ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

FACULTY OF SCIENCE

DEPARTEMANT OF STATISTICS

Determinants of Child Survival Chances in Rural Ethiopia

BY

Ermias Dessie

A Thesis submitted to the office of Graduate programs of Addis Ababa University in partial fulfillment of the requirement for the Degree of Master of Science in Statistics

JULY 2009

Table of Content

Acknowledgment

Many thanks to almighty God, Then my sincere and heartfelt gratitude to my adviser, Dr.Emmanuel G/Yohannes who was beside me in the entire study and also giving me wonderful, reliable and important comments.

My appreciation and thanks are extended to other people who have contribution to this thesis directly or indirectly. And lastly but not the least I would like to thank my family, friends for their special and continuous cooperation and encouragement throughout my study period.

Abstract

This study investigates the effect of socioeconomic, environmental, demographic and health related variables on child mortality in rural Ethiopia. The study data set used originates from the demographic and health survey (DHS) conducted in Ethiopia 2005.The analysis was conducted using Cox proportional hazards model which analyses the effects of covariates on child mortality. The study shows that source of drinking water, birth order number, sex of child, breast feeding status, wealth index of household, father's education, mother's education and family size have significant contribute on child mortality.

CHAPTER ONE

INTRODUCTION

1.1 Background

Ethiopia is the third largest populous country in Africa with population of 73.9 million of which more than 84.2% live in rural areas (CSA, 2007). The land area of Ethiopia is estimated at about 1.1 million square kilometers and it is a federal democratic republic composed of 9 national regional states (NRS) Tigray, Afar, Amahara, Oromia, Somalia, Beni-Shangul-Gumuz, South Nations Nationalities and People Region (SNNRP), Gambella and Harari, and two administrative states (Addis Ababa city administrative and DireDawa council).

Ethiopia is one of the poorest countries in the world. Subsistence agriculture is the dominant sector in the economy. According to the World Fact book (2008) the proportion of population at the age group 15-64 year is 53.8 %. Total fertility rate is estimated to be 5.4 children born /woman and also fertility rate are estimated to be 2.4 and 6.0 children born/woman in urban and rural, respectively. In Ethiopia Under five mortality rate stands at 98 and 135 in the year 2005 per 1000 live births in urban and rural area, respectively (CSA,2005).

There are numerous causes of infant and under five mortality that have to be mentioned. Among these birth related causes, low household income, low parental education, unsafe water; poor sanitation facility and limited access to health care service are the dominant factors. Women and children are often more vulnerable to health threatening factors which are associated with poverty or inadequate standard of living. Although in Ethiopia living conditions differ largely between urban and rural dwellers, households living in the rural area are relatively much poorer than household living in urban areas. Rural areas of the country have poor access to health facilities, schools and piped water etc (UNICEF, 2008). The problem of child mortality is therefore most severe in rural areas.

This study focuses on the determinants of child mortality in the rural areas. More specifically examining how child mortality is related to household's environmental situations, demographic and socio-economic characteristics and health related variables is the center of attention. This may help in identifying households with high child mortality risks and targeting resources towards those households.

1.2 Statement of the problem

Normally poverty is the problem of least developed countries. Poverty is a pronounced deprivation of wellbeing. It is defined in terms of not having enough to eat, a low life expectancy, high rate of infant mortality, low educational opportunity, poor drinking water, inadequate health care and unfit housing conditions (Wolday, 2001). Thus the downtrodden rural society of the low income countries is fated to experience unhealthy environment which may result in loss of lives of both adults and children. This study focuses on the impacts of the household’s environmental situation, demographic, socio-economic and health related variables on the likelihood of the child’s death in rural Ethiopia.

1.3 Objective of the study

The primary objective of the study is to examine the household’s environmental situation, demographic, socio-economic and health related variables that are associated with child mortality in rural Ethiopia.

Specifically the aims of the study are:

- To identify the factors that contribute to child mortality.
- To estimate the probability of child survival within five years life time.
- To develop a statistical model that determines child mortality.

1.4 Significance of the study

In a country like Ethiopia that has poor social conditions (Education, housing, sanitation, etc) and more than 50% of its population illiterate and under the poverty line ,the level of communicable diseases is found to be the highest magnitude (CIA World Fact book ,2009). In studying child mortality, socio-economic, demographic and household environment have been given little attention by biomedical scientists since they primarily focus on the disease causing agents. So the importance of this study is to reveal the impact of the socio- economic, demographic and environmental variables on child mortality risks associated with household's environment in rural Ethiopia. Such information is critical for prioritizing public investments in order to maximize the health benefit for given resources, particularly in the context of achieving the targets set by the Millennium Development Goals (MDG) on child mortality.

1.5 Scope of the study

The study will investigate the effects of environmental, demographic, socio-economic and health related variables on the child's survival chances in rural Ethiopia. The reason for selecting rural Ethiopia as study area is due to inadequate social services and the population is leading a life without adequate shelter or unacceptable housing condition, unsafe source of drinking water and no basic sanitation facilities.

1.6 Limitation of the study

This study has encountered an analytical problem that is related to the data set used in the study. That is, information on covariates refers to the time of the survey, but not necessarily to the exact time of exposure of the children. In other word, the analysis is restricted to the period 2000-2005 and it is assumed that no investments have been undertaken by the households in order to improve their living conditions. For example, the source of drinking water and sanitation facility are constructed based on the state of affairs at the time of the survey.

CHAPTER TWO

LITERATURE REVIEW

2.1 Theoretical literature

Most countries, whether industrialized or not, experience important demographic changes. One of the most important of which is the transition from a phase of rapid population growth to one in which population growth is low. Initially, countries experience a mortality decline and fertility increase, both contributing to a rise in population growth. This process is known as demographic transition (Houndroyiannis and Papaetrou, 2002).

Modern theories of population relate fertility, mortality and economic development with people's behavior. However, the modern economic theory of population emphasizes the interdependence between infant mortality and fertility in the context of economic theories of behavior (Cigno, 1998: Becker et al. 1999). In Cingo's model, reduction in child mortality may either rise or lower fertility. When the level of child mortality is high, reduction in it is likely to raise both fertility and survival enhancing expenditures on children, because it lowers the price of a surviving child. The effect of child mortality reduction at the early state of transition would be to increase the number of surviving children. Many more children, therefore, survive to productive ages, thus reducing the cost of their upbringing while also contributing to their families' production effort. Feyissa (1984) disclosed that high fertility strains budgets of poor families, reducing the family's resources, available to feed, educate and provide health care to children. Recent theories espoused the idea that there is an intermingled relation between fertility and mortality depending on the behavioral pattern of individuals.

Demographers view declines in mortality and fertility as components of a single "demographic transition". There are many competing theories about why fertility declines. One theory favored by demographers is that the fertility decline is due to the mortality decline, i.e., it is the response to the improved survival chances of the offspring. An alternative theory proposed by Becker

(1981) suggests that the demographic transition occurs since at high levels of income, the adverse effect of the opportunity cost of children on child rearing dominates the positive income effect. This theory, however, is inconsistent with the simultaneous occurrence of the fertility transition in countries that markedly differed in their levels of income (Kalemli-Ozcan, 2002). Proponents of modern population theories (Houndroyiannis and Papaetrou, 2002) argued that there might be a case that fertility may be induced by mortality.

More to the point, mortality has a direct relationship with fertility. Since, higher fertility can be a signal for narrow birth spacing, parents may not allot enough time to look after their children. Above all, a larger family size with meager resources forces the family to malnourishment, lower health care utilization, use of improper sanitation system and water supply. For this reason, there is a high risk of infant and child death.

And further the WHO Commission established that reducing child mortality is a key to economic growth, for a variety of reasons. Firstly, societies with high rates of infant and child mortality have higher rates of fertility, and large number of children reduce the ability of poor families to invest in health and education, resulting in an under trained, under- skilled productive work force.

In addition to its effect on fertility, child mortality is also important for the human capital investment decision of parents. Lower mortality implies a higher rate of return to education, and thus declining child and youth mortality provides an important incentive to increase investment in the education of each child (Kalemli-Ozcan, 2002). Heckman (2000) also argues that the return to human capital investment is highest before age five. Secondly, lowering infant mortality rates tends to lower, not raise, population growth over the long run, as people adjust to having smaller families (WHO, 2001).

Child mortality rate is one of the most important sensitive indicators of the socioeconomic and health status of a community. Child's survival depends on the socioeconomic conditions of their environment (Madise et al 2003). Thus poorest countries have an experience of high mortality rate (UN, 2007). Therefore, child mortality is related with poverty. Precisely, poor people are caught in a vicious cycle as follows:

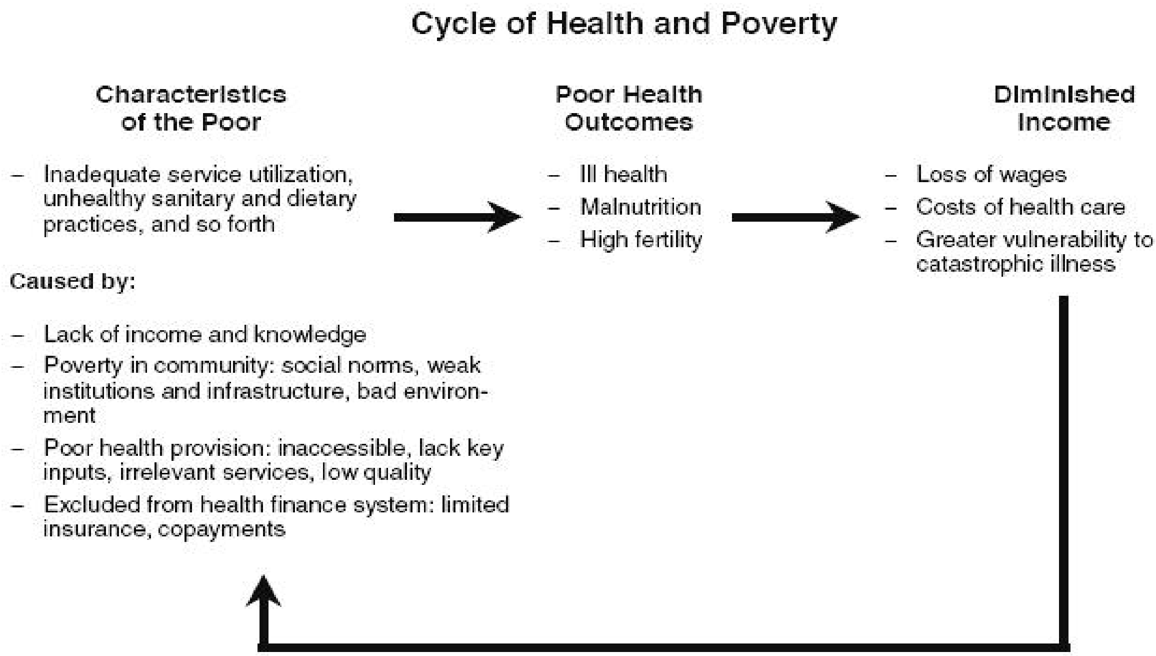


Figure 2.1. Health and Poverty Linkages

(Source: Claeson et al. (2005), Health, Nutrition and Population Chapter 18, p-203)

2.1 Empirical literature

The main importance of sorting out the linkages among the proxy factors which contribute to healthiness and the complementary nature of those factors would help to transform the unhealthy environment into a healthy one. Besides knowing the contribution of basic human necessities and their intricate relationship to health can be deemed as an indication for intervention areas that might improve the quality of life. For instance, improving the socioeconomic condition, economic growth, increase in per capita health expenditure, the appropriate level of sanitation facilities, access for clean drinking water and appropriate level of nutrition would contribute to the wellbeing of children and society at large. Considerable attention has been paid in recent years to the influence of socio-economic status of parents, household environmental characteristics and demographic factors on both the health and mortality of children.

Clive (2007) tried to examine environmental determinants of child mortality in urban Kenya. He constructs a duration model framework to capture socio-economic and environmental characteristics. The model is estimated using data from the Kenya demographic and health survey (KDHS). The estimation results show that socio-economic and environmental characteristics have significantly different impact on mortality rates at different ages. He used the estimated model for policy experiments by simulating the magnitude of child mortality reduction as a result of environmental improvement, particularly reduction in indoor air pollution due to dirty cooking fuels, provision of safe drinking water and sanitation.

Likewise Lyun (2000) tried to look at the effects of environmental situation of women on child mortality in south western Nigeria. This study was a follow-up analysis that attempts to further analyse and determine the relative significance of environmental and maternal factors on childhood mortality in two contrasting towns in southwestern Nigeria. The research design takes advantage of the integration of the medical and social sciences. The results of the current analysis reveal interesting insights into child mortality and maternal factors on one hand and domestic environmental conditions on the other. He gives credence to an ecological perspective as a way to understand the complexities behind child survival. Domestic environmental conditions (such as water and sanitation facility) were stronger predictors of child mortality in the more developed study town, Ota, than the more traditional town, Iseyin. However, in both sites maternal factors, in particular age of mother at marriage and age of mother at first childbirth were statistically significant predictors of child mortality. Mother’s education was only significant in the more urbanized center, and generally remains inconsistent in its relationship with child mortality. Furthermore, child mortality rates 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, especially the market environment.

Kim (2004) also examines the determinants of infant and child mortality in Korea. The study identifies the major factors which were associated with infant and child mortality in Korea using data from the 1974 Korean national fertility survey (KNFS). In urban and rural areas, mother's education, maternal age, number of rooms in household home, previous and successive birth interval were the most important determinant of infant mortality and child mortality. Infant mortality was also significantly affected by sex of child in urban areas and by birth order in rural areas. And also demographic factors are more important determinants of infant mortality in rural areas, where as socioeconomic factors play a major role for infant mortality in urban areas.

White (2006) examines the determinants of infant and child mortality in Andhra Pradesh (where the Young Lives project is taking place) and Kerala and the factors explaining their differential performance. The determinants of mortality are estimated using a Cox proportional hazards model. Infant mortality is found to depend on biological factors, including mother's age and birth order, and also factors related to health service provision such as tetanus injection and use of antenatal services. Economic well being is not significant once these other factors are taken into account. By contrast, economic well-being was found to be a significant determinant of child mortality, but substantially outweighed in importance by other factors such as maternal education and knowledge of health practices (ORS) and access to safe water. The data also show gender discrimination in Andhra Pradesh, notably toward girls with only female siblings, which is absent from Kerala. He conclude that raising service levels across India toward the levels found in Kerala is a necessary step towards meeting the MDGs, and that the success of these efforts is reinforced by female empowerment.

Ali (2002) tried to examine the effects of water and sanitation on child's mortality risk in Egypt. The central question of the study was whether improvements in water and sanitation services were leading to a decrease in mortality of children under the age of five in Egypt. The data set used in the study originated from the DHS conducted in Egypt between November 1995 and January 1996. The analysis was conducted using a three part model: a probit model specification for the neonatal case, nonparametric, semi-parametric, and parametric duration model specifications are used for different age intervals including infant and children less than five years. The findings of the study showed that access to municipal water and improved sanitation facilities had significant impact on child's mortality than other sources of water. Moreover, the study indicated mother education as an important factor to reduce child mortality. Based on the results she recommended that improving the society's knowledge about health care and hygiene is crucial to reduce risk of child's death.

Woldemichael (1998) tried to investigate the effects of household environment on child mortality in urban Eritrea, controlling for other socio-economic factors. The data used for the study originated from the 1995 Eritrea Demographic and Health Survey that include environmental and socioeconomic variables. The data was analyzed by classifying the under five children into three age cohorts. For the neonatal period he used a logistic regression model while the Cox Proportional Hazard Model was used so as to address the censored observation. The study also assumed that in the surveyed area there is no major change in water or sanitation system. The study found that improvement in the provision of water supply and toilet facilities are likely to reduce mortality especially beyond the neonatal period.

Jacoby and Wang (2004) also look on the environmental determinants of child's mortality risk using a competing risks approach. In their study they evaluated alternative empirical methodologies for estimating the impact of environmental factors on child mortality in micro data. Their primary question was whether taking into account causes of death using a competing risk model than an all cause hazard model affects conclusions about the effectiveness of policy intervention. Using the 1992 China Health Survey (CHS), they estimated three models for the length of time that the child survives. The models are the competing risk model (CRM) that allows for cause specific vector of coefficients, the piece wise weibull (PWW) which allows for age-specific coefficients and weibull model (WM) which allows for duration dependence. Looking at four hypothetical policies: (1) Universal Private access to safe water; (2) Universal access to basic sanitation facilities (3) Universal access to clean fuels; and (4) universal female primary education attainment, they found little difference in their impact across the three models. However, the largest and most significant impact comers from access to safe water. Increasing this dummy variable from its sample mean value of 0.33 to universal access would save more than 3 lives out of 1000 births, based on the CR and W models, and some what less than 3 lives according to the PWW model. And further they showed that policies that achieve universal female primary education attainment also have a significant impact on reducing under-five mortality (although only at the 10% level of significance for W and CR models).

Jacoby and Wang also showed in their analysis that interventions targeted at improving access to safe water in rural china have a statistically significant impact on reducing Under-5 mortality probability. Unlike the above studies, they found evidence suggesting that the impact of access to sanitation or clean cooking fuels on child mortality risk in rural china was insignificant. The policy simulation also discloses that targeting environmental policies (in particular private access to safe water) in poor localities or poor households can avert more under-five deaths than untargeted interventions. Most importantly, they emphasized that the potential analytical advantage of using the CR model rests on validating causal relationships using information on cause of death. Moreover, they uncovered that the probability of dying from diarrhea disease is the most responsive, at least in relative terms, to interventions that improve access to safe water.

Household's financial capacity to obtain appropriate health services and adopt healthy dietary and sanitary practices has an impact on child mortality. There are studies conducted to investigate the impacts of household income and consumption behavior on the likelihood of child's death. Lee et al (1997) examined the effects of improved nutrition, sanitation and water quality on child health in high mortality population. They paid particular attention on non-random allocation of household resources to children and to the selectivity effects of health interventions via their effects on child survival. Unlike the previous studies, they employed a simultaneous equation model with selectivity. The results show that child mortality was affected by source drinking water, sanitation facilities, a child specific nutritional intake and mothers' education. In contrast to other studies, they concluded that variation in water sources and improvement in sanitation facilities do not have significant impact on child mortality, but wealth and parental schooling levels were significantly and positively associated with higher survival. The rationale behind such deviated result was that they focused on the reduced allocation of household income on children's health care with better facilities and not to survival selectivity.

There are many ways of measuring poverty, of which the infant mortality rate is 'regarded as one of the most revealing measures of how well the society is meeting the needs of its people'. Ali (2001) made an attempt to investigate the interaction between child mortality and poverty in Pakistan. The analysis was relied on the Pakistan Socio-economic Survey conducted during April to July 1999. The study applied multiple classification analysis that requires dependent variable not to be badly skewed. The study found that mother's work participation, household crowding, unadjusted housing conditions and malnutrition have significant negative impact on the livelihood of a child, where maternal education showed a considerable reducing effect on child mortality. And further the study figured out heterogeneity among urban and rural dwellers. The study also mentioned that malnutrition had a positive impact on mortality, but its effect is negligible in urban areas.

Wang (2002) described the existence of significant differences both in level and changes in child mortality between urban and rural areas. At the beginning of the 1990s, IMR and U5MR were 87 and 143 per 1000 births in rural areas whereas both figures are much lower in urban areas, being 67 and 105 per 1000 live births in China. Over the course of the 1990s, the annual rates of reduction in IMR were 1.7 per cent and 2.1 per cent, and in U5MR were 2.1 per cent and 2.6 percent for rural and urban areas, respectively. According to Wang , these two pieces of empirical evidences suggest that across low income countries, health interventions implemented in the 1990s have not been sufficiently effective at targeting the poor. Low income status and hence lower health care utilization due to lack of finance and access to health care services could also be deemed as a prime factor for infant and child mortalities. Schellenberg et al. (2002) examined the risk factors for child mortality in rural Tanzania. They conducted a community based Nested Case Control Study of Post-neonatal death in children less than five years. They investigated the effects of demographic, socio-economic, health seeking behavior and household environment on enhancing or impeding infant or child mortality. The results have shown that maternal education, socioeconomic status and breastfeeding have significant impact on infant and child mortality.

Poverty, disease, malnutrition and adverse reproductive and demographic characteristics keep on as impediments to bring improvements in infant and child mortality in many developing countries. Aguirre (1996) has investigated the relationship between fertility and working patterns including demographic and socio-economic factors in Bolivia. The study was conducted using a Bio-demographic Hazard Model (which constitutes of three fitted Cox Proportional Hazard Models). The empirical findings of the study confirmed that pace of childbearing, breastfeeding, maternal age, birth order and use of contraceptives have important impact on infant mortality, even after controlling for maternal education. This entails that these reproductive pattern covariates have consistent effects even after the inclusion of education as a control.

Zakir and Wunnava (1997) disclose that female literacy rates, per capita income, fertility rates, women's work status and government expenditures on health do have an impact on infant mortality rates. Of these, fertility rates and female literacy rates have the strongest impact on infant mortality rates. But the former have negative significant impact on the livelihood of infants.

Hailemariam and Tesfaye (1997) conducted a study in a small urban community in Sebeta, a town 25 km west of Addis Ababa, Ethiopia, They showed that higher birth order, early pregnancy and late pregnancy do have significant negative impact on the livelihood of infants. They used Cox's Proportional Hazard Model and their findings shows that maternal education, occupation of the father, household income, source of drinking water, availability of latrine and survival status of older offspring has direct effect on infant mortality.

By and large, the empirical literature espoused the theoretical literature confirming that environmental, socio-economic, demographic, health related and maternal reproductive factors have significant impact on child mortality. Thus it can be said that child mortality is significantly influenced by source of drinking water, toilet facility, female education, household income, breast feeding, age of mother at birth, father's educational level and preceding birth interval.

CHAPTER THREE

DATA SOURCE AND METHODOLOGY

3.1 DATA SOURCE

The data set that used in this study originates from the Demographic and Health Survey (DHS) conducted in Ethiopia in 2005. DHS is a large cross-sectional data set that is comparable across countries. It contains information obtained from women aged 15-49 years regardless of whether they had any preschool children. Ethiopia DHS is nationally representative and hence it helps in the construction of wide range of basic population indicators (in addition to mortality rates). These indicators include basic household characteristics, fertility, household's access to services such as safe water and sanitation, utilization of basic health and education services, mother's education and knowledge of treatment of common child diseases. The Ethiopia DHS 2005 is the second survey of its kind that provides information on population and health, estimates of a kind that are comparable to similar surveys conducted in other developing countries and permit making international comparisons.

The 2005 sample was designed to provide estimates for the health and demographic variables of interest for the following domains: Ethiopia as a whole; urban and rural areas of Ethiopia (each as separate domain); and 11 geographic area (9 region and 2 city administration), In general, a DHS sample is stratified, cluster and selected in two stages. In the 2005 EDHS a representative sample of approximately 14,500 households from 540 clusters was selected. The sample was selected in two stages. In the first stage, 540 clusters (145 urban and 395 rural) were selected from the list of enumeration areas (EA) from the 1994 population and housing census sample frame. In the census frame, each of the 11 administration areas is subdivided into zones and each zone into weredas. In addition to these administrative units, each wereda was subdivided into convenient areas called census EAs. The 2005 Ethiopia DHS is a nationally representative survey covering 14,070 women aged 15-49 among these 4423 and 9647 women are living in urban and rural areas respectively.

In the survey extended information is collected for children that are born five years before the survey was conducted. This means that information on children born prior to this period is not included since the longer the recall period the more likely the respondents misreport the case.
A total of 9861 children less than 59 months were identified in the households of selected clusters and among these 1358 and 8503 are living in urban and rural areas respectively. Then after a certain rearrangement the present study is based on 7395 children in rural areas that are the center of attention.

3.2 THE VARIABLES IN THE STUDY

The dependent variable used in this study is child survival time, which is measured as the duration in **month** starting from birth to death (if the event occurred) or from birth to the survey date (censored data).

Predictor Variables (Independent Variables)

The independent variables in this study are classified into three groups.

a) Bio-demographic and health related variables
b) Socio-economic variables
c) Environmental variables

a) Bio demographic and health related variables

This group of variables consists of sex of child, birth order, breast feeding status, age of the mother at birth, education level of the mother and father, preceding birth interval, immunization, place of delivery and marital status.

b) Socio-economic variables

In this category two variables are considered. These are mother's work status and wealth index of the household. The wealth index of a household has three categories i.e. poor, medium and rich.

c) Environmental Variables

The environmental variables considered in this study are source of drinking water, sanitation facilities and child bed net.

Table 3.1: Variables of interest and their description

variables	levels	Coding
Demographics and health related variables		
Sex of child(sex)	male	0 (reference categories)
	female	1
Family size	1-3(groupfamily1)	0 (reference categories)
	4-6(groupfamily2)	1
	>6(groupfamily3)	2
Mother's education(mlitrancy)	illiterate	0 (reference categories)
	Literate	1
Current Breast feeding status	No	0 (reference categories)
	Yes	1
Birth order number	1 (groupbord1)	0 (reference categories)
	2-4 (groupbord2)	1
	> 4 (groupbord3)	2
Marital status (maritalstatus)	Not Separate	0 (reference categories)
	Separate	1
Place of delivery	Home	0(reference categories)
	Health center	1
Age of mother at birth	< 20years(groupage1)	0 (reference categories)
	20-34year(groupage2)	1
	35-49year(groupage3)	2
Father's education (flitracy)	Illiterate	0 (reference categories)
	Literate	1
Socio-economic variables		
Mother work status (mother work)	No	0 (reference categories)
	Yes	1
Wealth index of household (wealth)	poor	0 (reference categories)
	Medium	1
	Rich	2
Environmental variable		
Source of drinking water (water)	Surface water	0 (reference categories)
	Covered water	1
Sanitation facility (sanitation)	Pit-toilet	0 (reference categories)
	No toilet facility	1
Child has bed net for sleeping(bednet)	No	0 (reference categories)
	yes	1

3.3 METHODOLOGY

The method that will be used in the study is basically survival analysis i.e. duration model like Cox-proportional hazard model and other statistical methods appropriate for the given data set.

3.3.1 Theoretical background of survival analysis

Survival analysis is just another name for time to event analysis. The term survival analysis is used predominately in biomedical sciences where the interest is observing time to death either of patients or of laboratory animals. Time to event analysis has also been used widely in the social sciences where interest on analyzing time to events such as job changes, marriage, birth of children, child mortality and so forth. The engineering sciences have also contributed to the development of survival analysis which is called "reliability analysis" or "failure time analysis" in this field, since the main focus is in modeling the time it takes for machines or electronic components to break down.

The developments from these diverse fields have for the most part been consolidated into the field of "survival analysis". There are certain aspects of survival analysis data, such as censoring and non-normality that generate great difficulty when trying to analyze the data using traditional statistical models such as multiple linear regressions. The non-normality aspect of the data violates the normality assumption of most commonly used statistical model such as regression or ANOVA, etc. A censored observation is defined as an observation with incomplete information. There are four different types of censoring possibility: right truncation, left truncation, right censoring and left censoring.

We have focused exclusively on right censoring. When an observation is right censored it means that the information is incomplete because the subject did not have an event during the time that the subject was part of the study. The point of survival analysis is to follow subjects over time and observe at which point in time they experience the event of interest. It often happens that the study does not span enough time in order to observe the event for all the subjects in the study. This could be due to a number of reasons. Perhaps subjects drop out of the study for reasons unrelated to the study (i.e. patients moving to another area and leaving no forwarding address).

The common feature of all of these examples is that if the subject had been able to stay in the study then it would have been possible to observe the time of the event eventually. And also right censoring is further classified according to whether the time is fixed by design or considered as random such that type one, type two and random right censoring. This study utilizes random right censoring. Random right censoring is such that the subjects enter into the study at different time point and the study will be terminated at a pre-specified time. Of course we can rescale all the starting time to 0. Here in our case the child survival time is measured as the duration in months starting from infant birth to death (if the event occurred) or from infant birth to the survey date (censored data) .

3.3.2 Hazard function and Survival function

The other important concept in survival analysis is the hazard rate. From visual inspection of data with discrete time (time measured in large intervals such as month, years or even decades) we can get an intuitive idea of the hazard rate. For discrete time, the hazard rate is the probability that an individual will experience an event at time t while that individual is at risk for having an event. Thus, the hazard rate is just the unobserved rate at which events occur. If the hazard rate is constant over time, say 1.5 for example, this would mean that one would expect 1.5 events to occur in a time interval that is one unit long. Furthermore, if a person had a hazard rate of 1.2 at time t and a second person had a hazard rate of 2.4 at time t then it would be correct to say that the second person's risk of an event would be two times greater at time t. It is important to realize that the hazard rate is an un-observed variable yet it controls both the occurrence and the timing of the events. It is the fundamental dependent variable in survival analysis.

Another important aspect of the hazard function is to understand how the shape of the hazard function will influence the other function of interest such as the survival and density function. Once we have modeled the hazard rate we can easily obtained these other functions of interest.

Let us think of time as a continuous random variable t which is the duration of stay in the state. The population is assumed to be homogenous with respect to the systematic factors, regressor variables, which affect the distribution of *t*. This means that every one's duration of stay will be a realization of random variables and have the same probability distribution (Lancaster, 1990).

Our interest is to estimate the probability of a child dying within the next period after surviving for t months as a result of different factors. We will focus on children that are born alive and model their mortality probabilities until reaching age 5.We will use duration models to specify these mortality probabilities. The survival function denotes the probability that an individual survives up to a particular time t. The function is obtained from what is known in the survival analysis literature as the failure function $F(t) = prob(T \le t)$ which is the probability that an individual will die before time t .This implies that the survival function:

$$prob(T > t) = 1 - F(t) = S(t) \tag{3.1}$$

is the probability that an individual will survive at time *t* or beyond *t*. Then the probability density function $f(t)$ will be:

$$f(t) = \frac{d(1 - S(t))}{dt} = -\frac{d(S(t))}{dt} \tag{3.2}$$

The hazard function $h(t)$ is the instantaneous rate of failure and *time dependence function. If the hazard function for a particular distribution slopes upwards or downwards, then the distribution has positive or negative duration dependence. This implies that the likelihood of failure at time t conditional on duration up to t, is increasing (decreasing) with t.* $h(t)$ is given by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t / T > t)}{\Delta t}$$

Then after some manipulation it can be shown that:

$$h(t) = \frac{f(t)}{S(t)} = \frac{-dS(t)}{dt} . \frac{1}{S(t)} = -\frac{d}{dt} \ln S(t) \qquad (3.3)$$

Integrating both sides we have:

$$- \int_0^t h(u)du = \int_0^t \frac{d}{du} \ln S(u)du = \ln(S(t))$$

Taking the exponential of both sides we get:

$$S(t) = \exp\{-\int_0^t h(u)du\} = \exp\{-H(t)\} \qquad (3.4)$$

where H(t) is the cumulative distribution of the hazard function.

Equation (3.4) describes how one can calculate the probability distribution of duration of state occupancy given the hazard function. From equation (3.3) and (3.4) the density function of time (t) can be written as:

$$f(t) = h(t) * \exp\{-\int_0^t h(u)du\} \qquad (3.5)$$

Thus, the density function is determined by the hazard function. Modeling of survival data usually employs the hazard function or the log hazard. For example, assuming a constant hazard, $h(t) = v$, implies an exponential distribution of survival times with density function:

$$f(t) = ve^{-vt}$$

Other common hazard model include $\log(h(t)) = v + gt$, which leads to the Gompertz distribution of survival times, and $\log(h(t)) = v + g\log(t)$ which leads to the Weibull distribution of survival time.(see Cox and oakes (1984) for these and other hazard models). In both the Gompertz and Weibull distributions, the hazard can either increase or decrease with time. Moreover in both instances, setting g=0 yields the exponential model.

3.3.1 Estimation of survival function and models

Several methods have been developed for the analysis of survival data. Some of these are:

i. Descriptive statistics include life tables and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from a sample.

ii. Non-parametric techniques that are available for comparing the survival experience between two or more groups. The most common and widely used of these techniques include log-rank test, Generalized Wilcoxon test, Peto-Prentice test, among others.

iii. Multivariate Methods most widely used are the Cox-proportional hazards model and accelerated failure time model. They are used to analyze the effect of explanatory variables on the hazard rate. The Cox-proportional hazards model is the most popular regression method for analysis of censored survival data. The popularity is due to the following reason:

- Very flexible method of estimation since the base line hazard is estimated non-parametrically and eliminates the risk of corrupting the estimated hazard parameters while the effect of the covariate takes a particular functional form.
- Fairly easy to fit.
- Does not make assumption about the underlying survival distribution (does not require the knowledge of the shape of the survival distribution)
- Does not require estimation of the baseline hazard rate to estimate the regression parameters (the primary interest of the model is estimation of the regression parameters with no attention on the baseline hazard function).
- Standard software exist like SPSS,SAS,STATA,S-PLUS,GENSTAT etc are capable to handle proportional hazards model and Kaplan-Maier survival function estimation.

3.3.1.1 Kaplan-Meier survival function estimator

Kaplan-Meier estimator of the survivorship function (Kaplan and Meier, 1958) is also called the product limit estimator. Kaplan-Meier estimator is used to estimate survival time of child and construct survival curve to compare survival experience of children between different categorical variables.

The Kaplan-Meier estimator of the survivorship function is defined as:

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} = \prod_{t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{3.6}$$

where $t_{(1)},...,t_{(m)}$ is the set of m distinct death times observed in the sample.

d_i is the number of deaths at $t_{(i)}$

n_i is the number of individuals "at risk" right before the i^{th} death time.

3.3.3.2 Comparison of Survivorship Functions

After providing a description of the overall survival experience in the study, we usually turn our attention to a comparison of the survivorship experience in key subgroups in the data. Standard statistical procedures, such as t-test, rank sum test and analysis of variance may be used without censored observations. Modifications are required when censored observations are present in the data.

When 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. 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, Generalized wilcoxen, Taron-ware test and so on.

The calculation of each test is based on the contingency table of group by status at each observed survival time. Most software packages base their estimator of the expected number and variance of deaths on the hypergeometric distribution, defined as follows:

$$\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i} \quad \text{and} \qquad \hat{v}_{1i} = \frac{n_{1i} n_{oi} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

where: n_{oi} is the number at risk at observed survival time $t_{(i)}$ in group 0

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

d_{oi} is the number of observed deaths in group 0

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

n_i is the total number of individuals at risk at time $t_{(i)}$

d_i is the total number of deaths

The test statistic depends on the type of test is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times. These tests may be defined in general as follows:

$$Q = \frac{\left[\sum_{i=1}^{m} w_i (d_{1i} - \hat{e}_{1i}) \right]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (3.7)$$

Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, then the significance of Q may be tested using the chi-square distribution with one degree of freedom. We can exeted the above test to compare k groups.

- **Log rank test**

The log rank test, sometimes called the Cox-Mantel test, is the most well known and widely used test. This test is based on weights equal to one, i.e. $w_i = 1$. Thus, the log rank test statistic becomes:

$$Q_{LR} = \frac{\left[\sum_{i=1}^{m}\left(d_{1i} - \hat{e}_{1i}\right)\right]^2}{\sum_{i=1}^{m}\hat{v}_{1i}} \sim \chi^2(1) \qquad (3.8)$$

3.3.3.3 Proportional hazards regression model

The proportional hazards model, also called Cox model, is a classical semi-parametric method. It relates the time of an event, usually death or failure, to a number of explanatory variables known as covariates. The proportional hazard function for an individual i is defined as follows:

$$h(t, x_i) = h_0(t).c(x_i) \qquad i = 1, 2....n \qquad (3.9)$$

where $x_i = (x_{i1}, ...x_{ik})'$ is a column vector of k measured covariates for the i^{th} individual, $h_0(t)$ is the baseline hazard function and $c(x_i)$ is a link function.

$h_0(t)$ is the hazard for the respective individual when all independent variable values are equal to zero. That is, it is independent of the covariates.

Cox (1972) proposed a link function:

$$c(x_i) = \exp(x_i{}'\beta) \quad i = 1, 2.....n \qquad (3.10)$$

where $\beta = (\beta_1.....\beta_k)'$ is a vector of regression parameters. The Cox proportional hazard model is thus defined as :

$$h(t, x_i) = h_0(t).\exp(x_i'\beta) \quad i = 1, 2......n \tag{3.11}$$

Here $exe(x_i'\beta)$ is the relative risk associated with the regressor. It is the exponential form of the estimated coefficients of the hazard model. It can be interpreted as the number of times that the baseline hazard is multiplied by each unit change in the covariates. $h_0(t)$ is an arbitrary and unspecified baseline hazard function.

- ***Properties of Cox proportional hazards model***

i. The survival function is

$$S(t, x_i, \beta) = \exp(-H(t, x_i, \beta)$$

Under the Cox model, the survival ship function is given by

$$S(t, x_i, \beta) = [S_0(t)]^{\exp(x_i'\beta)} \tag{3.12}$$

where $S_0(t) = \exp(-\int_0^t h_0(u)d_u)$ is the baseline survivorship function.

ii. Consider two observations i and $i*$ that differ in their x-values with the corresponding linear predictors:

$$\Omega_\mathrm{i} = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$$

and

$$\Omega_{\mathrm{i*}} = \beta_1 x_{i*1} + \beta_2 x_{i*2} + ... + \beta_k x_{i*k}$$

The hazard ratio for these two observations,

$$\frac{h_{i(t)}}{h_{i^*(t)}} = \frac{h_{o(t)*e^{\Omega \mathrm{i}}}}{h_{o(t)*e^{\Omega i*}}} = \frac{e^{\Omega i}}{e^{\Omega i*}}$$

is independent of time t. This implies that the hazard ratio does not depend on time. This is to say that the risk of failure is the same no matter how long the subject has been followed (this is main assumption of the Cox-proportional hazard model).

iii. Interpretation of the regression coefficient β's. Consider a one unit increase in one covariate x_p and fix the remaining covariates; i.e

$$x_i = (x_{i1}, \ldots, x_{ip}, \ldots\ldots x_{ik}) \text{ and } x_{0i} = (x_{i1}, \ldots\ldots, x_{ip} + 1, \ldots\ldots x_{ik})$$

Then the log hazard ratio is:

$$\ln\left[\frac{h(t, x_{i0})}{h(t, x_i)}\right] = \ln\left[\frac{h_0(t).e^{\beta_1 x_{i1} + \ldots\ldots\ldots\ldots \beta_p (x_{ip}+1) + \ldots\ldots \beta_k x_{ik}}}{h_0(t).e^{\beta_1 x_{i1} + \ldots\ldots\ldots\ldots \beta_p x_{ip} + \ldots\ldots \beta_k x_{ik}}}\right] = \beta_p$$

Thus, the hazard ratio is :

$$\frac{h(t, x_{i0})}{h(t, x_i)} = e^{\beta_p}$$

Thus, β_p is the change in log hazard ratio at anytime with a unit change in the pth covariate x_p, adjusting for the effect of other covariates.

The rate of change in hazard ratio is:

$$\frac{h(t, x_{i0}) - h(t, x_i)}{h(t, x_i)} = e^{\beta_p} - 1$$

- $(e^{B_p} - 1)*100\%$ is the percentage change in the hazard for a unit change in x_p adjusting for the remaining covariates.

3.3.2 Estimation of the parameters of the proportional hazards model

Consider n independent individuals. The data that we need for the Cox's model is represented by (t_i, δ_i, x_i) where t_i is the length of time i[th] individual is observed, $x_i = (x_{i1}, ...x_{ik})'$ and δ_i is an indicator of i[th] individual's censoring status, i.e,

$$\delta_i = \begin{cases} 1 \;\; if \;\; the\, event \;\; occurres \\ o \;\; if \;\; the \;\; event \;\; does\, not\, occurre(censoring) \end{cases}$$

From the relationship of the density function, survival function and hazard function (equation3.3) we have: $f(t, \delta_i, x_i) = h(t, \delta_i, x_i) * S(t, \delta_i, x_i)$

The likelihood function can be constructed for right censored data as:

$$L(\beta) = \prod_{i=1}^{n} f(t, \delta_i, x_i)^{\delta_i} . S(t, \delta_i, x_i)^{1-\delta_i}$$

$$= \prod_{i=1}^{n} \left[h(t, \delta_i, x_i) \right]^{\delta_i} .S(t, \delta_i, x_i) \quad \text{(from equation 3.3)}$$

$$= \prod_{i=1}^{n} \left[h_0(t).\exp(x_i ' \beta) \right]^{\delta_i} . \left[S_0(t) \right]^{\exp(x_i ' \beta)}$$

since $h(t, x_i, \beta) = h_0(t).\exp(x_i ' \beta)$ and $S(t, x_i, \beta) = \left[S_0(t) \right]^{\exp(x_i ' \beta)}$

The log likelihood thus becomes:

$$l(\beta) = \sum_{i=1}^{n} \delta_i h_0(t) + \sum_{i=1}^{n} \delta_i x_i ' \beta + \sum_{i=1}^{n} \exp(x_i ' \beta) \ln(S_0(t)) \tag{3.13}$$

This is referred to as full likelihood. The full maximum likelihood estimator of β can be obtained by differentiating with respect to the parameters β 's and the baseline hazard $h_0(t)$ .Unless we explicitly specify the baseline hazard $h_0(t)$, we cannot obtain the maximum likelihood estimator of the full likelihood. To avoid the specification of baseline hazard, Cox (1972) proposed a partial likelihood approach that treats the baseline hazard as a nuisance parameter and removes from the estimating equations.

Partial likelihood

To estimate the parameters of the Cox proportional hazards model, we proceed as follows

- Consider the probability that an individual experiences an event at t_i given that an event occurs at that time, say $p_r(A/B)$
- Let R_i be the set of individuals at risk at time t_i and assume that there is only one failure at time t_i (that is, no ties). Then the probability that individual i with covariates x_i is the one who experience the event at time t_i is equal to:

$$p_r(A/B) = \frac{h(t_i, x_i, \beta).}{\sum_{j \in R_i} h(t_j, x_j, \beta).}$$

where A: Individual i has an event at time t_i and B: One event occurred at time t_i.

We then define the partial likelihood for n individuals as:

$$L_p(\beta) = \prod_{i=1}^{n} p_r(A/B) = \prod_{i=1}^{n} \left[\frac{h(t_i, x_i, \beta).}{\sum_{j \in R_i} h(t_j, x_j, \beta)} \right]^{\delta_i} \tag{3.14}$$

Now, order failure times into m distinct failure times (event times) $t_{(1)} \ldots\ldots t_{(m)}$ and let $x_{(i)}$ be the vector of covariates at ordered failure time $t_{(i)}$. Then the likelihood becomes:

$$L_p(\beta) = \prod_{i=1}^{m} \left[\frac{h(t_i, x_i).}{\sum_{i \in R_{t(i)}} h(t_i, x_i).} \right]$$

$$= \prod_{i=1}^{m} \left[\frac{h_0(t_{(i)}).\exp(x_{(i)}'\beta)}{\sum_{i \in R_{t(i)}} h_0(t_{(i)}).\exp(x_{(i)}'\beta)} \right]$$

The baseline hazard will cancel out and the partial likelihood is given by the expression:

$$L_p(\beta) = \prod_{i=1}^{m} \left[\frac{\exp(x_{(i)}'\beta)}{\sum_{i \in R_{t(i)}} \exp(x_{(i)}'\beta)} \right] \tag{3.15}$$

This partial likelihood is only a function of β's.

The log partial likelihood function is :

$$l_p(\beta) = \sum_{i=1}^{n} \left\{ x_{(i)}'\beta - \ln\left[\sum_{j \in R_{t(i)}} \exp(x_{(i)}'\beta) \right] \right\} \tag{3.16}$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of equation (3.13) with respect to β , setting the derivative equal to zero and solving for the unknown parameters.

The derivative with respect to β is

$$\frac{dl_p(\beta)}{d\beta} = \sum_{i=1}^{m} \left\{ x_{(i)} - \frac{\sum_{j \in R_{t(i)}} x_j \exp(x_{(i)}'\beta)}{\sum_{j \in R_{t(i)}} \exp(x_{(j)}'\beta)} \right\}$$

$$= \sum_{i=1}^{m} \left\{ x_{(i)} - \sum_{j \in R_{t(i)}} w_{ij}(\beta) x_{(i)} \right\}$$

$$= \sum_{i=1}^{m} \left\{ x_{(i)} - \overline{X}_{wi} \right\} \quad \text{where} \quad w_{ij}(\beta) = \frac{exe(x_j'\beta)}{\sum_{j \in R_{t(i)}} exe(x_{(j)}'\beta)} \text{ and } \overline{X}_{wi} = \sum_{j \in R_{t(i)}} w_{ij}(\beta) x_j$$

In fact, the estimator obtained when setting the derivative in question (3.13) equal to zero and solving for β yield the value such that the sum of the risk set weighted means of the covariates are equal to the sum of the covariates over the non-censored subjects. We denote the maximum partial likelihood estimator as $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)'$.

The elements of the k by k information matrix are obtained by second-order partial differentiation of $l_p(\beta)$ with respect to β:

$$I(\beta) = -\frac{\partial^2 l_p(\beta)}{\partial\beta\partial\beta'}$$

The diagonal elements are

$$\frac{\partial^2 l_p(\beta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{m}\sum_{j\in R_{t(i)}} w_{ij}(x_{ik}-\overline{X}_{w_{ik}})^2$$

And the off-diagonal elements are

$$\frac{\partial^2 l_p(\beta)}{\partial\beta_k\partial\beta_l} = -\sum_{i=1}^{m}\sum_{j\in R_{t(i)}} w_{ij}(x_{ik}-\overline{X}_{w_{ik}})(x_{jl}-\overline{X}_{w_{il}})$$

The estimator of the covariance matrix of the maximum partial likelihood estimator is obtained as the inverse of the observed information matrix evaluated at the maximum partial likelihood estimator.

$$\operatorname{var}(\widehat{\beta}) = I(\widehat{\beta})^{-1} \tag{3.17}$$

If there are ties in the data set, the true partial log-likelihood function involves permutations and can be time-consuming to compute. In this case either the Breslow (1974) or Efron (1977) approximation to the partial log-likelihood can be used.

3.3.3.4 Partial likelihood tests

There are three different tests to assess the significance of the coefficient i.e. the partial likelihood ratio test, the Wald test and the score test. The null hypothesis, $H_0 : \beta = 0$ (the over all model is not a good fit) .This shows that the model with the included covariates does not explain the survival duration of the child better as compared to with out covariates model (null model). This means that at least one of the covariates included in our model has not a significant contribution in explaining the survival duration of the child.

i. **The Partial likelihood ratio test**

The partial likelihood ratio test is calculated as twice the difference between the log partial likelihood of the model containing the covariates and the log partial likelihood for the model not containing the covariates:

$$G = 2\{l_p(\beta) - l_p(0)\} \tag{3.18}$$

where $l_p(\underset{\sim}{o}) = -\sum_{i=1}^{m} \ln(n_i)$ and n_i denotes the number of subjects in the risk set at observed survival time $t_{(i)}$.

Under the null hypothesis that all coefficients are equal to zero, this statistic will follow the chi-square distribution with k-degrees of freedom.

ii. The Score test statistic is

$$\chi_s^2 = u'(0)\left[I(0)\right]^{-1} u(0) \tag{3.19}$$

where $I(0)$ is the information matrix evaluated at the coefficient vector equal to zero and $u(0)$ is the score vector at the coefficient vector equal to zero. χ_s^2 is distributed asymptotically as chi- square with k degree of freedom.

iii. **The Wald test**

The Wald test is obtained from equivalent theory which states that, under the null hypothesis, the estimator of the coefficient $\hat{\beta}$ will be asymptotically normally distributed with mean vector equal to zero and a covariance matrix that is estimated by the expression in question (3.17).

The Wald test statistic is:

$$\hat{\beta}' I(\hat{\beta})\hat{\beta} \tag{3.20}$$

This is also distributed asymptotically as chi-square with k degree of freedom.

3.3.5 Practical issues in proportional hazard model

1. **Model development**

   i. **selection of the covariates**

There are three different methods of selection of best subset of the covariates, i.e. purposeful selection, stepwise selection, best subset selection. But for this particular study the purposeful selection is used.

Purposeful selection of covariates

Purposeful selection of the covariates begins with a multivariable model that contains all the variables those are significant in the bivariate analysis at the 20-25 percent level of significant, as well as any other variables not selected with the criterion, but which judged to be important confounder. This full model could be the beginning of multivariable model, then using the p-values from the Wald tests of the individual coefficients we could identify covariates that might be deleted from the model and retain some covariates that are "important" even though they are not significant. This selection process provides preliminary subsets of covariates.

ii. checking linearity of continuous covariates

To check the linearity of continuous covariates, we can use plot of martingale residuals against the continuous covariate. Then, if the resulting plot is random (no systematic pattern) or the resulting smooth plot is a straight line, it shows linearity.

2. Assessment of model adequacy

Model adequacy refers to how well the fitted regression surface describes the data cloud. The requirements for the model assessment in Cox-proportional hazards model are testing the assumption of proportional hazards and overall summery measures of goodness-of-fit.

i. checking model assumption

The check for the validity of the proportionality assumption is made based on the analysis of Schoenfeld residuals. The basic idea in the wake of the proportionality assumption test based on the Schoenfeld residuals is to retrieve the residuals from Cox model. In its simplest form when there are no tied observations, the Schoenfeld residual for covariate x_{ik} $k = 1, 2 \ldots. p$ is given by:

$$r_i = x_{(i)} - \frac{\sum_{j \in R_{t(i)}} x_j \exp(x_{(i)}'\beta)}{\sum_{j \in R_{t(i)}} \exp(x_{(j)}'\beta)}$$

That is, r_i is the difference between the covariate value for the events and the weighted average of the covariate values (weighted according to the estimated relative hazard from a Cox model). Suppose that the coefficient on x_{ik} does vary with time so that β_j can be expressed as:

$$\beta_j(t) = \beta_j + q_j g_j(t) \qquad j = 1 \ldots.. k$$

where g (t) is some function of time (for example, $g(t) = \ln t$ and g $g(t) = \ln H_o(t)$)

The proportional hazards assumption requires that $q_j = 0$. Grambsch and Therneau (1994) provide a method of scaling the Schoenfeld residual to form:

$$E(r_j + \beta_j) = \beta_j(t) \text{ and } E(r_j) = q_j g_j(t)$$

Consequently, a graph of r_j versus t_j provides an assessment of the proportional hazards assumption. The null hypothesis, $H_0 : q_j = 0$ for all j is equivalent to saying that the plotted curve has zero slope. The rejection of the null hypothesis indicates a deviation from the proportional hazards assumption. This is equivalent to testing that the log hazard ratio function is constant over time (Cleves and Gould, 2004).

ii. Over all goodness of fit and measures

- **Over all goodness fit**

Until quite recently, all of the proposed test statistics for the overall goodness of fit of proportional hazards model were difficult to compute in most software packages and overall goodness of fit was based on a significance test of the coefficients for the added variables. Hosmow and Lemeshow(1998) use the score test in proportional hazard model by introducing G-1 dummy/design variables based on estimated risk scores $x_i'\beta$. Using the counting process approach, they derive an expression for the covariance matrix of the vector of G sums. They show that their quadratic form test statistic has a chi-square distribution with G-1 degree of freedom.

The test statistics is:

$$X^2_{cal} = -2[l_0 - l_1] \sim \chi^2_{G-1}$$

where l_0 is including only the main and interactions covariates and l_1 is including main, interactions covariates and G-1 dummy variables.

- **Measure *of goodness of fit***

Unlike regression analysis, there is no analogous measure to R^2 in measuring the model performance for proportional hazards model that can be easily calculated and interpreted and which is a stand alone measure. In Proportional hazards model all measures of goodness of fit depend on the proportion of censored observations (example, perfect adequate model may have low R^2 due to a high percentage of censored data).

Based on the partial likelihood, coefficient of determination (R_p^2) is given by:

$$R_p^2 = 1 - \exp\{2/n[l_0 - l_p]\}$$

where: n is the number of observation, l_p is the log partial likelihood for the fitted model with p-covariate and l_o is a log partial likelihood for the fitted model without covariates.

CHAPTER FOUR

ANALYSIS AND DISCUSSION

4.1 Descriptive analysis

In any applied setting, a statistical analysis should begin with a thoughtful and thorough univariate descriptive of the data before proceeding to more complicated models. Thus, we shall start with a descriptive analysis that will be used to highlight our subsequent findings.

Table 4.1 **summary of demographic and health related variables**

Variables		Frequency	Percent
Sex of child	Male	3,761	50.9
	Female	3,634	49.1
Family size	1-3 members	1331	18.0
	4-6 members	2988	40.4
	>6 members	3,076	41.6
Mother literacy	illiterate	6,269	84.8
	literate	1,126	15.2
Breastfeeding status	Not breastfeed	1,971	26.7
	breastfeed	5,424	73.3
BORD number	1 number	1790	24.2
	2-4 number	2470	33.4
	> 4 number	3135	42.4
Marital status	separated	293	4.0
	Not separated	7,102	96.0
Age of mother at birth	< 20years	2078	28.1
	20-34year	3357	45.4
	35-49year	1960	26.5
Father literacy	illiterate	4,934	66.7
	literate	2,461	33.3

Place of delivery	home	7076	95.7
	health center	319	4.3

As shown in Table 4.1, 50.9% of children in the study are males while the remaining 49.1% are females. The table also shows that 28.1% of mothers are less than 20 years old, 45.4% of mothers are between 20-34 years old and the remaining 26.5% of the mothers are between 35-49 years old when they gave birth. Regarding parent education, 15.2% of children's mothers and 33.3% of fathers are literate where as the remaining 84.8% of mothers and 66.7% of fathers are illiterate. It is observed that 96% of children were living with both their mother and father, while the reaming children were living with only their mothers. As the birth order shows in the table, 24.2% of children were born as the first child, 33.4% were born as second, third or fourth child and 42.4% were born as fifth and above child. Family size in the table shows that, 18% of children were from a family size of 1 - 3 members, 40.4 % of them were form 4-6 members and

41.6% of them were from 6 and above members in their family. Results shows that 4.3% of children were delivered at the health center where as 95.7% of them were delivered at home.

From Annex1 we can see that 52.2% and 20.8% of children came from poor and medium households, respectively, while the remaining 27% of children came from rich households. The work status of mothers in the annex shows that 88.8% of mothers are not employed whereas the remaining are employed.

As shown on Annex 2, 48.3% of children's belongs to households were using protected (covered well) drinking water whereas the remaining were not using protected drinking water (i.e using surface water). Only 2.9% of children's belongs to households which have toilet facility but 97.1% of households did not have toilet facility. 90.4% of children had bed net for sleeping while the remaining 9.6% of children were not using bed net.

The fundamental building block of survival analysis is an estimate of the survival function (see annex 3). Thus, after providing a descriptive analysis of the overall survival experience of the child, we usually turn our attention to a comparison of the survivorship experience across the different levels of a categorical variable. When comparing group of children between categorical variables, it is always a good idea to use graphical display of the survival experience of each group. (See annex 4). For instance children whose mothers are literate have more survival experience than those from illiterate mother. In order to test whether the observed differences based on the graphical displays are statistically significant or not, we use log-rank test. That is, the log rank statistic helps to test the null hypothesis that there is no difference in the survival experience of children between categorical variables at any time point. The results of the log rank test are displayed in Table 4.2.

Table 4.2 Log-rank test Comparison of Survival function between different levels of categorical variables.

Covariates	**Chi-Square**	**DF**	**P-Value**
mother literacy	5.7524	1	0.0165**
Source of drinking water	166.8788	1	0.0000*
family size	171.1577	2	0.0000*

sanitation facility	0.6743	1	0.4116
Have bed net for sleeping	0.08290	1	0.7734
Current marital status	0.09429	1	0.7588
father literacy	9.3543	1	0.0022*
mother work status	0.3517	1	0.5532
Sex of child	5.7392	1	0.0167**
Place of delivery	1.6598	1	0.1976
Wealth index	5.4983	2	0.0164**
Birth order number	13.09182	2	0.0014*
age of mother at birth	16.3452	2	0.0003*
Breast feeding status	301.0367	1	0.0000*
*significant at 1%	**significant at 5%		

As indicated in Table 4.2, the different levels of the categorical variables such as mother literacy, source of drinking water, breast feeding status, family size, father literacy, sex of child, birth order number, wealth index and age of mother's at birth are statistically significant. This means that children mortality is not the same in each category of these covariates. However, sanitation facilities, mother work status, usage of bed net for sleeping, place of delivery and current marital status are not statistically significant. One possible explanation for this is that, these are homogeneous covariates in rural part of Ethiopia. For instance, as can be seen from Table 4.1, annex 1 and annex 2, only 4% of parents are separated and only 2.9% of household have sanitation facilities. Further, almost all deliveries took place at home (95.7%) and 90.4% of children do not have bed net.

4.2 Estimation of model parameters

The estimated model is the Cox proportional hazards model:

$$\lambda(t,x_i)=\lambda_0(t).\exp(x'_i\,\beta)\quad i=1.....k$$

The variables considered in the model are birth order number, mother's age at birth, mother's education, father's education, source of drinking water, breast feeding, wealth index, sex of child and family size. The result of the fitted proportional hazards model is given in Table 4.3.

Table 4.3 Partial likelihood estimates for fitted proportional hazards models

	B	SE	Wald	df	Sig.	Exp(B)
Mliteracy	-.266	.128	4.295	1	.038**	.767
water	-1.038	.088	138.440	1	.000*	.354
Fliteracy	-.172	.096	3.222	1	.073***	.842
sex	-.180	.078	5.249	1	.022**	.835
beastfeeding	-1.151	.082	199.326	1	.000*	.316
wealth			15.746	2	.000*	
wealth(2)	.352	.099	12.685	1	.000*	1.422
wealth(3)	.291	.101	8.325	1	.004*	1.338
groupage			5.983	2	.050***	
groupage(2)	-.270	.124	4.750	1	.029**	.764
groupage(3)	-.395	.169	5.468	1	.019**	.674
groupbord			23.929	2	.000*	
groupbord(2)	.138	.138	1.004	1	.316	1.148
groupbord(3)	.639	.165	15.070	1	.000*	1.894
Groupfamily			87.084	2	.000*	
groupfamily(2)	-.824	.118	48.732	1	.000*	.439
groupfamily(3)	-1.317	.141	86.683	1	.000*	.268

*significant at 1% **significant at 5% ***significant at 10%

4.3 **Test for proportional hazards Assumption**

To check whether the Cox model fit is valid, we must check the proportionality assumption. That is the effects of covariates on risk remain constant over time. We illustrated the Schoenfeld residuals (1982). The scaled Schoenfeld residuals, which help to perform a variable-by-variable test, can be obtained from the fitted Cox model. The test statistics based on scaled Schoenfeld residuals is given in Table 4.4 assesses whether the null hypothesis, $H_0 : q_j = 0$ (the coefficient of x_{ik} does not vary with time) is valid.

Table 4.4 Test of proportional hazards assumption

Variables	**rho**	**chi2**	**df**	**Prob>chi2**
mliteracy	-0.03268	0.45	1	0.5015
water	0.02362	0.23	1	0.6318
fliteracy	0.00415	0.01	1	0.9321

sex	0.09493	3.66	1	0.5557
beastfeeding	0.18014	13.48	1	0.2222
wealth2	0.10224	4.20	1	0.4405
wealth3	0.00144	0.00	1	0.9765
grupedage2	0.13131	7.90	1	0.5049
grupedage3	0.06424	1.71	1	0.1908
groupedbord2	0.00413	0.01	1	0.9275
groupedbord3	0.00582	0.02	1	0.8975
grupedfamily2	-0.13475	8.04	1	0.3046
grupedfamiily3	-0.07596	2.63	1	0.1049
Global test		38.09	13	0.3343

The results from the PH test for all variables in Cox model show that there are no statistically significant coefficients and, hence, β's are not time varying coefficients and the proportional hazard assumption is satisfied. Moreover, the graph of scaled schoenfeld residuals against time have a zero slope confirming that the assumption of proportionality hazards is not violated. (see annex 5)

4.4 The overall goodness of fit tests and measures

As indicate in table 4.5, we reject that the overall model is not a good fit. This shows that the model with the included covariates explains the survival duration of the child better as compared to without covariates model (null model). This means that at least one of the covariates included in our model has a significant contribution in explaining the survival duration of the child.

Table 4.5 Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr >Chi-square
Likelihood Ratio	553.1483	13	<.0001
Score	628.1832	13	<.0001
Wald	560.0524	13	<.0001

And also the measure of the model performance using R_p^2 defined in section (3.3.5) is calculated using table 4.6 as follows:

$$R_p^2 = 1 - \exp\{2/n[l_0 - l_1]\} = 0.14$$

R_p^2 is terribly very low due to high percentage of censored data . However, based on the tests as indicated in table 4.5, the estimated model given in table 4.3 is good fit for the child mortality data. Now we can discuss the interpretation of the model and the results.

Table 4.6 Models Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	11514.513	10961.365

4.5 Interpretation of results and discussion of child mortality

As indicate in section (3.3.3) the interpretation of the estimated coefficient of Cox proportional hazards regression is similar to logistic regression. Hence the relative risk association with the covariates (x_i) is exp(βx_i). In our model all the covariates are categorical variables and the hazard ratio of i^{th} group of the covariate compared to the reference group is exp(β_i), keeping all other covariates fixed. If exp(β_i) < 1, the hazard rate for death of child that belongs to the i^{th} group is smaller than the reference group. On the other hand, if exp(β_i) >1,the hazard rate (death of child) that belongs to the i^{th} group is greater than the reference group.

As indicated in Table 4.3, the covariates included in the model : source of drinking water, birth order number, sex of child, breast feeding status, wealth index of household , mother's education and family size have statistically significant contribution for child mortality (5% level of significant). Father's education and age of mother at birth are also significant at the 10% level.

Consistent with expectations and other studies (Ali, 2002; Jocoby and Wang, 2004), the Cox regression results predicting child mortality as indicated in Table 4.3, show that maternal education has a significant contribution for child mortality in rural Ethiopia. This result indicates that children whose mothers are literate have more survival chance than those with illiterate mother. This may be due to the fact that educated mothers are more likely to use the health services, feed their children better and act in various ways to improve traditional means of health care. Children born from educated mothers have a 23.3% less mortality risk as compared to those from uneducated mothers. The results also indicate that the mortality risk of children from educated fathers is 15.8% less than those from uneducated fathers.

Children whose mother's age at birth lies in groupage2 (21-34 years) and groupage3 (35 and above years) have more survival chance than mother's whose age at birth is less than 20 years (reference group), i.e, children from mothers whose age at birth lies in groupage1 and groupage2 have 23.6% and 32.6% less mortality risk as compared to mothers whose age at birth is less than 20 years, respectively. Therefore, child mortality has indirect relationship with the age of mother at birth.

Children who are breast fed have more survival experience than those who are not breast fed. That is, the likelihood of child mortality of children who are breast fed is 69% less as compared to those who are not breast fed. Child mortality is also influenced by demographic variables like sex. Male children are more likely to die than female children. That is, female children have a 16.5% less mortality risk than male children. Similar results are reported by Ali (2002) and Krzysztol (2007).

As indicated in table 4.3, children who belong to groupedfamily2 (2-4 members) and groupedfamily3 (>6 members) have more survival chance than that of groupfamily1 (1-3 members). One of the strange results we get was that children from families with 4-6 and more than 6 members have 56% and 73% less mortality risk, respectively, than children with 1-3 family members. Many recent studies have focused on the improvement in survival of children due to the presence of large family size. Children from large family size often get more care, since older children can assist in the care of younger children.

Previous studies show that the risk of mortality is positively related to birth order of the child (Ali, 2004; Krzysztol, 2007). As indicated in Table 4.3, children whose birth order number is greater than 4 have less survival chance than those who are first born (reference category). When expressed in percentage, children whose birth order is greater than four have 89.4% more mortality risk as compared to those who are the first children. But there is no statistically significant difference between the first birth order number (reference group) and birth order number (2-4).

From the result in table 4.3, a child whose parents use covered water has more survival chance than those who use surface water. That is, a child whose parents use covered water has a 64.6% less mortality risk as compared to a child whose parents use surface water. Therefore, the improvement of water supply is expected to be inversely related to mortality risk. As mentioned earlier in the literature review, water supply and higher quality sanitation facilities are epidemiologically directly related to lower mortality.

According to the results in Table 4.3, wealth index has a contradictory result to our expectation. That is, as the wealth index increases, the risk of child mortality will also increase. For instance, children whose families are in medium wealth range are 1.422 times more likely to die as compared to children from poor family. In rural part of Ethiopia agriculture is the dominant sector in the economy. Households who are relatively rich do have high labor burden (to increase their production), and hence, devoted much time of their time in productive activities and less time in child health care.

In general, the findings revealed that socio economic, environmental, demographic and health related variables are important determinants of child mortality.

-

CHAPTER FIVE

Conclusion and Recommendation

5.1 Conclusion

Child survival is an important aspect of public health and its epidemiological appraisal may provide important clues towards public health programs. Child survival may get affected through biological and behavioral channels which are socioeconomic, environmental, demographic and health related variables. The results of this study indicated factors that are associated with child mortality in rural Ethiopia through Cox proportional hazards model by analyzing the effect of the covariates and censored observations. The findings of the study demonstrate that different factors such as mother's education, father's education, birth order number of children, age of mother at birth, sex of child, family size, source of drinking water and breast feeding status have statistically significant impacts on the survival chance of a child.

The study shows that mother's education and father's education have a significant effect on the survival of child, that is, children born to literate mothers and father are less likely to face the risk of death. Also age of mother at birth and family size have a positive impact on the child survival, that is, as age of the mother at birth and family size increase, the survival chance of a child will also increase.

Sex of the child was found to be one of the determinants of child mortality. According to our results, female children have a relatively less risk of death than male children. Breastfeeding status and birth order number were also found to be have a contribution to child mortality, that is, children at a higher birth order number have higher risk of child mortality and children who are breast fed have less risk of death before the age of five.

The study also shows, that source of drinking water has a significant effect on child mortality. A child that comes from a household with access to covered/protected well water has lower risk of death than a child that belongs to a household which consume water from unprotected sources. Contrary to our expectation, the results showed inconsistent relationship between socioeconomic status (measured by wealth index) and child mortality, that is, children from poorest households have less child mortality than richest and middle household classes. This is difficult to explain and requires more detailed examination about the data collection and the wealth index as a valid tool for measuring socioeconomic status.

5.2 Recommendation

Child mortality reduction is an integral part of social protection, an efficient means of reducing poverty and increasing future growth. Poverty has a disabling impact on the minds, bodies and future potential of children and leads to poor human development and weak economic performance. Investing in children's health, development, and social well-being is a cost-efficient and social endeavor providing lifetime gains to the child and overall benefits to the society.

One of the Millennium Development Goals is the reduction of child mortality by two-thirds in 2015. In order to achieve this goal, identifying the important socioeconomic, environmental, demographic and health related factors that affect child mortality, and acting on them is mandatory. Based on our findings, we make the following recommendations:

- Since educated parents have enough Knowledge on how to care their children, parents should be educated to reduce child mortality.
- Since breast feeding has a great role for reduction of child mortality, mothers should be encouraged to breast feed their children.
- According to our finding, early marriage has a positive impact on child mortality rate. In order to reduce child mortality, marriage at the right age is recommended.
- Integrate child care and family planning programs. This can help by delaying the birth interval to ensure the survival of a child.
- Government should improve society services like providing clean water which contributes to better children health status.

REFERENCE:

- Aguirre, G.P. (1996), “Child Mortality and Reproductive Patterns in Urban Bolivia.” Center for Demography and Ecology (CDE) Working Paper No. 95-28.

- ALI, Hala Abou. (2002), “The Effects of Water and Sanitation on Child’s Mortality in Egypt”, Environmental Economics Unit, Department of Economics, Goteborg University.

- Ali, SM. (2001), “Poverty and Child Mortality in Pakistan” Micro Impact of Macroeconomic Adjustment Policies (MIMAP), Technical Paper Series No. 6.

- Becker, G. S. (1981), “Altruism in the Family and Selfishness in the Market Place,” Economica 48,1-15.

- Becker, G. S, Glaeser, E.L. And Murphy K.M.(1999), “Population and Economic Growth” American Economic Review, Papers and Proceedings; 89(2):145-149.

- Breslow.N (1974),”Covariance Analysis of Censored survival data”, Biometrics 30, 89-99.

- CIA World Fact Book .(2009), Ethiopia people 2009 from the internet (http://www.theodora.com/wfbcurrent/ethiopia/ethiopia_people.html).

- Cingo, A. (1998), “Fertility Decisions When Infant Survival is Endogenous” Journal of Population Economics; 11:21-28.

- Claeson, Griffin, Johnston, Mclachlan, Soucat, Wagstaff and Yazbeck (2005), Health, Nutrition and Population Chapter 18: P.203.

- Cleves, M.A., W. W.Gould, and R.G. Gutierrez (2004), "An Introduction to Survival Analysis Using Stata" Revised Edition; Stata Corporation, Texas.

- Clive J.Mutanga,(2007) ,"Evironmental Determinats of Child Moratlity in Kenya" United Nations University.

- Cox, D.R (1972), "Regression Models and Life Tables" Journal of the Royal Statistical Society; B (34): 187-220.

- Cox and Oakes (1984), "biometrical centenary:survival analysis" Biometrika 23 99-142.

- CSA (2005), DHS Ethiopia 2005, Ethiopia Demographic and Health Survey 2005. Addis Ababa, Ethiopia and Calverton, Maryland USA: Central Statistical Authority and ORC Macro.

- CSA (2007),Population and housing census result: Central Statistics Agency of Ethiopia.

- DHS/WB (2005), DHS Dimensions: A Semi-Annual Newsletter of the Demographic and Health Related Surveys Project; 4(2).

- Efron (1977)),"Covariance Analysis of Censored survival data", Biometrics 35, 42-43.

- Fayissa, B. (2001), “The Determinants of Infant and Child Mortality in Developing Countries: The Case of Sub-Saharan Africa” The Review of Black Political Economy.

- Grambsch, P. M. and T. M. Therneau (1994), “Proportional Hazards Tests and Diagnostics based on Weighted Residuals” Biometrika 81: 515-526.

- Hailemariam.A and Tesfaye.M(1997) “Determinants of Infant and Early Childhood Mortality in Small Urban Community of Ethiopia using Hazard Model Analysis” Ethiopia Journal of Health Department :11(3):189-200.

- Heckman, J. J. (2000), “Policies to Foster Human Capital,” Research in Economics 54, 3-56.

- Hondroyiannis, G. and E. Papaetrou (2002), “Demographic Transition in Europe” Economics Bulletin; 10(3): 1-8.

- Hosmer and lemeshow(1998),”Regression Modeling of time to event data” .

- Jacoby H., Limin Wang (2004), “Environmental Determinants of Child Mortality in Rural China: A competing Risks Approach”. World Bank Policy Research Working Paper 3241.

- Kalemli-Ozcan, Sbnem (2002), “Does the Mortality Decline Promote Growth?”. Journal of Economic Growth, July (2002).

- Kaplan, E. L. and P. Meier (1958), "Non-Parametric Estimation from Incomplete Observations." Journal of the American Statistical Association 53:457-481.

- Kim TH (2004), "determinants of Child and Infant Mortality in Korea 1955-1973".

- Krzystof.T (2007), "The Correlates of Infant and Childhood Mortality" A theoretical overview and new evidence from the analysis of longitudinal data from Bejsce parish register reconstitution study 18th-20th centuries, Poland.

- Lancaster, T. (1990), "The Econometric Analysis of Transition Data" Econometric Society Monographs; Cambridge University Press: New York.

- Lee L.F., M.R. Rosenzweig and M.M. Pitt (1997), "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High Mortality Populations". Journal of Applied Business Research; 13(1).

- Lyun B.Folasade(2000), " Environmental Factors,Situation of Women and Child Mortality in South West Nigeria".

- Madise, N.J., and I. Diamond (1995), "Determinants of infant mortality in Malawi: An analysis to control for death clustering within families". *Journal of Biosocial Sciences* 27: 95-106.

- PARK, K. (2005), "Preventive Medicine in Obstetrics, Pediatrics and Geriatrics", *PARK'S textbook of preventive and social medicine*, (18th edition) India: BHANOT (2005); Pp 414-422.

- Schellenberg, J.A., R. Nethan, S.Abdulla, O. Mukasa, T.J. Marchant, M. Tanner and C. Langeler (2002), "Risk Factors for Child Mortality in Rural Tanzania" Tropical Medicine and International Health; 7(6): 506-511.

- Schoenfeld, D.(1982), "Partial Residuals for the Proportional Hazards Regression Model" Biometrika; 69:239-241.

- UNICEF. (2006): *State of World's Children 2006*.

- UNICEF (2008), Population Statistics at A Glance: Ethiopia Statistics Browsed from the internet(http://www.unicef.org/specialsession/about/sgreportpdf/01_InfantAnd Under- FiveMortality_D734Insert_English.pdf).

- United nation (2007), "United Nations human development index".

- Wang, L. (2002), "Health Outcomes in Low income Countries and Policy Implications: Empirical Findings from Demographic and Health Surveys."

- White.L (2006), "determinates of child and infant mortality in Andre Pradesh", India

- WHO (2001), Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics and Health.

- Wolday Amha (2001), Addressing Poverty in Ethiopia: Concepts, Measurements, and Challenges. Training Manual, Addis Ababa.

- Woldemichael G. (1998), "The effects of water supply and sanitation on childhood Mortality in urban Eritrea", Stockholm Research Reports in Demography No. 127, Demography unit, Stockholm University.

- World fact Book (2008) , “Ethiopian population” Ethiopia.

- Zakir, M. and P. V., Wunnava (1997), “Factors Affecting Infant Mortality Rates: Evidence from Cross-Sectional Data”, Applied Economics Letters; 6:271-273.

Annex 1 Descriptive Statistics of socioeconomics variables

Variables		Frequency	Percent
Mother work status	Not working	6567	88.8
	working	828	11.2
Wealth index	poor	3,858	52.2
	medium	1,541	20.8
	rich	1,996	27.0

Annex 2 Descriptive statistics of environmental variables

		Frequency	Percent
Source of drinking water	surface water	3,570	48.3
	covered(protected well)	3,825	51.7
sanitation facility	toilet facility	211	2.9
	No toilet facility	7,184	97.1
use bed net for sleeping	not use bed net	6,684	90.4
	Use bed net	711	9.6

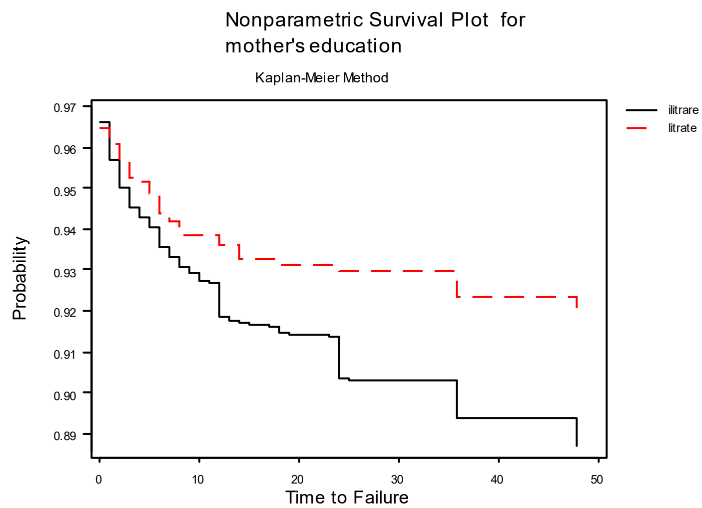
Annex 3: Kaplan-Meier Estimates of survival function

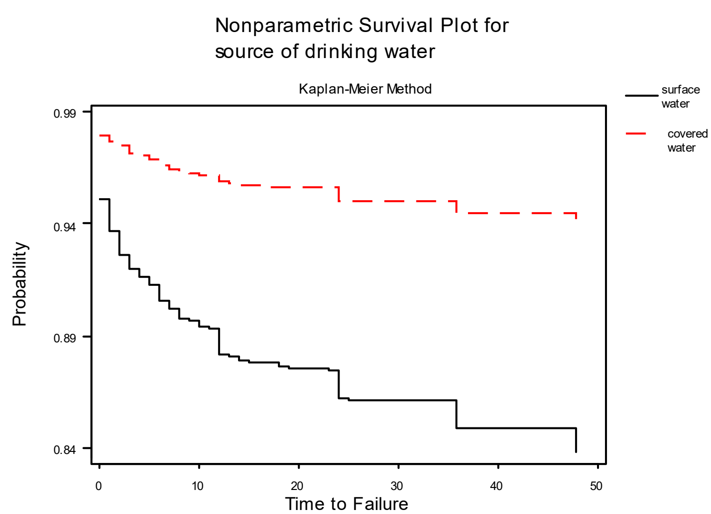
Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
0.000000	7394	313	0.9577	0.0023	0.9531	0.9623
1.000	7005	72	0.9478	0.0026	0.9428	0.9529
2.0000	6809	66	0.9386	0.0028	0.9331	0.9441
3.0000	6604	44	0.9324	0.0029	0.9266	0.9381
4.0000	6413	19	0.9296	0.0030	0.9238	0.9355
5.0000	6282	27	0.9256	0.0031	0.9196	0.9317
6.0000	6130	39	0.9197	0.0032	0.9135	0.9260
7.0000	5961	21	0.9165	0.0033	0.9101	0.9229
8.0000	5829	21	0.9132	0.0033	0.9067	0.9197
9.0000	5687	7	0.9121	0.0034	0.9055	0.9186
10.0000	5571	9	0.9106	0.0034	0.9040	0.9172
11.0000	5494	3	0.9101	0.0034	0.9034	0.9168
12.0000	5403	56	0.9007	0.0036	0.8936	0.9077
13.0000	5204	6	0.8996	0.0036	0.8926	0.9067
14.0000	5061	9	0.8980	0.0036	0.8909	0.9052
15.0000	4943	2	0.8977	0.0036	0.8905	0.9048
17.0000	4739	2	0.8973	0.0037	0.8901	0.9045
18.0000	4635	8	0.8957	0.0037	0.8885	0.9030
19.0000	4528	1	0.8955	0.0037	0.8883	0.9028
22.0000	4259	1	0.8953	0.0037	0.8881	0.9026
23.0000	4190	2	0.8949	0.0037	0.8876	0.9022
24.0000	4116	57	0.8825	0.0040	0.8747	0.8904
25.0000	3937	1	0.8823	0.0040	0.8744	0.8901
36.0000	2822	36	0.8710	0.0044	0.8625	0.8796
48.0000	1418	16	0.8612	0.0050	0.8515	0.8709

Mean of survival time of child

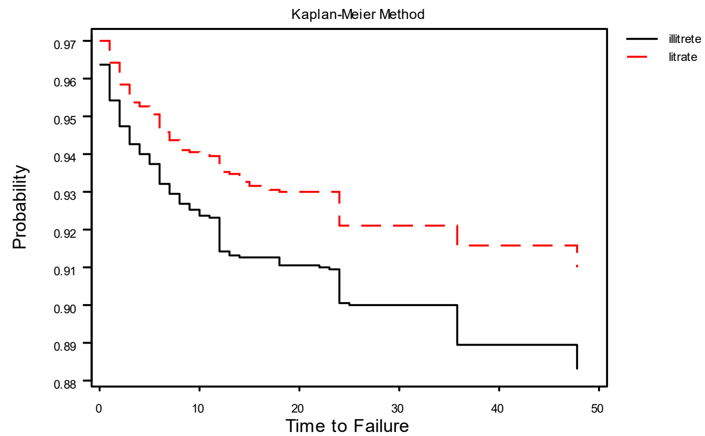
Mean time	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
42.9175	0.1678	42.5886	43.2464

Annex4: Graph of survival functions

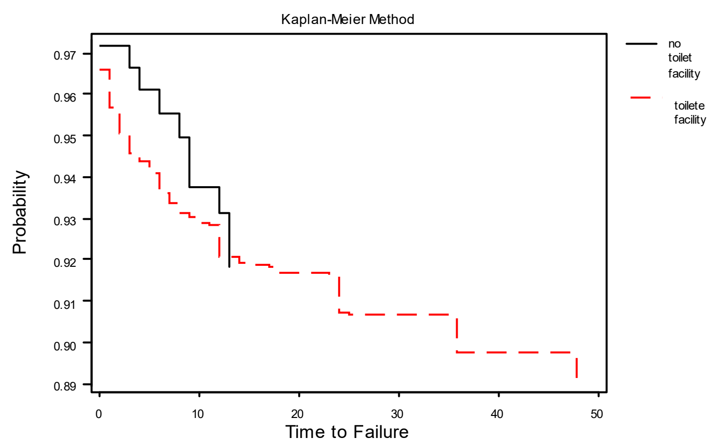




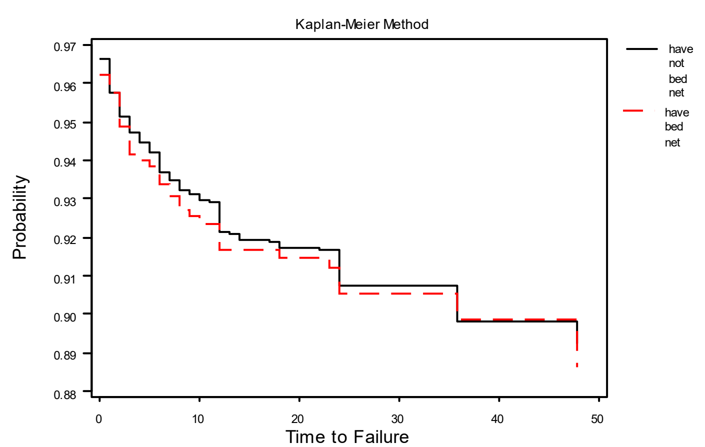
Nonparametric Survival Plot for
father education



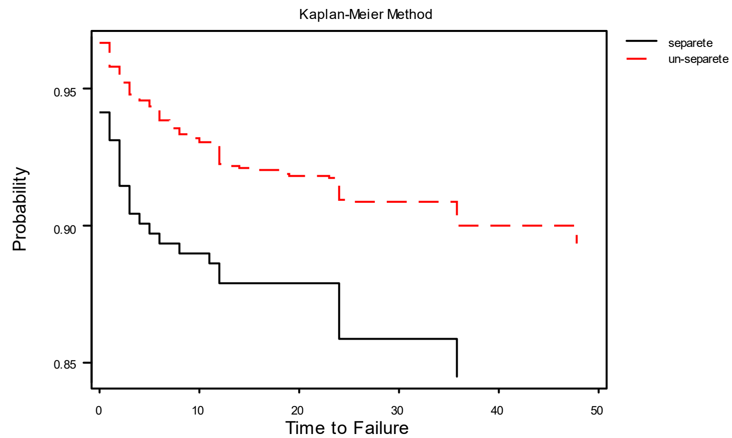
Nonparametric Survival Plot for
sanitation



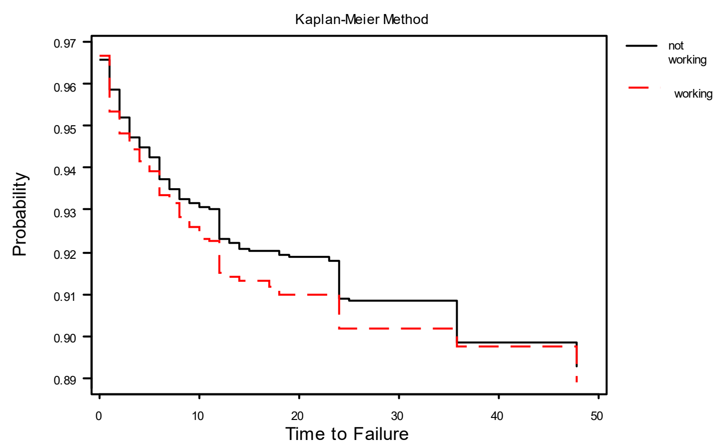
Nonparametric Survival Plot for
bednet



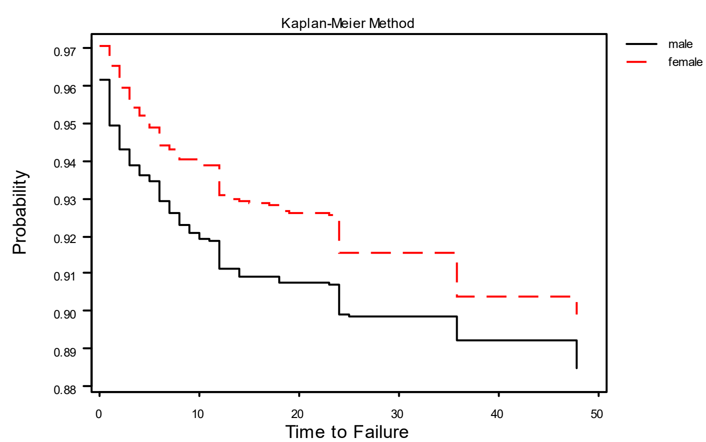
Nonparametric Survival Plot for marital status



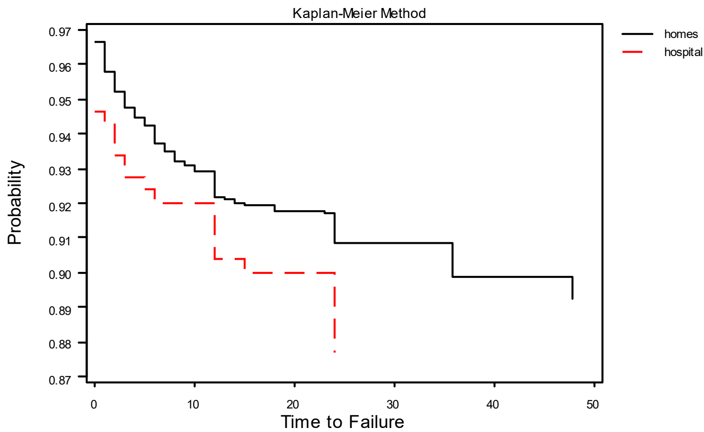
Nonparametric Survival Plot for mother work



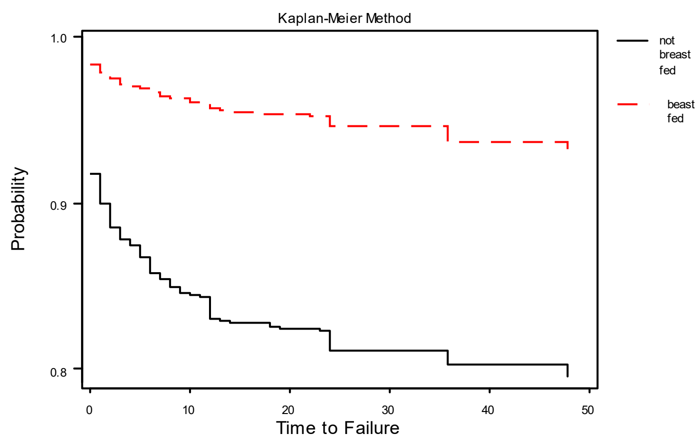
Nonparametric Survival Plot for sex of child



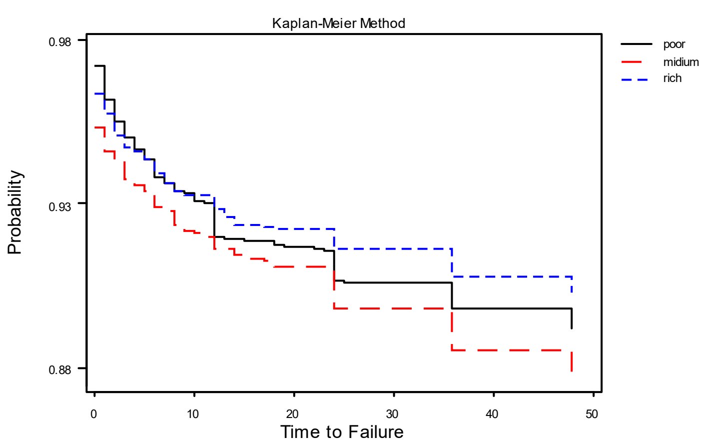
Nonparametric Survival Plot for
place of delivery



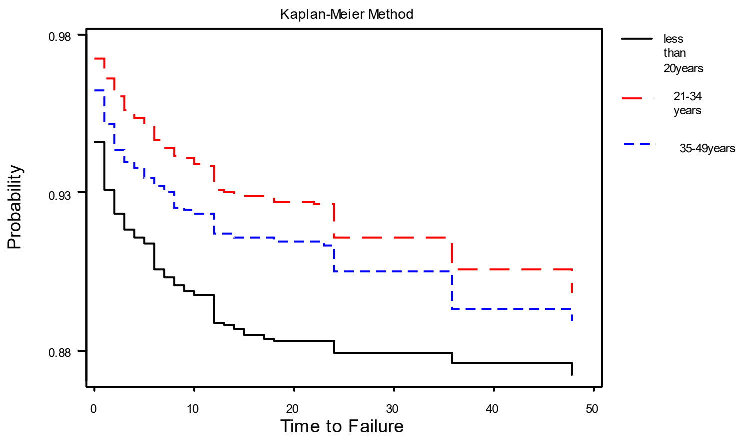
Nonparametric Survival Plot for
breast fed



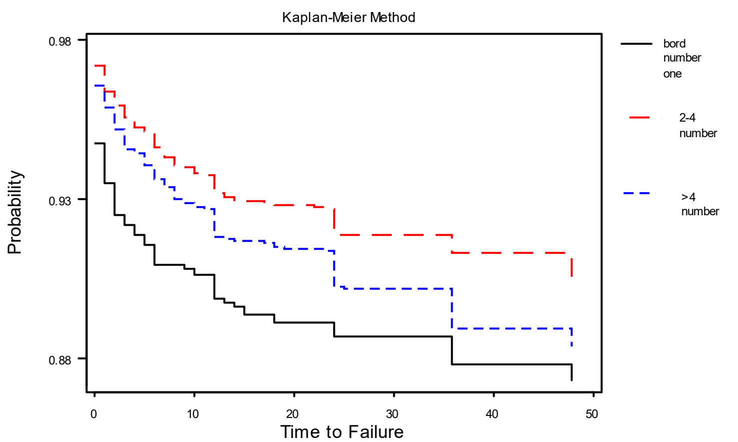
Nonparametric Survival Plot for
wealth



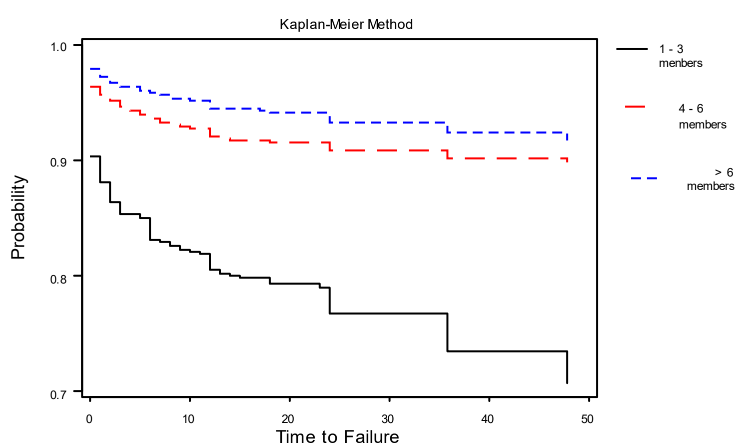
Nonparametric Survival Plot for age
of mother at birth



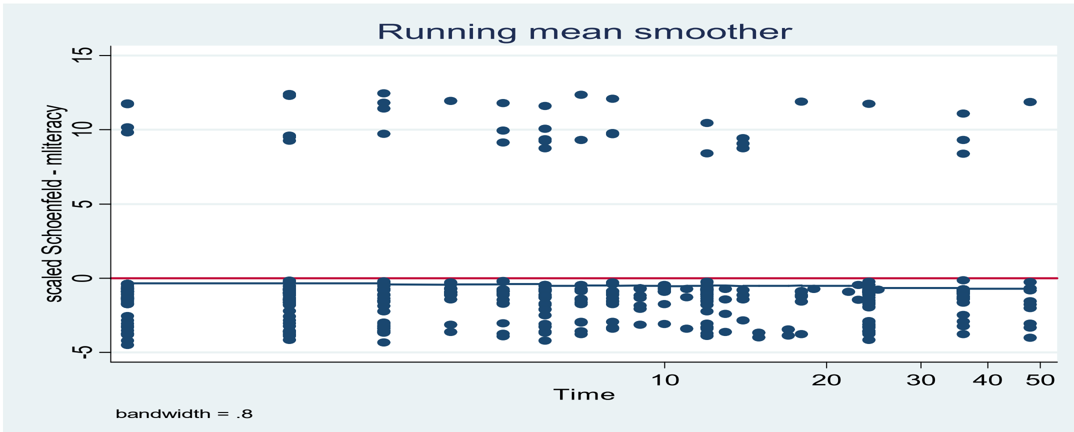
Nonparametric Survival Plot for
birth order number

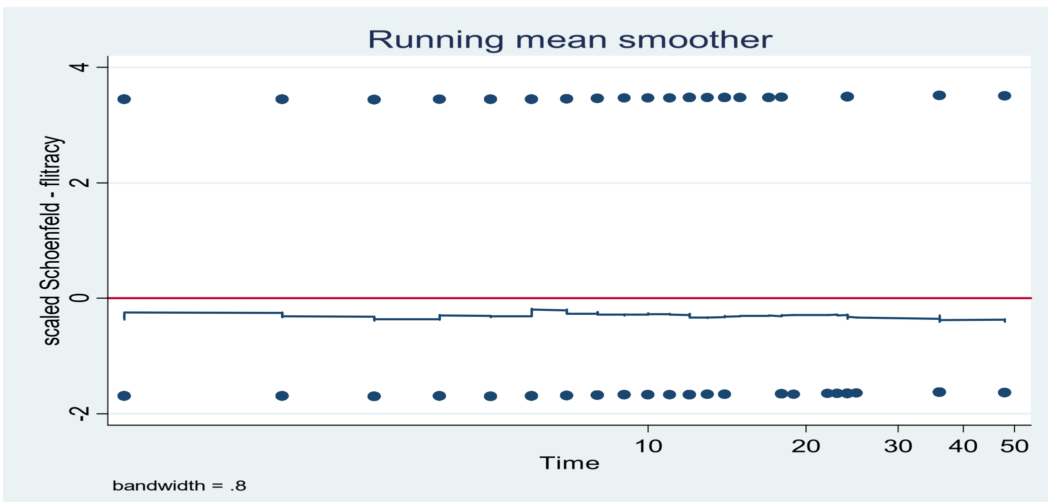


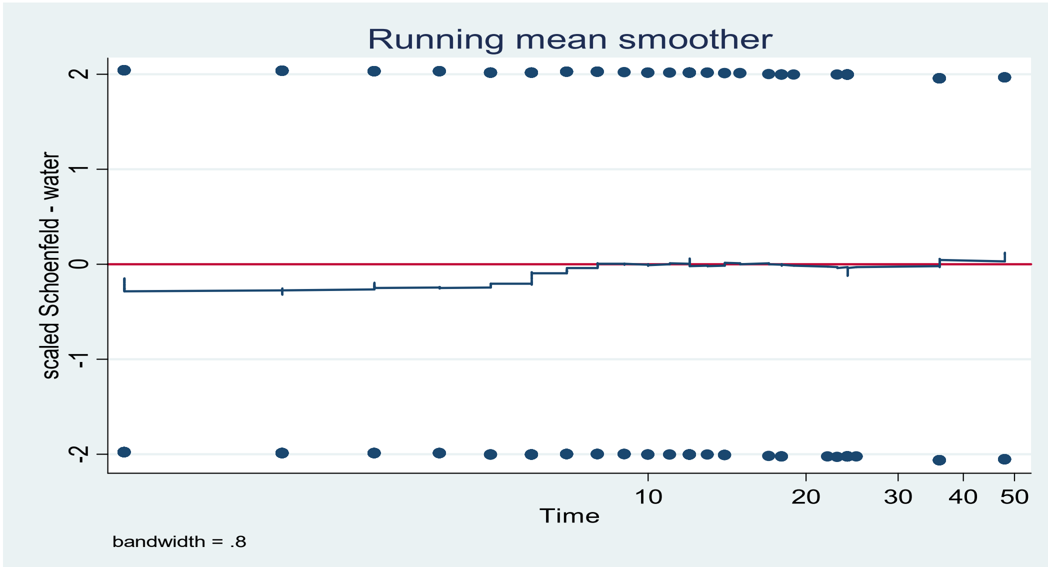
Nonparametric Survival Plot for
family size

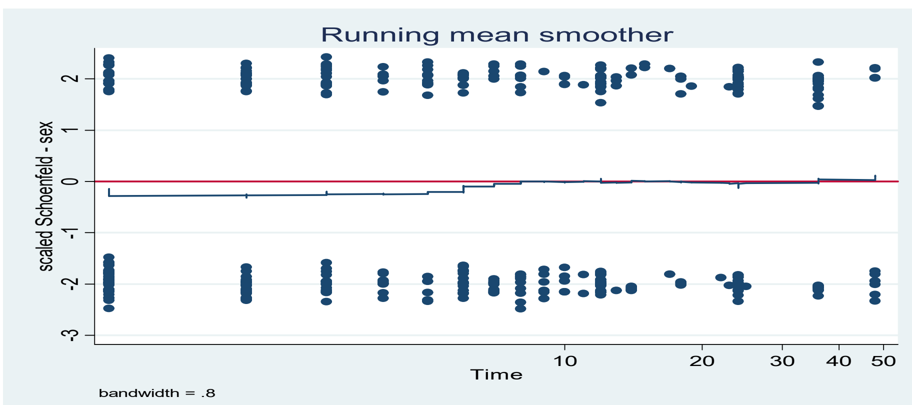


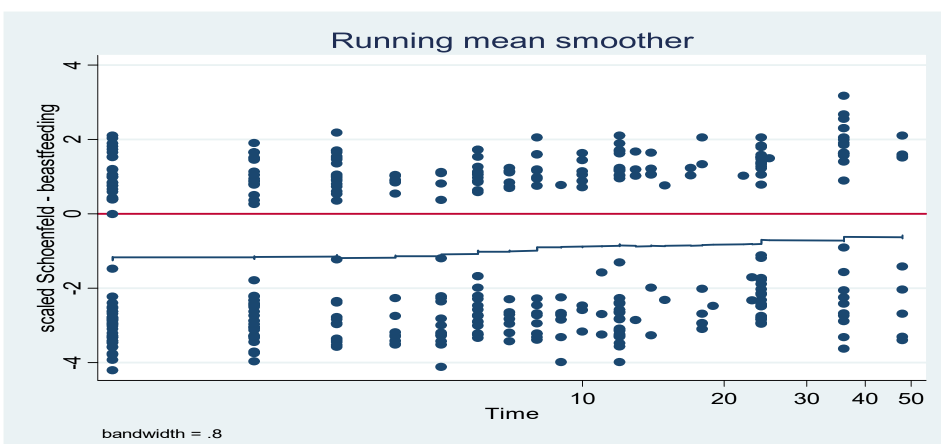
Annex 5: Graphical Test of Proportional Hazards

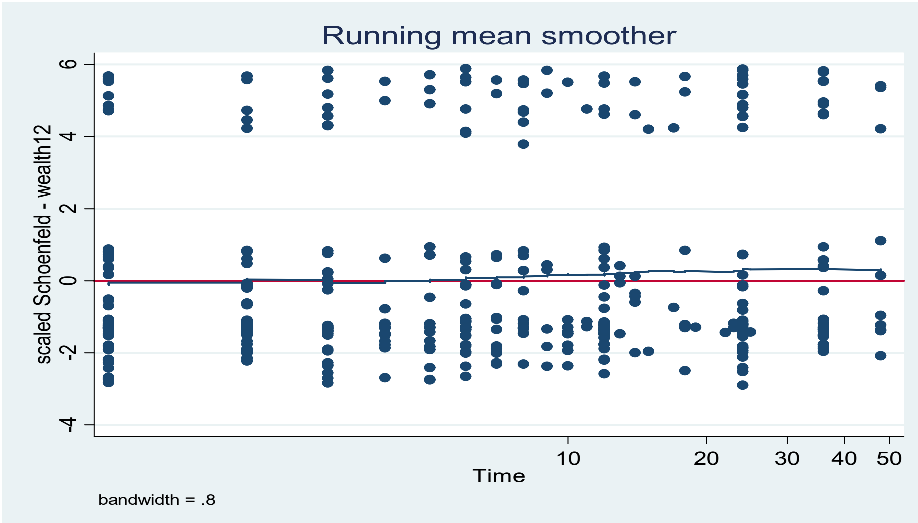


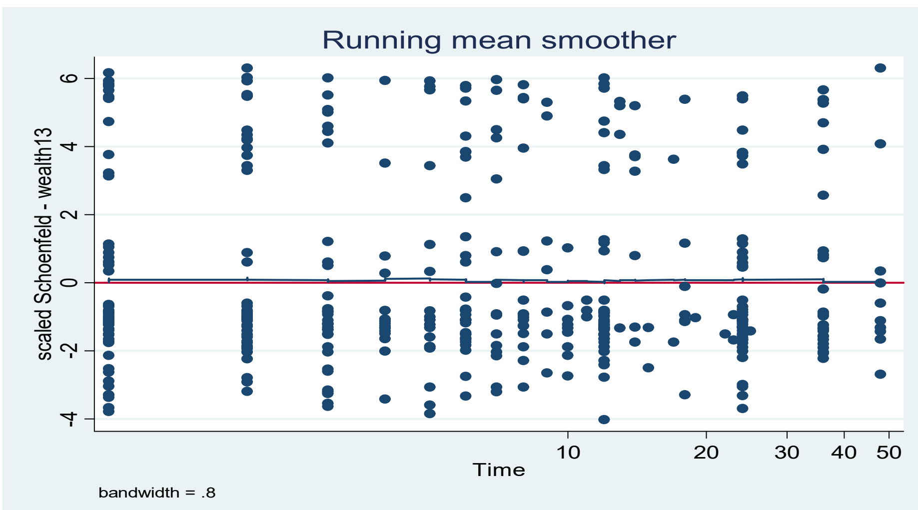



Running mean smoother
scaled Schoenfeld - water
2
1
0
-1
-2
10
20
30
40
50
Time
bandwidth = .8

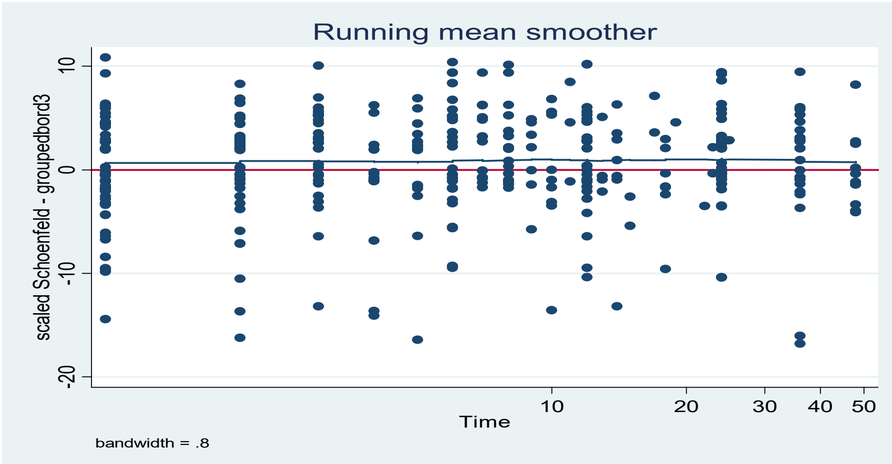

Running mean smoother
scaled Schoenfeld - sex
2
1
0
-1
-2
-3
10
20
30
40
50
Time
bandwidth = .8

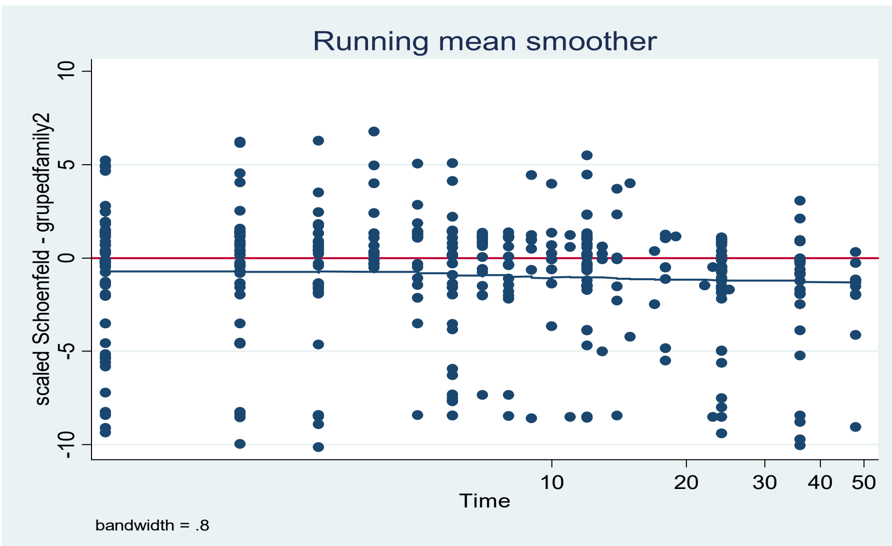

Running mean smoother
scaled Schoenfeld - beastfeeding
4
2
0
-2
-4
10
20
30
40
50
Time
bandwidth = .8


Running mean smoother
scaled Schoenfeld - wealth12
-4
-2
0
2
4
6
10
20
30
40
50
Time
bandwidth = .8

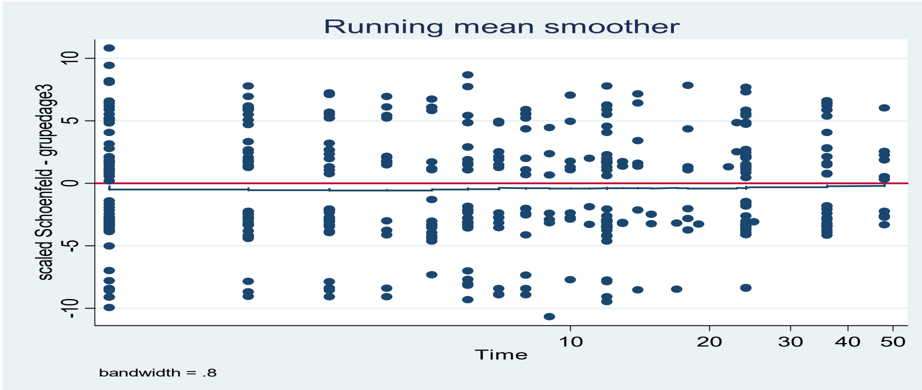

Running mean smoother
scaled Schoenfeld - wealth13
-4
-2
0
2
4
6
10
20
30
40
50
Time
bandwidth = .8


Running mean smoother
scaled Schoenfeld - grupedage2
-10
-5
0
5
10
10
20
30
40
50
Time
bandwidth = .8

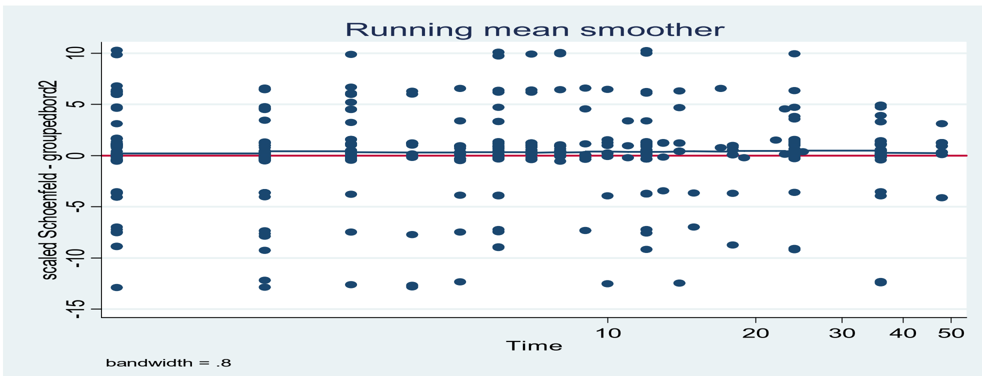

Running mean smoother
scaled Schoenfeld - groupedbord3
Time
bandwidth = .8


Running mean smoother
scaled Schoenfeld - grupedfamily2
Time
bandwidth = .8


Running mean smoother
scaled Schoenfeld - grupedfamily3
Time
bandwidth = .8



Running mean smoother
scaled Schoenfeld - grupedage3
10
5
0
-5
-10
10
20
30
40
50
Time
bandwidth = .8



Running mean smoother
scaled Schoenfeld - groupedbord2
10
5
0
-5
-10
-15
10
20
30
40
50
Time
bandwidth = .8