

**HOUSEHOLD ENVIRONMENTAL AND SOCIO-ECONOMIC
DETERMINANTS OF CHILD MORTALITY IN MALAWI: A COMPARISON
OF COX PROPORTIONAL HAZARDS AND FIVE PARAMETRIC
SURVIVAL MODELS**

MSc (BIOSTATISTICS) THESIS

BY

AGNESS THAWANI

UNIVERSITY OF MALAWI

CHANCELLOR COLLEGE

NOVEMBER 2012

**HOUSEHOLD ENVIRONMENTAL AND SOCIO-ECONOMIC
DETERMINANTS OF CHILD MORTALITY IN MALAWI: A COMPARISON
OF COX PROPORTIONAL HAZARDS AND FIVE PARAMETRIC
SURVIVAL MODELS**

MSc (BIOSTATISTICS)

By

Agness Thawani

BSc in (Mathematical sciences Education)-University of Malawi

A Thesis submitted to Faculty of Science in partial fulfilment of the requirements

For the degree of

Master of Science in Biostatistics

UNIVERSITY OF MALAWI

CHANCELLOR COLLEGE

NOVEMBER 2012

DECLARATION

I the undersigned hereby declare that the work in this dissertation was carried out in accordance with the regulations of the University Of Malawi. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any degree. Any views expressed in the dissertation are those of the author and in no way represent those of the University Of Malawi.

Signed.....Date.....

AGNESS THAWANI

Certificate of Approval

The undersigned certify that this thesis represents the students own work and effort and has been submitted with our approval

Signature.....Date.....

Sarah Ann White, PhD (Lecturer)

Main Supervisor

Signature.....Date.....

Jupiter Simbeye MSc. (Lecturer)

Course Coordinator of Master of Science in Biostatistics

To
The Almighty God for the Grace
And
My Husband
Dick Mwambula

ACKNOWLEDGEMENTS

I am deeply indebted to my Supervisor, Dr Sarah White for her invaluable advice and patient guidance. This thesis could not have been written without her help and support.

ABSTRACT

Background

The Cox proportional hazards regression model has become the most used tool in the analysis of censored survival data. However, some features of the Cox model may cause problems for the analyst or an interpreter of the data. They include the restrictive assumption of proportional hazards for covariate effects, and “loss” (non-estimation) of the baseline hazard function induced by conditioning on event times. Compared with Cox Proportional Hazard model, parametric models are different in the way they exploit the information contained in the data. Parametric models specify how the hazard varies over time (hazard shape), which may provide insights into and guidance on how best to compare outcomes. This study compares the goodness of fit of Cox proportional hazard model and parametric survival models in modelling the household environmental and social economical determinants of under-five child mortality

Methods

The study used the 2010 Malawi Demographic Health survey data. The Cox Proportional hazard model was used and the proportional hazard (PH) assumption was assessed using both the graphical method and by adding time-dependent covariate in the Cox model. Good-ness of fit of the Cox PH model was also assessed using Cox Snell residuals. The parametric proportional hazard as well as accelerated failure time models was also used. The Weibull, lognormal, log-logistic, exponential and Gompertz model were fit and to find the most appropriate model, these models were compared using Akaike Information Criterion (AIC) and the goodness of fit for all the parametric models was assessed using Cox-Snell residuals.

Results

The Cox Proportional hazard model violated the assumption of proportionality and was not fitting the data well. The lognormal model was found to fit the data well and since this model is expressed in terms of accelerated failure time model, the violation of the proportion hazards assumption was overcome. Mother education, father education, house hold size, source of cooking fuel, type of toilet facility and wealth index are found to be significantly associated with child mortality. On the other hand, area of residence, source of water and access to electricity are found to be not significantly associated with child mortality.

Conclusion

The results obtained from the Cox PH model are not as effective as those obtained from the parametric AFT model since the PH assumption was found to be violated in the Cox PH model. Log-normal AFT model is found to be the most appropriate parametric model to be used in the analysis of child survival. Hence, researchers in child mortality using survival analysis can use the log-normal model as this will give them the more accurate and efficient results.

Key words: under-five mortality, Cox-Snell residuals, Cox proportional hazard, Weibull distribution, log-normal distribution, log-logistic distribution, Gompertz distribution, exponential distribution.

TABLE OF CONTENTS

ABSTRACT	v
Chapter 1: Introduction	1
1.1 Background	1
1.2 Literature review	5
1.2.1. Causes of Infant Mortality	5
1.2.2. Models in the analysis of child mortality	7
1.2.3. Under-five children mortality studies	12
1.3 Problem Statement	25
1.4 Objectives of the study	25
1.4.1. Broad Objective	25
1.4.2. Specific objectives	26
1.5 Justification of the study	26
1.6 Definitions of terms	27
Chapter 2: Methods and Data sources	28
2.1 Data	28
2.1.1. Sample design	28
2.1.2. Limitations	28
2.2. Variables considered	30
2.2.1. Measurement of variables	30
2.3 Theoretical Model	32
2.3.1. Survival time distribution	33
2.3.2. Ordinary least squares	35
2.3.3. The Kaplan-Meier estimate of the survival function	36
2.3.4. The Cox proportional Hazards model	36
2.3.5. Partial likelihood estimate for Cox proportional hazards model	38
2.4 Parametric PH models	39
2.4.1. Weibull PH model	40
2.4.2. Exponential PH model	42
2.4.3. Gompertz PH model	45
2.5 Accelerated Failure Time model(AFT)	45
2.5.1. Log-Logistic model	47

2.5.2. Log-normal model	49
2.6 Time dependence properties of the hazard function	50
2.7 Frailty models.....	51
2.8 Model checking using statistical criteria.....	52
2.8.1. Residual plots.....	53
2.8.2. The Akaike's Information Criterion (AIC).....	53
2.9 Data analysis	54
Chapter 3 Results	57
3.1 Descriptive	57
3.2 Cox PH Model.....	63
3.3 Parametric Results	69
3.3.1. Exponential Model	64
3.3.2. Weibull Model.....	65
3.3.3. Gompertz model	65
3.3.4. Log-logistic model	67
3.3.5. Log-normal model	68
3.4 Determining the best model fit.....	70
3.4.1. Model selection/Comparison.....	70
3.4.2. Cox Snell Residuals	71
Chapter 4 Discussions	73
Chapter 5 Conclusion and recommendation	80
References.....	81
Appendices.....	90

LIST OF FIGURES

Figure 1 Major causes of death in neonates and children under five.....	5
Figure 2 Kaplan Meier Curves.....	61
Figure 3 test of PH assumptions graphically using scaled Schoenfeld.....	68
Figure 4 cumulative hazard plot of the Cox –Snell residuals for the Cox PH model.....	69
Figure 5 Hazard functions.....	69
Figure 6 Cumulative hazard plot of the Cox Snell residual for parametric models.....	72

LIST OF TABLES

Table 1 Models in the analysis of under-five child mortality.....	9
Table 2 Variable description.....	30
Table 3 Baseline characteristic and outcomes.....	58
Table 4 Log –rank test for equality of survival functions.....	59
Table 5 Univariate and multivariate Cox PH model for the relative hazard of child mortality.....	64
Table 6 Test of Test PH assumptions by including time dependent covariate in the model.....	66
Table 7 Test for proportional hazard assumption using schoenfeld residuals.....	67
Table 8 Coefficients and hazard rate from parametric models for under-five children	63
Table 9 Akaike Information Criterion.....	70

Appendices

Appendix 1 Parametric Univariate Estimates.....	88
Appendix 2 Multivariate parametric results	89
Appendix 3 Hazard functions for different parametric models fitted	90
Appendix 4 Relevant pages from the Women and Household 2010 MDHS Questionnaires.....	91

LIST OF ABBREVIATIONS AND ACCRONAMES

PH	Proportional Hazard
AFT	Accelerated Failure time
WHO	World Health Organisation
MDHS	Malawi Demographic Health Survey
MDG	Millennium Development Goal
IMR	Infant Mortality rate

Chapter 1: Introduction

1.1 Background

The Cox proportional hazards regression model has become the most used tool in the analysis of censored survival data (Moran 2002). Researchers in medical science often tend to prefer semi parametric models instead of the parametric models because the relationship between covariates and the hazard can be estimated without having to make assumptions about the nature and shape of the baseline hazard rate. However, some features of the Cox model may cause problems for the analyst or an interpreter of the data. They include the restrictive assumption of proportional hazards for covariate effects, and “loss” (non-estimation) of the baseline hazard function induced by conditioning on event times. A review of survival analysis in cancer journals reveals that only 5% of all studies using the Cox PH model considered the underline assumption (Altman et al., 1985). Where PH assumption is not met, it is improper to use standard Cox PH model as it may entail serious bias and loss of power when estimating or making inference about the effect of a given prognostic factor on mortality (Allison 1995).

Compared with Cox Proportional Hazard model, parametric models are different in the way they exploit the information contained in the data. Parametric models specify how the hazard varies over time (hazard shape), which may provide insights into and guidance on how best to compare outcomes (Cleves 2004).

Recently, AFT models as parametric models have attracted considerable attention, because not only they do not need PH assumption but also thanks to availability of

standard methods such as Maximum Likelihood (ML), parameter estimation and testing can be done readily (Klien 1997).

When survival time has a specific statistical distribution, the statistical power of parametric survival models is higher than nonparametric or semi-parametric survival models. The exponential, Gompertz, Weibull, log-logistic, and lognormal are among parametric distributions commonly used for studying survival time analysis. Survival estimates obtained from parametric survival models typically yield plots that are more consistent with a theoretical survival curve (Kleinbaum 2005).

Parametric accelerated failure time (AFT) models provide an alternative to the PH model for statistical modelling of survival data (Wei, 1992). Unlike the PH model, the AFT approach models survival times directly and generates a summary measure that is interpreted in terms of the survival curve (Hutton and Monaghan, 2002).

Since recently AFT models have not been used very often and the few usage of these models are found in kidney transplant studies (Saint-Marcoux 2005). Based on our knowledge, it has not been used to recognize the prognostic factors of under-five mortality in Malawi so far.

In this study, the model performance of the Cox proportional hazard model was compared with that of the parametric survival models like Weibull, Gompertz, log-normal, exponential and Log-logistic models which are used in modelling the household environmental and social economical determinants of under-five child mortality in Malawi. The results are expected to assist researchers in child mortality to get insight into the suitable model to use.

Child survival is a function of availability and access to basic needs to support life at both individual, household and neighbourhood levels (WHO 2004). Malawi is on track to achieve the millennium Development Goal (MGD #4) on reducing child mortality. The 2006 Malawi multiple Indicator Cluster Survey (MICS) showed that there was a sharp decline in the infant and under-five mortality rates, from 104 and 189 per 1000 live births respectively in 2000 to 72 and 122 in 2006. They attributed the decline to sustained high coverage of immunisation and vitamin A supplementation, elimination of neonatal tetanus, malaria control activities, and increased rates of exclusive breastfeeding and access to safe drinking water. Although accurate information on cause of death is lacking, the leading cause of under-five mortality in Malawi is neonatal conditions, pneumonia, diarrhoea, malaria, AIDS and malnutrition (UNICEF Malawi 2010).

The U.N. report (2006) indicated that, Malawi is faring better than many of its counterparts in addressing infant mortality despite its existence in the sub-Saharan Africa which is regarded as the most dangerous place in the world for new born. According to the 1992, 2000 and 2004 Malawi demographic and Health Surveys, infant mortality rate has been steadily decline in Malawi from a very high level of 134 in 1992 to 69 in 2006. Though there is this decline, under-five mortality is still high.

Children are exposed to serious health risks from the environmental hazards. Environmental risk factors are most often neglected and yet they are playing a great role in threatening the lives of under-five children. These environmental factors are particularly influenced by adverse social and economic conditions, particularly conflict, poverty and malnutrition. At least 3 million children under the age of five die each year due to environment-related diseases. Acute respiratory infections annually

kill an estimated 2 million children under the age of five. As much as 60 per cent of acute respiratory infections worldwide are related to environmental conditions. Diarrhoeal diseases claim the lives of nearly 1.5 million children every year. Eighty to 90 per cent of these diarrhoea cases are related to environmental conditions, in particular, contaminated water and inadequate sanitation. (WHO 2010)

Environmental risks to children vary from region to region. Children in many countries still face the major traditional environmental hazards, including unsafe water, lack of sanitation and contaminated food, injuries, indoor air pollution and use of solid fuel, outdoor air pollution and exposure to a myriad of toxic heavy metals, chemicals and hazardous wastes that may be brought home from work place. However, other children live in adverse environment that are vastly different from those of generations ago. In addition to the traditional environmental hazards, due to rapid changes in economic structures, technologies and demography, new or modern environmental hazards have appeared or been recognized, such as increased use of radiation. These may be linked to global challenges such as uncontrolled urbanisation, industrialisation in developing countries, ecosystem degradation, and the impacts of climate change (Mutunga, 2004).

1.2 Literature review

1.2.1. Causes of Infant Mortality

There are several causes of under-five children deaths. Figure 1 shows the major causes of under-five children deaths. Apart from the one indicated in the chart, there are some environmental and socio-economic factors that also affect child survival. These have been highlighted in the following discussions.

Source: World Health Statistics 2011, WHO

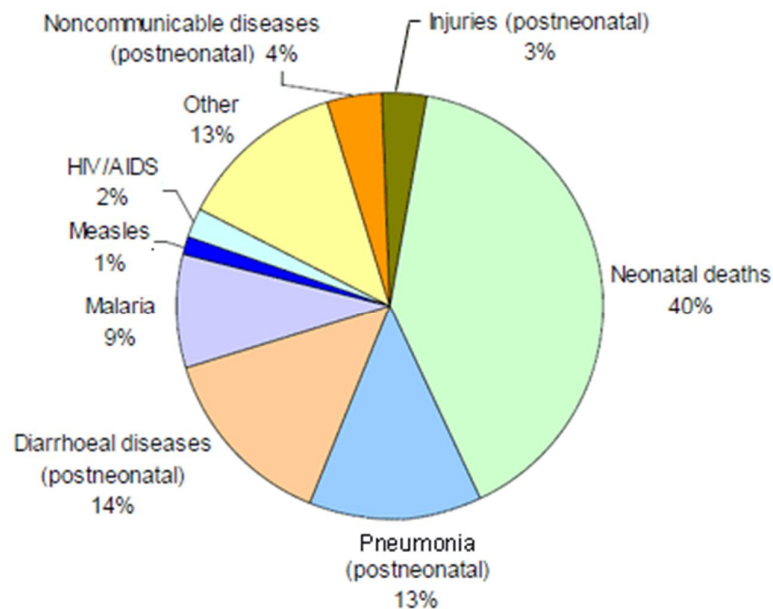


Figure 1 Major causes of death in neonates and under five children globally-2008

Mutunga (2004) argues that causes of infant mortality are multi-factorial, especially in developing countries, where there are great variations between social, economic and demographical groups of people even inside one country. Thus, in determining infant mortality one must take into account this diversity.

A three-tier model of causes of child mortality in developing countries was first put forward by Millard et al. (1989), as cited in Espo (2002) which mentioned three layers of causes as proximate, intermediate and ultimate. The proximate causes included the immediate biomedical conditions that result in death and typically involve the interactions of malnutrition and infection. Many public health programmes have addressed the proximate causes in an attempt to improve child health, for instance, through immunization campaigns. The intermediate layer includes child care practices and other behaviour that increases the exposure of children to causes of death on the proximate tier. Specific patterns that increase exposure to proximate causes include breast feeding habits, health-seeking behaviour etc. The ultimate tier encompasses the broad social, economic, and cultural processes and structures that lead to the differential distribution of basic necessities like food, shelter, and sanitation. The ultimate tier thus forms the context of causes located on the other tiers. In developing countries, mortality rates are influenced by socio-economic, demographic and environmental variables (Mosley and Chen 1984).

Mortality and its converse indicator, life expectancy are among the most important measures of well-being and development in poor countries. It is particularly important to analyse the determinants of child mortality in poor countries such as Malawi since child mortality has an overwhelming influence on life expectancy.

Several household environmental and socio-economic characteristics make children more vulnerable to the attack of various diseases. These determinants usually involve education of the parents, income or wealth situation of the household, access to water and sanitation services and access to health services. In this study we will specifically

examines how infant and child mortality is related to the household's environmental and socio-economic characteristics, such as mother's education, source of drinking water, Type of toilet facility, type of cooking fuels, wealth index, access to electricity, partner education, area of residence, and Household size.

1.2.2. Models in the analysis of child mortality

Cox PH models the relationship between explanatory variables and time to event. In the analysis of child mortality, length of time lived or survival time is an important indicator, so survival models need to be used when modelling child mortality. There are several models which have been used in the analysis of child mortality where Cox Proportion hazard model has been used popularly because its form is flexible enough to allow time-dependent covariates as well as stratification. There are few studies which have employed parametric models and most of them used Weibull because it can be presented in both PH and AFT metric hence can be used to estimate both relative rates and relative extension in survival time (Kelvin J, 2002) and others choose Weibull without any particular reasons (Mutunga, 2004).

Child mortality has so far been analysed using different models classes. Table 1 gives different models that have been used in the analysis of child mortality and their findings. The name of the author, the year of publication, the model used in the analysis, the data sources and the finding from the study are all included in the table. From the table, the commonest determinants of child mortality were wealth index, Mother education, partner education, birth interval, sanitation facility, source of water, access to electricity, infant immunization, area of residence, maternal age, multiple birth, and household size.

Low levels of education for the mother or the father, higher mortality rate are experienced in low income households. The mothers and fathers level of education is strongly linked to child survival. Higher levels of education attainment are generally associated with lower mortality rates. Safe source of drinking water has negative significant effects on child mortality risk. Those household using safe source of water like tapped water had lower risk of child mortality as compared to those households using unsafe water like water from the Well. The same applies for sanitation, where in most cases access to a flush toilet or a ventilated improved pit latrine was associated with lower mortality rate. Similarly, urban areas have more advantages and therefore better child survival prospects. The patterns of mortality by maternal age and birth order were typically U-shaped. Children born to both relatively old and young women have higher mortality rates than others. In this case the effect of maternal age at birth on infant mortality rate were biological i.e. it depends on reproductive maturity. Less child mortality risk were also experienced in children who were immunised. Higher child survival prospects were found in small households as compared to larger households.

Table 1 Models in the analysis of under-five child mortality(UV=Univariate and MV=multivariate)

Author	Title(Data sources):Outcome	Model class	Results/Factors associated with increased risk of child mortality
Mturi A.J. &Curtis S.L(1995)	Determinants of infant and child mortality in Tanzania(1991/92TDHS Data) Outcome: Hazard ratio	Cox PH Hazard model (maternal and socioeconomic factors	Mortality risk at age 1-59 months was significantly associated with partner's education and zone of residence. Uneducated partners faced increased risk of child mortality.
Zerai(1996)	Socio-economic determinants of infant mortality in Zimbabwe(1988 ZDHS Data) Outcome: Hazard ratio	Cox regression model	Women's average education level in their community exerts a greater influence on infant survival. Poor survival prospects are experienced in communities with low women average education levels.
Manda(1996)	Relationship between infant and child mortality and maternal factors in Malawi (1992 MDHS data) Outcome: Hazard ratio	Cox regression models	Birth interval and maternal age affects child mortality. Children born to both relatively old and young women have higher mortality rates than others.
Kembo and Ginneken(2009)	Determinants of infant and under-five mortality in Zimbabwe (used 2005-06 Zimbabwe DHS data) Outcome: Hazard ratio	MV proportional hazard models	Birth of order 6+ with short preceding interval had the highest risk of infant mortality Social economic variables did not have a distinct impact on child mortality
Uthman &Mubashir(2008)	Maternal determinants of child mortality in Nigeria(using 2003 Nigeria DHS data) Outcome: Hazard ratio	UV &MV survival regression models with weibull hazard function	Maternal education and household asset index was associated with lower risk of infant mortality. Multiple births were strongly negatively associated with lower risk of infant survival. Children from uneducated mothers and poor households had increased risk of child mortality.
Mutunga(2007)	Environmental determinants of child mortality in Kenya(using 2003 KDHS) Outcome: Hazard ratio	Cox PH models (The study used socioeconomic, Demographic, and Environmental variables)	Children born from youngest mothers and oldest women experience high risk due to biological factors. Better survival prospects were found in children born in wealthier families, household with electricity, household with access to safe drinking water and sanitation facilities, Those using less polluting fuels for cooking, and larger households.

Table 1 Models in the analysis of child mortality ;UV=Univariate and MV=multivariate (cont.)

Author	Title(Data sources):Outcome	Model class	Results/Factors associated with increased risk of child mortality
Mazbahul G. &Tashina K.(2009)	Examining the trend and annual rate of reduction in infant from 1998-2007 and correlates causal factors based on data from statistical year book of Bangladesh 2008 and sample vital registration system 2007 Outcome: Risk rate	Log-log specified ordinary least square and simultaneous quantile regression models	Infant immunization was the most significant factor that reduces infant mortality. Access to electricity, household with electricity had decreases infant mortality rates.
Mojekwu & Joseph Mnamdi(2012)	Environmental daterminants of child mortality in Nigeria(2008 NDHS) Outcome: Risk rate	Principal component analysis as a data reduction technique with Varimax rotation	Better survival prospects were found in home with high income, household that have access to immunization, those with sanitation facilities and those using low polluting fuels as their main source of cooking.
Raheem Usman A. &Segun-Agboola, B.T,(2009)	Exploring the social and environmental determinants of child health in Ilorin, Nigeria Outcome: Risk rate	Multiple logistic regression model	The study concluded that Health of children was considered in a typical urban Africa and residential quality as epitomised by availability of environmental services like kitchen, bathroom and toilet were more determinant of child health
Merimaaria V. & Teija K. (2000)	Antanatal and perinatal predictors of infant mortality in rural Malawi (used a cohort of 733 live born infants Outcome: Relative risk	UV and MV analysis used to determine relative risk	HIV epidemic was an important but not the main determinant of infant mortality. Maternal factors were the main determinants.

Table 1 Models in the analysis of child mortality; UV=Univariate and MV=multivariate (cont.)

Author	Title(Data sources):Outcome	Model class	Results/Factors associated with increased risk of child mortality
Halaabou-ali(2002)	The effect of water and sanitation on child mortality in Egypt(used 1995 DHS data) Outcome: risk rate	Three part model specification comprising discrete choice and transition models	Access to municipal water decreases the risk and sanitation was found to have a more pronounced impact on mortality than water
Cornelia K. and Ingo P.(2011)	Behavioural factors as emerging main determinants of child mortality in middle income countries: a case study of Jordan(2007 Jordan DHS data) Hazard rate: Risk rate	Logit model (biological, birth order, birth interval, breastfeeding, access to drinking water were some of the variables)	They observed higher mortality rates among infants and children who lack access to either safe drinking water or improved facilities.
Andreas J. Sian F.(2006)	Child mortality in rural Malawi(used 2000-2006 demographic surveillance system data in northern Malawi Risk rate: Relative risk	Poisson regression models (variables used are parental education, maternal factors)	Found loss in the usual gap in survival between the poor and the less poor because the less poor have been disproportionately affected by HIV rather than because of relative improvement in the survival of the poorest.
Sudhanshu H (2008)	Child mortality in the eastern and southern Africa (used DHS for Malawi (1997-2004), Mozambique (1997-2003), Tanzania (1996-2004) and Zambia (1992-2001))	MV regression model (The study used birth size, breastfeeding status, water source, and sanitation)	The largest protective factors found were parity, birth spacing and to a lesser degree household wealth.

1.2.3. Under-five children mortality studies

This section presents in more detail several studies that have been done in finding the factors that affect child mortality using various statistical methods.

In 1964, Bourgeois-Pichat identified two types of factors ‘endogenous’ and ‘exogenous’ that affect infant mortality. Exogenous factors of infant mortality are dependent on environment in which an infant is exposed and include deaths to infants due to infections, parasite and respiratory diseases. Such causes normally occur in the post-neonatal period (1 to 11 months of age of infant) and they are easier to control. On the other hand, endogenous causes of mortality are more biological in nature and include deaths due to congenital malformations and birth process. They occur in the neonatal period (less than 1 month of age of infant) and are rather difficult to control.

Gandotra and Das (1988) later categorised the underlying factors behind the immediate causes of infant deaths into five broad groups: demographic factors, socio-economic factors; environmental, sanitation and hygienic factors; nutrient availability factors; and medical care factors.

Pandey et.al. (1998) while analysing the NFHS-1 data considered child’s year of birth, child’s sex, mother’s age at child birth, residence, mother’s literacy, religion-caste/tribe membership, mothers exposure to mass media, availability of toilet facility, type of cooking fuel and ownership of goods scores as the covariates of infant mortality.

The social economic variables that have been used in previous as well as in this study include parental education levels, type of place of residence (urban/rural), Access to electricity, wealth status and household size. Socioeconomic variables as well as

household environmental variables play a major role in determining the survival of a child.

The 2004 MDHS data indicated that urban mortality rates are generally lower than rural rates; the under-five mortality rate is 116 per 1,000 in urban areas compared to 164 per 1,000 in rural areas. The MDHS compared the three regions in Malawi, and the Northern Region was found to have lower under-five mortality (120 per 1,000 live births), than either the Central (162 per 1,000) or the Southern Regions (164 per 1,000). Similarly, the infant mortality rate was lowest in the Northern Region (82 per 1,000), compared with either the Central Region (90 per 1,000) or the Southern Regions (98 per 1,000). These regional differences in mortality were also observed in the 1992 MDHS and the 2000 MDHS. The 2004 MDHS shows a relationship between mother's education and child survival as the 2000 MDHS. For every age interval, higher levels of education are generally strongly associated with lower mortality risks. The same is true for the wealth index.

Mturi, and Curtis (1995) investigated the determinants of infant and child mortality in Tanzania using the 1991/92 Tanzania Demographic and Health Survey. A hazards model was used to assess the relative effect of the variables hypothesized to influence under-five mortality. The study showed that there was a remarkable lack of infant and child mortality differentials by socioeconomic subgroups of the population, which may reflect post-independence health policy and development strategies. Mortality risk at age 1-59 months was significantly associated, partner's education, and zone of residence.

Zerai (1996) examined socio-economic and demographic variables in a multi-level framework to determine conditions influencing infant survival in Zimbabwe. He

employed Cox regression analysis to the 1988 Zimbabwe DHS data to study socioeconomic determinants of infant mortality. The unique finding was that women's average educational levels in their community exert a greater influence on infant survival than the mother's educational level.

Manda (1999) used data from the 1992 DHS in Malawi to study the relationship between infant and child mortality and birth interval, maternal age at birth and, birth order, with and without controlling for other relevant explanatory variables. He also investigated the direct and indirect (through its relationship with birth intervals) effects of breastfeeding on childhood mortality. The study employed proportional hazards models. The results show that birth interval and maternal age effects are largely limited to the period of infancy.

Manda further found that as the child increases in age, the influence of social and economic variables on the mortality risk is enhanced, and the relationship between bio-demographic variables and mortality risk is strengthened. The study further shows that breastfeeding status does not significantly alter the effects of preceding birth interval length on mortality risk, but does partially diminish the succeeding birth interval effect.

Kalipeni (1993) in his paper examines the spatial variation of infant mortality in Malawi between 1977 and 1987. Data from the 1977 and 1987 censuses were used in simple correlation and forward stepwise regression analysis to explain and/or predict the variation and change of infant mortality at district (county) level. The results indicated that, at the macro-level, the variation of infant mortality was strongly associated with a number of demographic and socioeconomic variables like Age at first marriage, total fertility rates. Female literacy rates, number of home craft centres

per 10,000 women and females in agriculture (%) were among the most important demographic and socioeconomic variables. For example, districts with high levels of literacy rates and educational facilities had lower rates of infant mortality. Districts with more females engaged in agriculture had higher levels of infant mortality. As a result, there are distinct regional differentials due to the inequitable distribution of educational opportunities, health care facilities, and non-agricultural activities. But even after controlling for these socioeconomic variables, regional differentials in infant mortality still persist. They also find that the region in which a district finds itself also played a role as far as levels of infant mortality were concerned.

Kalipeni (1993) further estimated several multivariate regressions to show how the relationship between region and child mortality changes when controls for other factors. To discern how the different variables of interest operate to affect mortality, the variables were introduced into the regression in stages. The first model only included dummy variables for the region the mother was living in at the time of the survey. This establishes regional differentiation of mortality. Due to the vast differences in educational infrastructure among the three regions, the second model tested whether the regional differentiation was explained by education. The third model adds controls for medical and health environment and the final model added other socioeconomic controls in attempts to explain away the differential.

The results of the bivariate and regression analyses showed some interesting relationships between the regional variation of child mortality and several of the variables. Without controls, there was a significant difference in child mortality between the North and the South, and between the North and the Centre. But not

between the South and the central controlling for education status, the difference between the North and the South were insignificant. This suggested that the differential in child mortality between the two regions is explained by education.

Kembo and Ginneken (2009) used multivariate proportional hazard models to determine the impact of maternal, socioeconomic and sanitation variables on infant and under- five child mortality in Zimbabwe. Results showed that births of order 6+ with a short preceding interval had the highest risk of infant mortality. Socioeconomic variables did not have a distinct impact on infant mortality. Determinants of child mortality were different in relative importance from those of infant mortality.

Uthman and Mubashir (2008) examined the relationship between multiple births and infant mortality in Nigeria. They used univariable and multivariable survival regression procedure with Weibull hazard function, controlling for child's sex, birth order, prenatal care, delivery assistance; mother's age at child birth, nutritional status, education level; household living conditions and several other risk factors. Maternal education and household asset index were associated with lower risk of infant mortality. They concluded that multiple births are strongly negatively associated with infant survival in Nigeria independent of other risk factors. Mother's education played a protective role against infant death. This evidence suggests that improving maternal education may be a key to improving child survival in Nigeria. A well-educated mother has a better chance of satisfying important factors that can improve infant survival: the quality of infant feeding, general care, household sanitation, and adequate use of preventive and curative health services.

Mutunga (2007) examined how infant and child mortality is related to the households environmental and socio-economic characteristics, such as mother's education, source of drinking water, sanitation facility, type of cooking fuels, and access to electricity. A hazard rate framework was used to analyse the determinants of child mortality. They applied duration models to the problem of child mortality since this class of models straightforwardly accounts for problems like right-censoring, structural modelling and time varying covariates which traditional econometric techniques cannot handle adequately. In this study, Mutunga found that of the demographic variables, children born of the youngest and oldest women experience high risk rates of death. All of these are mainly due to biological factors. As for the socio-economic variables, better survival prospects were found to exist for children born in wealthier families. Lower mortality rates had also been found in households with electricity. Household size was negatively related to child mortality, meaning that lower child survival prospects were experienced in smaller households. Similarly, environmental characteristics of households were found to be significantly related to child mortality. Lower mortality rate were experienced in households that have access to safe drinking water, those with access to sanitation facilities and those using low polluting fuels as their main source of cooking.

Kazembe and Mpetekula (2010) in their study quantifying spatial disparities in neonatal mortality used structured additive regression model and found that both fixed and spatial effects were associated neonatal mortality. The results showed that infants with birth weight above average (> 2500 grams), born as single tons, born of mothers who sought antenatal care and those whose mothers were all associated with lower probability of dying in the neonatal period. The effect of being a boy child, first born,

born in rural area, and born to mothers who attained primary education was positively associated with neonatal deaths.

Mazbahul and Tashina (2007) examine the trend and annual rate of reduction in infant mortality from 1998-2007 time periods and correlates causal factors based on different data from Statistical Yearbook of Bangladesh 2008 and Sample Vital Registration System 2007. Seven explanatory variables were considered and the log-log specified ordinary least square and simultaneous quantile regression models were employed to investigate and compare the stochastic impacts of these predictors on changing infant mortality. Infant immunization was the most effective factor that reduces infant mortality especially at lower quantile districts. Most notably, lower poverty line implies increasing trend with upper quantile, indicates that districts with low infant mortality rate has low effect for any positive rate of change of it. The least square as well as simultaneous quantile regression result disclose that share of population lived in electricity accessed houses, road density, no. of female per family planning personnel has potential and statistically significant impacts on infant mortality rate. Likewise, infant mortality decreased with the increased percentage of household having television.

Mojekwu and Nnamdi (2012) in their study environmental determinants of child mortality in Nigeria, used principle component analysis as a data reduction technique with varimax rotation to assess the underlying structure for sixty-five measured variables, explaining the covariance relationship amongst the large correlated variables in a more parsimonious way and simultaneous multiple regression for child mortality modelling in Nigeria. For purpose of robustness, a model selection

technique was implemented. Estimation from stepwise regression model showed that household environmental characteristics do have significant impact on mortality. Better survival prospects were found to exist in homes with high income. Lower mortality rates were experienced in household that have access to immunization, those with sanitation facilities, those with proper refuse and solid waste disposal facilities as well as those with good healthy roofing and flooring materials and those using low polluting fuels as their main source of cooking.

Usman. and Agboola (2009) in their paper “exploring the social and environmental determinants of child health in Ilorin, Nigeria, used a multiple regression model to analyse the nature and degree of explanation offered by each of the variables unsafe drinking water and poor sanitation and hygiene’ mother socio-economic characteristics. The study found inverse relationship between mother’s socio-economic characteristics and the health of their children. In this regard, mothers in informal occupation were shown to possess 23% likelihood of their children falling sick compared to mothers in other categories of occupation($r = 0.48$). On the other hand, children of illiterate mothers had 15.4%likelihood of diarrhoea occurrence than children of educated mothers ($r = 0.39$). In households with large family sizes, the likelihood was higher by 16.9% compared to household with smaller family sizes. The study concluded that health of children is considered in a typical urban Africa. The result shows that residential quality as epitomised by availability of environmental services like kitchen, bathrooms and toilet were more determinant of child health.

Omolaubi (1995) had earlier emphasized that children of educated women, especially if the latter had completed secondary education, have much higher survival chances than children of illiterate mothers. This is because maternal education works through three different pathways. It influences the socio-economic level of household, governs mothers' attitude and influences her behaviour (including health seeking) on issues relating to the health of their children.

Vaahtera and Kulmala(2000), in their study of antenatal and perinatal predictors of infant mortality in rural Malawi used a cohort of 733 live born infant from approximately 24 gestation weeks onwards. Univariate analysis was used to determine relative risks for infant mortality after selected antenatal and perinatal exposures. Multivariate modelling was used to control for potential confounders, The HIV epidemic was an important but not the main determinant of infant mortality. The infant mortality rate was 136 deaths/1000 live births. Among singleton newborns, the strongest antenatal and perinatal predictors of mortality were birth between May and July, maternal prim parity, birth before 38th gestation week, and maternal HIV infection. Theoretically, exposure to these variables accounted for 22%, 22%, 17%, and 15% of the population attributable risk for infant mortality, respectively.

Masangwi (2010) in their study of Behavioural and environmental determinants of child diarrhoea in Chikwawa Malawi, used a Bayesian logistic regression analysis to analyse domestic water sources, sanitation and hygiene practice and their impact on diarrhoea which is also one of the causes of under-five child death. Results showed that children from households with no toilet facilities were more likely to have suffered from diarrhoea than those who own such facilities. (Odds ratio: 1.72, 95%

CI: 1.18, 2.51). On the other hand, children from households that use private taps were less likely to have suffered diarrhoea than those that use public taps (OR=0.16, 95% CI: 0.08, 0.32). Those where each member use a basin (OR=0.37, 95% CI: 0.20, 0.70) or running water from a tap (OR=0.10, 95%CI: 0.02, 0.53) for washing hands were less likely to have suffered diarrhoea than those that use cups to pour water from containers.

Sudhanshu and Wen (2008) in their study child mortality in the eastern and southern Africa, provide an in depth analysis of the micro-level determinants of child survival in four eastern and southern Africa (ESA) countries over time: Malawi, Mozambique, Tanzania and Zambia. Two of these countries were thought to have made important gains in reducing mortality (Malawi and Mozambique), mortality increased in Tanzania in the late 1990s but then dropped in the early part of this decade, while in Zambia the pattern of child mortality during the 1990s was unclear. Each of these countries has two comparable national household surveys (DHS) at least five years apart which provide a sufficiently long enough window to observe sustained changes if they occurred. For each country they provided three sets of analyses. First, they pool the data and estimate survival functions with cohort effects to test whether the probability of survival has changed significantly over time. They estimated the change in the relative risk of death over time (the death hazard) as well as statistical tests for differences over time and by age cohort. Second, they estimated full survival models by age group to assess whether there are patterns in the main determinants of child survival across the countries. Finally, they quantify the change in child survival over the study period that is attributable to each of the variables included in the regression

model, thus providing a picture of some of the key drivers of mortality changes in the time period being studied.

Cornelia and Ingo (2011) run different logit estimations to allow for different set of variables. They concluded that once a country has passed a certain threshold in household income, education and access to health care and safe drinking water, policies targeting behavioural changes are the most promising for achieving further reductions in mortality rates. That is among the household and community characteristics, they observed higher mortality rates among infants and children who lack access to either safe drinking water or improved facilities.

Abou-ali (2002) assessed water and sanitation's impacts on child mortality in Egypt. The analysis was conducted using a three-part model specification, comprising discrete choice to model the child prospects of dying during the neonatal period. The remaining parts uses transition models to model infant and childhood risk of death where unobserved heterogeneity is accounted for. The results show that access to municipal water decreases the risk and sanitation is found to have a more pronounced impact on mortality than water. The results suggest that increasing awareness of the Egyptian population relative to health care and hygiene is an important feature to decrease child's mortality risk. Moreover, gender discrimination is found to be of an important effect beyond the neonatal period.

Jahn et.al (2010) in their study Child Mortality in rural Malawi, as part of a demographic surveillance system in northern Malawi in 2002-6, covering a population of 32,000, information was collected on socio-economic status of the

households. Deaths were classified as HIV/AIDS-related or not by verbal autopsy. Poisson regression models were used to assess the association of socio-economic indicators with all-cause mortality, AIDS-mortality and non-AIDS mortality among children. There were 195 deaths in infants, 109 in children aged 1–4 years, and 38 in children aged 5–15. All-cause child mortality in infants and 1–4 year olds was similar in households with higher and lower socio-economic status. In infants 13% of deaths were attributed to AIDS, and there were no clear trends with socio-economic status for AIDS or non-AIDS causes. For 1–4 year olds 27% of deaths were attributed to AIDS. AIDS mortality was higher among those with better built houses, and lowest in those with income from farming and fishing, whereas non-AIDS mortality was higher in those with worse built houses, lowest in those with income from employment, and decreased with increasing household assets. They concluded that in this population, since HIV infection among adults was initially more common among the less poor, childhood mortality patterns have changed. The usual gap in survival between the poor and the less poor has been lost, but because the less poor have been disproportionately affected by HIV, rather than because of relative improvement in the survival of the poorest.

Mohamad and Ebrahim (2007) compared two survival regression methods – Cox regression and parametric models - in patients with gastric adenocarcinomas who registered at Taleghani hospital. They retrospectively studied 746 cases from February 2003 through January 2007. Gender, age at diagnosis, family history of cancer, tumor size and pathologic distant of metastasis were selected as potential prognostic factors and entered into the parametric and semi parametric models. Weibull, exponential and lognormal regression were performed as parametric models with the Akaike Information Criterion (AIC) and standardized of parameter estimates to compare the

efficiency of models. The survival results from both Cox and Parametric models showed that patients who were older than 45 years at diagnosis had an increased risk for death, followed by greater tumor size and presence of pathologic distant metastasis. In multivariate analysis Cox and Exponential were similar. Although it seems that there may not be a single model that is substantially better than others, in univariate analysis the data strongly supported the log normal regression among parametric models and they concluded that it can be lead to more precise results as an alternative to Cox.

Kourosh and Mohammad (2008) evaluated the prognostic factors of overall survival after haematopoietic stem cell transplant in acute lymphoblastic leukaemia patients using accelerated failure time (AFT), Cox proportional hazard (PH) and Cox time varying coefficient model. In this study, the predicted power of weibull AFT models was superior to Cox PH model and Cox with time varying coefficients. Cox-Snell residual show weibull AFT fitted to data better than the other distributions in multivariate analysis and they concluded that AFT distribution can be useful tool for recognising prognostic factors of overall survival.

From the review of the literature above, it is clear that there are few studies that used parametric survival models in determining infant and child mortality in Malawi as well as other countries. It is against this background that in this study different parametric models have been used and their performance been compared in order to find the best fitted model used in determining factors that affect child survival. Our results will offer an in-depth use of DHS data and are expected to improve the understanding of the mortality situation of under-five children in Malawi. Since the

best model has been established, we expect the results to be more accurate, thereby be of interest to people working on other studies that analyse child mortality risks in Malawi.

1.3 Problem Statement

Malawi infant and under-five mortality is still high though there is evidence of a sharp decline in the infant and under-five mortality (2006 MICS survey). Accurate information on cause of death is lacking, what is mostly known is that the leading cause of under-five mortality in Malawi is neonatal conditions, pneumonia, diarrhoea, malaria, AIDS and malnutrition (UNICEF Malawi 2010). As such, many studies on child mortality have concentrated on disease as the main cause of child mortality. If under-five mortality has to be reduced, there is a great need to find all other factors that may cause under-five mortality. This will guide as to which interventions to focus on in order to reduce child mortality.

Similarly most child survival studies have employed Cox proportional hazard model. If we have to come up with effective and efficient results which can guide policy, there is a great need to identify the suitable model which can be used in the analysis of child mortality. Hence, the need to compare the performance of Cox PH model and parametric models.

1.4 Objectives of the study

1.4.1 Broad Objective

The general aim of the study was to compare the performance of Cox Proportional hazard model and parametric models like exponential, Weibull, Gompertz, Log-logistic, and log-normal model in modelling child mortality

1.4.2 Specific objectives

1. To identify social economic and household environmental factors that influence under-five child mortality using Cox proportion hazard model.
2. To Identify the best fitting model between parametric and semi parametric and among the parametric models.
3. To identify social economic and household environmental factors that influence under-five child mortality using five parametric survival models

1.5 Justification of the study

Since children are basic for every development aspect of a country there is a need to: identify the best model that determines factors that affect under-five mortality. Although the Cox parameter estimations are well known to the researchers in the field of medical sciences, the results in accelerated failure times can be interpreted as the relative risk that is known to medical scientists. Thus, these parameters can be interpreted as factors accelerating or decelerating similarly in the interpretation of cox' hazard ratio.

This study is expected to contribute to methodological innovation in infant and child mortality studies in Malawi by introducing parametric survival analysis into child mortality modelling. Survival models are the most suited for such analysis because they account for problems like right-censoring, and structural modelling which traditional econometric techniques cannot handle adequately (Mutunga 2004).

The results of this study are expected to shed light on the linkage between the household's environmental status and child mortality, and consequently inform policy on the importance of improving households' environmental and socio-economic characteristics in a bid to reduce child mortality. This is in line with the current government's effort of mainstreaming the environment into sustainable development planning and commitment to the achievement of the Millennium Development Goals

1.6. Definitions of terms

- Neonatal mortality : the rate of dying within the first month of life
- Infant mortality : The rate of dying between births and the first birthday
- Child mortality: The rate of dying between exact ages one and five
- Under-five mortality: rate of dying between birth and the fifth birthday
- Wealth index: proxy measure of the wealth of households which is based on household characteristics, ownership of assets (house ownership, source of drinking water, electricity, sanitation facility (toilet), floor material type, roof material type etc.)
- Household: This is a social group of one or more individual members. They are usually but not always related.

Chapter 2: Methods and Data sources

2.1. Data

The analysis of child mortality used data from the 2010 Malawi Demographic and Health survey (MDHS). The 2010 Malawi Demographic Health Survey (2010 MDHS) was implemented by the National Statistical Office (NSO) and the community Health Sciences Unit (CHSU) from June through November 2010.

2.1.1 Sample design

The 2010 MDHS sample included 849 clusters: 158 in urban areas and 691 in rural areas. A complete listing of households was done in each of the MDHS clusters from May to June 2009. A minimum sample size of 950 households was required per district to provide an acceptable level of precision for the indicators measured in the survey. The survey interviewed a representative sample of 19,967 women aged between 15 and 49 years. A two stage stratified sampling design was implemented to collect the data. The data were realized through a questionnaire that included questions on marriage and reproductive histories of which histories of all birth they ever had, area of residence, age of the mother, type of sanitation facilities, whether the child is alive or not, age at death and many more were gathered (Appendix 3 shows the relevant pages of the questionnaire used in this study). Survival time of each child was computed in months, all children whose survival time was less than 60 months were classified as under-five deaths. All children above 60 months were censored.

2.1.2. Limitations

The Data collected using the birth histories in the 2010 MDHS were subject to a number of potential errors. First, the data reflect only surviving women age 15-49 years; no data were available for children of women who died. To the extent that child

mortality of surviving and non-surviving women differs substantially and that young children of non-surviving women make up a good portion of all young children, the mortality indicators derived from the birth history would be biased.

Another possible error is underreporting of events; respondents may forget events that occurred in the more remote past. Omission of infants' deaths may take place, especially in cases where deaths occur early in infancy. There was also a potential of forgetting the death of neonates and of those who had born recently due to cultural emotional response which does not regard the young baby as having lived. If such deaths are selectively reported, consequences will not only be a lower infant mortality rate (IMR) and neonatal mortality rate (NNMR), but also a low ratio of neonatal deaths to infant deaths. On the other hand, mis-reporting of the date of birth and age at death would sometimes result in distortion of the age pattern of death. This may affect the final indices obtained because of shifting of ages above or below the cut-offs for the different mortality categories. Another aspect that affects the childhood mortality estimates is the quality of reporting of age at death. Here they just reported the age and not month and year or exact date when the child died. To minimise errors in the reporting of age at death, the interviewers were instructed to record the age at death in days if the death took place within one month after birth, in months if the child died within 24 months, and in years if the child was two years or older. In general if ages at death are misreported, it may bias the estimates, especially if the net effect of age misreporting results in transference of deaths from one age bracket to another. For the purpose of the analysis, the age at death were imputed into months.

2.2. Variables considered

Mosley and Chen (1984) were among the first to study the intermediate biomedical factors affecting child mortality, labelled ‘proximate determinants’ They distinguished fourteen proximate determinants and categorized them into four groups: maternal [fertility] factors, environmental sanitation factors, availability of nutrients to the foetus and infant, injuries, and personal illness control factors.

Independent variables

Based on the Mosley and Chen (1984) determinants of under-five child mortality framework, the independent variables that were studied in this research were:

Socioeconomic variables: (mother education, Father Education, wealth index, area of residence, household size and Access to electricity)

Household environmental /Sanitation variables: (source of drinking water, toilet facility and source of cooking fuel).

The Response Variable

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

2.2.1 Measurement of variables

Wealth index was calculated by the MDHS on the basis of ownership of household assets. Wealth index had five categories, which are richest, rich, medium, poor and poorest. The poor status category was used as the reference category.

Mother and partner education, had four categories; no education, primary education, secondary education and higher education. In this analysis those with no education and with primary education were put in one category and secondary and higher education were also recoded in one category. This was done to balance the sample size as we have highest education with smallest sample size.

In this study, households that have either a flush toilet or a pit latrine, whether private or shared are regarded as having sanitation as opposed to those without any facility. Similarly, households with access to private or public tap water, as well as covered well water are considered to have safe water.

Area of residence had two categories; rural and urban. Similarly, access to electricity had two categories those with electricity and those without. Using electricity and different kinds of gas were considered as using less polluting fuel. On the other hand those using wood, charcoal were regarded as using high polluting source of fuels for cooking. The categorical predictor household size had four levels; households with two to three members, household with four to five members, household with six to seven members and household with eight or more members. And household with two to three members was used as a reference category. The variables and codes used in the analysis of this study are provided in the Table 2 below.

Table 2 variable description and codes for categories

Variable	Description	Codes/Values for categories
V025	Type of place of Residence	0= urban*, 1= rural
highedu	Mother education	0 = no or primary education*, 1= secondary or Higher
Fuel	Type of cooking fuel	0=Electricity or gas*, 1=firewood or charcoal
partedu	Partner education	0 = no or primary education*, 1= secondary or Higher
Water2	Source of drinking water	0= piped water*, 1=Well water
Toilet1	Type of toilet facility	0= flush or pit latrine*, 1=no toilet facility
Hsize	Household size	0= 2 to 3 members* 1= 4 to 5 members, 2= 6 to 7 members, 3=8 to 23 members
Status	Wealth index	0 = poor 1= poor* 2 = medium 3 = rich 4 = richest
elect	Access to electricity	0= No*, 1= yes

Note: * were used as the reference categories)

2.3 Theoretical Model

This section describes the statistical models that were used, and have been implemented elsewhere, to study infant and child mortality. Section 2.3.1 is dedicated to distributional properties of time-to-event data and discusses the concept of survival and hazard function. Section 2.3.3 and 2.3.4 discuss the Cox PH and parametric approaches, respectively, to modelling the relationship between child mortality and the covariate identified.

This study employed survival analysis. Survival models relate the time that passes before some event occurs to one or more covariates that may be associated with that quantity. The main concepts of which are the hazard function and the survivor

function. The underlying hazard function, often denoted $h_0(t)$, described how the hazard (risk) changes over time at baseline levels of covariates; and the effect parameters, describing how the hazard varies in response to explanatory covariates. The hazard rate (the term was first used by Barlow (1963) is defined as the probability per time unit that a case that has survived to the beginning of the respective interval will fail in that interval. Specifically, it is computed as the number of failures per time units in the respective interval, divided by the average number of surviving cases at the mid-point of the interval (Kay R, 2004).

2.3.1 Survival time distribution

Let T be a random variable denoting the survival time. The distribution of survival times is characterized by any of three functions: the survival function, the probability density function or the hazard function.

The survival function defined as the probability that the survival time is greater or equal to t .

$$S(t) = P(T \geq t), t \geq 0 \quad (2.1)$$

The failure function $F(t)$ is given as

$$F(t) = P(T \leq t) = \int_0^t f(x)dx, \quad (2.2)$$

where t is the elapsed time since entry into the study (child was born) at time 0 and $f(\cdot)$ is the probability density function of T .

Thus, we can express the survival function in terms of the failure function as

$$S(t) = pr(T > t) = 1 - F(t) \quad (2.3)$$

The survivor function $S(t)$ and the failure function $F(t)$ are each probability, and therefore inherit the properties of probabilities. The survivor function lies between zero and one, and is a non-increasing function of t . The survivor function is equal to one at the start of the spell ($t = 0$)

and is zero at infinity.

Closely related is the concept of hazard rate, which is given as:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \delta t | T > t)}{\delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}. \quad (2.4)$$

There is a one-to-one relationship between a specification for the hazard rate and the survivor function, which after some manipulation is given as:

$$S(t) = \exp[-H(t)] \quad (2.5)$$

Where

$$H(t) = \int_0^t h(u) du = \ln[s(t)] \geq 0$$

$H(t)$ is referred to as the cumulative hazard function or integrated hazard function.

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

2.3.2 Ordinary least squares

Ordinary least squares are used in survival data analysis. However, they cannot handle censoring/and truncation, time varying covariates and structural modeling (Stephen 2005).

To illustrate the (right) censoring issue, let us suppose that the “true” model is such that there is a single explanatory variable, X_i for each individual $i = 1, \dots, n$, who has a true survival time of T_i^* . In addition, in the population, a higher X is associated with a shorter survival time. In the sample, we observe T_i where $T_i = T_i^*$ for observations with completed spells, and $T_i < T_i^*$ for right censored observations.

Suppose too that the incidence of censoring is higher at longer survival times relative to shorter survival times. (This does not necessarily conflict with the assumption of independence of the censoring and survival processes .it simply reflects the passage of time. The longer the observation period, the greater the proportions of spells for which events are observed.)

By OLS, we mean: regress T_i , or better still $\log T_i$ (noting that survival times are all non-negative and distributions of survival times are typically skewed), on X_i , fitting the linear relationship

$$\log(T_i) = \alpha + \beta x_i + e_i$$

The OLS parameter estimates are the solution to $\min_{a,b} \sum_{i=1}^n (e_i)^2$. $\hat{\alpha}$ is the vertical intercept; $\hat{\beta}$ is the slope of the least squares line.

2.3.3. The Kaplan-Meier estimate of the survival function

The Life table is the earliest statistical method to study human mortality rigorously, but its importance has been reduced by the modern methods, like the Kaplan-Meier (K-M) method. The K-M estimator for the survival curves is usually used to analyse individual data, whereas the life table method applies to grouped data.

Suppose that r individuals have failures in a group of individuals.

$$\text{Let } 0 \leq \dots < t_{(1)} < \infty$$

be the observed ordered death times. Let r_j be the size of the risk set at t_j , where risk set denotes the collection of individuals alive and uncensored just before t_j ,

Let d_j be the number of observed deaths at t_j , $j = 1, \dots, r$

Then the K-M estimator of $s(t)$ is defined by

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

This estimator is a step function that changes values only at the time of each death.

2.3.4. The Cox proportional Hazards model

The Cox proportional hazard model is given by:

$$h(t|x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = h_0(t) \exp(\beta'_x) \quad (2.6)$$

Where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero. $X = x_1, x_2, \dots, x_p$ is the value of the vector of explanatory variables for a particular individual, and $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ is the vector of regression coefficients.

The corresponding survival functions are related as follows:

$$S(t|x) = S_0(t) \exp(\sum_{i=1}^p \beta_i x_i) \quad (2.7)$$

This model, also known as the Cox regression model, makes no assumptions about the form of $h_0(t)$ (non-parametric model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model. The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients β , hazard ratios, and adjusted hazard curves.

The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates x and x^* is

$$\widehat{HR} = \frac{h_0 t \exp(\hat{\beta}' x)}{h_0 t \exp(\hat{\beta}' x^*)} = \exp(\sum \hat{\beta}' (x - x^*)) \quad (2.8)$$

This hazard ratio assumes that covariate effects act proportionally on the baseline hazard function (the values of the functions when all covariates are set to 0), independent of time. This is why this model is called the proportional hazards model. The proportional hazard means that the risk of death at any given time for an

individual in one group is proportional to the risk at that time for similar individual in the other group.

2.3.5 Partial likelihood estimate for Cox proportional hazards model

Fitting the Cox proportional hazards model, we wish to estimate $h_0(t)$ and β . One approach

is to attempt to maximize the likelihood function for the observed data simultaneously with respect to $h_0(t)$ and β . A more popular approach is proposed by Cox [13] in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β . Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters ($h_0(t)$ in the Cox PH model). In this section, we will construct the partial likelihood function based on the proportional hazards model.

Let $t_1, t_2, \dots, t_n$ be the observed survival time for n individuals. Let the ordered death time of r individuals be $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ and let $R(t_{(j)})$ be the risk set just before $t_{(j)}$ and r_j for its size. So that $R(t_{(j)})$ is the group of individuals who are alive and uncensored

at a time just prior to $t_{(j)}$. The conditional probability that the i th individual dies at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ dies at $t_{(j)}$ is

$P(\text{individual } i \text{ dies at } t_{(j)} \mid \text{one death from the risk set } R(t_{(j)}) \text{ at } t_{(j)})$

$$= \frac{p(\text{individual } i \text{ dies at } t_{(j)})}{p(\text{One death at } t_{(j)})}$$

$$= \frac{p(\text{individual } i \text{ dies at } t_{(j)})}{\sum_{K \in R(t_{(j)})} P(\text{individual } k \text{ dies at } t_{(j)})}$$

$$\begin{aligned}
&\cong \frac{P\{\text{individual } K \text{ dies at } (t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t}{\sum_{K \in R(t_{(j)})} P\{\text{individual } k \text{ dies at } (t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t} \\
&= \frac{\lim_{\Delta t \rightarrow 0} P\{\text{individual } i \text{ dies at } (t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t}{\lim_{\Delta t \rightarrow 0} \sum_{K \in R(t_{(j)})} P\{\text{individual } k \text{ dies at } (t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t} \\
&= \frac{h_i(t_{(j)})}{\sum_{K \in R(t_{(j)})} h_k(t_{(j)})} \\
&= \frac{h_0(t_{(j)}) \exp(\beta' x_i(t_{(j)}))}{\sum_{K \in R(t_{(j)})} h_0(t_{(j)}) \exp(\beta' x_k(t_{(j)}))} \\
&= \frac{\exp(\beta' x_i(t_{(j)}))}{\sum_{K \in R(t_{(j)})} \exp(\beta' x_k(t_{(j)}))}
\end{aligned}$$

Then the partial likelihood function for the Cox PH model is given by

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta' x_i(t_{(j)}))}{\sum_{K \in R(t_{(j)})} \exp(\beta' x_k(t_{(j)}))} \quad (2.9)$$

In which $x_i(t_{(j)})$ is the vector of covariate values for individual i who dies at $t_{(j)}$.

Note that this likelihood function is only for the uncensored individuals. The partial likelihood is valid when there are no ties in the dataset. That means there is no two subjects who have the same event time.

2.4 Parametric PH models

The parametric proportional hazards model is the parametric versions of the Cox proportional hazards model. It is given with the similar form to the Cox PH models.

The hazard function at time t for a particular individual with a set of p covariates $x_1, x_2, \dots, x_p$ is given as follows:

$$H(t|x) = h_0(t)\exp(\beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p) = h_0(t)\exp(\beta'_x) \quad (2.10)$$

The key difference between the two kinds of models is that the baseline hazard function is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data, whereas the Cox model has no such constraint. The coefficients are estimated by partial likelihood in Cox model but maximum likelihood in parametric PH model.

Partial likelihood differs from maximum likelihood because it does not use the likelihoods for all subjects, it only considers likelihoods for subjects that experience the event and it considers subjects as part of the risk set until they are censored

Other than this the two types of models are equivalent. Hazard ratios have the same interpretation and proportionality of hazard is still assumed. A number of different parametric PH models may be derived by choosing different hazard functions. The commonly applied models are exponential, Weibull or Gompertz model.

2.4.1 Weibull PH model

The generalization of the exponential distribution to include the shape parameter is the Weibull distribution. The cumulative distribution function of the Weibull distribution is

$$F(t) = 1 - \exp\{-\theta t^\gamma\}, t > 0 \quad (2.11)$$

where θ is the shape parameter and γ is the scale parameter, and the probability density

function of the Weibull distribution is

$$f(t) = \gamma \theta t^{\gamma-1} \exp\{-\theta t^\gamma\}, t > 0 \quad (2.12)$$

The survival function and hazard function of the Weibull distribution are

$$s(t) = \exp\{-\theta t^\gamma\},$$

$$h(t) = \gamma \theta t^{\gamma-1}$$

respectively

It is easy to see just how flexible the Weibull distribution can be. When $\gamma=1$, the Weibull distribution becomes the exponential distribution with $\theta = \lambda$ and the hazard rate remains constant as time increases, For $3 \leq \gamma \leq 4$, it is close to the normal distribution and when γ is large, say $\gamma \geq 10$ it is close to the smallest extreme value distribution (Nelson, 1982). When $\gamma > 1$ the hazard rate increases as time increases, and for $\gamma < 1$ the hazard rate decreases. Under the weibull PH model, the hazard function of a particular individual with covariates $(x_1, x_2, \dots, x_p)$ is given by:

$$h(t|x) = \theta \gamma(t)^{\gamma-1} \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_p x_p) = \theta \gamma(t)^{\gamma-1} \exp(\beta'x). \quad (2.13)$$

In this case, the survival time of this child has the weibull distribution with shape parameter $\theta \exp(\beta'x)$ and scale parameter γ . Therefore the weibull family with fixed γ possesses PH property. This shows that the effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains

constant. The Weibull model nests the exponential model. We used the Weibull model to test if the exponential model is appropriate.

2.4.2 Exponential PH model

The exponential PH model is a special case of the weibull model when $y=1$. The hazard function under this model is to assume that it is constant over time. This implies that the conditional ‘probability’ that a child with covariate values x_j dies given that the child survived to the beginning of the interval is constant over time (and that events occur according to a Poisson process). In other words, the risk of an event occurring (child dying due to environmental factors) is flat with respect to time assuming an exponential distribution for survival time, the hazard function is actually constant. Modelling the dependency of the hazard rate on covariates entails constructing a model that ensures a non-negative hazard rate (or non-negative expected duration time). The exponential density function is

$$f(t) = \lambda \exp(-\lambda t) \text{ for } \lambda > 0 \quad (2.14)$$

$$\text{and } t > 0$$

It has a constant hazard

$$h(t) = \lambda$$

and its survival function is

$$s(t) = \exp(-\lambda t)$$

Thus, a large λ implies a high risk and a short survival. Conversely, a small λ indicates a low risk and a long survival. This distribution has the memoryless property meaning that how long an individual has survived does not affect its future survival (Lee, 1992). The exponential distribution is limited in applicability because it has only

one parameter, the scale parameter λ . By adding a shape parameter the distribution becomes more flexible and can fit more kinds of data.

The exponential model is the simplest of the parametric survival models because it assumes that the baseline hazard is constant, (Lawless, 2003).

$$\begin{aligned}
 h(t|x_j) &= h_0(t) \exp(x_j \beta_x) \\
 &= \exp(\beta_0 \exp(x_j \beta_x)) \\
 &= \exp(\beta_0 + x_j \beta_x)
 \end{aligned}
 \tag{2.15}$$

For some constant β_0 , the notation β_0 has been used to emphasize that the constant may also be thought of as an intercept term from the linear predictor. Using the well-known relationships for the exponential model,

$$H(t|x_j) = \exp(\beta_0 + x_j \beta_x) t \tag{2.16}$$

$$S(t|x_j) = \exp\{-\exp(\beta_0 + x_j \beta_x) t\} \tag{2.17}$$

Therefore, under the exponential PH model, the hazard function of a particular child is given by:

$$h(t|x) = \lambda \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = \lambda \exp(\beta' x) \tag{2.18}$$

3.4.2.1 Maximum likelihood estimator

Here we are interested in fitting a model to the data. We will use the MLE to find the most likely parameters of our model. In this Section, we show how to fit the exponential model to the data.

The parameter that we want to estimate is $\beta = (\beta_0, \beta_1) \in R^2$. We first need to find the log-likelihood function:

$$f_\tau(t) = \frac{1}{\exp(\beta_0 + \beta_1 x_i)} \exp\left(\frac{-t}{\exp(\beta_0 + \beta_1 x_i)}\right), \quad t > 0, \quad i = 1, \dots, n$$

So

$$l(\beta_0, \beta_1) = \log\left(\prod_{i=1}^n f_\tau(t)\right) = \sum_{i=1}^n \left[-(\beta_0 + \beta_1 x_i) - \frac{t_i}{\exp(\beta_0 + \beta_1 x_i)}\right]$$

We will see that there is a special case in solving the partial derivatives equations which gives a closed form, and there is no closed form in the general case

A special case : $\beta_1 = 0$

If we assume that β_1 is fixed to 0, there is a closed form for $\hat{\beta}_0$, given by:

$$\frac{dl}{d\beta_0}(\beta_0) = 0 \leftrightarrow \sum_{i=1}^n \frac{t_i}{\exp(\beta_0)} = n \leftrightarrow \sum_{i=1}^n \frac{t_i}{\exp(\beta_0)} = n \leftrightarrow \beta_0 = \log(\bar{t})$$

General case : $\beta_1 \neq 0$

In order to find the maximum of the log-likelihood function, we try to set the partial derivatives to 0:

$$\left\{ \begin{array}{l} \frac{dl}{d\beta_0}(\hat{\beta}) = 0 \\ \frac{dl}{d\beta_1}(\hat{\beta}) = 0 \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} -n + \sum_{i=1}^n \frac{t_i}{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)} \\ \sum_{i=1}^n \frac{x_i t_i}{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)} - \sum_{i=1}^n x_i = 0 \end{array} \right\} = 0$$

We note that we have no closed form for the solution of the system of equations.

Hence, an iterative method to find the solution can be used. We propose Newton's method. For more information (see Davison 2006).

2.4.3 Gompertz PH model

The Gompertz model is available only in PH metric and assumes a baseline hazard

$$h_0(t) = \exp(\theta t) \exp(\beta_0) \quad (2.19)$$

So that the PH model

ard

$$\begin{aligned} h(t|x_j) &= h_0(t) \exp(x_j \beta_x) \\ &= \exp(\theta t) \exp(\beta_0 + x_j \beta_x) \end{aligned} \quad (2.20)$$

The survival and hazard function of the Gompertz distribution are given by

$$s(t) = \exp\left(\frac{\lambda}{\theta} (1 - e^{\theta t})\right), \quad h(t) = \lambda \exp(\theta t), \quad (2.21)$$

For $0 \leq t < \infty$ and $\lambda > 0$. The parameter θ determines the shape of the hazard function.

When $\theta = 0$, the survival time then have an exponential distribution, i.e. the

exponential distribution is also a special case of the Gompertz distribution. Like the

weibull hazard function, the Gompertz hazard increases or decreases monotonically.

For the Gompertz distribution, $\log(h(t))$ is linear with t .

Under the Gompertz PH model, the hazard function of a particular child is given

$$h(t|x) = \lambda \exp(\theta t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = \lambda \exp(\beta' x) \exp(\theta t) \quad (2.22)$$

It is straightforward to see that the Gompertz distribution has the PH property. But the

Gompertz PH model is rarely used in practice.

2.5 Accelerated Failure Time model(AFT)

Although parametric PH models are very applicable to analyse survival data, there are

relatively few probability distribution for the survival time that can be used with these

models. In these situations, the accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models, we measure the direct effect of the explanatory variable on the survival time instead of hazard, as we do in the PH model. This characteristic allows for easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. Currently, the AFT model is not commonly used for the analysis of clinical trial data, although it is fairly common in the field of manufacturing. Similar to PH model, the AFT model describes the relationship between survival probabilities and a set of covariates.

The members of the AFT model class include the exponential AFT model, Weibull AFT model, log-logistic AFT model, log-normal AFT model and gamma AFT model.

Accelerated failure-time models, also known as accelerated–time models or $\ln(\text{time})$ models, follow the parameterization

$$\ln(t_j) = x_j\beta_x + \epsilon_j, \epsilon_j \sim \text{Oddly, but not odd given the context}$$

In this case β_x are the coefficients on x and ϵ_j is the error term.

The word “accelerated” is used in describing these models because rather than assuming that failure t_j is exponential, Weibull, or some other form- a distribution is instead assumed for

$$\tau_j = \exp(-x_j\beta_x)t_j$$

and $\exp(-x_j\beta_x)$ is called the accelerated parameter. If $\exp(-x_j\beta_x) = 1$, then $\tau_j = t_j$, and time passes its “normal” rate. If $\exp(-x_j\beta_x) > 1$, then time passes more quickly for the subject (time is accelerated), so failure would be expected to occur sooner. If $\exp(-x_j\beta_x) < 1$, then time passes more slowly for the subject (time is decelerated), so

failure would be expected to occur later. The random quantity $\ln(\tau_j)$ has a distribution determined by what is assumed about the distribution of τ_j , and in the usual nomenclature of these models, it is the distribution of τ_j that is specified (Cleves, 2010).

2.5.1 Log-Logistic model

One limitation of the Weibull hazard function is that it is a monotonic function of time. However, the hazard function can change direction in some situations. The log-logistic distribution is an accelerated failure time model and has a hazard function which can be non-monotonic with respect to time, but if $k > 1$ the hazard has a single mode whereby there is initially an increasing hazard followed by a decreasing hazard. The log-logistic survival and hazard function are given by

$$s(t) = \{1 + e^{\theta} t^k\}^{-1}, \quad (2.23)$$

$$h(t) = \frac{e^{\theta} k t^{k-1}}{1 + e^{\theta} t^k} \quad \text{for } 0 \leq t < \infty, k > 0 \quad (2.24)$$

Where θ and k are unknown parameters and $k > 0$. When $k \leq 1$, the hazard rate decreases monotonically and when $k > 1$, it increases from zero to a maximum and then decreases to zero.

In the AFT metric,

$$\tau_j = \exp(-x_j \beta_x) t_j,$$

$$\tau_j \sim \text{loglogistic}(\beta_0, \gamma)$$

τ_j is distributed as loglogistic with parameters (β_0, γ) with cumulative distribution function

$$F(\tau) = 1 - \left[1 + \{\exp(-\beta_0) \tau\}^{\frac{1}{\gamma}} \right]^{-1} \quad (2.25)$$

Thus

$$\begin{aligned}\ln(t_j) &= x_j\beta_x + \ln(\tau_j) \\ &= \beta_0 + x_j\beta_x + u_j\end{aligned}$$

Where u_j follows a logistic distribution with mean 0 and standard deviation $\pi\gamma/\sqrt{3}$. As a result,

$$E\{\ln(t_j) | x_j\} = \beta_0 + x_j\beta_x$$

We can also derive the AFT formulation by accelerating the effect of time on survival experience. At baseline values of the covariates x , $T_j = t_j$ because all covariates are equal to zero. Thus the baseline survivor function of t_j is obtained from (2.24) to be

$$S_0(t_j) = [1 + \{1 + \{\exp(-\beta_0) t_j\}^{\frac{1}{\gamma}}\}]^{-1}$$

In the AFT model, the effect of the covariates is to accelerate time by a factor of $\exp(-x_j\beta_x)$. Thus for the AFT model,

$$\begin{aligned}S(t_j | x_j) &= S_0\{\exp(-x_j\beta_x) t_j\} \\ &= [1 + \{\exp(-\beta_0) \exp(-x_j\beta_x) t_j\}^{\frac{1}{\gamma}}]^{-1} \\ &= [1 + \{\exp(-\beta_0 - x_j\beta_x) t_j\}^{\frac{1}{\gamma}}]^{-1}\end{aligned}$$

The log logistic distribution closely resembles the lognormal distribution. Like log-normal model, the log logistic model has no natural PH interpretation. One advantage of the log-logistic model over the lognormal model is that the log logistic model has simpler mathematic expressions of the hazard and survivor function. If $\gamma < 1$, the log logistic hazard increases and then decreases. If $\gamma \geq 1$, then the hazard is monotone decreasing.

2.5.2 Log-normal model

This model has no PH interpretation. As such it is interpreted in the AFT metric (Cleves, 2010). For the lognormal regression model, it is assumed that

$$\tau_j \sim \text{lognormal}(\beta_0, \sigma)$$

τ_j is distributed as Lognormal with parameters (β_0, σ) with cumulative distribution functions

$$F(\tau) = \varphi\left(\frac{\ln \tau - \beta_0}{\sigma}\right)$$

where $\varphi()$ is the cumulative distribution function for the standard Gaussian (normal) distribution. Thus

$$\ln(t_j) = x_j \beta_x + \ln(\tau_j) \quad (2.26)$$

$$= \beta_0 + x_j \beta_x + u_j$$

where u_j follows a standard normal distribution with mean 0 and standard deviation σ .

That is, for the lognormal model, transforming time into $\ln(\text{time})$ converts the problem into simple linear regression (with possible censoring). As a result,

$$E\{\ln(t_j) | x_j\} = \beta_0 + x_j \beta_x \quad (2.27)$$

We can also derive the AFT formulation by accelerating the effect of time on survival experience at baseline, where all covariates are equal to zero. Thus the baseline survivor function of t_j is obtained as

$$S_0(t_j) = 1 - \varphi\left(\frac{\ln t_j - \beta_0}{\sigma}\right) \quad (2.28)$$

The attractive feature (for some problems) of this distribution is that the hazard function is non-monotonic in that it increases and then decreases.

2.6. Time dependence properties of the hazard function

The hazard rate is the probability of experiencing an event at time t_i . The precise shape of the hazard rate-the way it changes with time- is likely to vary from one situation to the next. For example, the hazard rate might increase with time in some situations:

$$\frac{dh(t)}{dt} > 0$$

This means that the risk of an event occurring (child dying) increases with time. This type of situation exhibits what is often referred to as positive duration dependence.

The hazard rate might decrease with time in other situations:

$$\frac{dh(t)}{dt} < 0$$

This means that the risk of an event occurring (child dying) decreases with time. This type of situation exhibits what is often referred to as negative duration dependence.

The hazard rate might be constant across time in still other situations:

$$\frac{dh(t)}{dt} = 0$$

This means that the risk of an event occurring is constant over time. There are also other situations with more complicated hazards rates that increase and decrease over time or that increase or decrease at faster or slower rates. Exactly how the hazard rate varies with time is generally referred to as time dependency.

The main property of parametric survival models is that they assume a particular shape for the hazard rate. For example, the exponential assumes a flat hazard, the weibull assumes a monotonic hazard; the log-normal and log-logistic assume a non-

monotonic hazard. If the characterisation of the underlying time-dependency is accurate, the particular distributional function is picked then parameter estimates will generally be more precise than estimates from semi-parametric and nonparametric models where the underlying time dependency is left unspecified. So, there can be advantages to using parametric models. That is it is more informative, predicted hazard functions, predicted survival functions and median survival times can be obtained. The effect of covariates is to accelerate or delay the duration of illness by a constant amount (acceleration factor or time ratio), the effect size is time ratio which is easier to interpret and more relevant to clinician. Problems arise in such a way that the AFT assumption must hold and also there is a need to specify the distribution of the survival time, but an appropriate distribution may be difficult to identify.

2.7 Frailty models

Studies on determinants of child mortality have mainly used either logistic regression or Cox proportional hazards model assuming that the outcomes are independent. To find more accurate estimates for the determinants of child mortality that has critical implications for resource allocation for improving child survival, sibling structures in child mortality data from demographic surveys have been treated as multivariate failure time data (Guo 1993). As failure time data, many attempts have been made to extend the Cox proportional hazards model. In this context, the variance-corrected Cox model has received much attention (Spiekerman 1998). In the variance-corrected Cox model, regression parameters of the determinants are estimated by ignoring intra-family correlation but adjusted for in the inference procedure; however, it ignores the variation of underlying risk among families. To overcome this, multivariate failure time data are modelled by an unobserved random quantity called frailties (Vaupel

1979). These frailties are common to observations from the same cluster and assumed to follow a given statistical distribution, known as multivariate random effects model or Cox frailty model.

In Malawi, studies on child mortality have mainly addressed the role of maternal, socioeconomic and health-related determinants. These studies were restricted to the analysis of mortality risks in children at individual level and not considered the correlation among children of the same family. There is need to emphasize those determinants which are nearer in time to the outcome and can be modified by program than those which are remote or far apart in time to the outcome of concern. The former covariates are referred to as programmable determinants and the latter as background variables. Therefore, there is a need to identify the programmable determinants of under-five mortality using Cox frailty model to account for sibling-level correlation for providing valid estimates needed for policy-decision making. In order to appreciate the influence of sibling-level correlation over the estimates of the determinants of under-five mortality, the results of Cox frailty model can be compared with the Cox proportional hazards model and variance-corrected Cox model. It is beyond the scope of this study to look at frailty models.

2.8. Model checking using statistical criteria

After fitting the Cox model and all the parametric models, the adequacy of model fit was assessed using residuals. In linear regression methods, residuals are defined as the difference between the observed and predicted values of the dependent variable. However, when censored observations are present and partial likelihood function in the Cox PH model, the usual concept of residuals is not applicable. Three major

residuals are Cox-Snell residual, the deviance residual and the Schoenfeld residual. In this study, Cox- Snell residual was used to assess the adequacy of model fit.

2.8.1. Residual plots

Residual plots were used to check the goodness of fit of the model. One of the most useful plots is based on comparing the distribution of the Cox-Snell residuals with the unit exponential distribution. The Cox Snell residual for the i^{th} individual with observed time t_i is defined as

$$r_{c_i} = \hat{H}(t_i|x_i) = -\log[\hat{S}(t_i|x_i)],$$

(2.29)

where t_i is the observed survival time for individual i , x_i is the vector of covariate values for individual i , and $\hat{S}(t_i)$ is the estimated survival function of the fitted model. If the model fits the data well then the true cumulative hazard function conditional on the covariate vector has an exponential distribution with a hazard rate of one. In this study the fitted parametric models were evaluated and compared using the Cox-Snell residuals. For each model, the Cox-Snell residuals were calculated, their survival function were estimated using Kaplan –Meier method and then, their cumulative hazard functions of these estimations were calculated. Finally, according to the Cox-Snell residuals, the hazard function graphs were drawn and the better fitted model was the one whose graph was closer to the bisector.

2.8.2 The Akaike Information Criterion (AIC)

The AIC was used to compare the performance of different parametric models. Typically, we would like models whose log-likelihood is big. The AIC is a measure of the goodness of fit of an estimated statistical model. The AIC is an operational way of trading off the complexity of an estimated model against how well the model fits the

data. Akaike's method penalizes each model's log likelihood to reflect the number of parameter that are being estimated and then compares them.

For our models discussed, the AIC is given by

$$AIC = -2 \log (\text{likelihood}) + 2(p+k) \quad (2.30)$$

Where p is the number of model covariates and k is the number of model specific distributional parameters. That is $k=1$ for the exponential model, $k = 2$ for the weibull, log logistic and log normal models, and $k = 3$ for generalised gamma (Klein and Moeschberger, 1997). Essentially, you compare the AIC scores for different parametric models and then select the one with the smallest AIC score.

2.9. Data analysis

Using cross - tabulations, descriptive statistics were obtained to give more information about the distribution of the variables. For each category of each variable, number of observation and number of failures as well as percentages were obtained. Kaplan-Meier estimation was done for each variable to determine the survival curves of each categorical predictor. This provided an insight into the shape of the survival function for each category. For the categorical variables, log rank test was used to compare the survival of two or more groups.

Univariate analysis was used to identify all the risk factors before proceeding to more complicated model. The univariate Cox Proportional hazards models were fitted and the hazard ratios as well as the coefficients for child survival from different factors were obtained. Then a full multivariate Cox PH model including all the risk factors was also fitted regardless of results of univariate analyses to establish if there were some variables which were not significant in the univariate analysis but are significant in the multivariate so as to find variables to be included in the final model. The

variables were identified as significant using 1%, 5% and 10% significant level. The study used up to 10% significance level in order to avoid excluding some important variables in the final model. Then the final model which excluded variables which were insignificant in both the univariate and multivariate models was fitted.

The proportionality assumption was checked with graphical method and two statistical methods (including time-dependent covariates in the cox model by using the tvc and the texp options in the stcox command, and tests based on Schoenfeld residuals). Time dependent covariates are interactions of the predictor and time. In this analysis the interaction with log (time) was used because this is the most common function of time used in time –dependent covariates. If a time –dependent covariate is significant, this indicates a violation of the proportionality assumption for that specific predictor. The goodness of fit of the Cox PH model was evaluated using Cox-Snell residual. If the model fits the data well the true cumulative hazard function conditional on the covariate vector has an exponential distribution with a hazard rate of one.

Parametric models such as weibull, exponential, gompertz, log normal, and log logistic models were fitted to obtain hazard ratios and coefficients. For each kind of model, the univariate and multivariate models were fitted. The accelerated failure time (AFT) model of weibull, exponential, log-normal and log-logistic is another alternative of the Cox PH model and was used when the PH assumption was violated

Residual plots which are also the transformation of the Q-Q plot were used to check the AFT assumption and to check the goodness of fit of each parametric model. Performance between AFT models was compared using statistical criteria likelihood

ratio (LR) test and Akaike information criterion (AIC), a measure of the goodness of fit for statistical models. The AIC is a measure of the goodness of fit of regression models that is based on the concept of entropy. It can be viewed as the amount of information lost when a model is used to describe a set of observations. The AIC includes a penalty for number of model parameters and thus represents the trade-off between bias and variance. Lower AIC values indicate a better model fit. Furthermore, we checked the goodness of fit of the model using residual plots. Post estimation test for parametric models was also conducted using Cox-Snell residuals to check the goodness of the model fit. The data was analysed using STATA software.

Chapter 3 Results

3.1 Descriptive

First descriptive statistics were used to give information about the distribution of the variables. In this case, the baseline characteristics for each predictor and the outcome variable in 19,947 participants were tabulated using the descriptive statistics. Table 3 gives the descriptive statistics.

The study involved 19,967 under-five children of which 1,607 were reported to be dead and 18,360 were alive. Similar proportions of children from urban and rural areas were reported to have died. On the other hand, 8.6% of the children from mothers with no education, 8.3% from mothers with primary education, 6.4% from mothers with secondary education and 4.7% children from mothers with higher education were reported dead.

Almost similar proportions of children from households which use electricity, charcoal or wood as source of cooking fuel were also reported dead. And also, the mortality rate was 8.9%, 8.3%, 7.4%, and 5.6% for children from fathers with no education, primary education, secondary education, and higher education respectively. The mortality rate was 7.6%, 8%, and 8.3% for children with flush toilet, pit latrines, no toilet facility respectively. Similarly, the death rate was 6.6% for households with two to three members and 8.9% for households with more than eight members. The proportions of children who were reported dead were almost similar for children from poor, medium and rich households. The proportions of children who were reported dead were also approximately similar for children from households with piped water and those without piped water.

Table 3 Baseline characteristics and outcomes

Predictor		Child is alive		Total
		No	Yes	
Type of place of Residence	Urban	157(8.28)	1,739(91.7)	1,896
	Rural	1,450(8.02)	16,621(91.98)	18,071
	Total	1,607	18,360	
Mother education	No education	289(8.6)	3,083(91.5)	3,372
	Primary edu	1,145(8.3)	12,720(91.7)	13865
	Secondary	168(6.4)	2455(93.6)	2623
	higher	5(4.7)	102(95.3)	107
	Total	1,607	18,360	
Type of cooking fuel	Electricity and gas	16(10.6)	135(89.4)	151
	Wood and charcoal	1591(8.03)	18,221(91.9)	18356
	Total	1,607	18,356	
Partner education	No education	187(8.9)	1,904(91.1)	2,091
	Primary edu	1,009(8.3)	11,149(91.7)	12,158
	Secondary	360(7.4)	4,523(92.6)	4,883
	higher	19(5.6)	322(94.4)	341
	Don't know	7(7.1)	92(92.9)	99
	missing	6(12)	44(88)	50
	Total	1,588	18,032	
Source of drinking water	Piped water	274(7.86)	3,212(92.14)	3,486
	Tube well water	904	10,088	10992
	Dug well water	306	3534	3840
	Surface water	91	1189	1280
	Total	1,607	18,360	
Type of toilet facility	Flush	18(7.6)	219(92.4)	1,836
	Pit latrine	1,348(8.02)	15,468(91.9)	16616
	No toilet facility	241(8.3)	2673(91.73)	2914
	Total	1,607	18,360	
Household size	2 to 3 members	153(6.59)	2,170(93.41)	2,323
	4 to 5 members	564(7.67)	6,789(92.33)	7,353
	6 to 7 members	527(8.59)	5,661(91.33)	6,188
	8 to 23 members	363(8.87)	5,661(91.48)	4,094
	Total	1607	18,351	
Wealth index	Poorest	339(7.5)	4,195(92.5)	4,534
	Poor	378(8.5)	4,093(91.6)	4,471
	medium	377(8.4)	4,133(91.6)	4,510
	Rich	338(8.9)	3,447(91.1)	3,785
	richest	175(6.6)	2,492(93.4)	2,667
	Total	1,607	18,360	
Access to electricity	No	1,527(8.1)	17,321(91.9)	18,848
	Yes	80(7.2)	1,039(92.9)	1,119

Survival time distribution was estimated for each category of the predictor using the K-M method and compared using the log-rank test. This provided an insight into the shape of the survival function for each group and gave an idea of whether or not groups are proportional (i.e. survival functions are approximately parallel). Tests of equality across strata, to explore whether or not to include the predictor in the final model, were also considered. For the categorical variables, the log-rank test of equality across strata which is a non-parametric test was used. The predictor was included in the final model if the test has a p-value of 0.25 or less. This elimination scheme was used because all the predictors in the data set are variables that could be relevant to the model. The predictor which is insignificant in both the univariate analysis and multivariate analysis was also not included in the final model. Table 4 shows results from the log-rank test.

Table 4 Log –rank test for equality of survival function

Variable	Chi(1)	p-value
Area of residence	0.19	0.66
Mother education	10.4	0.001
Partner education	5.74	0.01
Access to electricity	1.05	0.30
Wealth index	14.95	0.004
Source of water	2.29	0.51
Type of toilet facility	4.80	0.02
Household size	12.26	0.007
Source of fuel	1.06	0.30

From the log-rank test of equality across strata, There is a significant difference in child death between mothers and fathers who had secondary education and higher and those without any education or with primary education only. There was also a significant difference in child mortality between those who were poorest, poor, medium, rich, and richest. Similarly there was also a significant difference in child death between those who were using at least a toilet facility whether flush or pit latrine and those without any toilet facility. Another significant difference in child death was due to household size. However, there was no significance difference in child mortality between those children living in rural areas and urban areas. There was also no significant difference between those using electricity and those using charcoal .The table also show no significant difference between those using tapped water and untapped water. In addition there was no significance difference in child mortality between those household with access to electricity and those without.

The log-rank test of equality across strata for the predictor mother education has a p-value of 0.0013, thus mother education was included as a potential candidate in the final model. The log-rank test of equality across strata for the predictor source of water has a p-value of 0.51 hence was not included in the final model because it was also insignificant in both the univariate as well as the multivariate model of Cox and all parametric models.

The log rank-test of equality for the predictor access to electricity has a p-value of 0.30 thus access to electricity was not included in the final model since it is also not significant in both the univariate and multivariate analysis of both Cox and other parametric models.

The log-rank test of equality for wealth index has a p-value of 0.0048, thus was included in the final model. Similarly, the log-rank test of equality for source of cooking fuel has a P-value of 0.30, but was still included in the final model because it was significant in the multivariate analysis of all the parametric models. In addition, the log-rank test of equality for household size is 0.007, thus household size was included in the final model.

The log-rank test of equality across strata for the predictor partner education has a p-value of 0.017, thus partner education was included in the final model. Similarly type of toilet facility was included in the final model because has a p-value of 0.028. On the other hand, area of residence was not included in the final model because its log-rank test of equality across strata has a p-value of 0.66 which is greater than 0.25 and was also not significant in both the univariate and multivariate analysis of both Cox and parametric survival models. Figure 2 shows the K-M estimates for different predictors and Appendix one shows results of the fitted parametric models with all covariates.

Figure 3.1 Kaplan Meier Curves comparing categories for each risk factor (Analysis time is in months)

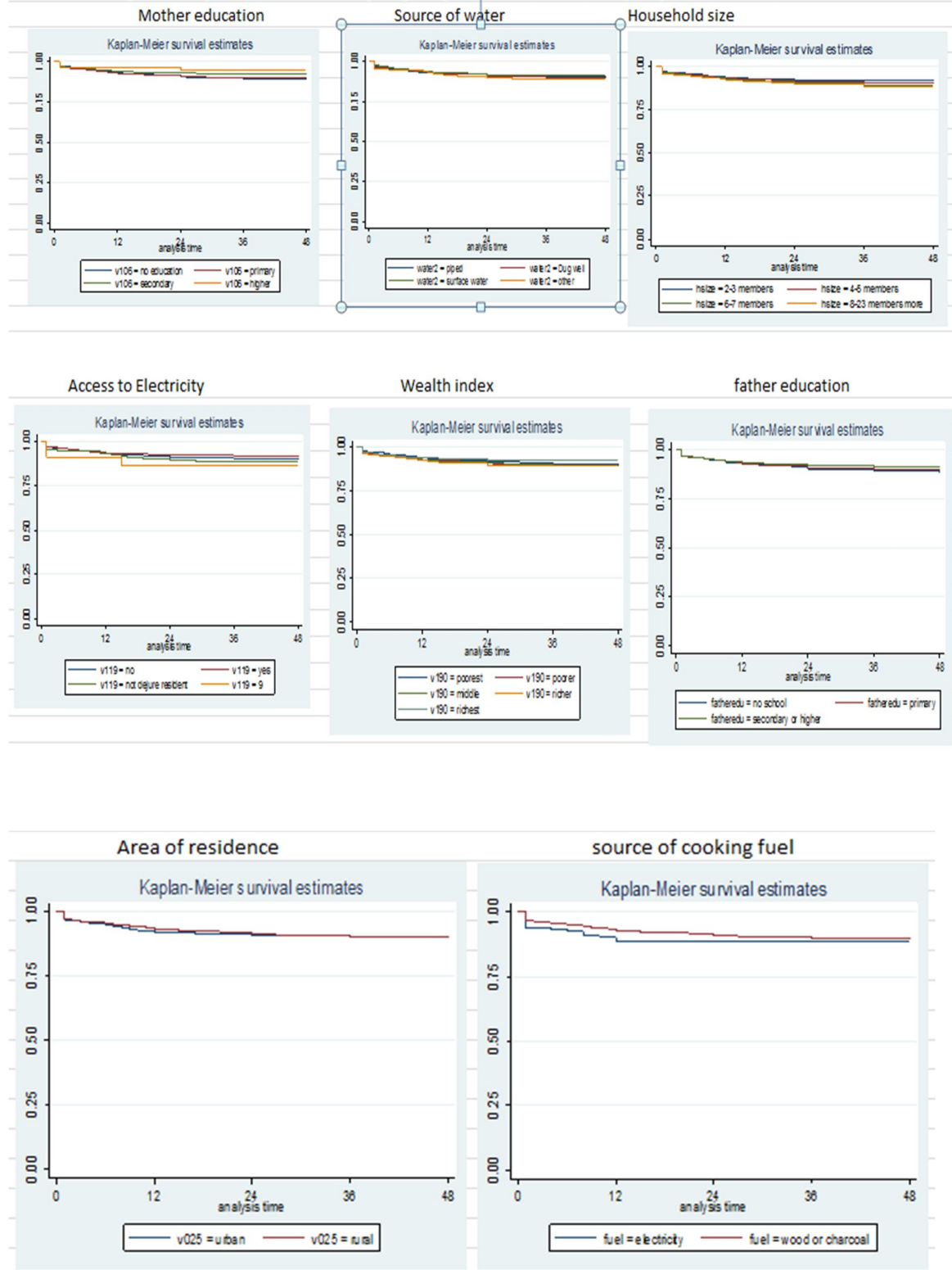


Figure 2 Kaplan- Meir curves comparing survival curves for each risk factor

3.2. Cox PH Model

Univariate analysis was used to check all risk factors before proceeding to more complicated models. Univariate and multivariate Cox Proportional hazard regression model for mother education, area of residence, partner education, Type of toilet facility, source of cooking fuel, access to electricity, Source of water, household size, and wealth index was used. The likelihood ratio test was considered in both the univariate and multivariate Cox PH model.

From both the univariate and multivariate analysis, Low levels of mother education, larger household size, and poor households were significantly associated with high under-five mortality rates. Using the multivariate analysis results, there was decreased risk for children from mothers who had secondary and higher education (HR=0.81; CI= 0.67, 0.97). Larger household sizes of about 8 or more members were significantly associated with increased risk of child mortality (p-value=0.005) as compared to smaller household size (HR=1.31; 95%CI = (1.08, 1.59)). Similarly, children from richest households faces lower hazard (p-value=0.01) as compared to children from poor households (HR=0.74, 95% CI=(0.59, 0.93)). In the multivariate analysis of Cox PH model use of electricity as the source of cooking fuel was significantly associated with higher child mortality rate (P-value=0.02). This is different from the Cox PH univariate analysis results where source of cooking fuel was not significantly associated with child mortality. From the univariate analysis, Households with educated fathers had lower hazard (HR= 0.87) compared with household with fathers who had primary or no education (p-value=0.02, 95% CI= 0.77 to 0.97). Similarly, in the univariate analysis, children from household with no toilet facility were at a higher risk of child mortality (HR=1.23) as compared to those children from household with a toilet facility (p-value=0.03). From both univariate

and multivariate analysis, area of residence, source of water, and access to electricity were not significantly associated with child mortality. Table 5 shows both the univariate and multivariate results of Cox PH model. Note that urban setting, mother with no or primary, use of electricity and gas as source of cooking fuel, partner with no or primary education, use of piped water, availability of toilet facility, household size of two to three members poor and no access to electricity were used as the reference categories.

Table 5 Univariate and multivariate Cox PH model for the relative hazard of child mortality

	Univariate analysis			Multivariate analysis		
	Haz. Ratio	95%CI	p-value	Haz.Ratio	95%CI	p-value
Rural area	0.96	(0.81,1.13)	0.67	.86	(0.71, 1.06)	0.16
Mother with secondary and higher education	0.773	(0.66,0.91)	0.001	.81	(0.67, 0.97)	0.02
Access to electricity	0.890	(0.71,1.11)	0.31	1.06	(0.79, 1.41)	0.68
Wood and charcoal as source of fuel	0.774	(0.47,1.26)	0.30	.515	(0.29,0.90)	0.02
Source of water						
Non piped water	1.15	(0.780,1.69)	0.48	.95	(0.86, 1.05)	0.34
Household size						
4-5 members	1.10	(0.92,1.31)	0.29	1.09	(0.90, 1.30)	0.36
6-7 members	1.23	(1.02,1.47)	0.02	1.19	(0.99, 1.43)	0.06
8-23 members	1.31	(1.08,1.58)	0.005	1.31	(1.08, 1.59)	0.006
Having toilet facility	1.225	(1.02,1.47)	0.03	1.18	(.97, 1.44)	0.10
Partner with secondary and higher	0.87	(0.77,0.97)	0.02	.94	(0.83, 1.07)	0.36
Wealth index						
poorest	0.87	(0.75,1.01)	0.08	0.86	(0.75,1.00)	0.05
medium	0.98	(0.85,1.13)	0.83	0.99	(0.86,1.14)	0.89
rich	1.04	(0.90,1.21)	0.54	1.04	(0.89,1.21)	0.49
richest	0.77	(0.64,0.91)	0.004	0.74	(0.59,0.93)	0.01

A multivariate Cox PH model was then fit using only covariates which were significantly associated with child mortality (Mother education, household size, wealth index, type of toilet facility, and source of cooking fuel) at 5% significance level. The final multivariate Cox Proportional hazard model which was fitted using six is then given by:

$$h_i(t) = h_0(t)\exp(-0.20\text{Educated mother} - 0.79\text{Cooking fuel} \\ + 0.08\text{Hsize}_{4 \text{ to } 5 \text{ members}} + 0.16\text{Hsize}_{6 \text{ to } 7 \text{ members}} \\ + 0.26\text{Hsize}_{8 \text{ or more members}} + 0.15\text{No Toilet facility} \\ - 0.15\text{Poorest status} - 0.01\text{medium status} + 0.06\text{Richstatus} \\ - 0.21\text{richest status} - 0.06\text{educated father})$$

After a Cox PH model which included significant variables only is fitted in the analysis, the PH assumption assessed. To check proportionality, time –dependent covariates were included in the model. Time dependent covariates are interactions of the predictors and time. In this study log (time) was used because this is the most common function of time used in time-dependent covariates but any function of time could be used. If time dependent covariate is significant it indicates violation of the proportionality assumption for that specific predictor. From Table 6, the results indicate that the PH assumption for wealth index is violated in some of the categories (P-value for medium status times t is 0.01).

Table 6 Test of Test PH assumptions by including time dependent covariate in the model

predictor	Coef.	Std. Err.	z	P-value	95% CI
main					
Mother education	-0.14	0.13	-1.08	0.28	-0.39,0.11
Source of fuel	-0.94	0.35	-2.71	0.01	-1.62,-0.26
Wealth index					
poorest status	-0.30	0.11	-2.66	0.01	-0.52,-0.08
medium	0.65	0.20	3.26	0.00	0.26,1.05
rich	0.13	0.11	1.16	0.25	-0.09,0.34
richest	-0.09	0.15	-0.59	0.56	-0.37,0.20
Household size					
4 to 5 members	0.03	0.13	0.26	0.80	-0.22,0.28
6 to 7 members	0.06	0.13	0.44	0.66	-0.20,0.31
8 or more members	0.19	0.14	1.40	0.16	-0.08,0.46
Type of toilet facility	0.14	0.14	1.01	0.31	-0.13,0.42
Partner education	-0.03	0.09	-0.33	0.74	-0.21,0.15
tvc					
Mother education	-0.05	0.07	-0.73	0.47	-0.20,0.09
Source of fuel	0.15	0.23	0.65	0.51	-0.30,0.59
Wealth index					
poorest status	0.10	0.06	1.79	0.07	-0.01,0.22
medium	-0.25	0.10	-2.46	0.01	-0.45,-0.05
rich	-0.05	0.06	-0.90	0.37	-0.17,0.06
richest	-0.10	0.08	-1.24	0.22	-0.26,0.06
household size					
4 to 5 members	0.04	0.07	0.54	0.59	-0.10,0.18
6 to 7 members	0.08	0.07	1.14	0.26	-0.06,0.23
8 or more members	0.05	0.08	0.70	0.49	-0.10,0.21
Type of toilet facility	0.01	0.08	0.11	0.91	-0.14,0.16
Partner education	-0.02	0.05	-0.38	0.70	-0.12,0.08

Note: variables in tvc equation interacted with $\ln(t)$

The Schoenfeld and scaled Schoenfeld residuals were also used to test the PH assumption. The p-value for testing whether the correlation between Schoenfeld residual for this covariate

and ranked survival time was zero was checked. For wealth index, the P-value was less than 0.05 for the rich category which suggest that the PH assumption is violated

for wealth index, but reasonable for all the other covariates. Table 7 shows the schoenfeld results.

Table 7 Test for proportional hazard assumption using schoenfeld residuals

	rho	Chi2	df	p-value
Mother education	-0.004	0.03	1	0.85
Source of fuel	0.02	1.19	1	0.27
Poorest	0.02	0.79		0.37
Medium	-0.03	2.39		0.12
Rich	-0.04	3.67		0.04
richest	-0.04	2.63		0.10
Household size				
4 to 5 members	0.02	0.81	1	0.36
6 to 7 members	0.04	2.77	1	0.09
8 or more members	0.03	2.01	1	0.15
toilet facility	-0.0005	0.001		0.98
Partner education	-0.002	0.01		0.92
Global test		17.14	11	0.104

For each predictor, graph of the scaled schoenfeld assumption was also obtained. A horizontal line in the graphs is further indication that there is no violation of the proportionality assumption. Using Figure 3, two categories for wealth index; poorest category (status1) and medium category (status2) seem to violate the proportionality assumption. Thus, using all the three methods to check the PH assumption, wealth index violate the assumption of proportionality.

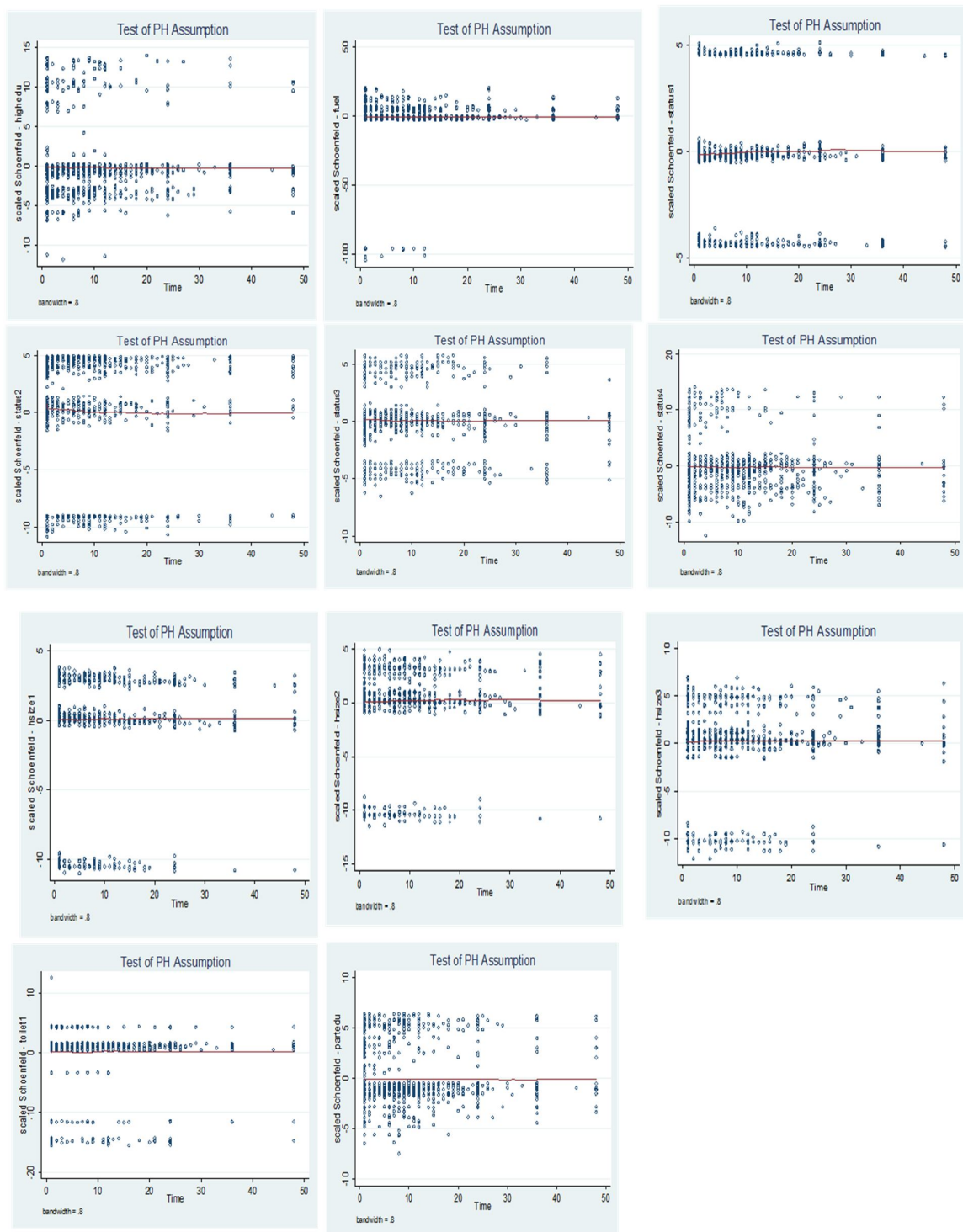


Figure3 Test of PH assumptions graphically using scaled Schoenfeld

The goodness of fit was also assessed by Cox-Snell residual plot. Cox –Snell residual plot is presented in Figure 4. There is some evidence of a systematic deviation from the straight line, which gives us some concern about the adequacy of the fitted Cox model.

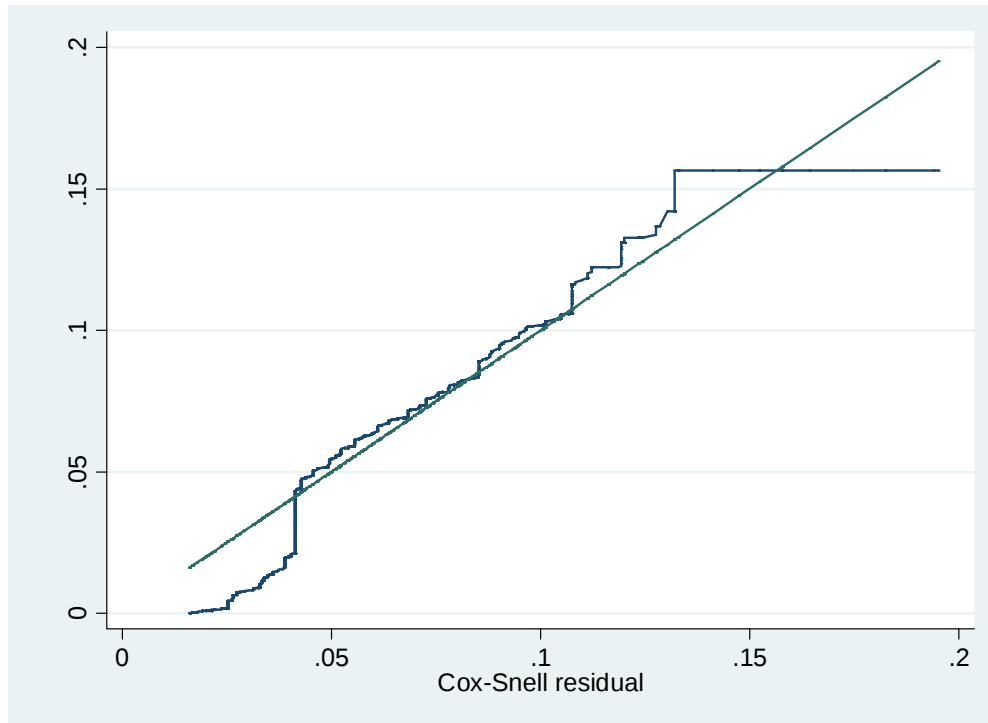


Figure 4 Cumulative hazard plot of the Cox –Snell residuals for the Cox PH model with significant predictors

3.3 Parametric Results

This section presents results for each of the parametric models that were used in this study. Both the univariate and multivariate analysis were used for all the fitted parametric models appendix 2 shows the results of multivariate analysis using all the nine predictors. Table 8 shows the estimates of the parametric models fitted using only the variables which were significant at 5% significance level in either the univariate or the multivariate of the fitted parametric models using all the nine predictors.

Table 8 Coefficients and hazard rate from parametric models for under-five children Time to death (n=19460)

	exponential			weibull			Gompertz			Log-logistic			log-normal		
Predictor	Coef.	Std. Err.	p-value	Coef.	Std error	P-value	Coef.	Std. Err.	P-value	Coef.	Std. Err.	P-value	Coef	Std. Err.	P-value
Mother with secondary or higher education	-0.20	0.09	0.04	-0.20	0.09	0.03	-0.19	0.09	0.04	0.38	0.18	0.03	0.36	0.18	0.04
Wood and charcoal	-0.77	0.27	0.01	-0.78	0.27	0.00	-0.77	0.27	0.00	1.53	0.53	0.00	1.58	0.56	0.01
wealth index															
poorest	-0.16	0.08	0.04	-0.15	0.08	0.04	-0.15	0.08	0.04	0.30	0.15	0.04	0.35	0.15	0.02
medium	-0.02	0.07	0.81	-0.01	0.07	0.89	-0.01	0.07	0.93	0.01	0.14	0.92	0.03	0.15	0.98
rich	0.04	0.08	0.63	0.05	0.08	0.54	0.05	0.08	0.52	-0.10	0.15	0.51	-0.11	0.15	0.48
richest	-0.25	0.10	0.02	-0.23	0.10	0.03	-0.22	0.10	0.04	0.43	0.20	0.03	0.39	0.20	0.04
household size															
4 to 5 members	-0.03	0.09	0.74	0.04	0.09	0.69	0.07	0.09	0.48	-0.07	0.18	0.68	-0.08	0.18	0.64
6 to 7 members	0.06	0.09	0.55	0.12	0.09	0.19	0.16	0.09	0.10	-0.24	0.18	0.18	-0.24	0.18	0.19
8 to 23 members	0.16	0.10	0.10	0.23	0.10	0.02	0.26	0.10	0.01	-0.44	0.19	0.02	-0.45	0.19	0.02
No toilet facility available	0.17	0.10	0.09	0.16	0.10	0.11	0.15	0.10	0.12	-0.31	0.19	0.11	-0.30	0.19	0.13
partner with secondary and higher education	-0.05	0.07	0.42	-0.05	0.07	0.41	-0.05	0.07	0.43	0.11	0.13	0.40	0.11	0.13	0.40
_cons	-4.97	0.29	0.00	-3.45	0.29	0.00	-3.96	0.29	0.00	6.18	0.56	0.00	6.99	0.59	0.00
/ln_p				-0.63	0.02	0.00									
p				0.53	0.01	0.51									
1/p				1.88	0.04										
/ln_gam										0.60	0.02	0.00			
/gamma							-0.09	0.00	0.00	1.82	0.04				
/ln_sig													1.34	0.02	0.00
sigma													3.82	0.08	

3.3.1 Exponential Model

In the exponential model which was fitted using all the nine variables, mother education, and wealth index were found to be significantly associated with child mortality in both the univariate and multivariate analysis. In the exponential model, Partner education and type of toilet facility and household size was significantly associated with child mortality in the univariate analysis but not in the multivariate analysis. On the other hand, source of cooking fuel was significantly associated with child mortality in the multivariate analysis but not in the univariate analysis. As found in the Cox model, area of residence, source of water and access to electricity were not significantly associated with child mortality. The fitted hazard function for the exponential model is:

$$\begin{aligned}x_j \hat{\beta}_x = & -0.20 \text{educated mother} - 0.77 \text{Cooking fuel} - 0.16 \text{poorest} \\ & - 0.02 \text{medium} + 0.04 \text{rich} - 0.25 \text{richest} - 0.03 \text{Hsize}_{4 \text{ to } 5 \text{ members}} \\ & + 0.06 \text{Hsize}_{6 \text{ to } 7 \text{ members}} + 0.16_{8 \text{ or more}} + 0.17 \text{No toilet facility} \\ & - 0.05 \text{Educated father}\end{aligned}$$

$$\text{And } \hat{\beta}_0 = -4.97$$

the estimate of the baseline hazard is

$$\hat{h}_0(t) = \exp(-4.97) = 0.007,$$

And the estimate of the overall hazard is

$$\begin{aligned}h(t|x_j) = & 0.007 \exp(-0.20 \text{educated mother} - 0.77 \text{Cooking fuel} - 0.16 \text{poorest} \\ & - 0.02 \text{medium} + 0.04 \text{rich} - 0.25 \text{richest} - 0.03 \text{Hsize}_{4 \text{ to } 5 \text{ members}} \\ & + 0.06 \text{Hsize}_{6 \text{ to } 7 \text{ members}} + 0.16_{8 \text{ or more}} + 0.17 \text{No toilet facility} \\ & - 0.05 \text{Educated father})\end{aligned}$$

3.3.2. Weibull Model

In the weibull model which was fitted using all the nine predictors, mother education, wealth index and household size were all found to be significantly associated with child mortality in both the univariate and multivariate analysis. Source of cooking fuel was found to be significantly associated with child mortality in the multivariate analysis but not in the univariate analysis. On the other hand toilet facility and partner education was found to be significantly associated with child mortality in the univariate analysis but not in the multivariate analysis. As found in the Cox and Exponential models, area of residence, source of water and access to electricity were all insignificantly associated with child mortality. The estimate of shape parameter in Weibull was 0.53 which is less than 1 implying that the hazard is monotone decreasing and the 95% CI is (0.52, 0.56) which does not cover the null value 1. Hence the Weibull model is better than the exponential model. The estimated hazard function for the i^{th} individual is:

$$\hat{h}_i(t) = \lambda \gamma t^{\gamma-1} \exp(\beta' x_i)$$

Where $\lambda \gamma t^{\gamma-1}$ is the baseline hazard function.

$$\begin{aligned} = & -3.45 * 0.53 t^{0.53-1} \exp(-0.20 \text{Educated mother} - 0.78 \text{Cooking fuel} \\ & + 0.04 \text{Hsize}_{4 \text{ to } 5 \text{ members}} + 0.12 \text{Hsize}_{6 \text{ to } 7 \text{ members}} \\ & + 0.23 \text{Hsize}_{8 \text{ or more members}} + 0.16 \text{No Toilet facility} \\ & - 0.15 \text{Poorest status} - 0.01 \text{medium status} + 0.05 \text{Rich status} \\ & - 0.23 \text{richest status} - 0.05 \text{educated father}) \end{aligned}$$

3.3.3. Gompertz model

As with the other models described, a model with all the nine predictors was fit and mother education, household size, and wealth index were found to be associated with child mortality in both the univariate and multivariate analysis. Source of cooking fuel

was also found to be significantly associated with child mortality in the multivariate analysis but not in the univariate analysis. On the other hand, toilet facility and partner education was also found to be significantly associated with child mortality in the univariate analysis but not in the multivariate analysis. As depicted in Table 8, gamma which is the shape parameter is -0.09 , implying that the hazard decreases with time ($p\text{-value} < 0.001$).

In the Gompertz model, there is decreased mortality risk for children born from educated mothers ($HR=0.82$, $CI=0.68$ to 0.98). Children from the educated mothers (those with secondary and higher education) face 82% of the hazard that children from uneducated mothers (no and primary education only) face. There is a decreased risk of child mortality for children from poorest and richest families as compared to those from poor families. Children born from poorest households face about 85% of the hazards that children from the poor family face ($P\text{-value} = 0.04$, $CI=0.73$ to 0.99). Also children from the richest family face about 80% of the hazard that children from the poor family face ($P\text{-value} = 0.65$ to 0.98).

There is increased child mortality risk for children from households with many members as compared to households with few members. In this case, children from household with 8 to 23 members faces 30% more hazard than children from households with two to three members ($P\text{-value} = 0.008$, $CI = 1.07$ to 1.57). With regard to the source of cooking fuel, children from households which used fire wood or charcoal as source of cooking fuel faces 46% of the hazards children from households using electricity face ($p\text{-value}=0.004$, $CI=0.27$ to 0.78).

3.3.4. Log-logistic model

As found in the weibull and Gompertz model, Mother education, household size, and wealth index were found to be significantly associated with child mortality in the univariate and multivariate analysis of Gompertz model. Source of Cooking fuel was significant in the multivariate analysis only and type of toilet facility and partner education were significant in univariate analysis only. At 10% level of significance, type of toilet facility was significant. However, area of residence, source of water and access to electricity were not significantly associated with child mortality in both the univariate and multivariate analysis of log-logistic model. A log-logistic model with the significant predictors as well as significant categories was finally fitted. The fitted survival function for the i th individual is:

$$\hat{S}_i(t) = \{1 + t^{\frac{1}{\sigma}} \exp(\hat{\omega}_i)\}^{-1}$$

$$= \{1 + t^{\frac{1}{1.82}} \exp(\hat{\omega}_i)\}^{-1}$$

Where

$$\hat{\omega}_i = \frac{-\mu - \hat{\alpha}x_i}{\hat{\delta}}$$

$$\begin{aligned} = \frac{1}{1.82} \{ & (-6.18) - 0.38 \text{Educated mother} - 1.53 \text{Cooking fuel} + 0.07 \text{Hsize}_{4 \text{ to } 5 \text{ members}} \\ & + 0.24 \text{Hsize}_{6 \text{ to } 7 \text{ members}} + 0.44 \text{Hsize}_{8 \text{ or more members}} \\ & + 0.31 \text{No Toilet facility} - 0.30 \text{Poorest status} - 0.01 \text{medium status} \\ & + 0.10 \text{Rich status} - 0.43 \text{richest status} - 0.11 \text{educated father} \} \end{aligned}$$

The estimated hazard function for the i^{th} individual is:

$$\hat{h}_i(t) = \frac{1}{\hat{\sigma}t} \{1 + t^{-\frac{1}{\hat{\sigma}}} \exp(-\hat{\omega}_i)\}^{-1}$$

$$= \frac{1}{1.82t} \{1 + t^{-\frac{1}{1.82}} \exp(-\hat{\omega}_i)\}^{-1}.$$

Where $\hat{\omega}_i$ is as defined above

3.3.5. Log-normal model

The results of the univariate and multivariate analysis of log-normal model are similar to those of log-logistic model. Education of the mother, using fire wood and charcoal as source of cooking fuel, small household size, and having more wealth increases the survival time of children. From the univariate analysis living in households with no toilet facility and living in household with uneducated father, decreases the survival time thus increases the hazard rate. From the log-normal model, the hazard is high in the first early months of life and then it decreases with time. Figure 5 shows the mean hazard for each of the six significant variables separately.

Mathematically, log normal AFT model is given by:

$$\ln(t_j) = x_j\beta_x + \ln(\tau_j)$$

Thus

$$\begin{aligned} \ln(t_j) = & 6.99 + 0.36\text{Educated mother} + 1.58\text{Cooking fuel} - 0.08\text{Hsize}_{4 \text{ to } 5 \text{ members}} \\ & - 0.24\text{Hsize}_{6 \text{ to } 7 \text{ members}} - 0.45\text{Hsize}_{8 \text{ or more members}} \\ & - 0.30\text{No Toilet facility} + 0.35\text{Poorest status} + 0.01\text{medium status} \\ & - 0.11\text{Richstatus} + 0.39\text{richest status} + 0.11\text{educated father} \\ & + \ln(\tau_j) \end{aligned}$$

Where the random quantity $\ln(\tau_j)$ has a distribution determined by what is assumed about the distribution of τ_j . In this case τ_j follows a log-normal distribution, which implies that $\ln(\tau_j)$ follows a normal distribution. Figure 5 show that the hazard function decreases with time.

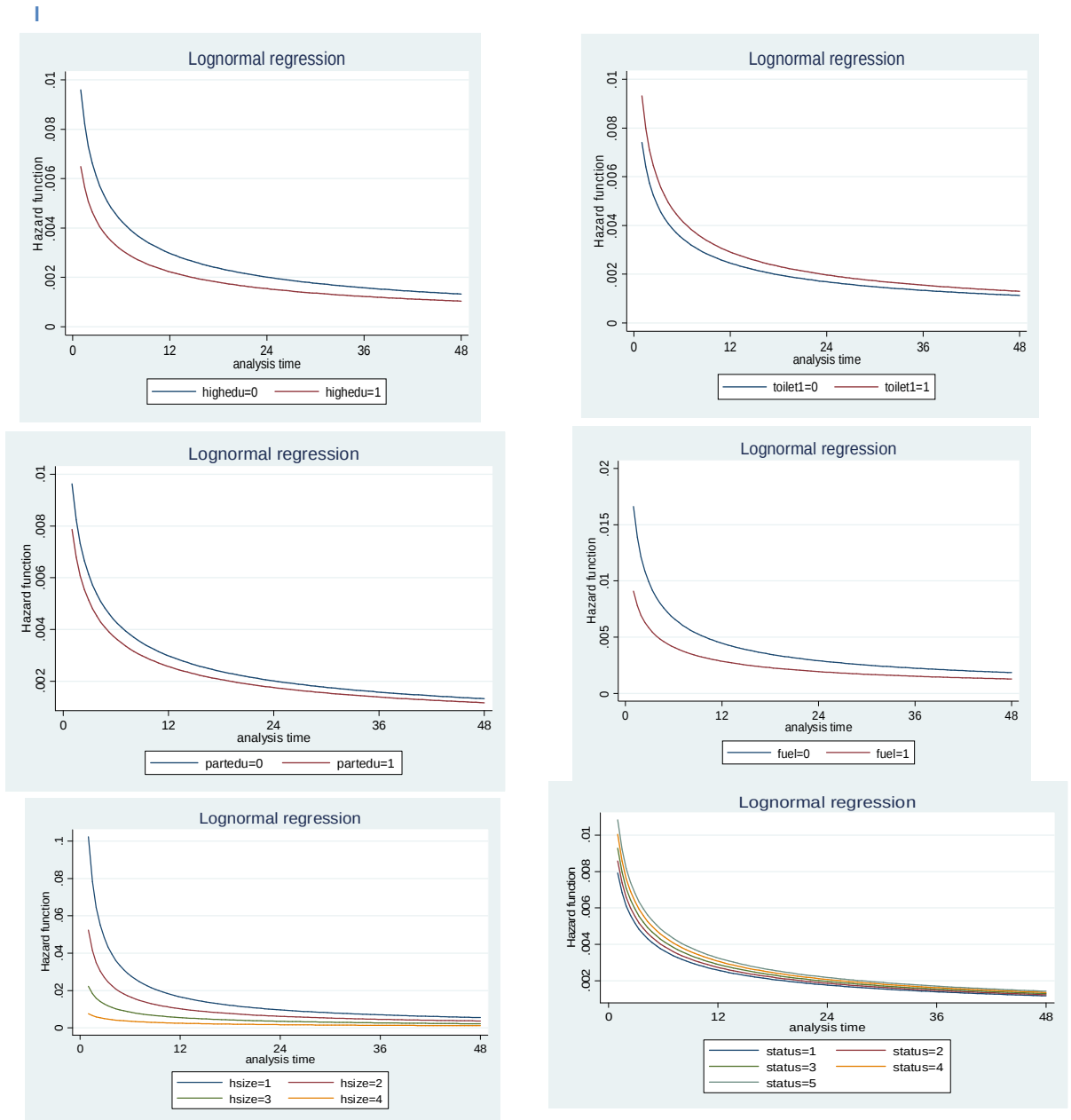


Figure 5 Hazard functions for lognormal model

3.4. Determining the best model fit

After fitting the Cox and the parametric models, the models were compared using the Akaike Information Criterion. In addition, the goodness of fit for each of the parametric model was also established.

3.4.1. Model selection

The AIC was used to compare the models and each model was fitted using mother education, partner education, source of cooking fuel, type of toilet facility, household size, and wealth index. Table 9 shows the AIC for each model. Comparing the AIC of parametric models, the log-normal model has the smallest AIC scores (AIC=15,114.96) hence it appears to be an appropriate parametric model according to AIC. It is followed by Gompertz (AIC=15,143.9) model and then log-logistic (AIC=15232.71).

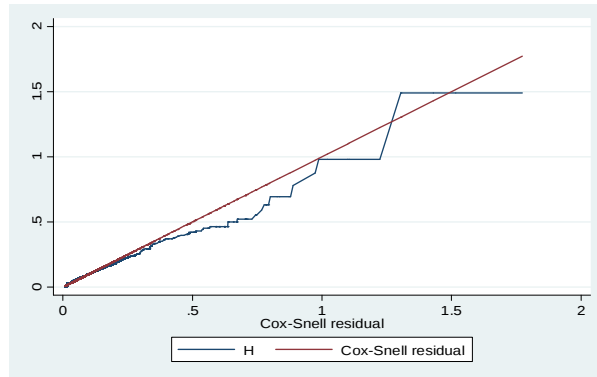
Table 9 Akaike Information Criterion(AIC) in the parametric models (n=19460)

Distribution	LL(null)	LL(model)	df	AIC
Exponential	-8,350.26	-8,329.9	12	16,683.89
Weibull	-7,857.76	-7,837.28	13	15,700.57
Gompertz	-7,793.57	-7,772.6	13	15,571.2
Loglogistic	-7,846.67	-7,825.89	13	15,677.79
Lognormal	-7,765.62	-7,745.70	13	15,517.4

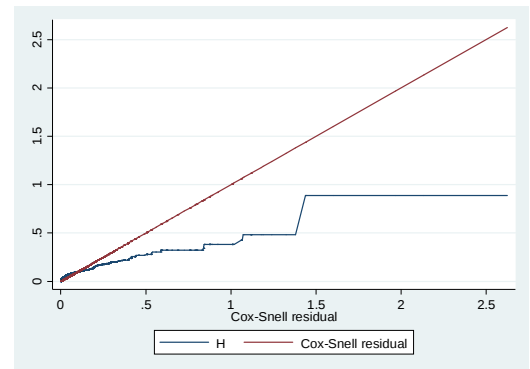
3.4.2. Cox Snell Residual

Furthermore, the goodness of fit of the model was checked using residual plots. The cumulative hazard plot of the Cox-Snell residuals in all the parametric models is presented in Figure 6. From the plot, the Cox-Snell residual plot for the log-normal model seems to fit the data well as compared to the plots of the other models since the plotted point lies closer to the line that has a unit slope and zero intercept. On the other hand, the exponential model is the poorest fit since there is so much deviation of the Cox-Snell residual plot from the 45-degree straight line. So based on the AIC and Cox-Snell residual plots, the log-normal AFT is the most suitable model as compared to the other parametric models.

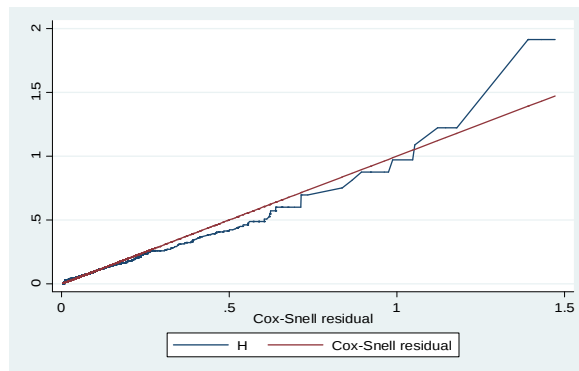
Weibull



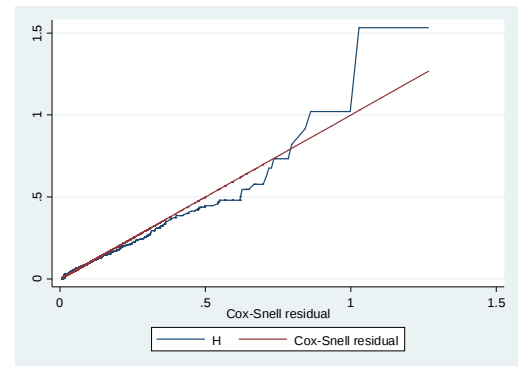
Exponential



Gompertz



Loglogistic



Log-normal

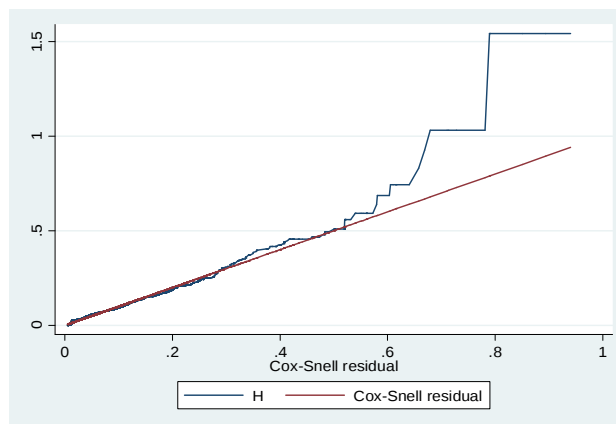


Figure 6 Cumulative hazard plot of the Cox Snell residual for parametric models

Chapter 4 Discussions

The aim of this study was to investigate the comparative performance of Cox and parametric models in the survival of under-five children exposed to different household environments. The Cox model as well as parametric models was fitted in this study using nine variables. The PH assumption of Cox model was assessed using three different methods; adding time dependent covariate to the model, using Schoenfeld residuals and using graphs of scaled schoenfeld residuals. In all the methods, the PH assumption was violated for wealth status. And yet the PH assumption must always hold if the results from the Cox model are to be reliable. The study further assessed the adequacy of the fitted Cox model using Cox snell residuals and the Cox model shows some lack of fit, and also comparing the Cox Snell residual of Cox and of Log-normal, the Cox-Snell residual for lognormal was better, hence there are some doubts about the suitability of the fitted Cox model.

The Akaike Information criterion was used to evaluate among the parametric models and log-normal model has the smallest Akaike scores hence was found to be the suitable model followed by Gompertz and then Log logistic. The study also evaluated the adequacy of the fitted parametric models using Cox Snell residuals and again lognormal indicated a better fit followed by Gompertz and then log logistic. Both log-normal and log-logistic does not have PH metric, they are in AFT metric. In this study, the PH assumption was somehow violated and these AFT models have proved to be fruitful. Although Weibull is the most widely used parametric proportional hazards model (Collette, 2003), Gompertz PH model has been found to be a better model than weibull. Exponential model had the highest akaike score amongst the

parametric models and its Cox Snell residual plot showed lack of fit hence it is also not the best model to use in the analysis of child survival. Therefore, based on AIC criteria and Cox-Snell residuals the study finally concludes that the log-normal model is the best fitting parametric model. The hazard rate for the log-normal model was computed for each of the six variables and the hazard rates are generally declining over time which is consistent with the well-known fact that the risk of death in Africa is high at very young ages and then declines rapidly beyond one and two years of age. For all the covariates, the results show that the hazard is high in the first six month of life but it is very high for children from mothers or fathers with no or primary education, for children from households with no toilet facility, for children from households which use electricity as a source of cooking fuel, for children from households with more members. Appendix two shows the baseline hazard for all parametric models.

The Cox PH model is the most widely used way of analysing survival data in the clinical research. From the review of literature of survival analysis, there are few studies using the Cox PH models that check PH assumption (Altman et al, 1985). However, PH assumption is not always satisfied in the data. There are various solutions to consider if the PH assumption does not hold. One of the alternative methods for the analysis of survival data even when the hazards are not proportional is use of the AFT model that does not assume PH metric like log logistic and log normal model AFT model. Based on asymptotic results, AFT models lead to more efficient parameter estimates than Cox model when the PH assumption is violated (Cleves 2010).

This study explored the impact of household socioeconomic and environmental determinants on child mortality and the univariate and multivariate analysis of both

Cox and parametric models was used. Mother education, wealth index, and household size were found to be significantly associated with child mortality in both the univariate and multivariate analysis of both Cox and parametric models except in the univariate and multivariate analysis of exponential model where household size was not significantly associated with child mortality. Source of cooking fuel was significantly associated with child mortality in the multivariate analysis of both Cox and Parametric models but not in the univariate analysis. On the other hand, partner education and type of toilet facility were all significantly associated with child mortality in the univariate analysis only of both Cox and Parametric models. On the other hand, area of residence, source of water and access to electricity were not significantly associated with child mortality in both the univariate and multivariate analysis of both Cox and Parametric models.

Mother education was significantly associated with child mortality. The results indicate that risk of child mortality are lower among women having secondary and higher education than those having no or primary education. These results are consistent with the findings from previous research by Zerai(1996), Mutunga (2004), Omulaubi(1995) and uthman(2008). Mutunga (2007) argues that maternal education works through three different pathways. It influences the socio-economic level of households, govern mothers attitude, and influences her behaviour (including health seeking) on issues relating to health of their children. Uthman and Mubashir B. (2008) concluded that mother education played a protective role against death and suggest that improving maternal education may be a key to improving child survival in Nigeria since educated mother had a better chance of satisfying important factors that can improve infant survival: the quality of feeding, general care, household sanitation, and adequate use of preventive and curative health services.

There was decreased child mortality risk for children born from poorest and richest households as compared to those children born from poor households. The reduction in child mortality among the poorest families may be due to underreporting of under-five deaths. These results are consistent with the findings of Mojekwu (2012) and Mutunga (2004) where better survival prospects were also found to exist for children born in wealthier families but not poorest families. Wealth status is associated with the availability of nutritional resources, which is especially important for the survival of a child because once infants reach the age of six months; they can no longer depend on nourishment from breast milk alone. Children from poor households are exposed to risk of diseases through inadequate water and sanitation, Crowding and poor housing conditions. They are also, more likely to have lower resistance to infectious diseases because they are undernourished (WHO 2002).

There is increased child mortality risk for children born from households with many family members as compared to children born from households with few members. Burstrom (1999) found similar results and he suggested that small household size contribute a lot to the reduction of infant and child mortality due to the family ability to afford better facilities and nutrition. Thus increase the probability of survival in children. These results are different from the findings of Mutunga (2004) where household size was negatively related to child mortality. That is lower child survival prospects were experienced in smaller households.

With regard to the source of cooking fuel, children born in households using less polluting fuels such as electricity and gas as their main source of cooking fuel have higher mortality rates as compared to those using high polluting fuels like Charcoal and fire wood. This result is like that because the number of households using electricity was relatively small as compared to those using charcoal or wood, and also

most of them were urban dwellers. The results are contrary to the findings of Mutunga (2007) and Mojekwu (2012) where high polluting fuels was associated with high child mortality rates.

Children from household with no toilet facility faces increased mortality risk as compared to those children from a household with either flush toilet or pit latrine. Children born in households with either flush toilet or pit latrines have lower mortality rate than those born in households without any toilet facility. This underscores the importance of good quality sanitation in the prevention of diseases such as cholera, diarrhoea and dysentery. Modern sanitation technology ensures the proper disposal of human waste, which is important in preventing the spread of these diseases. Mojekwu (2012), and Mutunga(2004) also found similar results

Children from educated fathers (secondary and higher) experiences decreased mortality rate as compared to children from uneducated fathers. These results are consistent with the findings of Mturi and Curtis(1995), and Agha (2010) who found that father's literacy was associated with under-five mortality. He suggested that an educated husband can make better decisions and seek timely and appropriate treatment for his children. Father education was significant in the univariate analysis but not in the multivariate analysis. This may be so because mother education was also affected by partner education since more educated women may be able to marry men who are educated and can care more about children.

Area of residence was not significantly associated with child mortality. There was no significant difference in the risk of child mortality for children from rural and urban areas. These results agrees with the argument of Akoto and Tabutin (1989) who

argued that it is not much the fact that living in urban setting that provides the advantage in terms of mortality to children born of urban mothers, but Socioeconomic factors instead. This includes high concentration of salaried workers (who generally have higher incomes) in urban centres, better education in urban areas and concentration of public infrastructure in urban areas that provides sanitation services including water supply, household waste and excreta removal and disinfection and better hospital infrastructure in the urban areas. On the contrary, Mturi and Curtis (1995) found that under-five mortality risk was associated with area of residence. In this case Children from rural areas were at a higher risk than those children born in urban areas. But the rapid growth of the urban population has reduced the child mortality gap which was there between children from urban and rural. This rapid population growth in the urban has strained the ability of local areas to provide adequate levels of infrastructure and public services, resulting in environmental threats on child health such as poor sanitation, water supply, and access to health care.

Access to electricity is not statistically significantly associated with child mortality. There is no significant difference in child mortality between children from households with electricity and those without. These results are different from the findings of Mutunga (2004) who found lower mortality rates in the households with electricity.

Source of drinking water was not significantly associated with child mortality rates. This could be due to the consistent use of water guard or boiled water by households without access to tapped water (Kumwenda 2009). It might also be due to the use of boreholes in the villages since ground water is relatively a safe source of portable water in rural areas as compared with other unprotected water sources like river, spring, well water etc. Similarly those households which use piped or tapper water, they do share the taps as a result, containers used for collection and transportation of

water from boreholes are mostly without covers. It is always observed that when lifting and balancing the collection vessel on the head, fingertip-dipping is common and unavoidable resulting in contaminated water. This, leads to no significant difference in child mortality between those using tapped and untapped water. These results are different from the findings of Mutunga (2004), Cornelia K. and Ingo P. (2011) who observed higher mortality rates among infants and children who lack access to safe drinking water.

Chapter 5 Conclusion and recommendation

In this study, low levels of mother education, poor wealth status of the household, large household size and Source of Cooking fuel, have been found to be significantly associated with higher under-five children mortality risk. From the univariate analysis, low father education levels and no toilet facility have been also found to be significantly associated with high child mortality risk. On the other hand, area of residence, source of drinking water and access to electricity were not significantly associated with child mortality. Source of cooking fuel was related to child mortality. In this case, those using less polluting fuels like electricity had higher mortality rates as compared to those using more polluting fuels like charcoal and wood. This is in contrast with the findings of other similar studies, as such, there is a need for further research.

In addition, previous studies have found area of residence to be significantly associated with child mortality. This is different from the findings of this study, thus calling for further studies as to why the gap in child mortality is now minimal between rural and urban dwellers. One main disadvantage of using the parametric model is that the specific distribution of survival time is unknown in many cases. Further study of this data could attempt using a non-parametric version of the AFT model which does not require the specification of the distribution that can be applied in child mortality data (Wei, 1992). The results from this model could then be compared with the standard AFT model and Cox PH models.

Despite our intention in recording all covariates relevant to a specific analysis, we might encounter heterogeneity in the sample that cannot be explained by the observed covariate alone. Further similar studies should consider using frailty models as these

can be informative. Then compare the performance of Cox frailty models with that of the standard Cox model.

In summary, the results of the current study suggest that when implementing survival analysis in under-five child mortality, using the PH model may not be the optimum approach. It is important to identify the distribution of Overall survival and to seek for an appropriate model like AFT models for data analysis. The results from an AFT model are easily interpreted and provide a more appropriate description of survival time in many researches, and should be considered as an alternative to the Cox PH model.

The choice of the appropriate model will certainly lead to identify real factors that are associated with child mortality, thereby help to have a more effective interventions.

References

- Agha, A., AJimal, F., & Asam, I. (2010). Father's support and literacy-factors associated with child mortality in Gambit, Sindh- Pakistan, *Journal of the Pakistan medical association*, 60(2), 81-85.
- Akoto, E., & Tauten, D. (1989). Socio-economic inequalities and child mortality in sub-Saharan Africa. *UIESP, Liege*, 36.
- Allison, D. (1995). *Survival Analysis using the SAS system a practical guide*. North Carolina, USA: SAS Institutes.
- Altman, D., Stavola, E., & Love B. (1985). Review of survival analysis published in cancer journals. *British Journal of Cancer*, 72, 511-518.
- Baker, R. (1999). *Differential in Child Mortality in Malawi*, (Social Networks Project, Working Papers No 3). Pittsburg: University of Pennsylvania.
- Bourgeois-Pichat, J. (1964). *The Determinants and Consequences of Population Trends*, New York: United Nations, Department of Economics and Social Affairs.
- Barnstorm, B., Diderichsen, F., & Smedman, L. (1999). The impact of household size and number of children in the family on the risk of death from measles. *American Journal of Epidemiology*, 149, 1134-1141.
- Caldwell, C., MacDonald, F. (1982). Influence of maternal education on infant and child mortality: levels and causes. *Health Policy and Education*, 2, 251-267.
- Cleves, M., Gould, W., Gutierrez, G. (2010). *An introduction to survival analysis using STATA*. Texas: Stata Press.
- Cleves, M., Gould, W., Gutierrez, R. (2004). *An Introduction to Survival Analysis Using STATA*. Texas: Stata Press.

Collett, D. (2003) *Modelling survival data in medical research* (2nd ed.). London: Chapman & Hall.

Cox, D.R. (2002). *The effect of water and sanitation on child mortality in Egypt*. Gothenburg: Environmental Economic unit, Department of Economics, Gothenburg University, Sweden.

Duerden, M. (2009). *What are hazard ratios?* New York: Hayward Medical Communications.

Espo, M. (2002). *Infant mortality and its underlying determinants in rural Malawi*. Unpublished Master's thesis, University of Tampere Medical school, Finland.

Gandotra, M., & Nayaran, D. (1988). *Infant Mortality: An Analysis of recent Births in Infant Mortality in India: Differentials and Determinants*. New Delhi: Sage Publications.

Guo, G. (1993). Use of sibling data to eliminate family mortality effects in Guatemala. *Demography*, 30(1), 15-32.

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

Hutton, J., Monaghan, P. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: asymptotic results. *Lifetime data analysis*. 8, 375–393. Retrieved on 19th January, 2006 from: <http://www.ncbi.nlm.nih.gov/pubmed/12471946>.

Hung, Wen-Shal, (2008). *Survival analysis for unobserved heterogeneity on estimated mortality in Taiwan*. *Economics Bulletin*, 9(25), 1-10.

Hobcraft, J. (1985). Demographic determinants of infant and early childhood mortality: A comparative analysis. *Population Studies*, 39, 363-383

Houweling, T., Kunst, A., & Looman C. (2005). Determinants of under-five mortality among the poor and the rich; a cross national analysis of 43 developing countries. *International Journal of Epidemiology*, 1257–1265

Jacoby, H., & Wang, L. (2003). *Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach*. Washington DC: World Bank.

Jahn, A., Floyd, S., McGrath, N. et al. (2010). Child mortality in rural Malawi: HIV closes the survival gap between the Socio-economic strata. *PLOS ONE* 5(6): e11320.

Kaldwei, C., & Pitterle, I. (2011). Behavioural factors as emerging main determinants of child mortality in middle- income countries: a case study of Jordan. *Department of Economic and social affairs*, 103.

Kalipeni, E. (1993). Determinants of infant mortality in Malawi: A spatial perspective. *Social Science and Medicine*, 37(2), 183-198

Kay, R. (2004). An explanation of the hazard ratio. *Pharm stat*, 3, 295-297.

Kaurosh, S., Mohammed, R. (2008). Prognostic factors of survival time after hematopoietic cell transplant in acute lymphoblastic leukemia patients: Cox PH versus AFT models. *Journal of experiments and clinical cancer research*, 27, 74. doi:10.1186/1765-9966-27-74.

Kazembe, L., Mpeketula, P. (2010). Quantifying spatial disparities in neonatal mortality using a structured additive regression model. *PLOS ONE* 5(6):e11180. doi:10.1371/journal.pone.0011180

Kembo, J., Joroen, K., Ginnekem, V. (2008). *Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis*. Online journal, Max planck institute for Demographic Research Konrad-Zuse str.1 D-18057 Restock. Germany.

Koch, S., Handa, S., & Wen Ng, S. (2008). *Child mortality in Eastern and southern Africa*, North Carolina: University of North Carolina, Chapel Hill.

Kleinbaum, D.G., Klein, M. (2005). *Survival Analysis – A Self-Learning Text*. Berlin: Springer-Verlag.

Klien, P. (1997). *Survival analysis techniques for censored and truncated data*. Berlin: Springer-Verlag, 83-400.

Kumwenda, S. (2009). *Assessment of water guard used at household level in Chikwawa district*. Unpublished Masters thesis, University Of Malawi, College of Medicine. Accessed on www.medcol.mw/commhealth/mph/dissertations/kumwenda_save_Dissertation_%2009072009.pdf

Lambert, C., Royston, P. (2009). Further development of flexible parametric models for survival analysis, *The STATA journal*, 9(2), 265-329.

- Lee, T. (1992). *Statistical Methods for survival Data Analysis*. New York: John Wiley & Sons, Inc.
- Lawless, J. (2003). *Statistical models and methods for lifetime Data*. Hoboken: John Wiley & Sons, Inc.
- Manda, M. (1999). Birth intervals, breastfeeding, and determinant of childhood mortality in Malawi. *Social science and Medicine*, 48(3), 301-312
- Masangwi, S., Morse, T., Ferguson, S. et al (2010). Behavioural and environmental determinants of childhood diarrhoea in Chikwawa, Malawi. *Desalination*, 252, 267-274.
- Millard, A. (1994). A causal model of high rate of child mortality. *Social Science and Medicine*, 38(2), 253-268.
- Moran, J.L., Bersten, A.D., Solomon, P.J., Edibam, C. et al (2008): Modelling survival in acute severe illness: Cox versus accelerated failure time models . *Journal of Evaluation in Clinical Practice*, 14, 83-93. Retrieved from <http://www.jeccr.com/pubmed/18211649>.
- Mosley, W., & Chen, L. (1984). An analytic framework for the study of child survival in developing countries. *Population and development review*, 10, 25-45.
- Mohamad, A., Hajizadeh, E. (2007). Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. *Asian Pacific Journal of Cancer Prevention*, 8, 412-416.
- Mturi, A., & Curtis, L. (1995). Determinants of infant and child mortality in Tanzania. *Health policy plan*, 10, 384-394.
- Mutunga, C. (2004). *Environmental determinants of child mortality in Kenya*. Kenya Institute for public policy research and analysis (KIPPRA), Nairobi, Kenya. Accessed on 21/09/2006 from: <http://www.webmeets.com/files/papers/ERE/WC3/405/world%20Congress%Paper.pdf>
- Mutunga, J. (2007). Environmental determinants of child mortality in urban Kenya. (*World Institute for Development Economics research, research paper*, 83). [Nairobi]: WIDE.
- Nnamdi, M., Godsom, M.C.(2012). Environmental determinants of child mortality in Nigeria *Journal of sustainable development*, 5(1). do:10.5539/jsd.v5n1pe5

Nelson, W. (1982). *Applied life Data Analysis*. New York: John Wiley & Sons, Inc.

Omoluabi, E. (1995). *Child Health in Ondo State: Seasonal fluctuation of weight velocity in B.F lyun: Health of Nations; medicine, disease and development in the third world*. Aldershot, England: Ashgate publishing Group.

Orbe, J., Ferreira, E., and Nunez-Anton ,V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med*, **21**, 3493-3510.

Pandey, A. et al, (1998). *Infant and Child Mortality in India, (National Family Health Survey Subject Report Number 11)*. Honolulu, Hawaii, U.S.A: IIPS Mumbai and East-West Centre Program on Population.

Raheen, A., Sheu, A., & Segun- Agboola, T. (2008). Explaining the Social and environmental determinants of child health in Ilorin, Nigeria. *Ethiopian journal of environmental studies and management*, 2, 32009.

Saint-Marcoux, F. , Knoop, C., Debord, J. et al. (2005). Pharmacokinetic study of tacrolimus in cystic fibrosis and non-cystic fibrosis lung transplant patients and design of Bayesian estimators using limited sampling strategies 14. *Clin Pharmacokinet* , 44,1317-1328.

Spiekerman, C., Lin, D. (1998). Marginal regression models for multivariate failure time data. *Journal of the American statistical association*. 93(443), 1174-1175.

Sudhanshu, H., Steve, K., Shu-Wen, (2008). *Child mortality in Eastern and southern Africa*. Retrieved from:
http://www.unc.edu/~shanda/research/child_mortality_in_ESA_Handa_Koch_Ng_Sept_08.pdf

UNICEF Malawi, (2007). *Health and nutrition issue National nutrition Guidelines for Malawi*. Government of Rwanda, Ministry of Health. United Nations Development Program (online). Millennium Development Goals (MGDs). Retrieved on 12/10/2011 from <http://http://www.unmillenium>

Uthaman & Mubashir (2008). Effect of low birth weight on infant mortality: analysis using weibull model. *The internet journal of epidemiology*, 6(1).

Vaupel, J., Manton, K., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3), 439-454.

Vaahtera, M., Kulmala, T., Ndekha, M. (2000). Antenatal and perinatal predictors of infant mortality in Rural Malawi. *Arch Dis child fatal Neonatal*, 82, 200-204.

Wang, L. (2003). *Environmental determinants of child mortality: Empirical results from the 2000 Ethiopia DHS*. Washington D.C.: World Bank.

Wei, J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in the survival analysis. *Statistics in Medicine*, 11, 1871-1879.

World Health Organisation. (2002). *World Bank Working Group on child health and poverty. Better health for poor children: A special report*. Geneva: World Health Organisation.

World Health Organisation. (2004). *World Health Report 2004: Changing History*. Geneva: World Health Organisation,

World Health Organisation. (2010). *World health statistics*, Geneva: World Health Organisation.

Zerai, A. (1996). Preventive health strategies and infant survival in Zimbabwe. *African population studies*, 11(1), 29-62.

Appendix 1 Parametric Univariate Estimates

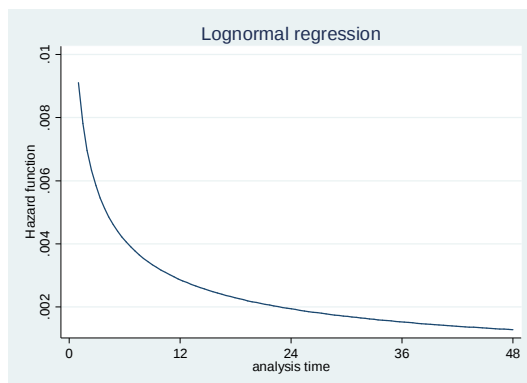
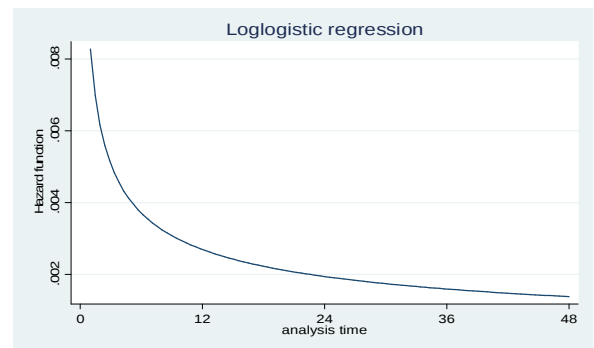
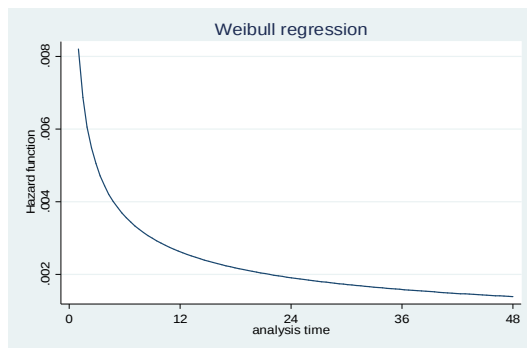
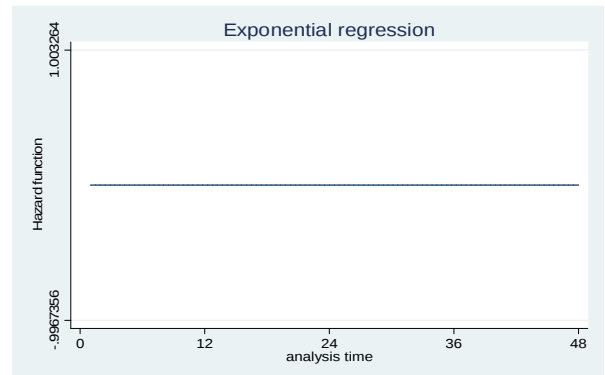
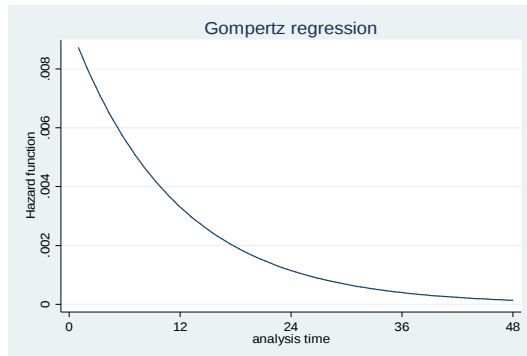
Appendix 1 shows the Coefficients and Hazard ratios for univariate analysis for exponential PH, weibull PH, Gompertz PH, lognormal AFT and log-logistic AFT model. From the table, Mother Education, household size, type of toilet facility, and partner education were significantly associated with child mortality in all the five parametric models.

Predictor	exponential			weibull			Gompertz			Log-logistic			Log-normal		
	Haz. Ratio	Std. Err.	p-value	Haz. Ratio	Std. Err.	p-value	Haz. Ratio	Std. Err.	P-value	Coef.	Std. Err	P-value	Coef.	Std. Err	P-value
Mother education	0.78	0.06	0.001	0.78	0.06	0.001	0.78	0.06	0.001	0.48	0.15	0.001	0.43	0.15	0.005
Area of residence	0.97	0.08	0.7	0.97	0.08	0.68	0.96	0.08	0.64	0.08	0.16	0.64	0.09	0.17	0.57
electricity	0.88	0.1	0.25	0.89	0.1	0.3	0.9	0.1	0.35	0.22	0.22	0.31	0.19	0.22	0.4
Source of fuel	0.8	0.2	0.38	0.79	0.2	0.34	0.78	0.2	0.32	0.5	0.49	0.31	0.59	0.52	0.26
Source of water	1.00	0.05	0.99	1.00	0.05	0.99	1.00	0.05	0.99	0.01	0.09	0.98	0.01	0.09	0.96
Household size															
4 and 5 members	0.99	0.09	0.90	1.06	0.10	0.55	1.09	0.10	0.37	-0.11	0.17	0.54	-0.11	0.17	0.53
6 and 7 members	1.11	0.10	0.26	1.18	0.11	0.07	1.22	0.11	0.03	-0.32	0.18	0.07	-0.32	0.18	0.07
8 and 23 members	1.20	0.12	0.06	1.27	0.12	0.01	1.32	0.13	0.00	-0.47	0.18	0.01	-0.48	0.19	0.01
Toilet facility	1.26	0.12	0.02	1.24	0.12	0.02	1.22	0.11	0.03	-0.4	0.18	0.03	-0.37	0.18	0.04
Partner education	0.87	0.05	0.02	0.87	0.05	0.02	0.87	0.05	0.02	0.26	0.11	0.02	0.24	0.11	0.04
Wealth index															
poor	0.86	0.06	0.06	0.87	0.06	0.06	0.87	0.06	0.06	0.27	0.14	0.06	0.30	0.15	0.04
medium	0.97	0.07	0.71	0.98	0.07	0.78	0.98	0.07	0.82	0.03	0.14	0.81	0.02	0.15	0.82
rich	1.02	0.07	0.73	1.03	0.07	0.82	1.04	0.07	0.57	-0.07	0.14	0.58	-0.08	0.15	0.56
richest	0.78	0.06	0.001	0.76	0.06	0.003	0.76	0.07	0.004	0.51	0.17	0.03	0.003	0.18	0.02

Appendix 2 Multivariate parametric results

	exponential			weibull			gompertz			loglogistic			lognormal		
predictor	Haz. Ratio	Std. Err.	P>z	Haz. Ratio	Std. Err.	P>z	Haz. Ratio	Std. Err.	P>z	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
Area Of residence	0.86	0.09	0.15	0.87	0.09	0.16	0.87	0.09	0.16	0.28	0.20	0.16	0.27	0.20	0.19
highedu	0.81	0.08	0.03	0.81	0.08	0.03	0.82	0.08	0.03	0.39	0.18	0.03	0.37	0.18	0.04
Access to electricity	1.07	0.16	0.66	1.07	0.16	0.66	1.08	0.16	0.62	-0.12	0.28	0.66	-0.11	0.28	0.70
fuel	0.53	0.15	0.03	0.52	0.15	0.02	0.53	0.15	0.03	1.29	0.56	0.02	1.35	0.59	0.02
poorest	0.86	0.06	0.04	0.86	0.07	0.05	0.86	0.07	0.05	0.30	0.15	0.04	0.34	0.15	0.02
medium	0.98	0.07	0.75	0.98	0.07	0.83	0.99	0.07	0.86	0.03	0.14	0.86	0.01	0.15	0.93
rich	1.02	0.08	0.81	1.03	0.08	0.71	1.03	0.08	0.68	-0.06	0.15	0.68	-0.07	0.16	0.64
richest	0.72	0.08	0.01	0.73	0.09	0.01	0.73	0.09	0.01	0.60	0.23	0.01	0.56	0.23	0.02
water2	0.95	0.05	0.32	0.95	0.05	0.34	0.96	0.05	0.36	0.09	0.10	0.35	0.09	0.10	0.39
hsize															
4 to 5 members	0.97	0.09	0.78	1.04	0.10	0.66	1.07	0.10	0.46	-0.08	0.18	0.65	-0.09	0.18	0.61
6 to 7 members	1.07	0.10	0.48	1.14	0.11	0.15	1.18	0.11	0.08	-0.25	0.18	0.15	-0.25	0.18	0.16
8 or more members	1.19	0.12	0.07	1.27	0.13	0.02	1.32	0.13	0.01	-0.47	0.19	0.01	-0.48	0.19	0.01
Type of toilet facility	1.20	0.12	0.07	1.19	0.12	0.09	1.18	0.12	0.10	-0.33	0.19	0.09	-0.31	0.19	0.11
Partner education	0.94	0.06	0.39	0.94	0.06	0.38	0.95	0.06	0.40	0.11	0.13	0.37	0.12	0.13	0.37
Constant										6.07	1.00	0.00	6.85	1.02	0.00
ln_p				-0.63	0.02	0.00			0.00						
p				0.53	0.01	0.51						0.00			
1/p				1.88	0.04	1.80									
/ln_gam										0.60	0.02	1.75			
gamma							-0.09	0.00		1.82	0.04				
/ln_sig													1.34	0.02	0.00
sigma													3.82	0.08	3.67

Appendix 3 Hazard functions for different parametric models fitted



Appendix 4 Relevant pages from the Women and Household 2010 MDHS Questionnaires

22 August 2008

DEMOGRAPHIC AND HEALTH SURVEYS
MODEL WOMAN'S QUESTIONNAIRE
WITH HIV/AIDS AND MALARIA MODULES

[NAME OF COUNTRY]
[NAME OF ORGANIZATION]

IDENTIFICATION (1)																			
PLACE NAME _____ NAME OF HOUSEHOLD HEAD _____ CLUSTER NUMBER HOUSEHOLD NUMBER REGION LARGE CITY/SMALL CITY/TOWN/RURAL (2) (LARGE CITY=1, SMALL CITY=2, TOWN=3, RURAL=4) NAME AND LINE NUMBER OF WOMAN _____	<table border="1" style="margin: auto;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>																		
INTERVIEWER VISITS																			
	1	2	3	FINAL VISIT															
DATE	_____	_____	_____	DAY <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>															
INTERVIEWER'S NAME	_____	_____	_____	MONTH <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>															
RESULT*	_____	_____	_____	YEAR <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>															
NEXT VISIT: DATE	_____	_____		INT. NUMBER <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>															
TIME	_____	_____		RESULT <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>															
*RESULT CODES: 1 COMPLETED 2 NOT AT HOME 3 POSTPONED 4 REFUSED 5 PARTLY COMPLETED 6 INCAPACITATED 7 OTHER _____ (SPECIFY)																			
COUNTRY-SPECIFIC INFORMATION: LANGUAGE OF QUESTIONNAIRE, LANGUAGE OF INTERVIEW, NATIVE LANGUAGE OF RESPONDENT, AND WHETHER TRANSLATOR USED																			
SUPERVISOR NAME _____ DATE _____	FIELD EDITOR NAME _____ DATE _____	OFFICE EDITOR 	KEYED BY 																

¹ This section should be adapted for country-specific survey design.

² The following guidelines should be used to categorize urban sample points: "Large cities" are national capitals and places with over 1 million population; "small cities" are places with between 50,000 and 1 million population; remaining urban sample points are "towns".

SECTION 1. RESPONDENT'S BACKGROUND

INTRODUCTION AND CONSENT

INFORMED CONSENT	
Hello. My name is _____ and I am working with (NAME OF ORGANIZATION). We are conducting a national survey that asks women (and men) about various health issues. We would very much appreciate your participation in this survey. This information will help the government to plan health services. The survey usually takes between 30 and 60 minutes to complete. Whatever information you provide will be kept strictly confidential and will not be shared with anyone other than members of our survey team. Participation in this survey is voluntary, and if we should come to any question you don't want to answer, just let me know and I will go on to the next question; or you can stop the interview at any time. However, we hope that you will participate in this survey since your views are important. At this time, do you want to ask me anything about the survey? May I begin the interview now? Signature of interviewer: _____ Date: _____ RESPONDENT AGREES TO BE INTERVIEWED 1 RESPONDENT DOES NOT AGREE TO BE INTERVIEWED ... 2 → END	

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
101	RECORD THE TIME.	HOUR MINUTES	
102	How long have you been living continuously in (NAME OF CURRENT PLACE OF RESIDENCE)? IF LESS THAN ONE YEAR, RECORD '00' YEARS.	YEARS ALWAYS 95 VISITOR 96	→ 104
103	Just before you moved here, did you live in a city, in a town, or in the countryside?	CITY 1 TOWN 2 COUNTRYSIDE 3	Replace 104 with 106 if Questions 104-105 are deleted.
104 (1)	In the last 12 months, on how many separate occasions have you traveled away from your home community and slept away?.	NUMBER OF TRIPS NONE 00	→ 106
105 (1)	In the last 12 months, have you been away from your home community for more than one month at a time?	YES 1 NO 2	
106	In what month and year were you born?	MONTH DON'T KNOW MONTH 98 YEAR DON'T KNOW YEAR 9998	
107	How old were you at your last birthday? COMPARE AND CORRECT 106 AND/OR 107 IF INCONSISTENT.	AGE IN COMPLETED YEARS	
108	Have you ever attended school?	YES 1 NO 2	→ 112
109	What is the highest level of school you attended: primary, secondary, or higher? (2)	PRIMARY 1 SECONDARY 2 HIGHER 3	
110	What is the highest (grade/form/year) you completed at that level? (2)	GRADE/FORM/YEAR	

211 Now I would like to record the names of all your births, whether still alive or not, starting with the first one you had.
 RECORD NAMES OF ALL THE BIRTHS IN 212. RECORD TWINS AND TRIPLETS ON SEPARATE LINES.
 (IF THERE ARE MORE THAN 12 BIRTHS, USE AN ADDITIONAL QUESTIONNAIRE, STARTING WITH THE SECOND ROW).

212	213	214	215	216	217 IF ALIVE:	218 IF ALIVE:	219 IF ALIVE:	220 IF DEAD:	221
What name was given to your (first/next) baby? (NAME)	Were any of these births twins?	Is (NAME) a boy or a girl?	In what month and year was (NAME) born? PROBE: What is his/her birthday?	Is (NAME) still alive?	How old was (NAME) at his/her last birthday? RECORD AGE IN COMPLETED YEARS.	Is (NAME) living with you?	RECORD HOUSE-HOLD LINE NUMBER OF CHILD (RECORD '00' IF CHILD NOT LISTED IN HOUSE-HOLD).	How old was (NAME) when he/she died? IF '1 YR', PROBE: How many months old was (NAME)? RECORD DAYS IF LESS THAN 1 MONTH; MONTHS IF LESS THAN TWO YEARS; OR YEARS.	Were there any other 'live' births between (NAME OF PREVIOUS BIRTH) and (NAME), including any children who died after birth?
01	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (NEXT BIRTH)	DAYS... 1 MONTHS 2 YEARS... 3	
02	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH
03	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH
04	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH
05	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH
06	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH
07	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↙ BIRTH NO... 2 NEXT ↘ BIRTH

212	213	214	215	216	217	218	219	220	221
What name was given to your next baby? (NAME)	Were any of these births twins?	Is (NAME) a boy or a girl?	In what month and year was (NAME) born? PROBE: What is his/her birthday?	Is (NAME) still alive?	How old was (NAME) at his/her last birthday? RECORD AGE IN COMPLETED YEARS.	Is (NAME) living with you?	RECORD HOUSEHOLD LINE NUMBER OF CHILD (RECORD '00' IF CHILD NOT LISTED IN HOUSEHOLD).	How old was (NAME) when he/she died? IF '1 YR', PROBE: How many months old was (NAME)? RECORD DAYS IF LESS THAN 1 MONTH; MONTHS IF LESS THAN TWO YEARS; OR YEARS.	Were there any other live births between (NAME OF PREVIOUS BIRTH) and (NAME), including any children who died after birth?
08	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↗ BIRTH NO... 2 NEXT ↘ BIRTH
09	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↗ BIRTH NO... 2 NEXT ↘ BIRTH
10	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↗ BIRTH NO... 2 NEXT ↘ BIRTH
11	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↗ BIRTH NO... 2 NEXT ↘ BIRTH
12	SING 1 MULT 2	BOY 1 GIRL 2	MONTH YEAR	YES... 1 NO... 2 ↓ 220	AGE IN YEARS	YES... 1 NO... 2	LINE NUMBER (GO TO 221)	DAYS... 1 MONTHS 2 YEARS... 3	YES... 1 ADD ↗ BIRTH NO... 2 NEXT ↘ BIRTH
222	Have you had any live births since the birth of (NAME OF LAST BIRTH)? IF YES, RECORD BIRTH(S) IN TABLE.					YES 1 NO 2			
223	COMPARE 208 WITH NUMBER OF BIRTHS IN HISTORY ABOVE AND MARK: NUMBERS ARE SAME ☐ NUMBERS ARE DIFFERENT ☐ (PROBE AND RECONCILE) CHECK: FOR EACH BIRTH: YEAR OF BIRTH IS RECORDED. FOR EACH BIRTH SINCE JANUARY 2001(1): MONTH AND YEAR OF BIRTH ARE RECORDED. FOR EACH LIVING CHILD: CURRENT AGE IS RECORDED. FOR EACH DEAD CHILD: AGE AT DEATH IS RECORDED. FOR AGE AT DEATH 12 MONTHS OR 1 YEAR: PROBE TO DETERMINE EXACT NUMBER OF MONTHS.								
224	CHECK 215 AND ENTER THE NUMBER OF BIRTHS IN 2001 (1) OR LATER. IF NONE, RECORD '0' AND SKIP TO 226.								

HOUSEHOLD CHARACTERISTICS

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
101	What is the main source of drinking water for members of your household?	PIPED WATER PIPED INTO DWELLING 11 PIPED TO YARD/PLOT 12 PUBLIC TAP/STANDPIPE 13 TUBE WELL OR BOREHOLE 21 DUG WELL PROTECTED WELL 31 UNPROTECTED WELL 32 WATER FROM SPRING PROTECTED SPRING 41 UNPROTECTED SPRING 42 RAINWATER 51 TANKER TRUCK 61 CART WITH SMALL TANK 71 SURFACE WATER (RIVER/DAM/ LAKE/POND/STREAM/CANAL/ IRRIGATION CHANNEL) 81 BOTTLED WATER 91 OTHER 96 	

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP																					
106	Do you do anything to the water to make it safer to drink?	YES 1 NO 2 DON'T KNOW 8	→ 108																					
107	What do you usually do to make the water safer to drink? Anything else? RECORD ALL MENTIONED.	BOIL A ADD BLEACH/CHLORINE B STRAIN THROUGH A CLOTH C USE WATER FILTER (CERAMIC/ SAND/COMPOSITE/ETC.) D SOLAR DISINFECTION E LET IT STAND AND SETTLE F OTHER X (SPECIFY) DON'T KNOW Z																						
108	What kind of toilet facility do members of your household usually use? (4)	FLUSH OR POUR FLUSH TOILET FLUSH TO PIPED SEWER 11 SYSTEM FLUSH TO SEPTIC TANK 12 FLUSH TO PIT LATRINE 13 FLUSH TO SOMEWHERE ELSE 14 FLUSH, DON'T KNOW WHERE 15 PIT LATRINE VENTILATED IMPROVED PIT LATRINE 21 PIT LATRINE WITH SLAB 22 PIT LATRINE WITHOUT SLAB/ OPEN PIT 23 COMPOSTING TOILET 31 BUCKET TOILET 41 HANGING TOILET/HANGING LATRINE 51 NO FACILITY/BUSH/FIELD 61 OTHER 96 (SPECIFY)	→ 111																					
109	Do you share this toilet facility with other households?	YES 1 NO 2	→ 111																					
110	How many households use this toilet facility?	NO. OF HOUSEHOLDS IF LESS THAN 10 0 10 OR MORE HOUSEHOLDS 95 DON'T KNOW 98																						
111	Does your household have: (5)	<table border="0"> <thead> <tr> <th></th> <th>YES</th> <th>NO</th> </tr> </thead> <tbody> <tr> <td>Electricity?</td> <td>1</td> <td>2</td> </tr> <tr> <td>A radio?</td> <td>1</td> <td>2</td> </tr> <tr> <td>A television?</td> <td>1</td> <td>2</td> </tr> <tr> <td>A mobile telephone?</td> <td>1</td> <td>2</td> </tr> <tr> <td>A non-mobile telephone?</td> <td>1</td> <td>2</td> </tr> <tr> <td>A refrigerator?</td> <td>1</td> <td>2</td> </tr> </tbody> </table> [ADD ADDITIONAL ITEMS. SEE FOOTNOTE 5.]		YES	NO	Electricity?	1	2	A radio?	1	2	A television?	1	2	A mobile telephone?	1	2	A non-mobile telephone?	1	2	A refrigerator?	1	2	
	YES	NO																						
Electricity?	1	2																						
A radio?	1	2																						
A television?	1	2																						
A mobile telephone?	1	2																						
A non-mobile telephone?	1	2																						
A refrigerator?	1	2																						

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP																					
118	MAIN MATERIAL OF THE ROOF. (4) RECORD OBSERVATION.	NATURAL ROOFING NO ROOF 11 THATCH/PALM LEAF 12 SOD 13 RUDIMENTARY ROOFING RUSTIC MAT 21 PALM/BAMBOO 22 WOOD PLANKS 23 CARDBOARD 24 FINISHED ROOFING METAL 31 WOOD 32 CALAMINE/CEMENT FIBER 33 CERAMIC TILES 34 CEMENT 35 ROOFING SHINGLES 36 OTHER 96 (SPECIFY)																						
119	MAIN MATERIAL OF THE EXTERIOR WALLS. (4) RECORD OBSERVATION.	NATURAL WALLS NO WALLS 11 CANE/PALM/TRUNKS 12 DIRT 13 RUDIMENTARY WALLS BAMBOO WITH MUD 21 STONE WITH MUD 22 UNCOVERED ADOBE 23 PLYWOOD 24 CARDBOARD 25 REUSED WOOD 26 FINISHED WALLS CEMENT 31 STONE WITH LIME/CEMENT 32 BRICKS 33 CEMENT BLOCKS 34 COVERED ADOBE 35 WOOD PLANKS/SHINGLES 36 OTHER 96 (SPECIFY)																						
120	How many rooms in this household are used for sleeping?	ROOMS																						
121	Does any member of this household own: A watch? A bicycle? A motorcycle or motor scooter? An animal-drawn cart? A car or truck? A boat with a motor?	<table> <thead> <tr> <th></th> <th>YES</th> <th>NO</th> </tr> </thead> <tbody> <tr> <td>WATCH</td> <td>1</td> <td>2</td> </tr> <tr> <td>BICYCLE</td> <td>1</td> <td>2</td> </tr> <tr> <td>MOTORCYCLE/SCOOTER</td> <td>1</td> <td>2</td> </tr> <tr> <td>ANIMAL-DRAWN CART</td> <td>1</td> <td>2</td> </tr> <tr> <td>CAR/TRUCK</td> <td>1</td> <td>2</td> </tr> <tr> <td>BOAT WITH MOTOR</td> <td>1</td> <td>2</td> </tr> </tbody> </table>		YES	NO	WATCH	1	2	BICYCLE	1	2	MOTORCYCLE/SCOOTER	1	2	ANIMAL-DRAWN CART	1	2	CAR/TRUCK	1	2	BOAT WITH MOTOR	1	2	
	YES	NO																						
WATCH	1	2																						
BICYCLE	1	2																						
MOTORCYCLE/SCOOTER	1	2																						
ANIMAL-DRAWN CART	1	2																						
CAR/TRUCK	1	2																						
BOAT WITH MOTOR	1	2																						
122	Does any member of this household own any agricultural land?	YES 1 NO 2	→ 124																					

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
123	How many hectares of agricultural land do members of this household own?	HECTARES 95 OR MORE HECTARES 95 DON'T KNOW 98	
124	Does this household own any livestock, herds, other farm animals, or poultry?	YES 1 NO 2	→ 126
125	How many of the following animals does this household own? (6) IF NONE, ENTER '00'. IF MORE THAN 95, ENTER '95'. IF UNKNOWN, ENTER '98'. Cattle? Milk cows or bulls? Horses, donkeys, or mules? Goats? Sheep? Chickens?	Questions 127-137 are part of the malaria module and should be omitted in countries that do not adopt the module. The malaria module should be used in all malaria endemic countries. CATTLE COWS/BULLS HORSES/DONKEYS/MULES ... GOATS SHEEP CHICKENS	
126	Does any member of this household have a bank account?	YES 1 NO 2	
127	Does your household have any mosquito nets that can be used while sleeping?	YES 1 NO 2	→ 138
128	How many mosquito nets does your household have? IF 7 OR MORE NETS, RECORD '7'.	NUMBER OF NETS	

SECTION 8. HUSBAND'S BACKGROUND AND WOMAN'S WORK

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
801	CHECK 601 AND 602: CURRENTLY MARRIED/ LIVING WITH A MAN ☐ FORMERLY MARRIED/ LIVED WITH A MAN ☐ NEVER MARRIED AND NEVER LIVED WITH A MAN ☐	→ 803 → 807	
802	How old was your husband/partner on his last birthday?	AGE IN COMPLETED YEARS	
803	Did your (last) husband/partner ever attend school?	YES 1 NO 2	→ 806
804	What was the highest level of school he attended: primary, secondary, or higher? (1)	PRIMARY 1 SECONDARY 2 HIGHER 3 DON'T KNOW 8	→ 806
805	What was the highest (grade/form/year) he completed at that level? (1)	GRADE DON'T KNOW 98	
806	CHECK 801: CURRENTLY MARRIED/ LIVING WITH A MAN ☐ FORMERLY MARRIED/ LIVED WITH A MAN ☐ What is your husband's/partner's occupation? That is, what kind of work does he mainly do? What was your (last) husband's/ partner's occupation? That is, what kind of work did he mainly do?	_____ _____ _____	
807	Aside from your own housework, have you done any work in the last seven days?	YES 1 NO 2	→ 811
808	As you know, some women take up jobs for which they are paid in cash or kind. Others sell things, have a small business or work on the family farm or in the family business. In the last seven days, have you done any of these things or any other work?	YES 1 NO 2	→ 811
809	Although you did not work in the last seven days, do you have any job or business from which you were absent for leave, illness, vacation, maternity leave or any other such reason?	YES 1 NO 2	→ 811
810	Have you done any work in the last 12 months?	YES 1 NO 2	→ 818
811	What is your occupation, that is, what kind of work do you mainly do?	_____ _____ _____	
812	CHECK 811: WORKS IN AGRICULTURE ☐ DOES NOT WORK IN AGRICULTURE ☐		→ 814
813	Do you work mainly on your own land or on family land, or do you work on land that you rent from someone else, or do you work on someone else's land?	OWN LAND 1 FAMILY LAND 2 RENTED LAND 3 SOMEONE ELSE'S LAND 4	

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
814	Do you do this work for a member of your family, for someone else, or are you self-employed?	FOR FAMILY MEMBER 1 FOR SOMEONE ELSE 2 SELF-EMPLOYED 3	
815	Do you usually work at home or away from home?	HOME 1 AWAY 2	
816	Do you usually work throughout the year, or do you work seasonally, or only once in a while?	THROUGHOUT THE YEAR 1 SEASONALLY/PART OF THE YEAR 2 ONCE IN A WHILE 3	
817	Are you paid in cash or kind for this work or are you not paid at all?	CASH ONLY 1 CASH AND KIND 2 IN KIND ONLY 3 NOT PAID 4	
818	CHECK 601: CURRENTLY MARRIED/LIVING WITH A MAN ☐ NOT IN UNION ☐		→ 827
819	CHECK 817: CODE 1 OR 2 CIRCLED ☐ OTHER ☐		→ 822
820	Who usually decides how the money you earn will be used: mainly you, mainly your husband/partner, or you and your husband/partner jointly?	RESPONDENT 1 HUSBAND/PARTNER 2 RESPONDENT AND HUSBAND/PARTNER JOINTLY 3 OTHER 6 (SPECIFY)	
821	Would you say that the money that you earn is more than what your husband/partner earns, less than what he earns, or about the same?	MORE THAN HIM 1 LESS THAN HIM 2 ABOUT THE SAME 3 HUSBAND/PARTNER DOESN'T BRING IN ANY MONEY 4 DON'T KNOW 8	→ 823
822	Who usually decides how your husband's/partner's earnings will be used: you, your husband/partner, or you and your husband/partner jointly?	RESPONDENT 1 HUSBAND/PARTNER 2 RESPONDENT AND HUSBAND/PARTNER JOINTLY 3 HUSBAND/PARTNER HAS NO EARNINGS 4 OTHER 6 (SPECIFY)	
823	Who usually makes decisions about health care for yourself: you, your husband/partner, you and your husband/partner jointly, or someone else?	RESPONDENT = 1 HUSBAND/PARTNER = 2 RESPONDENT & HUSBAND/PARTNER JOINTLY = 3 SOMEONE ELSE = 4 OTHER = 6 1 2 3 4 6	
824	Who usually makes decisions about making major household purchases?	1 2 3 4 6	
825	Who usually makes decisions about making purchases for daily household needs?	1 2 3 4 6	
826	Who usually makes decisions about visits to your family or relatives?	1 2 3 4 6	