birth order is six and above (reference group) controlling for other covariates. Children of birth order two to five have 77.2% higher risk of dying than the reference group. The confidence interval implies that the risk of dying for a child born first is as low as 1.861 and as high as 2.414 times the risk of a child whose birth order is six and above and for a child with birth order two up to five the risk of dying is as low as 1.586 and as high as 1.979 times the risk of dying for a child whose birth order is six and above.

4.3. Model diagnostics

After a model has been fitted to an observed set of survival data the adequacy of the fitted model needs to be assessed. The use of diagnostic procedures for model checking is an essential part of the model in process.

Assessing proportionality assumption

The Cox regression model is applied using the assumption of proportional hazards for two or more groups. For checking this assumption we fit a proportional hazards model of the covariates and the interaction between the covariates and log (time). Table 7 shows results about the interaction of covariates and log-time.

Table 7: Multivariable proportional hazards model for covariates and interaction between covariates and log (time).

Covariates	DF	Wald Chi-Square	Pr > Chi Sq
Residence	1	1.9891	0.1584
Mothers age	1	0.3187	0.5724
Marital status	1	0.0453	0.8315
Mothers education	1	0.8053	0.3695
Birth order	2	25.6229	<0.001
Residence*log(*t*)	1	3.3653	0.0666
Mothers age*log(*t*)	1	1.1275	0.2883
Marital status*log(*t*)	1	2.3724	0.1235
Mothers education*log(*t*)	1	1.3873	0.2389
Birth order* log(*t*)	2	0.0176	0.8943
Test of proportionality	6	9.1707	0.1024

The results show that the time varying covariates are not significant implying that the covariates are time independent. The overall proportionality test is also not significant. Therefore the proportionality assumption is not violated. Figures in the appendix show the plot of Schoenfeld residuals for each covariate in the model against time. In all the figures there are no eminent departures from proportionality.

To check for the linearity of continuous covariates martingale residuals plot is used. Figure 17 in the Appendix shows the plot of martingale residuals versus ungrouped mother‟s age. The figure doesn‟t show any pattern implying that the linearity assumption is not violated.

Checking for influential and outlier observations

For the assessment of the outlier observations the martingale residuals versus the linear predictor are given in Figure 18 in the Appendix. No outlier observation is detected in the figure.

Evaluation of regression diagnostic statistic to identify any influential observation in the estimates of the Cox regression parameters or on the fit of the model is carried out using DFBETA statistic which is used to examine the awkward effect of each observation on the j^{th} parameter estimate. The five smallest and largest changes in the parameter estimates are presented in Table 8.

Table 8: The five smallest and highest differences of the parameter estimates of the variables included in the final model when the data value for each child is deleted from the model.

	Smallest Difference	Obs	Highest Difference	Obs
Residence	-0.00371355	2388	0.00208436	10308
	-0.00371355	2199	0.00208436	10408
	-0.00371355	1907	0.00284920	10888
	-0.00371355	1906	0.00284920	11555
	-0.00371355	1822	0.00284920	11613
Mother age	-0.000885281	12686	0.00170054	2087
	-0.000885281	12593	0.00170054	2187
	-0.000885281	12546	0.00171500	3079
	-0.000885281	12460	0.00174633	155
	-0.000885281	12421	0.00174633	1182
Marital status	-0.00205663	12678	0.00216706	534
	-0.00205663	12668	0.00216706	628
	-0.00205663	12417	0.00216706	629
	-0.00205663	12401	0.00216706	1817
	-0.00205663	12366	0.00216706	2254
Mother education	-0.00142558	1491	0.00320540	3079
	-0.00137445	2310	0.00321417	786
	-0.00137445	2120	0.00321417	1699

	-0.00137445	2025	0.00325946	155
	-0.00137445	1821	0.00325946	1182
Birth order	-0.000915505	2490	0.00134818	9663
	-0.000915505	2454	0.00134818	10157
	-0.000915505	2447	0.00135299	2560
	-0.000915505	2437	0.00139679	4251
	-0.000915505	2383	0.00142087	1996

The results in Table 8 indicate that the differences for all the covariates are not significant.

Overall goodness of fit test

Table 9 below shows the model fit statistics for the final model. By comparing the value –2 Log L for the final proportional hazards model with that of the null model (a model with no covariates) we found that the final model reduces this value by a substantial amount.

Table 9: Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	44535.369	44077.140

By calculating likelihood ratio, Wald and score tests we can assess the adequacy of the model by testing the hypothesis that all the parameters of the final model are zero. Table 10 shows the results of the three tests for the final model.

Table 10: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	458.2290	6	<.0001
Score	435.3212	6	<.0001
Wald	425.7016	6	<.0001

The results speak for rejection of the null hypothesis that all the parameters in the final proportional hazards model are zero. Therefore, we can say that there is no evidence that the null hypothesis is not rejected.

4.4. Discussion

The results of the Cox proportional hazards regression analysis show that place of residence, birth order, mother‟s age at the time of giving birth, mother‟s education and marital status were significantly associated with child mortality.

The study shows that children living in rural areas face higher risk of mortality than children living in urban area. A similar study by Kumar and Gemechis (2010) in Ethiopia shows that most of the deaths of children occurred in rural area. A study by Sahn and Stifle (2003) shows that child mortality in urban areas is lower relative to rural areas. A similar study in West Africa by Balk et al (2003) also found that children residing in

urban areas have a better chance of survival than those residing in rural areas. A study by Dashtseren (2002) shows that child mortality rates are higher in rural areas than urban areas.

This study suggests that first born children experience higher risk of dying than children whose birth order is six and above; children with birth order two up to five have a higher risk of dying than a child whose birth order is six and above. A study by Desta (2011) found that birth order is one of the determinants of child mortality showing that a first born child was exposed to a high risk of mortality. Another study by Balk et al (2003) indicates that first births are less likely to survive than higher order births.

The current study found out that those children whose mother‟s age is 15 to 19 years have the higher risk of dying relative to children whose mother age is 20 and above. A similar study in Malawi by Manda (1999) also found that mother age at first birth has a significant effect on child mortality showing that a child born to a younger mother experienced the highest risk of dying. Another study in Bangladesh by Mondal et al (2009) indicates that mother‟s age at birth is the most significant predictor of child mortality levels showing that the mortality level is higher for children whose mothers are under 20 years of age. A study by Balk et al (2003) also shows the same result. A study by Girson and Maurice (2010) found that children born to a mother of age below 20 have the highest risk of dying than the children born to a mother of age above 20.

This study shows that children born to single mothers are expected to experience a higher risk of dying than children born to a married mother. Desta (2011) shows that children born to women not currently married were exposed to higher risk of mortality than children born to women who are currently married. Goro (2007) also found that marital status is a significant determinant of child mortality showing that children born to women who are currently not in union are almost two times more likely to die than children born to women who are currently in union.

The current study indicates that mother‟s education is a significant determinant of child mortality showing that children born to illiterate mother experience higher risk of mortality than children born to mothers with primary and higher education. A study by Ezra and Gurum (2002) found a similar result. Kumar and Gemechis (2010) show that the mortality risk of children born to illiterate mothers is higher as compared to a child born to an educated mother. A similar study in Kenya by Hill (2001) shows that mother‟s education level has a significant impact on child mortality showing that children born to literate mother have the highest chance of survival. A study in China by Jacoby and Wang (2003) found that higher education level is negatively associated with child mortality risk. Another study by Klaauw and Wang (2003) in India and Chowdhury et al (2010) in Bangladesh also found that children born to illiterate mothers have the highest risk of mortality. A study by Ombok et al (2010) found that higher mortality is associated with illiteracy. A study by Jalandhar and Arokiasamy (2006) shows that the risk of early childhood mortality is more than twice for illiterate women compared with literate women.

Chapter Five

Conclusion and Recommendation

5.1. Conclusion

The Cox proportional hazards model results show that the most contributing factors of child mortality are bio-demographic and socio-economic factors. From those factors mother‟s age, marital status, birth order, mother‟s level of education and place of residence were found to be significantly associated with child mortality. Younger age (15-19 years) of mothers, being a single mother, first born, illiteracy of mothers and residing in rural areas all affect child survival negatively. The study shows that children born to mothers of age 20 and above have a higher chance of survival than children born to a mother in the age group 15 to 19 at first birth. Children born to currently married women have the highest chance of surviving than children born to women currently not married. Children with birth order six and above have the highest chance of survival as compared to children with birth order below six. Also children born to mothers with primary and higher education have the higher chance of survival than children born to a mother with no education. Children born in urban areas have a higher chance of survival than children born in rural areas.

5.2. Recommendations

Ethiopia is trying to achieve the millennium development goals by 2015. One of the goals is the reduction of child mortality. The government has implemented health oriented interventions to reduce child mortality. In order to reduce the rate of child mortality this study recommends: (i) improving the level of education of women is vital; (ii) awareness

about the danger of giving birth at early ages must be created; and (iii) because the survival experience of children in rural areas is much lower than in urban areas there is a need to improve the health infrastructures in rural areas.

References

1. Akhtar Hussain, Keramat Ali and Gunnar Kvale(1999): Determinants of mortality among children in the urban slums of Dhaka city, Bangladesh. Tropical Medicine and International Health volume 4, pages 758-764.
2. Amare Deribew, Fasil Tessema, Belaineh Girma (2007). Determinants of under-five mortality in Gilgel Gibe Field Research Center, Southwest Ethiopia. Ethiopian Journal of Health Development. 21(2). Pages 1-8.
3. Baker, R. (1999). Differential in child mortality in Malawi. Social networks project working Papers, No. 3. University of Pennsylvania.
4. Balk Deborah, Tom Pullum, Adam Storeygard, Fern Greenwell, and Melissa Neuman (2003) *Spatial Analysis of Childhood Mortality in West Africa.* Calverton, Maryland,USA: ORC Macro and Center for International Earth Science Information Network (CIESIN), Columbia University.
5. Chowdhury QH., Rafiqul I. and Kamal H.(2010): Socio-economic determinants of neonatal, post neonatal, infant and child mortality, in Bangladesh. International Journal of Sociology and Anthropology vol. 2(6), pages 118-125.
6. Collett, D. (2003). Modeling survival data in medical research (Second edition). Chapman and Hall/CRC, London.
7. Cornelia Kaldewei and Ingo Pitterle (2011): Behavioral Factors as Emerging Main Determinants of Child Mortality in Middle-Income Countries: A Case Study of Jordan. DESA working paper No. 103.
8. Dashtseren A. (2002). Determinants of infant and child mortality in Mongolia. Paper presented at the IUSSP Regional Conference, Bangkok, Thailand.

9. Desta Mekonnen (2011). Infant and Child Mortality in Ethiopia: The role of socioeconomic, demographic and biological factors in the previous five years period of 2000 and 2005. Lund University.

10. Ezra, M. and Gurum, E. (2002). Breastfeeding, birth intervals and child survival: analysis of the 1997 community and family survey data in southern Ethiopia.

11. Frank Kaharuz, Svend Sabroe and Flemming Scheutz (2000): Determinants of child mortality in rural eastern Uganda. Department of epidemiology and Social Medicine, University of Aarhus,Vennelyst Boulevard 6,DK-8000 Aarhus, Denmark.

12. Girson N. and Maurice M. (2010). Some socio-economic and demographic determinants of infant and child mortality in Tanzania: A case study of Karagwe District, Kagerar region.

13. Goro, M. (2007). The stalling child mortality: the case of three northern regions. The 5th conference of union for Africa population, Tanzania.

14. Hala A. (2002): The effect of water and sanitation on child mortality in Egypt. Environmental Economics Unit, Department of Economics, Gothenburg University, Sweden.

15. Hill, K., Bicego, J. and Mahy M. (2001). Childhood mortality in Kenya: An examination of trends and determinants in the late 1980s to mid 1990s. Macro International Report, Maryland, USA.

16. Hosmer DW, Lemeshow S. and May S. (2008): Applied survival analysis regression modeling of time to event data. John Wiley and Sons, Inc., New York.

17. Jacoby, H. and L. Wang (2003). "Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach" World Bank, Washington D.C.

18. Jalandhar P. and Arokiasamy P. (2006). High infant and child mortality rates in Orissa: An assessment of major reasons. Population Space and Place 12, pages 187-200.

19. Joshua K. and Jeroen G. (2009). Determinants of infant and child mortality in Zimbabwe: Result of multivariate hazard analysis. Demographic Research: Volume 21, pages 367-384.

20. Klaauw, V.B. and Wang L. (2003) „Child Mortality in Rural India", World Bank Working Paper, Washington DC: World Bank.

21. Kumar, P. and Gemechis File. (2010). Infant and child mortality in Ethiopia: As statistical analysis approach. Ethiopian Journal of Science and Education, Volume 5, no 2, pages 51-57.

22. Manda, S.O.M. (1999). Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. GenusLIV 143-164

23. Maurice Ombok, Kubaje Adazu,, Frank Odhiambo, Nabie Bayoh, Rose Kiriinya, Laurence Slutsker Mary J. Hamel, John Williamson, Allen Hightower, Kayla F. Laserson and Daniel R. Feikin. Geospatial distribution and determinants of child mortality in rural western Kenya 2002–2005. Tropical Medicine and International Health. Vol 15 No 4, pages 423-433.

24. Ministry of Finance and Economic Development (MoFED). Millennium Development Goals Report on Ethiopia, 2010.

25. Mohamed S.Mahfouz, Adil A.Surur, David A.Ajak and Eihab A.Eldawi (2009): Level and Determinants of Infant and Child Mortality in Malakal Town – Southern Sudan. Sudanese journal of public health, April 2009, vol.4 No.2:250-255.

26. Mutunga, C. J. (2004). Environmental determinants of child mortality in Kenya. Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi, Kenya.

27. Nazrul I. Mondal, Kamal Hossain and Korban Ali (2009): Factors Influencing Infant and Child Mortality: A Case Study of Rajshahi District, Bangladesh. J Hum Ecol, 26(1):31-39.

28. Sahn, D. E. and Stifel, D. C (2003). Exploring alternative measures of welfare in the absence of expenditure data. Review of Income and Wealth 49(4): 463-489.

29. Uddin, Jamal (2009): Child mortality in a developing country: A statistical analysis. Journal of Applied Quantitative Methods, Volume 4, pages 270-283.

30. UN Millennium Campaign. Millennium Development Goals report card, 2010.

31. UNICEF (2008), Five million child deaths every year in Africa. http://sanitationupdates.wordpress.com/2008/05/30/africa-unicef-reports-five-million-child-deaths-every-year/. Posted on May 30, 2008.

32. UNICEF (2008). Child mortality rate in Ethiopia falls by 40 percent. http://www.medindia.net/news/Child-Mortality-Rate-in-Ethiopia-Falls-by-40-Percent-UNICEF-32194-1.htm. Seen on 9/10/2011.

33. UNICEF: Monitoring the situation of children and women, 2010. http://www.childinfo.org/mortality.html. posted on September 2011.

34. United Nations. Child mortality report 2011: Level and trend of child mortality. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation.

35. Wang L. (2003). "Environmental Determinants of Child Mortality: Empirical Results from the 2000 Ethiopia DHS" World Bank, Washington D.C.

Appendix

Figure 1: Kaplan-Meier survival curve for the whole data

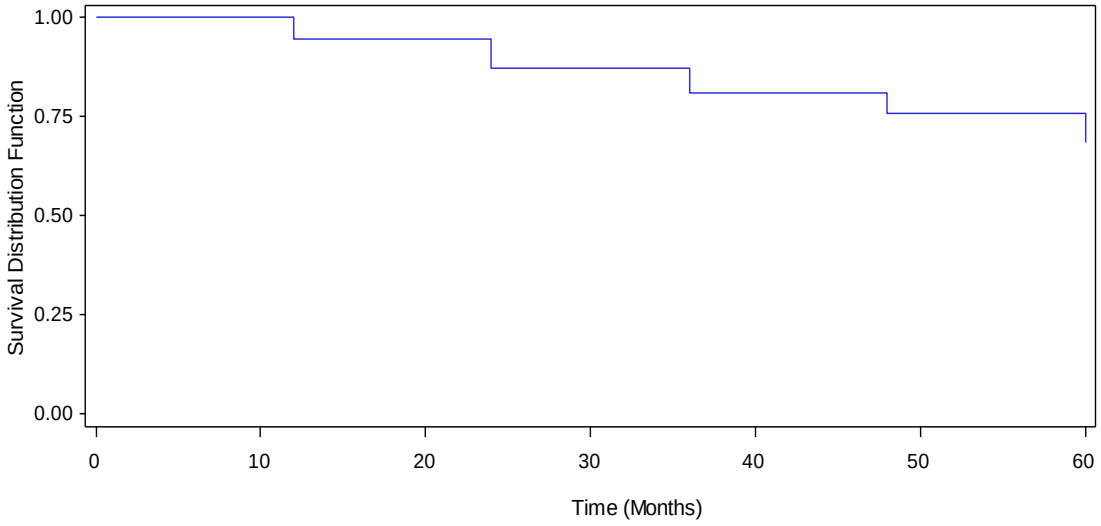

Figure 2: Survival curve by mother‟s education

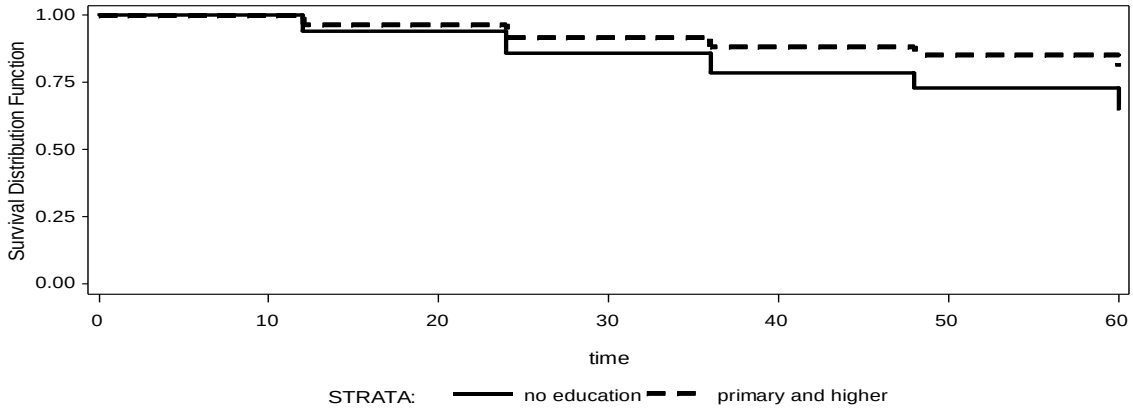

Figure 3: Survival curve by place of residence

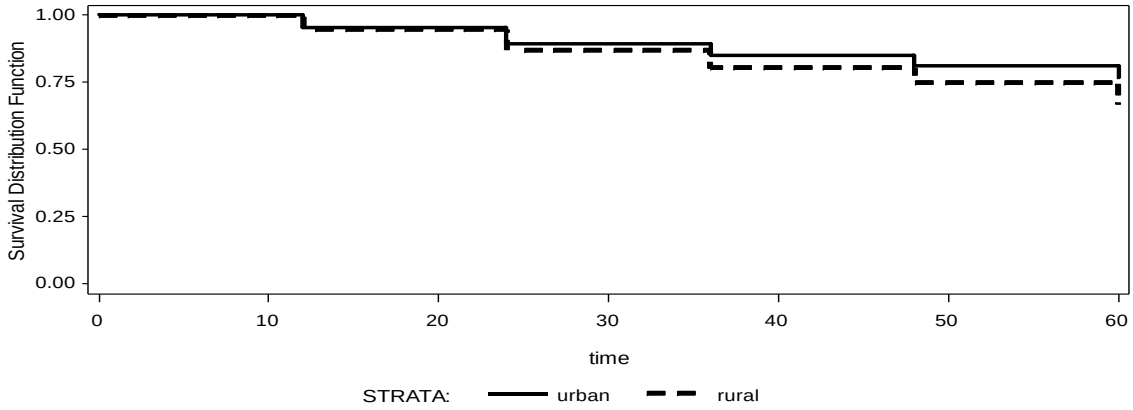

Figure 4: Survival curve by source of water

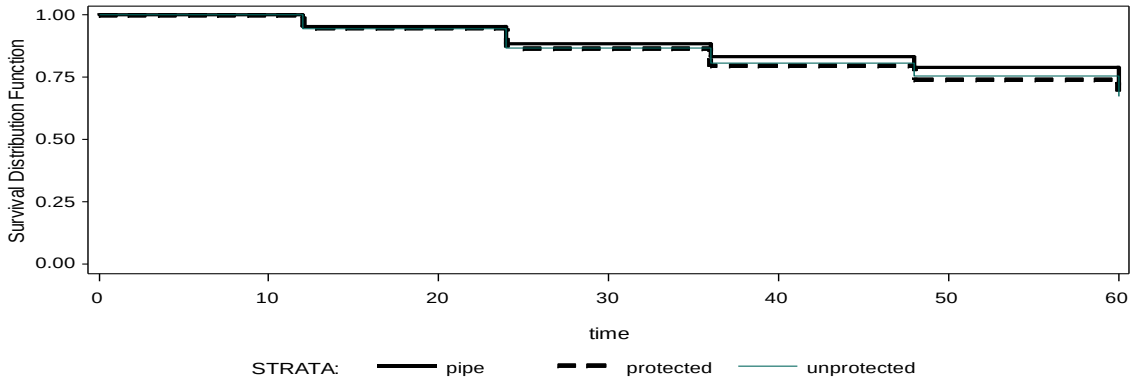

Figure 5: Survival curve by toilet facility

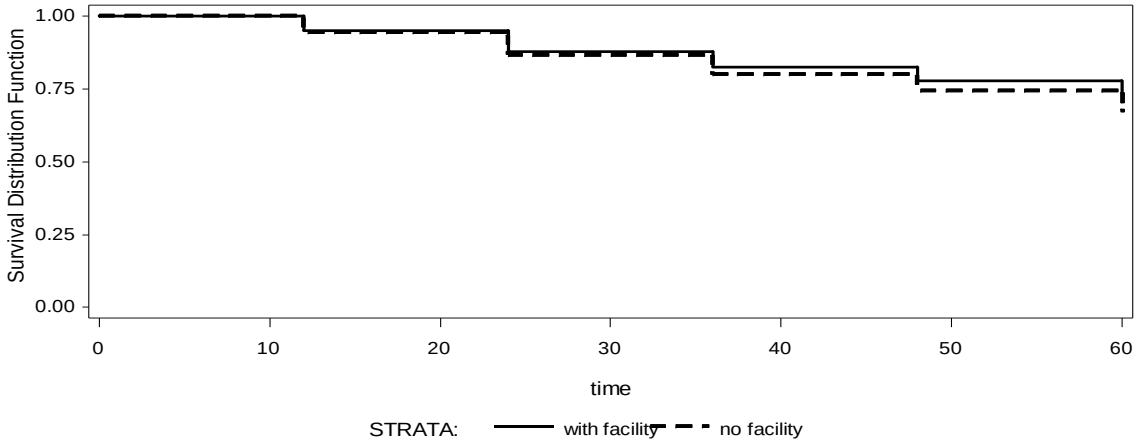

Figure 6: Survival curve by family size

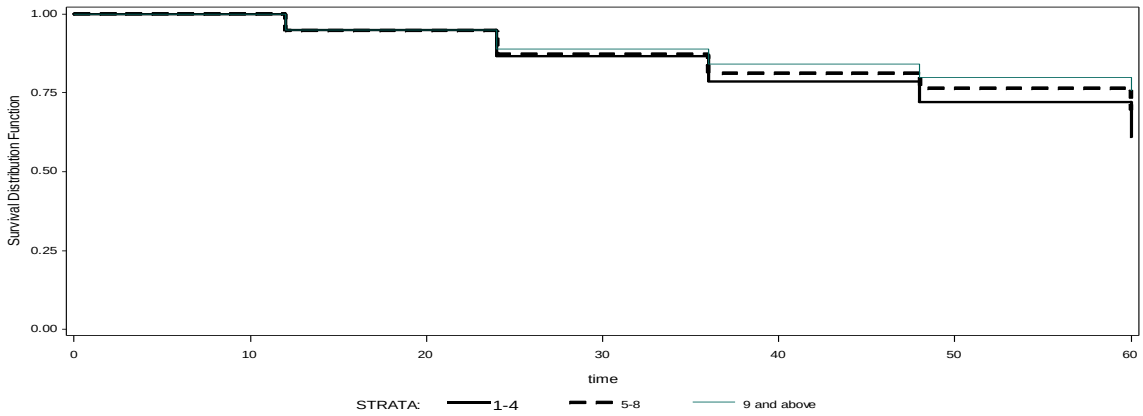

Figure 7: Survival curve by wealth index

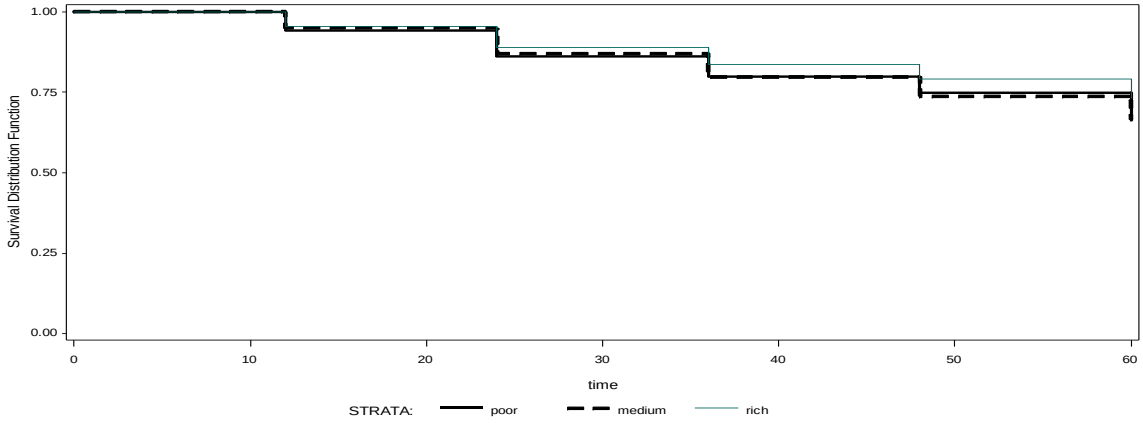

Figure 8: Survival curve by mother age at birth

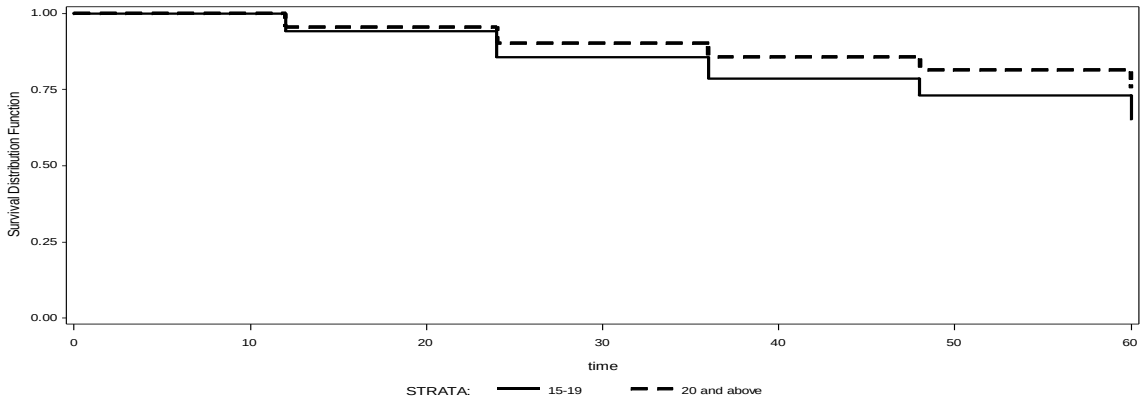

Figure 9: Survival curve by marital status

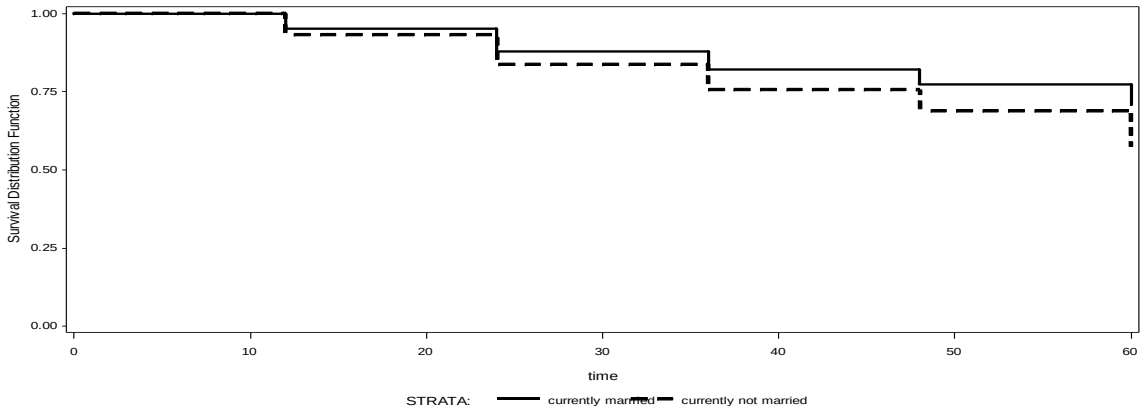

Figure 10: Survival curve by birth order

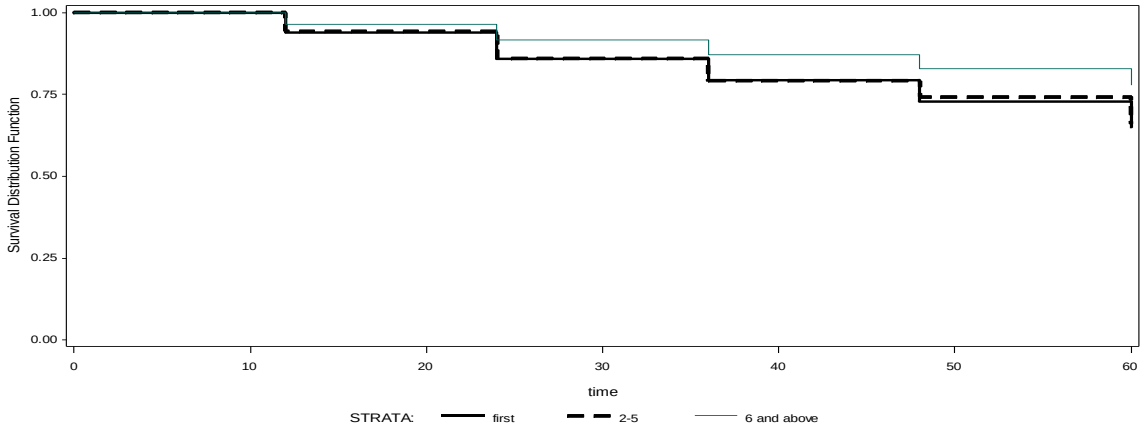

Figure 11: Survival curve by sex of a child

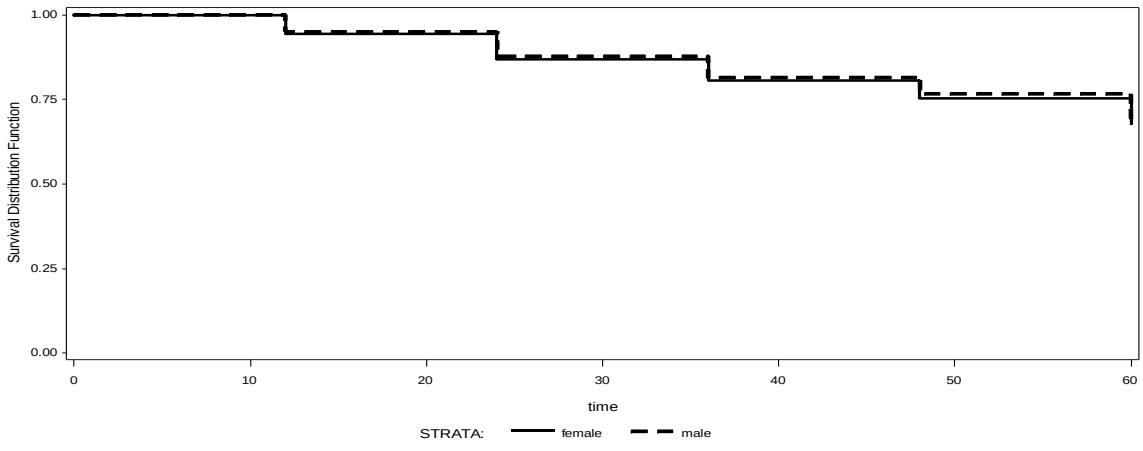

Figure 12: Schoenfeld residual plot by mother"s age

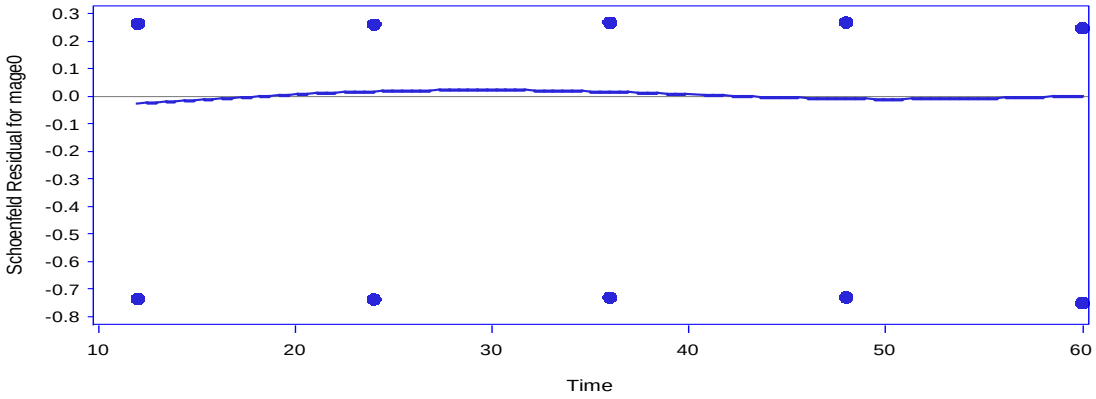

Figure 13: Schoenfeld residual plot by mother"s education

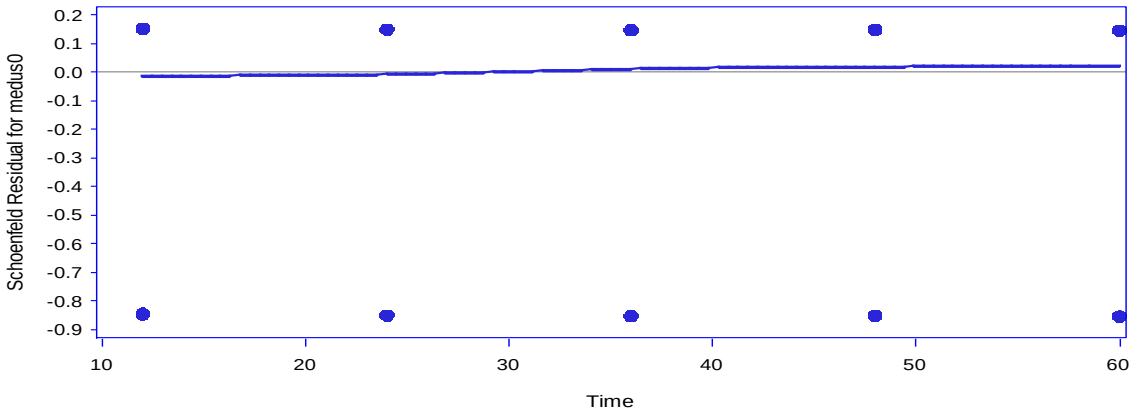

Figure 14: Schoenfeld residual plot by marital status

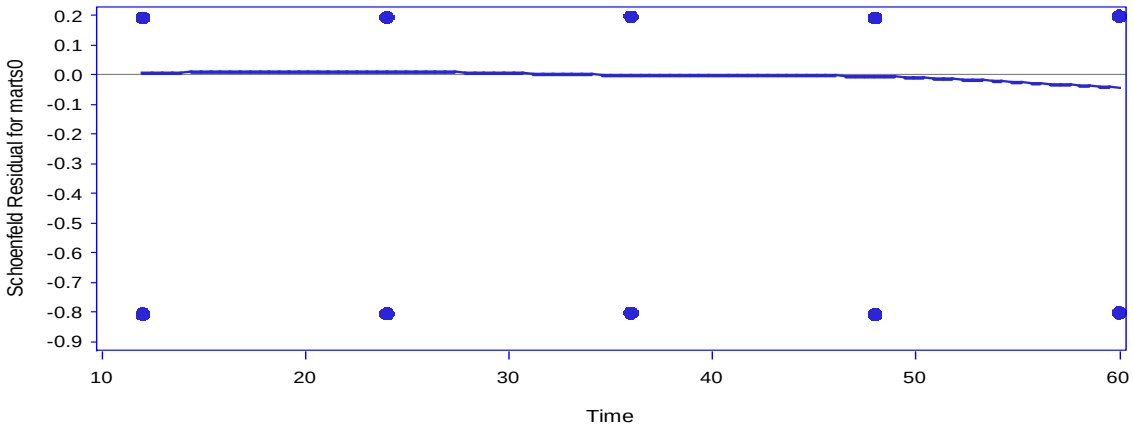

Figure 15: Schoenfeld residual plot by birth order

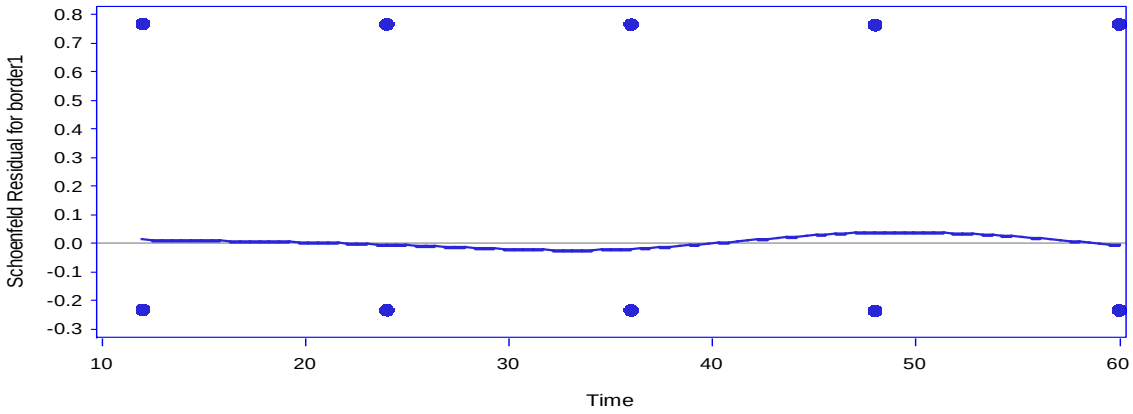

Figure 16: Shoenfeld residual plot by place of residence

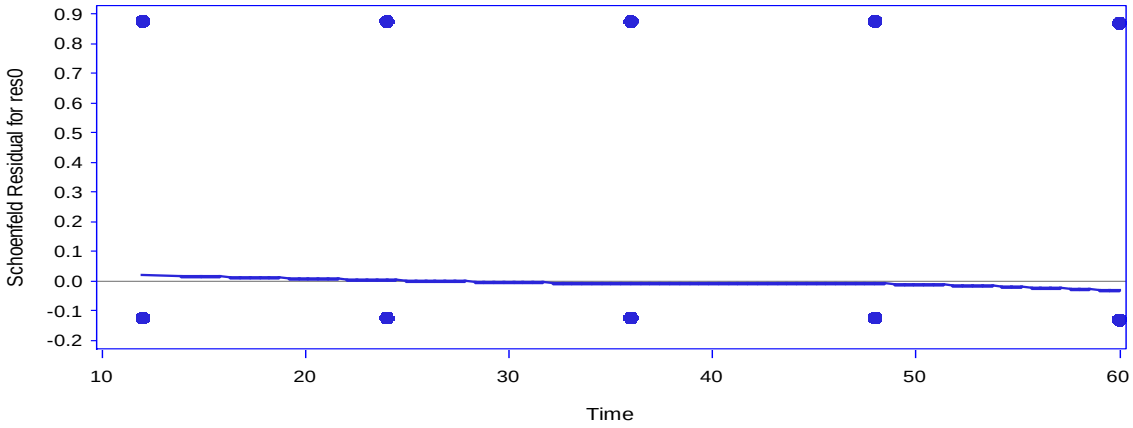

Figure 17: Plot of martingale residuals for assessing linearity of ungrouped mother‟s age

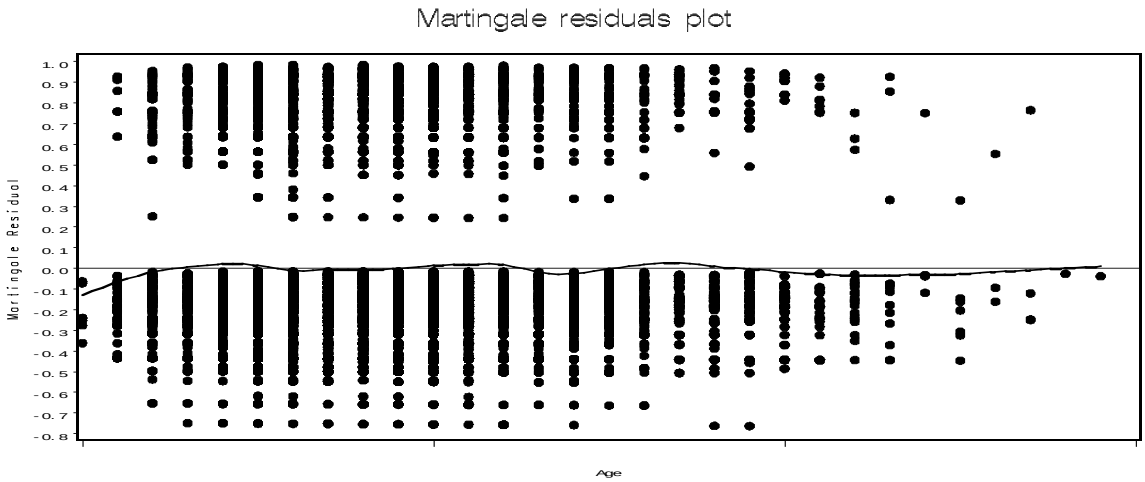

Figure 18: Martingale residual plot for checking outlier‟s