Research Article: Survival Analysis of Under Five Mortality in Rural Parts of Ethiopia

Yared Seyoum** and M.K. Sharma*

Addis Ababa University, Addis Ababa, Ethiopia

Abstract: Child mortality is a factor that is associated with the well-being of a population and it is taken as an indicator of health development and socioeconomic status. According to the 2011 UN report during the last 10 years, the death rate for children under five has decreased by 35% worldwide. UNICEF in 2008 reported that Ethiopia has reduced under-five mortality by 40 percent over the past 15 years. From the EDHS 2011 report child mortality rate in Ethiopia was reduced from 50/1000 deaths in 2005 to 31/1000 deaths in 2011. The Ethiopian Demographic and Health Survey data are used for the study. In this paper we have attempted to find out the impact of socioeconomic, demographic and environmental factors in the context of under five mortality. In this attempt we first analyzed our data using Kaplan-Meier non-parametric method of estimation of survival function and also using lifetable. We have also used Log-Rank test to compare different survival functions and found that sex, type of birth, religion, mothers' education, birth order, maternity age, source of drinking water and region have statistically significant difference in the under five survival time. We have also used Cox proportional hazard model to identify the covariates which influence the under five mortality. But we found that our data do not fulfill the proportionality assumption of Cox proportional model in case of infant and child mortality. Then we applied stratified Cox proportional model to our data to find out the potential covariates which influence under five mortality and found birth order, mothers' education level, sex, type of birth and the interaction of birth order and sex as vital factors for the deaths occurring under the age of five. The Cox proportional hazard models which were used separately for each stratum also identified mothers' educational level, sex, type of birth, and the interaction of sex and water supply as the risk factors for the death of infants. Whereas for child stratum; type of birth, mothers' education, sex and the interaction of water supply and sex were the risk factors associated with the death of children.

Keywords: Under five mortality, maternal, socioeconomic and environmental factor.

1. INTRODUCTION

1.1. Background

Under five mortality (U5MR) is a factor that can be associated with the well-being of a population and taken as one of the development indicators of health and socioeconomic status and also indicates a life quality of a given population, as measured by life expectancy. That is why reduction of under five mortality (U5MR) is a worldwide target and one of the most important key indexes among Millennium Development Goals (MDGs). Hence its indication is very important for evaluation and public health strategy. Thus, it is an area that many researchers focus and it has also attracted the attention of policy-makers and program implementers worldwide. One of the most important targets of Millennium Development Goals (MDGs) that introduced in 2000 at the United Nations Millennium Summit was reducing under-five child mortality (U5MR) rates by two-thirds from the 1990 levels by 2015. In 2000 the Ethiopian government announced the intention by signing the millennium declaration committing to achieve the Millennium Development Goals (MDGs) by 2015, many of which overlap with the 2015 national policy goals, introduced by the federal government in 1991 and a policy action has been continued, for instance, in 2004 the Ethiopian government prepared child survival strategy and implementation plan to reduce under five mortality (U5MR) of 140/1000 live births to 67/1000 live births by 2015, this means a reduction of two-thirds of from the 1990 rates about 200/1000 live births or a 52 percent reduction from 2004 rate about 140/1000 live births [1].

Ethiopia has the second largest population in Africa and is one of the least developing country with high fertility and rapid population growth rates. The country's population is estimated nearly 79 million and a growth rate of 2.73 percent per year [2]. The Ethiopian population is a predominantly rural and young society and the majority of the population has traditionally been concentrated in the highlands, with nearly 85 percent of the population living in rural areas while the rest lives in urban areas. The population growth rates about 2.73 percent which is slightly greater than the sub-Saharan Africa countries an average growth of 2.5 percent. Urban populations growing nearly 2.3 percent per year while the rural population growing at about 4 percent. The age structures suggest nearly 45 percent of the populations are under age 15 and the percentages of the population above age 65 are only about 3.2 percent [3] (Ringheim *et al.*, 2009). Since the late 1950s the country economic performance indicates that the average GDP growth has reduced from 5 percent in the 1950s to less than one percent in the 1980s. Emerging

*Address correspondence to these authors at the Addis Ababa University, Addis Ababa, Ethiopia; Tel: +251920887186; E-mail: mk_subash@yahoo.co.in
**Tel: +251911383550; Fax: +251 116670662; E-mail: yaredsw@gmail.com

E-ISSN: 1929-6029/14

from long civil war and several famine, Ethiopia replaced the Derg regime, which negatively affected the country's economy, with a federal structure of government, introducing a new national policy that address population growth and population dynamics and was implemented in 1993, which allowed Ethiopian to take off from this economic stagnation and to start a sustainable and enduring political, economic and social development. Economic growth during this period was impressive that the per capita GDP grew at average rates of 5.6 per annum (from 1991 to 2000) and the recent growth register showed 10.7 percent per annum. Nonetheless, the economic performance has been fragile and uneven due to internal and external shocks including Ethio-Eritrea war during the period 1998 to 2000 [4].

In Ethiopia, in coincidence with the population growth, starting in the late 1950s onwards, crude death rates substantially declined, it reduced from 30 to 15 per 1000 people from 1960-2005 due to the introduction of basic health services, particularly in the rural areas and at the same time life expectancy (about 42 years) slightly lower as compared to sub-Saharan countries (about 45 years) [5].

This study investigated U5MR and identifies the more important factors associated with the U5MR, since in recent decades demographic researchers indicated that there is a strong relationship between infant and child mortality and socioeconomic and bio-demographic factors.

In this thesis we have used stratified Cox proportional hazards model to analyze the determinants of U5MR as well as we compared the result of the estimate of the model and distinguish which factors are more pronouncedly contribute to the decline level of U5MR. We estimated the model using data from Ethiopia Demographic and Health surveys, 2011.

1.2. Statement of the Problem

The population of Ethiopia is a predominantly rural, with nearly 85% percent of population living in rural areas. This gives a reason to make the rural population our focus of study. Since U5MR is one of the basic indicators of a country's socio-economic status and quality of life. According to [6] report the current level of child mortality in rural parts of Ethiopia is around 114 deaths per 1000 live births. Therefore working towards the factors affecting U5MR in rural Ethiopia is a great addition in achieving the targets of MDG and in improving the quality of life in the country.

Despite the fact most researches are focused on the study of U5MR for the whole country, rural areas of the country needs special attention. While the targets of the MDGs is reducing the mortality rate of children aged from birth to their 5^{th} year, yet again most researches are focused on addressing the issue on a separate age group (infant: aged from birth to 11 months old and child: aged from 12 months to 59 months). This paper will fill these two gaps and address the mortality case in rural parts of the country for those children aged from birth to 59^{th} months old.

As mentioned earlier U5MR is very high particularly in rural parts of the country. If Ethiopia is committed to achieving the MDG targets and further reduction of U5MR, it is necessary to understand clearly the factors that are contributing to the high levels of mortality in the rural parts of the country. This study therefore explores the demographic, socio-economic, environmental factors and their effect on U5MR in rural parts of Ethiopia.

1.3. Objective of the Study

The general objective of this study is to identify determinants U5MR in rural parts of Ethiopia.

The specific objectives are to:

- Assess the relationship between the risk factors and U5MR
- To compare under five survival time among different groups

1.4. Significance of the Study

- The result of this study could provide information to government and other concerned bodies in setting policies, strategies, and further investigation for reducing U5MR.
- The results could be of interest to other studies related to U5MR risks in Ethiopia.
- It is expected to improve the understanding of the mortality situation of under the age of five in rural parts of Ethiopia.

2. DATA AND METHODOLOGY

2.1. The Data

The source of the data used in this study is the 2011 Ethiopia Demographic and Health Survey [6] conducted in Ethiopia as part of the worldwide demographic and health survey project. The 2011 Ethiopia Demographic and Health Survey were conducted by the Central Statistical Agency (CSA) with the support of the Ministry of Health. This is the third Demographic and Health Survey (DHS) conducted in Ethiopia, under the worldwide MEASURE DHS project, a USAID-funded project providing support and technical assistance in the implementation of population and health surveys in countries worldwide.

The survey interviewed a nationally representative population in about 18,500 households. Indicators relating to family planning, fertility levels and determinants, fertility preferences, infant, child, adult and maternal mortality, maternal and child health, nutrition, women's empowerment, and knowledge of HIV/AIDS are provided for the nine regional states and two city administrations. In addition, data by urban and rural residence at the country level are provided.

Information on child mortality was found from the birth history of women who were included in the survey. The interest of this study is about children from birth until age five. This age group is selected because the risk is expected to be higher in this age group and the risk for children above age five is less since they are older.

2.2. Variables in the Study

The explanatory (independent) variables are classified into three groups: maternal, socioeconomic and environmental. The response (dependent) variables for this study is the survival time of a child measured in months from birth until death/censored.

2.3. The Response Variable

The dependent variable is the age of a child until the occurrence of an event. Observations are censored, in case, for some units, the event of interest has not occurred at the time the data are analyzed.

2.4. Predictor Variables

Predictor variables or explanatory variables are those whose effect on the age of a child we wish to assess. The predictor variables which are assumed to influence the survival the under five mortality included in the model are (i) child's sex, (ii) mother's age, (iii) child's birth order, (iv) type of birth, (v) mother's education, (vi) father's education, (vii) region, (viii) religion, (ix) water supply.

2.5. Methodology

2.5.1. Survival Analysis

The study focused on time to event, which makes the employment of survival analysis more appropriate. We have used Kaplan-Meier estimator and Cox proportional hazard model for the analysis. For the comparison of survival functions, we used logrank test. Kaplan-Meier analysis was used to study survival pattern; the KM plot, which is a step function, gives some indications about the shape of the survival distribution. The figure in general shows if the pattern of one survivorship function lies above another which means the group defined by the upper curve lived longer, or had a more favorable survival experience than the group defined be the lower curve.

2.5.2. The Cox Proportional Hazard Model

The Cox proportional Hazards model [7] is probably the most widely used method for modeling survival data. For data with one explanatory variable, i.e. one covariate, non-parametric methods like plotting of Kaplan-Meier survival probabilities may be adequate if the groups being compared are reasonably similar. Frequently however, the groups being compared differ in many respects. They may have different age distributions, different proportion of men and women, different smoking habits etc. These differences come in addition to the covariate we are really interested in, and the analysis must be adjusted to compensate for these other differences, which may otherwise confound the analysis. The Cox proportional hazards model is a semi- parametric model for fitting survival data.

Let t denote a continuous non-negative random variable representing survival time.

The basic Cox model is as follows:

$$h(t\,|\,\boldsymbol{X}) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}) \qquad (1)$$

Where:

$h(t|X)$ hazard function at time t with covariates $\boldsymbol{X}=(X_1, X_2, \ldots X_P)$

- h_0 is the arbitrary baseline hazard function that characterizes how the hazard function changes as a function of survival time.
- $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_p)$ is a column vector of p regression parameters associated with explanatory variables.

- $\exp(\boldsymbol{\beta}'\mathbf{X})$ characterizes how the hazard function changes as a function of subject covariates.

2.5.3. Stratified Cox Proportional Hazard Model

Now to accommodate non-proportionality assumption one can apply stratified Cox proportional hazard model in which the stratification in most cases is done by using a covariate fixed by design. Suppose we have s= 1,2,….,S strata, and then allow the baseline unspecified hazard function to vary among the strata.

The hazard function for stratum, s is

$$h_s(t|\boldsymbol{X}) = h_{s0}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}) \qquad (2)$$

2.5.4. Model Selection

The methods available to select a subset of covariates to include in a proportional hazards regression model are essentially the same as those used in any other regression model. There are three methods of selection of influential covariates. These are purposeful selection, stepwise selection (forward selection and backward elimination) and best subset selection. Survival analysis using Cox regression method begins with a thorough univariable analysis of the association between survival time and all important covariates [8].

2.5.5. Model Diagnostic

Scaled Schoenfeld residuals were used to assess the goodness of fit of the model. And if the smoothed curve is an approximately horizontal line around zero, it means the covariates have satisfied the assumption of proportional hazards.

3. RESULTS

3.1. Descriptive Analysis

The study included 7621(0-11 months old: 2704 and 12-59 months old: 4917) children under the age of five years preceding the date of the survey in the rural part of Ethiopia. Summary results for socio-demographic and environmental variables included in this study are presented in Table **1** below. Of the total of 7621 children included, 3718 were females.

When we look at the mothers' age, 4739 of the children are born from mothers whose age is below 20 years, 2768 of the children are born from mothers of age range from 20 to 29 years, and the rest 114 children are born from mothers of age above 30 years.

Table 4.1 has also provided the educational status of both mothers and fathers. According to the table 5253 mothers and 3855 fathers have no any kind of formal education, 1994 mothers and 2909 fathers have at least primary level education and 374 mothers and 857 fathers from the sample have a formal educational level of at least secondary.

The other factor which we will investigate the significance on U5MR is the wealth index. Table **1** shows the distribution of families in rural areas across wealth index. We can find on the table that 3690 children are from poor family, 2638 children are from a family of wealth index medium, whereas the rest 1293 children are from rich family.

3.2. Summary Statistics

It is vital to do some statistical tests that will be used as initiation to our subsequent finding. We start with the test of whether the observed differences in data summary among different factors are statistically significant or not. We use log-rank test for both infant and child strata and Kaplan-Meier survival estimates to look into the significance of the difference among different factors.

The log-rank test is used at 5% level of significance to validate the differences in the survival time of each factor. The null hypothesis to be tested claims, that there is no difference between the probabilities of an event occurring at any time point. The output from computer is summarized in Tables **2** and **3** below.

According to Table **2** the variables which have statistically significant difference in experiencing death event are sex, type of birth, religion, mothers' education, birth order, maternity age, source of drinking water and region.

Table **3** exhibits that sex, type of birth; mothers' education level and region are the variables which have statistically significant difference in experiencing death event.

The Kaplan-Meier estimator survival curve can be used to estimate survivor function among different strata/group of covariates so that one can make comparison. Separate graphs of the estimates of the Kaplan-Meier survivor functions are produced for different categorical covariates. In general, the survivorship pattern of one is lying above another means the group defined by the upper curve has a

Table 1: Distribution of Important Socioeconomic and Environment Characteristics of U5MR in Ethiopia

Covariate/Factor	Category	Censored	Dead (%)	Total
Supply of drinking water	Pipe	460	36 (7.3)	496
	Other	6420	705 (9.9)	7125
Wealth Index	Poor	3296	394(10.7)	3690
	Medium	2413	225(8.5)	2638
	Rich	1171	122(9.4)	1293
Mother Age	<20	4254	485(10.2)	4739
	20-29	2525	243(8.8)	2768
	>=30	101	13(11.4)	114
Mother Education	No education	4716	537(10.2)	5253
	Primary	1811	183(9.2)	1994
	Secondary or higher	353	21(5.6)	374
Father Education	No education	3445	410(10.6)	3855
	Primary	2648	261(9.0)	2909
	Secondary or higher	787	70(8.2)	857
Type of birth	Single	6717	653(8.9)	7370
	Multiple	163	88(35.1)	251
Sex of child	Male	3491	412(10.6)	3903
	Female	3389	329(8.8)	3718
Child Birth Order	1^{st}	1299	151(10.4)	1450
	2-5	3872	393(9.2)	4250
	6+	1709	197(10.3)	1906
Religion	Orthodox	2089	229(9.9)	2318
	Catholic	61	13(17.6)	74
	Protestant	1464	147(9.1)	1611
	Muslim	3093	335(9.8)	3428
	Others	150	40(21.0)	1900
Region	Tigray	839	81(8.8)	920
	Afar	705	84(10.6)	789
	Amahara	746	86(10.3)	832
	Oromia	1274	125(8.9)	1399
	Somali	662	69(9.5)	731
	Benshangul-Gumuz	545	84(13.3)	629
	SNNP	943	112(10.6)	1055
	Gambela	524	63(10.7)	587
	Harari	414	42(9.2)	456
	Dire Dawa	212	11(4.9)	223

Table 2: Log Rank Test for Equality of Survival Time Among the Different Groups of Covariates for Children Aged from Birth to 11 Months Stratum

Group	Chi-square	df	Pr>Chi-square
Sex	8.132	1	0.004
Type of Birth	134.678	1	<0.001
Religion	10.389	4	0.034
Father Education Level	3.127	2	0.209
Mother Education Level	6.264	2	0.044
Child Birth Order	4.885	2	0.087
Maternity Age	7.811	2	0.020
Wealth Index	1.473	2	0.473
Water supply	5.072	1	0.024
Region	19.563	9	0.013

Table 3: Log Rank Test for Equality of Survival Time Among the Different Groups of Covariates for Children Aged 12 Months to 59 Months Stratum

Group	Chi-square	df	Pr>Chi-square
Sex	10.3571	1	0.003
Type of Birth	7.520	1	0.004
Religion	3.992	4	0.407
Father Education Level	4.845	2	0.089
Mother Education Level	6.272	2	0.043
Child Birth Order	3.190	2	0.203
Maternity Age	1.722	2	0.423
Wealth Index	4.236	2	0.120
Water supply	2.352	1	0.125
Region	20.754	9	0.011

better survival than the group defined by the lower curve.

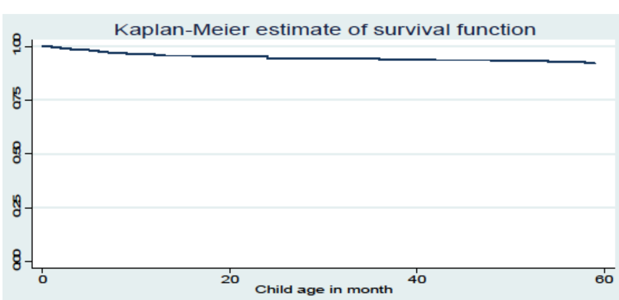

Figure 1: The Plot of Kaplan-Meier estimate of survivor function of children under the age of five in rural part of Ethiopia.

Graphs from Figure **2** there is clear difference among the various group of level Type of Birth, Religion, Parents Educational Level (for both cases: mother and father) Water supply and Region.

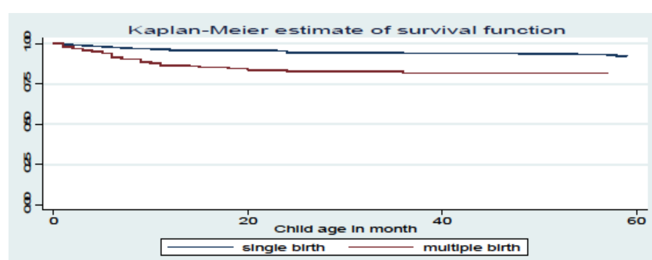

Figure 2: survival curves by Types of Birth.

The survival curves in Figure **2** shows that children who were born singlet have better survival time than children who were born as twins, triplet, and quadruples and so on.

From Figure **3** one can guess that other than those four major religions which is the “others” category has less survival rate.

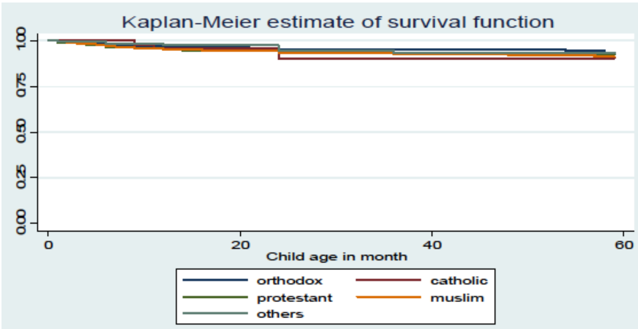

Figure 3: survival curves by Religion.

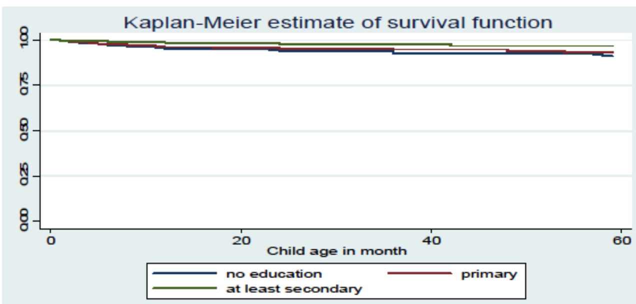

Figure 4: survival curves by Mother Education.

In general, it is expected that children from a better educated mother to have a better survival rate and Figure **4** gives a confirmation for that.

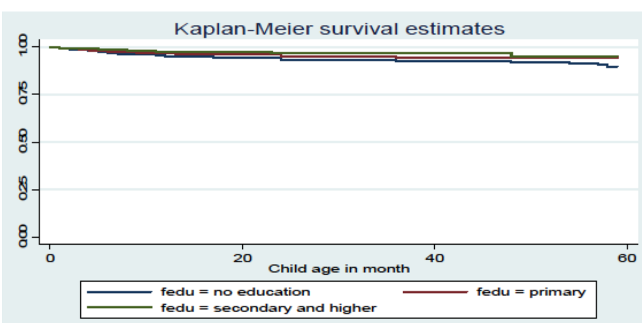

Figure 5: survival curves by Father Education Level.

The above Figure **5** tells the same thing as the case of mothers, children whose fathers’ education level is categorized in the higher education level have better survival rate.

3.3. Result of the Cox Proportional Hazard Model

In any applied setting, a statistical analysis shall begin with a thoughtful and thorough univariate model analysis of the data before proceeding to more complicated models.

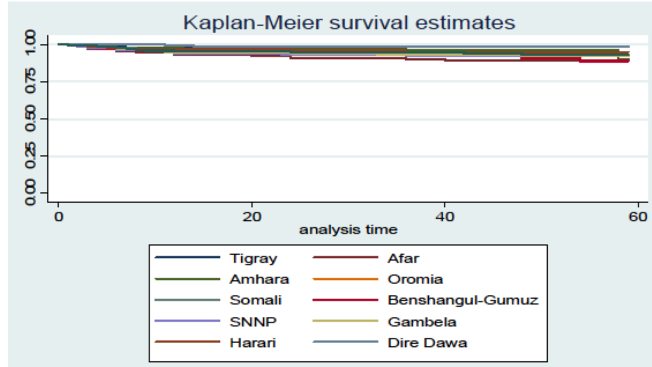

Figure 6: survival curves by Region.

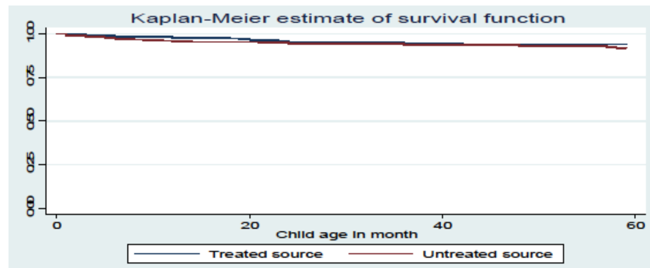

Figure 7: Survival curve by water supply.

[8] recommended the following procedure in variable selection.

I. Include all variables that are significant in the univariable analysis at the 25 percent level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.

II. The variables that appear to be important from step (I) are then fitted together in a multivariable model. In the presence of certain variables others may cease to be important. Consequently, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

III. Variables, that were not important on their own, and so were not under consideration in step (II), may become important in the presence of others. These variables are therefore added to the model from step (II), with forward selection method. This process may result in terms in the model determined at step (II) ceasing to be significant.

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

As we continue to the analysis we choose stratifying variable infant/child indicator so that we can fit Stratified Cox proportional hazard model. Figure **8** is graph of log(-log(survival)) Vs log(time) which shows that the infant and child mortality does not have parallel movement. It means that we should stratify over covariate infant/child indicator. For this reason ordinary Cox proportional hazard model will not be suitable for our data. Therefore we used stratified Cox proportional model to analyze our data. The result of stratified Cox proportional model is given in Table **4**.

The result in Table **4** shows that not all explanatory variables are important to fit multivariate stratified Cox proportional hazard model. Thus, the most appropriate covariates will be selected based on their contribution to the maximized log partial likelihood of the model -2Log L. The value of -2Log L for the null or empty model is 11353.488. Therefore, inclusion of covariates will be based on the amount of reduction of this value. From among the explanatory variables, type of birth covariate leads to the highest reduction in -2Log L, reducing its value to 91.707. This reduction is highly significant (p-value < 0.001) when compared with points of the X^2-distribution. The next highest change comes from region which is 23.600. Proceeding in this manner covariates will be eliminated. Thus, using the Wald chi-square test according to Table **4**, the predictors that are found to be significant, and will be considered for the next multivariable analysis at p-value 0.25 are sex, type of birth, father education, mother education, maternal age, wealth index, region and water supply including these covariates that are significant at univariate analysis.

The variables that appear to be important from Table **4** are fitted together. In the presence of certain variables, others may cease to be important. Consequently, we have five variables left in Table **5** that significantly reduces -2 Log L to 130.790. But those variables which are dropped or not significance on Tables **4** and **5** should be examined further. This examination is done because some variables could become significant in the presence of others.

Variables that were not important on their own, and so were not under consideration in Table **4** may become important in the presence of others. These variables are therefore added to the model from Table **4**, one at a time, and any that reduces -2Log L significantly are retained in the model. The model after forward selection of covariates at a significance level of 10% is presented below.

At last we need to do final pruning of main effects of the model by omitting variables that are non-significant, using significance level 0.05. At this stage, we may also consider adding interaction between any of the main effects currently in the model. Moreover we need to assess some realistic situation to see if two interaction effects can significantly change the survival time of children under the age five years. We will be using partial likelihood ratio test to see if the interactions are significant to be included in the model.

The hypothesis which we will test is going to be

H_o: The model with only main effect fits the model equally as well as the model having the main effect and their interaction.

H_A: The null hypothesis (H_o) is not true.

Test Statistics= $-2(\text{Log } L_{main} - \text{Log } L_{int})$

Decision Rule: Reject the null hypothesis (H_o) at significance level of a=0.05 if the test statistics is greater than or equal to $X^2_{1(a=0.05)}$=3.84, in other words include interaction in the multivariable analysis; otherwise we will retain the null hypothesis.

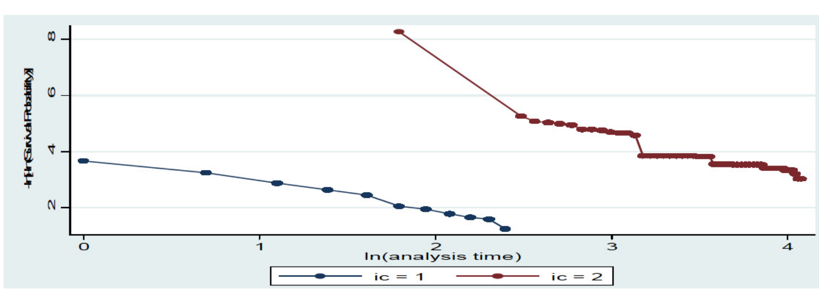

Figure 8: log(-log(survival)) Vs log(time).

Table 4: Summary of the Stratified Univariable Cox Proportional Hazard Model

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	-2 Log L
Sex(male)	-0.190	0.074	6.593	1	0.010	0.827	11346.858
Type of Birth (Single)	1.277	0.114	125.508	1	0.000	3.586	11261.781
Relgion(Orthodox)			4.697	4	0.320		1349.419
Catholic	0.526	0.285	3.401	1	0.065	1.692	
Protestant	-0.048	0.106	0.204	1	0.651	0.953	
Muslim	-0.038	0.086	0.197	1	0.657	0.963	
Others	0.158	0.252	0.394	1	0.530	1.171	
Father Education(No Edu)							11343.603
1.Primary	-0.187	0.129	9.673	2	0.008	0.821	
2.Secondary or above	-0.327	0.079	6.3845.591	1	0.012	0.829	
				1	0.018		
Mother Education(No Edu)							11339.816
1.Primary	-0.173	0.086	11.908	2	0.003	0.841	
2.Secondary or above	-0.663	0.222	4.092	1	0.043	0.515	
			8.886	1	0.003		
Birth Order(First)			2.503	2	0.286		11351.001
2-5	-0.137	0.096	2.049	1	0.152	0.872	
Late Born	-0.043	0.108	0.158	1	0.691	0.958	
Maternal Age(<20)			4.198	2	0.123		11349.231
20-29	-0.155	0.079	3.879	1	0.049	0.856	
>29	0.106	0.281	0.143	1	0.705	1.112	
Wealth Index(Poor)			4.457	2	0.108		11348.997
Medium	-0.108	0.104	1.081	1	0.299	0.898	
Rich	-0.172	0.084	4.248	1	0.039	0.842	
Water Supply(Piped)	0.328	0.171	3.689	1	0.055	1.389	11349.409
Region(Tigray)			21.270	9	0.019		11329.888
Afar	0.223	0.166	1.795	1	0.180	1.249	
Amhara	0.143	0.160	0.792	1	0.374	1.153	
Orimia	-0.040	0.150	0.070	1	0.792	0.961	
Somali	-0.004	0.175	0.000	1	0.983	0.996	
Benshangul-Gumuz SNNP	0.302	0.164	3.394	1	0.065	1.353	
Gambela	0.067	0.151	0.195	1	0.659	1.069	
Harrari	0.204	0.173	1.394	1	0.238	1.226	
Dire Dawa	0.009	0.194	0.002	1	0.962	1.009	
	-0.728	0.324	5.056	1	0.025	0.483	

Table **7** gives result for including interaction term in the model. According to it, all interaction terms are very significant, as the result of this we took all interaction terms with the main effects and have an automatic selection of the model *via* stepwise selection at significance level for entering 5% and remaining in the model 10%.

According to stepwise automatic variable selection process (with probability for entry and removal is 0.05 and 0.10 respectively) we have found that all the interaction terms except for birth order and sex cancelled each other out. Hence we proceed to the model diagnostics with model specified in Table **8**. This means we have come to the best fit model.

Without going into details and losing the theme of this paper we have suggested a separate model for each stratum. The following Tables **10** & **11** provide what the model would look like for each if each stratum were considered separately.

Table 5: After Backward Selection of Covariate on Stratified Cox Proportional Hazard Model

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Sex(Male)	-0.182	0.074	6.036	1	0.014	0.834	(0.721, 0.964)
Type of Birth(Single)	1.275	0.117	119.388	1	0.000	3.578	(2.847,4.498)
Father Education(No Edu)							
Primary	0.102	0.156	0.431	2	0.179	1.108	(0.816, 1.503)
Secondary or Above	-0.058	0.151	0.149	1	0.511	0.943	(0.701, 1.269)
				1	0.700		
Mother Education(No Edu)							
Primary	-0.052	0.095	4.482	2	0.106	0.949	(0.787, 1.144)
Secondary or Above	-0.535	0.253	0.299	1	0.584	0.586	(0.357, 0.961)
			4.477	1	0.034		
Region(Tigray)			13.977	9	0.174		
Afar	0.186	0.168	1.263	1	0.261	1.208	(0.869, 1.679)
Amhara	0.176	0.161	1.195	1	0.274	1.193	(0.870, 1.636)
Oromiya	0.015	0.151	0.010	1	0.921	1.015	(0.755, 1.365)
Somali	-0.041	0.176	0.055	1	0.814	0.960	(0.680, 1.354)
Benshangul-Gumuz	0.282	0.164	2.952	1	0.086	1.326	(0.961, 1.830)
SNNP	0.162	0.154	1.109	1	0.292	1.175	(0.870, 1.588)
Gambela	0.162	0.178	0.830	1	0.362	1.176	(0.830, 1.666)
Harrari	0.076	0.196	0.152	1	0.697	1.079	(0.735, 1.584)
Dire Dawa	-0.448	0.329	1.857	1	0.173	0.639	(0.335, 1.217)
-2 Log Likelihood=1122.698							

Table 6: After Forward Selection of Covariate on Stratified Cox Proportional Hazard Model

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Type of Birth(Single)	1.318	0.116	129.888	1	0.000	3.737	(2.979, 4.688)
Mother Education(No Edu)							
Primary	-0.182	0.089	15.227	2	0.000	0.834	(0.701, 0.992)
Secondary or Above	-0.810	0.227	4.185	1	0.041	0.445	(0.285, 0.693)
			12.780	1	0.000		
Birth Order(First)			9.042	2	0.011		
2-5	-0.285	0.099	8.310	1	0.004	0.752	(0.619, 0.913)
Late Born	-0.293	0.115	6.545	1	0.011	0.746	(0.596, 0.934)
Sex(male)	-0.191	0.074	6.628	1	0.000	3.737	(2.979, 4.688)
-2 Log Likelihood=11232.991							

Table 7: Partial Likelihood Ratio Test for Testing the Significance of Interaction Terms

Interaction term	$-2\text{Log } L_{main}$	$-2\text{Log } L_{int}$	$-2(\text{Log } L_{main} - \text{Log } L_{int})$	Decision
Mothers education*Birth Order	11353.488	11229.679	123.809	Reject H_o
Mothers education*Type of Birth	11353.488	11229.589	123.899	Reject H_o
Mothers education*Sex	11353.488	11231.589	121.689	Reject H_o
Birth Order*Type of Birth	11353.488	11230.243	123.245	Reject H_o
Birth Order*Sex	11353.488	11228.273	125.215	Reject H_o
Type of birth *Sex	11353.488	11232.991	120.495	Reject H_o

Table 8: Suggested Model after Final Stepwise Automatic Variable Selection

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Type of Birth(Single)	1.327	0.116	131.297	1	0.000	3.771	(3.005, 4.732)
Mother Education(No Edu)							
Primary	-0.182	0.089	15.212	2	0.000	0.834	(0.700, 0.992)
Secondary or Above	-0.809	0.227	4.200	1	0.040	0.445	(0.285, 0.694)
			12.750	1	0.000		
Birth Order(First)			14.159	2	0.001		
2-5	-0.464	0.127	13.384	1	0.000	0.629	(0.491, 0.806)
Late Born	-0.429	0.144	8.861	1	0.003	0.651	(0.491, 0.864)
Sex(male)	-0.506	0.169	8.954	1	0.003	0.603	(0.433, 0.840)
(Birth Order)*Sex			4.631	2	0.099		
(2-5)*Sex	0.423	0.197	4.613	1	0.032	1.527	(1.038, 2.248)
(Late Born)*Sex	0.335	0.222	2.269	1	0.132	1.398	(0.904, 2.161)

Table 9: Suggested Model for Infant Stratum after Stepwise Automatic Variable Selection

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Type of Birth(Single)	1.383	0.121	130.966	1	0.000	3.988	(3.147, 5.055)
Mother Education(No Edu)							
Primary	-0.145	0.056	9.351	2	0.009	0.865	(0.755, 0.975)
Secondary or Above	-0.716	0.251	6.704	1	0.014	0.489	(0.299, 0.799)
			8.161	1	0.004		
Birth Order(First)			14.202	2	0.001		
2-5	-0.404	0.124	10.657	1	0.000	0.668	(0.524, 0.851)
Late Born	-0.377	0.106	12.656	1	0.001	0.686	(0.558, 0.844)
Sex(male)	-0.191	0.074	6.628	1	0.000	3.737	(2.979, 4.688)
Supp Water(Piped)*Sex(male)	0.781	0.392	3.960	1	0.047	2.183	(1.012, 4.710)

Table 10: Suggested Model for Child Stratum after Stepwise Automatic Variable Selection

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Type of Birth(Single)	1.323	0.119	123.837	1	0.000	3.753	(2.973, 4.738)
Mother Education(No Edu)							
Primary	-0.174	0.085	11.908	2	0.009	0.840	(0.710, 1.042)
Secondary or Above	-0.667	0.235	4.190	1	0.103	0.513	(0.268, 0.792)
			4.175	1	0.041		
Sex(male)	-1.166	0.382	9.318	1	0.002	0.312	(0.148, 0.659)
Supp Water(Piped)*Sex(male)	0.985	0.383	6.618	1	0.010	2.679	(1.264, 5.675)

3.4. Model Diagnostics

In this section we will:

- Test the assumption of proportional hazards
- Check influence and poorly fit subjects and
- Over all summary of goodness of fit

3.4.1. Test of the Assumption of Proportional Hazards

The proportional hazard assumption, which asserts that the hazard ratios remain constants over time, is very important for interpretation and use of a fitted proportional hazards model. In other words, the risk of failure should be the same no matter how long subjects are studied. To test this assumption, graphical

Table 11: Result after Fitting Time Dependent Covariates

Covariate (Reference)	β	s.e(β)	Wald	d.f	Sig	HR	95% CI for HR
Type of Birth(Single)	1.336	0.125	114.094	1	0.000	3.806	(2.978, 4.863)
(Type of Birth)*Time	-0.17	0.017	1.075	1	0.300	0.983	(0.951, 1.016)
Mother Education(No Edu)							
Primary	-0.144	0.094	7.915	2	0.019	0.866	(0.720, 1.041)
Secondary or Above	-0.611	0.233	2.349	1	0.125	0.543	(0.344, 0.857)
			6.892	1	0.009		
(Mother Edu)*Time	-0.06	0.008	0.525	1	0.469	0.994	(0.978, 1.010)
Birth Order(First)			3.367	2	0.186		
2-5	-0.181	0.099	3.357	1	0.067	0.835	(0.688, 1.013)
Late Born	-0.139	0.123	1.288	1	0.256	0.870	(0.684, 1.106)
(Birth Order)*Time	0.010	0.006	0.2743	1	0.098	1.010	(0.998, 1.021)
Sex(male)	-0.244	0.084	8.350	1	0.004	0.784	(0.664, 0.925)
(Sex)*Time	0.010	0.008	1.790	1	0.181	1.010	(0.995, 1.026)
(Birth Order*Sex)	-0.101	0.037	7.389	1	0.007	0.904	(0.841, 0.972)
(Birth Order*Sex)*Time	0.13	0.011	1.400	1	0.311	1.139	(0.946, 1.144)

diagnoses of scaled Schoenfeld residuals and likelihood based tests, like Wald test can be employed to assess the proportional hazard assumption for covariates that are significant in the multivariable analysis.

Given that the assumption of proportionality of the proportional hazards model is satisfied, the distribution of residuals over time should be random and LOWES smoothing line will also be a straight line around zero.

To test proportionality assumption of proportional hazards assumption, generate time varying covariate by creating interactions of the predictors and a function of survival times, and include these in the model. If any of time covariates is significant then those predictors do not have a constant proportionality over the study period. In this case we can conclude the proportional hazards assumption is violated.

Table **9** shows Wald chi-square value and corresponding p-value for each covariate. From the table we have the smallest p-value for *time dependent covariate* is 0.098 which is still greater than 0.05. From this we can conclude that all covariates involved in the model are independent of time. In other words we have a model that satisfies the proportional hazards model.

<u>Graphical Assessment of the Proportional Hazard Assumptions</u>

The graphical displays show plots of the scaled Schoenfeld residuals against the survival time for each covariate namely mothers level of education, type of birth, birth order and sex of a child. In Figure **9-13** plots of scaled Schoenfeld residuals show randomness. Furthermore the smoothed curve is an approximately horizontal line around zero; as the result of this we can say the above five covariates have satisfied the assumption of proportional hazards.

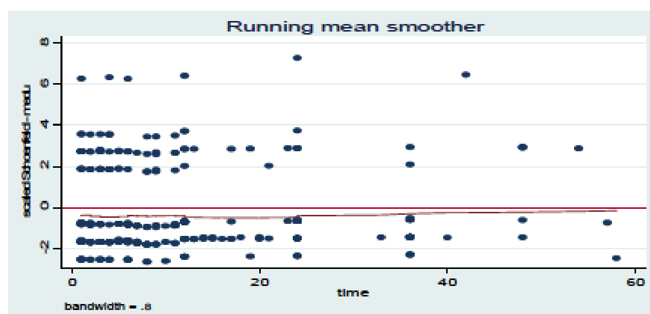

Figure 9: Scaled Schoenfeld Residuals for mother educational level.

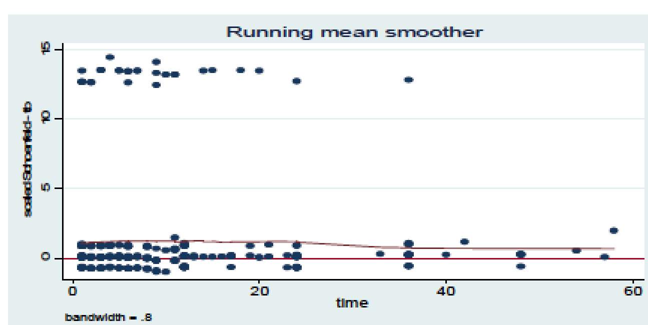

Figure 10: Scaled Schoenfeld Residuals for type of birth.

Figure 11: Scaled Schoenfeld Residuals for birth order.

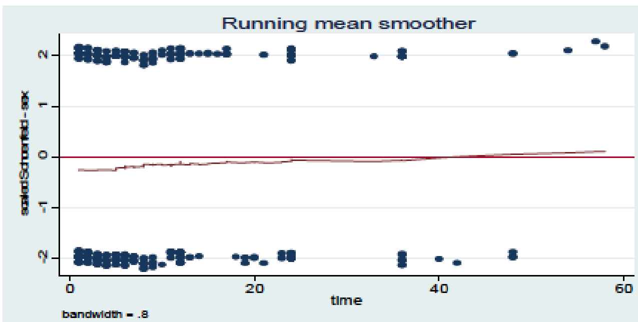

Figure 12: Scaled Schoenfeld Residuals for sex of child.

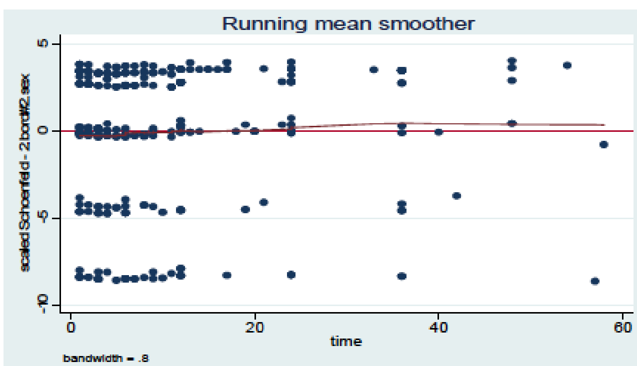

Figure 13: Scaled Schoenfeld Birth Order by Sex Interaction.

3.5. Interpretation of the Results

The model fitted to the under five data given in Table **8** contains four categorical covariates and one interaction term. All covariates have three level categories except for sex and type of birth indicator.

The estimated covariate coefficient $\hat{\beta} = 1.327$ for type of birth (reference group being single birth) indicator implies that the hazards ratio is $exp(\hat{\beta}) = 1.771$. This shows children with multiple births (includes twins, triplets and quadriples) have 3.771 times higher risk of death than children with single births. The implication of 95% confidence interval is that the hazard ratio can get as low as 3.005 and as high as 4.732.

For the education level of mothers, there are three categories (no education, primary level and at least secondary). Taking no education as reference group we have the coefficients $\hat{\beta} = -0.182$ and -0.809 for primary and at least secondary education level respectively. These figures imply that the hazard ratios for these categories are 0.834 and 0.445 for primary and at least secondary education level, respectively. A child born from a mother with primary education level has 16.6% lower risk of death than a child born from a mother with no education. The 95% confidence interval implies the rate could be as low as 0.700 and as high as 0.992. And a child born from a mother with at least secondary level of education has 55.5% lower risk of death than a child born from a mother with no education. The implication from 95% confidence interval is that the rate could get as low as 0.285 and as high as 0.694.

Birth order and sex indicator have been included in the model, as main effects and interaction. Since sex indicator is dichotomous, we fix male as reference group, we will examine the birth order hazard ratio at each level of sex indicator. The estimated log hazard as function of sex, birth order and their interaction (interaction effect of sex and birth order) holding the other variables constant is given as follows:

$$Log\,\widehat{HR}(sex, birth\ order, x) = \hat{\beta}_1 sex + \hat{\beta}_2 birth\ order + \hat{\beta}_3 sex * birth\ order + \hat{\beta}' x$$

The next expression for the difference of an increase of say for the birth order 2-5 holding sex constant is

$$Log\,\widehat{HR}\{sex, Birth\ Order + 1, x\}$$
$$= \hat{\beta}_1 sex + \hat{\beta}_2 (birth\ order + 1) + \hat{\beta}_3 sex * (birth\ order + 1) + \hat{\beta}' x$$

Finally, the estimated hazard ratio for being 2nd-5th in the birth order and male (i.e sex=0) is

$$\widehat{HR}(sex = 0, Birth\ Order) = \exp(-0.464)$$
$$= 0.628$$

And for female (i.e. sex=1) the hazard ratio becomes

$$\widehat{HR}(sex = 1, Birth\ Order) = \exp(-0.464 + 0.423)$$
$$= 0.959$$

The interpretation is that being 2nd -5th in the order of birth for male child has 37.2% lower risk of death than being the first born male child. And for female, being 2^{nd} - 5^{th} in the order of birth has 4% lower risk of death than being the first born.

We believe we need to say some about the interpretation of the models given in Tables **9** and **10**. According to Table **9** we found four variables and one interaction term significantly affecting the mortality of infants.

The estimated hazard ratios of dying of infants who have multiple birth type were as compared to those who had single birth type is 3.988. This means infants who had multiple birth order have 3.988 times higher risk of death than those infants with single birth type. And this risk of death can get 3.147 times as low and 5.055 times as high as compared to those infants with single birth type.

Taking mothers with no education as reference group we have the coefficients $\hat{\beta} = -0.145$ and -0.716 for primary and at least secondary education level respectively. These figures imply that the hazard ratios for these categories are 0.865 and 0.489 for primary and at least secondary education level, respectively. An infant born from a mother with primary education level has 13.5% lower risk of death than a child born from a mother with no education. The 95% confidence interval implies the hazard ratio could be as low as 0.717 and as high as 0.975. And for an infant born from a mother with at least secondary level of education has 51.1% lower risk of death than a child born from a mother with no education. The implication from 95% confidence interval is that the hazard ratio could get as low as 0.299 and as high as 0.799.

Having the first born as reference we have the hazard ratios 0.668 and 0.683 for 2-5 order and late born respectively. An infant with birth order in 2-5 has 33.2% less risk of death than an infant with first in the birth order. The hazard ratio for infant in 2-5 in order of birth can get as low as 0.755 and as high as 0.975. And for an infant who were late born has 31.7% less risk of death than an infant with first in the birth order. The 95% confidence interval imply that the hazard ratio can get as low as 0.558 and as high as 0.844.

The other two variables which are present in the model are sex and water supply. What is unique about them is that both are dichotomous and interacted to each others. Our references are male and treated water supply for sex and water supply covariates respectively. The estimated log hazard as function of sex and water supply holding other variables constant is give as follows:

$$\log \widehat{HR}(sex, water\ supp, x)$$
$$= \hat{\beta}_1 sex + \hat{\beta}_2 sex * water\ supp + \hat{\beta}' x$$

Next we have provided the expression for increase one level from the reference of water supply (from piped to others category)

$$\log \widehat{HR}(sex, water\ supp + 1, x)$$
$$= \hat{\beta}_1 sex + + \hat{\beta}_2 sex * (water\ supp) + \hat{\beta}' x - [\hat{\beta}_1 sex + \hat{\beta}' x]$$
$$= \hat{\beta}_2 sex * (water\ supp)$$

The estimated hazard ratio for others water supply category and female infant is

$$\widehat{HR}(sex = 1, water\ supp + 1) = \exp(0.781)$$
$$=2.183$$

The risk of death for a female infant whose source of water in the untreated source is 2.183 times than those female infant who use water from treated source.

Our final model which is presented in Table 4.11 is for child (this is for child age group 12 months to 59 months) stratum. It is composed of four variables and an interaction.

The hazard ratio for multiple birth type taking single birth type as reference is 3.753. This shows children with multiple births have 3.753 times higher risk of death than children with single births. The 95% confidence interval implies that the hazard ratio could be as low as 2.973 and as high as 4.738.

Taking mothers with no education as reference group we have the coefficients $\hat{\beta}$ =-0.174 and -0.667 for primary and at least secondary education level respectively. These figures imply that the hazard ratios for these categories are 0.840 and 0.513 for primary and at least secondary education level, respectively. Given the hazard ratio for mothers with primary level education is not significantly different from the category for mothers with no education as the p-value is greater than 0.05. But for a child from a mother with at least secondary level of education has 48.7% lower risk of death than a child born from a mother with no education. The implication from 95% confidence interval is that the rate could get as low as 0.268 and as high as 0.792.

Sex and water supply also are there in the model in Table **10**. Their log hazard function is same as in the case of infants and given below.

$$\log \widehat{HR}(sex, water\ supp, x)$$
$$= \hat{\beta}_1 sex + \hat{\beta}_2 sex * water\ supp + \hat{\beta}' x$$

Next we have provided the expression for increase one level from the reference of water supply (from treated to untreated category)

$$\log \widehat{HR}(sex, water\ supp + 1, x)$$
$$= \hat{\beta}_1 sex + + \hat{\beta}_2 sex * (water\ supp) + \hat{\beta}' x - [\hat{\beta}_1 sex + \hat{\beta}' x]$$
$$= \hat{\beta}_2 sex * (water\ supp)$$

The estimated hazard ratio for others water supply category and female children relative to female children who use water from pipe is

$$\widehat{HR}(sex = 1, water\ supp + 1) = \exp(0.985)$$
$$=2.679$$

The risk of death for a female child whose source of water in the untreated source is 2.183 times than those female children who use water from treated source.

4. DISCUSSION OF THE RESULTS

This study identified factors that are significantly associated with increased risk of under five mortality. The results of the Cox proportional hazards regression analysis show mother's educational level, birth order, types of birth, sex and interaction between sex and birth order were significantly associated with under five mortality.

The impact of mother's educational level on under five survival has been assessed by several studies indicating that non educated mothers are associated with high risk of death of infant and child. [10] identified that the mothers 'education is the most important factor that directly affects child mortality. A similar study in Malakal town, southern [11] found that mother's education, have significant influence on infant and under-five mortality. A study by [12] found that infant and child mortality are highly associated with mother's education. A study conducted in Ethiopia by [13] and by [14] also found that the mortality risk of children born to non educated mothers is higher as compared to a child born to an educated mother. Another study by [15] examined the effect of maternal education on infant and child mortality and found that there was a consistent negative relationship between maternal education and the probability of infant death. Children of mothers who attended primary school are less likely to die than are children of mothers with no education. Children of mothers with a secondary-school education are the least likely to experience infant and child deaths. The finding of the current study also agrees with the above sited findings.

The current study identifies that birth order as a risk factor of infant and child mortality. This study suggests that first born infants and children experience higher risk of dying than infants and children whose birth order is two up to five and six and above; infants and children with birth order two up to five have a higher risk of dying than a child whose birth order is six and above. A study conducted in Ethiopia by [14] and by [4] also found that birth order is one of the determinants of child mortality showing that a first born child was exposed to a high risk of mortality. Another study by [16] indicates that first births are less likely to survive than higher order births.

The current study also identifies that sex of child as a risk factor of infant and child mortality. This study suggests that the risk of dying for female is lower than males. Similarly in Kenya, [17] found that sex of child is important determinant for the risk of infant and child mortality.

The current study found out that the infant and child mortality risk associated with multiple births was 3.771 times higher relative to singleton births ($p<0.001$) the study also showed that death due to types of birth is highly dependent on birth order. A similar study by [17] in Zimbabwe found that multiple births tend to increase infant and child mortality. Another study by [18] also found that multiple births are associated with much higher risk of death.

Limitations of the study are as follows. The study is conducted based on secondary data which might have incomplete and biased information. Also information might have been missed in case of many censored observations. In many tuberculosis patients, multiple causes of death may act simultaneously, so the cause of death may not be determined accurately.

REFERENCE

[1] Federal Ministry of Health. Essential Health Service Package for Ethiopia. Addis Ababa; 2005 August 2005.

[2] Ministry of Finance and Economic Development (MoFED). 'Ethiopia: Progress toward achieving the MDGs', 2008.

[3] Ringheim K, Teller C, Sines E. Ethiopia at a crossroads: Demography, Gender, and Development, Addis Ababa, Ethiopia 2009.

[4] Desta M. Infant and Child Mortality in Ethiopia: The role of Socioeconomic, Demographic and Biological factors in the previous five years period of 2000 and 2005, Lund University 2011.

[5] World Bank. World Development Indicators (WDI). The World Bank Online Database, Washington, DC: World Bank 2004.

[6] Ethiopia Demographic and Health Survey, Central Statistical Agency, Addis Ababa 2011.

[7] Cox DR. 'Regression models and life tables', Journal of the Royal Statistical Society; series B 1972; 34: 187-220.

[8] Hosmer DW, Lemeshow S, May S. Applied survival analysis regression modeling of time to event data. John Wiley and Sons, Inc., New York 2008.
http://dx.doi.org/10.1002/9780470258019

[9] Collett D. Modeling survival data in medical research, Second edn, Chapman and Hall/CRC, London 2003.

[10] Aguirre PG. 'Child mortality and reproductive patters in urban', Bolivia, CDE working paper No. 95-28, Center for Demography and Ecology, University of Wisconsin-Madison 1995.

[11] Mahfouz MS, Surur AA, Ajak DAA, Eldawi EA. 'Level and Determinants of Infant and Child Mortality in Malakal Town – Southern Sudan, Sudanese Journal of Public Health 2009; 4(2): 250-255.

[12] Caldwell JC. Education as a factor in mortality decline: An examination of Nigerian data. Population Studies 1979; 33(3): 395-413.
http://dx.doi.org/10.2307/2173888

[13] Ezra M, Gurum E. Breastfeeding, birth intervals and child survival: analysis of the 1997 community and family survey data in southern Ethiopia'. Ethiopian Journal of Health Development 2002; 16: 41-51.
http://dx.doi.org/10.4314/ejhd.v16i1.9825

[14] Kumar P, Gemechis F. Infant and child mortality in Ethiopia: As statistical analysis approach. Ethiopian Journal of Science and Education 2010; 5(2): 51-57.

[15] Desai S, Alva S. Maternal education and child health: Is there a strong causal relationship? Demography 1998; 35(1): 71-81.
http://dx.doi.org/10.2307/3004028

[16] Balk D, Tom P, Adam S, Fern G, Melissa N. Spatial Analysis of Childhood Mortality in West Africa', Calverton, Maryland, USA: ORC Macro and Center for International Earth Science Information Network (CIESIN), Columbia University 2003.

[17] Mustafa E, Odimegwu C. Socioeconomic determinants of infant mortality in Kenya: Analysis of Kenya DHS 2003, 2008.

[18] Kombo J, Ginneken V. Determinants of infant and child mortality in Zimbabwe: Result of multivariate hazard analysis 2009.

Received on 14-04-2014 Accepted on 09-06-2014 Published on 05-08-2014

http://dx.doi.org/10.6000/1929-6029.2014.03.03.6