Time-to-event Analysis Models Including Frailty Effects in Understanding Infant and Child Mortality in Lesotho

by

Sithobile P. Zungu

Thesis Submitted
in Partial Fulfillment of the Requirements for the Degree of

Master of Science
in
Statistics

School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal
Pietermaritzburg
South Africa

May, 2016

Declaration

This dissertation is submitted to the School of Mathematics, Statistics and Computer Science at the University of KwaZulu-Natal, Pietermaritzburg, in fulfillment of the requirements for the degree of Master of Science in Statistics. The thesis presents the original work of the author and has not been otherwise been submitted in any form for any degree to any University. Where use has been made of the work of others it is duly acknowledged in the text.

Author: Miss. Sithobile Zungu	Date
Supervisor: Dr. S. Ramroop	Date
Co-Supervisor: Prof. H.G. Mwambi	Date

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal

Pietermaritzburg

Acknowledgements

Firstly, I would like to acknowledge the permission granted by Measure DHS to use the 2009 Lesotho Demographic Health Survey Data. Secondly, I am thankful to my supervisors Prof. H.G. Mwambi and Dr. S. Ramroop for their tireless guidance, support and patience which led to substantial improvement of this thesis. Last but not least I would like to extend my deepest gratitude to my family and to all my colleagues who supported me in one way or the other through the course of this study. May the almighty God bless you all.

Abstract

This thesis focuses on the determinants of infant and child mortality in Lesotho. It specifically examines how infant and child mortality is related to environmental, demographic and socio-economic factors. A survival analysis approach is used to analyze the determinants of child mortality. Duration or time-to-event models are easily applicable to the problem of child mortality as this class of models is able to account for problems like right-censoring, structural modeling and time varying covariates which other classes of models, such as logistic regression, cannot handle adequately. In this application the age at the child's death is used as the time to event.

Household, environmental, demographic and socio-economic factors are found to have significant impact on child mortality. Policies aimed at achieving the goal of reduced child mortality should be directed on improving the households environmental and / or socio-economic status of a child for this goal to be realized.

Keywords: child mortality, infant mortality, neonatal mortality, duration model, survival analysis, failure function, hazard rate

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Child mortality, commonly on the agenda of public health and international development agencies, has received renewed attention as part of the United Nations Millennium Development Goals (MDGs). Approximately ten million infants and children under five years of age die each year, with large variations in under-five mortality rates, across regions and countries [21].

Childhood mortality rates have declined all over the world in the last fifty-five years. Between the mid 1940s and early 1970s, child death rates, even in the developing countries, reduced significantly (see for example, [5]). A great deal of these gains was achieved through interventions targeted at communicable diseases (diarrhoea, respiratory infections, malaria, measles and other immunizable childhood infections).

However these health gains were short lived. In the mid 1970s the worldwide progress was not maintained and infant mortality rates rose especially in Africa because disease oriented vertical programmes were not effective alone. Maternal, environmental, behavioral and socio-economic factors were recognized as additional important determinants of infant survival. According to UNICEF [7], the decline in child mortality in Africa has been slower since 1980 than in the 1960s and 1970s. Of the thirty countries with the world's highest child mortality rates, twenty-seven are in sub-Saharan Africa. The regions under-five mortality in 1998 was 173 per 1000 live borns [76] compared to the minimum goal of 70/1000 internationally adopted in the 1990 World Summit for Children. 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.

Although enormous literature exist on child mortality, evidence on why infant and child mortality rates remain high in many sub-Saharan African countries despite action plans and interventions made is still scanty.

Environmental risk factors account for about one-fifth of the total burden of disease in low income countries according to recent estimates [68]. Authors in [57] report that among the ten identified leading high-mortality developing countries unsafe water, sanitation and hygiene ranked second, while indoor smoke from solid fuels ranked fourth. About 3 per cent of these deaths (1.7 million) are attributable to environmental risk factors and child deaths account for about 90 per cent of the total.

According to [67], environmental health risks fall into two broad categories. The first are the traditional hazards related to poverty and lack of development, such as lack of safe water, inadequate sanitation and waste disposal, indoor air pollution, and vector-borne diseases. The second category is the modern hazards such as rural air pollution and exposure to agro-industrial chemicals and wastes that are caused by development that lacks environmental safeguards. As the world covers the twenty-first century, debate on childhood mortality remains a big issue for developing countries. Their commitment is reflected in their desire to reduce the level of child mortality by two-thirds of their 1990 levels by the year 2015, as expressed in the Millennium Development Goals. To achieve this goal, it is imperative to determine what factors contribute to the high levels of child mortality in the different developing countries. This study focusses on Lesotho which is one of the high child-mortality countries in the sub-Saharan Africa region.

Several studies have been conducted on infant and child mortality in Lesotho, most of which have used indirect methods like the Trussell technique to estimate child mortality [73]. Some of these studies have also employed multivariate linear and logistic regression to identify the determinants of infant child mortality. However, Ordinary Least Squares (OLS) or binary dependent variable regression models cannot handle the modelling problem of child mortality well because the event of interest is time dependent and the observations may be subject to censoring (and truncation), time varying covariates and structural modeling [35]. This study introduces survival analysis into child mortality modeling in Lesotho. Time to event with age time models are the most suited for such analysis because they account for problems like right-censoring, structural modeling and time varying covariates which traditional econometric techniques cannot handle adequately.

1.1 Background

Globally, reducing child mortality and hence improving child health are two of the major concerns of development agencies and the international public health community [2]; [7]; [2]. The infant mortality rate is a good indicator of a country's health status and socioeconomic development [36]; [60]. According to [60], this is due to its sensitivity to structural transformations that affect the health of the entire population, such as disease epidemics and economic

development, and to other changes that affect general living conditions, such as social well-being and the quality of the environment. There is evidence that there has been a large decline in under-five mortality for all regions in the world, but sub-Saharan Africa still has the highest rates of both infant and under-five mortality, and this is associated with low levels of development in the region [74]. Another factor that explains why child mortality is high in sub-Saharan Africa is that the health transition in the region began later than in other parts of the world [24]. Identification of factors that contribute most to the high rates of childhood mortality is of great importance both to policy makers and to development agencies.

1.2 Country Background

Lesotho is a small landlocked, mountainous country that is completely surrounded by the Republic of South Africa (RSA). With a total population of about 2.2 million, the country is about $3,000$ square kilometres. Three-quarters of the land is made up of highlands and the remaining one-quarter is lowlands. However, the low land is home to over 55% of the population. Lesotho is a constitutional monarchy with the King as the Head of State and the Prime Minster as Head of Government and a dual legal system consisting of customary law and the common law. Political stability has been achieved through the adoption of a relatively more inclusive electoral system as of 2002. At present, the three major challenges facing the country are extremely high unemployment and HIV/AIDS prevalence rate coupled with a high degree of food insecurity all of which exacerbate poverty, gender inequality and erode the considerable gains that Lesotho has made in human development.

Until recently, Lesotho had registered relatively high rates of social development as illustrated by a relatively high life expectancy, high literacy and net primary school enrolment but with a reversed gender gap indicating disparity to the advantage of girls. The country was ranked 127 out of 174 countries in 1978, but its ranking declined to 145th out of 175 by the 2004 Human Development Index, indicating significant erosion in human development achievements. The number of people living below the poverty line is estimated to be around 60%. With a population growth rate of 2.1%, economic growth rate must at least be as much, if Lesotho is to maintain the existing standard of living and be able to respond to the growing demand for social services. The key socio-economic indicators are: GNP per capita was 570 dollars in 2003; life expectancy 45 years for men and 45.6 for women (2000); fertility rate of 4.1 children born per woman (2001); HIV/AIDs prevalence rate of 31% (2002); 63% of the population with access to safe drinking water; a net primary enrolment rate of 85%; under five morality ratio of $132/100,000$ and maternal mortality ratio of 419(2001).

The 1990*s* presented both a looming threat and new opportunities to Lesotho. For a long time, nearly half of Lesotho's male labour force was employed in the

mines of the RSA and contributed more than 20% of the country's gross national product (GNP) in the form of remittances. In late 1990s, the country experienced a short-lived economic boom that was triggered by the construction of the Lesotho Highland Water Project (LHWP) that was geared to exporting water to the RSA and providing hydroelectricity to the domestic market. Thirdly, Lesotho's ability to attract foreign direct investment (FDI) and engage in manufactured textile exports created a sizable employment. The combination of these favourable factors led to an average GDP growth of 3.5% between $1991 - 2000$. As a result, the country has witnessed an annual growth of 7% in rural-urban migration during the last decade, a trend that has put considerable pressure on social services.

1.2.1 Poverty Profile and Gender

Out of an estimated population of 2.2 million, women are 50.6% of the population. More than half of the population of Lesotho is under 18 years with 42% being under 14. Lesotho is one of the poorest countries in the world with an estimated 60% of the population living below the poverty line. All available information indicates that poverty is concentrated in rural areas and the mountain areas of Lesotho, which are home to approximately one third of the population, are significantly poorer. The Premium Rate Service (PRS) highlights geography as the greatest determinant of poverty followed by gender. Altitude, climate, soil quality and ease of communication affect the geographical distribution of poverty. Three quarters of the land is made up of highlands rising to nearly 3500 meters in the Drakensburg Maluti Mountain range. The overall climate is harsh but much more severe in the highlands with heavy snowfalls that cut off access to most of the mountain areas. Such conditions deprive the population of access to social services such as education, health, markets, and other inputs and resources for economic activities. Poverty is also on the increase in urban areas. Sectors of the population most at risk of poverty include those who depend heavily on subsistence agriculture for their livelihood, the youth, orphans and old women. Unemployment, which is estimated to have reached 40%, is also a major cause of poverty in Lesotho. Unlike other countries, it is not limited to urban areas due to the migratory labor system and the recent retrenchment of Basotho mine workers.

Following the rationalization of the mining production in the RSA, a process of retrenchment of Basotho labor began in the late 1980s but accelerated at a much more rapid rate during the 1990s culminating in an estimated unemployment rate of 40.5% by the end of 2000, then accompanied by a significant decline in miners' remittances. The export-led manufacturing strategy was bolstered by access to international textile and clothing markets, particularly in the United States under the Africa Growth and Opportunities Act (AGOA). Lesotho became a top exporter of apparel to the US in 2003. However, the recent period has been marked by declining rate of FDI and the closing down of a number of factories following the expiry of the Multi-Fibre Agreement under the World

Trade Organization (WTO). Consequently, Lesotho has been hit with another wave of unemployment, this time mostly female labor. The second and most alarming threat is the estimated 31% HIV/AIDS prevalence placing Lesotho as a country with one of the highest prevalence rates of HIV/AIDS in the world while facing a growing threat of food insecurity partly as a result of prolonged drought

Extreme poverty is concentrated in the rural areas, particularly the Mountain areas where 71% of the population live below the poverty line. Women in both rural and urban areas make up a predominant proportion of the poor. It is estimated that 30.7% of households are headed by women. The incidence of poverty among female-headed households is persistently high with approximately 64% well above the national average of 58% and a male headed average of 57 per cent. A large proportion of female-headed households are vulnerable to poverty because they lack agricultural assets due mostly to cultural beliefs and practices coupled with limited access to social services that increases their workload. While the available information tends to limit the gender profile of poverty to female-headed households, not all female-headed households are poor, while large numbers of women in male headed households are poor because they lack access and control of household resources and decisions. Women are also impoverished by discriminatory laws and most rural women are not aware of their legal rights.

Poor nutrition is also a major feature of poverty. The poor in Lesotho have a high degree of dependence on food purchase amounting to $45 - 60\%$ of their annual kilocalorie needs. About 25% of the total population are undernourished, 15% of children are underweight while 31% are stunted indicating the existence of a significant chronic food insecurity. Poverty and household food insecurity in Lesotho reveal a strong gender dimension. Women are reported to have earned 30.9% of the total national income despite their higher mean years of school while men earned 69.1%. While the advent of HIV/AIDS appears to be the most devastating and impoverishing force facing Lesotho, it has exacerbated the vulnerability of poor households especially the women to both income and non-income poverty.

Unlike other African countries, there is a reverse gender gap that favors girls and women when it comes to education. Women are more literate compared to men, while girls and women also enjoy higher net primary, and secondary and tertiary enrolment rates. The relatively higher female educational attainment has not automatically translated into higher income for women because of cultural and social norms that prevent them from having access and control to productive resources and the type of skills that they acquire. A majority of women also experience time poverty due to their heavy work load that combines household management, child care as well as income earning activities. Household management includes the time and energy intensive tasks of fetching water and fuel and food processing in a context where these services are either inadequate or do not exist. These multiple gender disadvantages also trigger the intergenerational

transfer of poverty as evidenced by poor social development indicators such as high child mortality and morbidity and low educational attainment. Gender aware poverty reduction is one that adopts a multi-pronged approach geared towards addressing cultural and legal barriers that prevent women's access to strategic resources while also reducing their work burden.

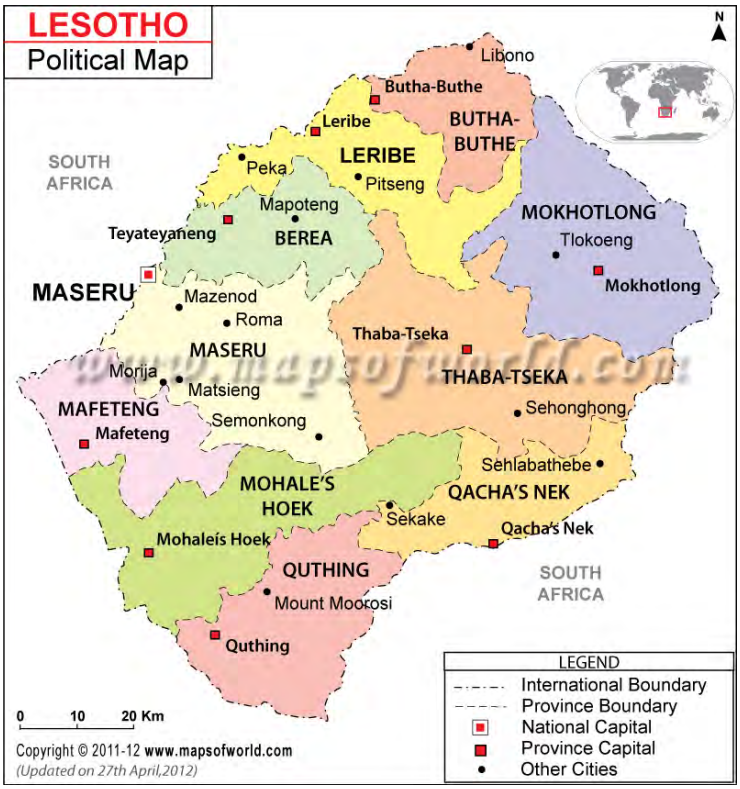


Figure 1.1: The map of Lesotho

1.3 Problem Statement

The environment, which sustains human life, is also a profound source of ill health for many of the world's people. In the least developed countries, one in five children does not live to see their fifth birthday, mostly because of avoidable environmental threats to health. This translates into approximately 11 million avoidable childhood deaths each year [83]; [56]. Hundreds of millions of others, both children and adults, suffer ill health and disability that undermine their quality of life and hopes for the future. These environmental health threats, arguably the most serious environmental health threats facing the world population today, stem mostly from traditional problems long since solved in the wealthier countries, such as a lack of clean water, sanitation, adequate housing,

and protection from mosquitoes and other insect and animal disease vectors.

Poverty also influences health because it largely determines an individual's environmental risks, as well as access to resources to deal with those risks. Throughout the developing world, the greatest environmental health threats tend to be those closest to home. Many in these countries live in situations that imperil their health through steady exposure to biological pathogens in the immediate environment. More than 1 billion people in developing countries live without adequate shelter or in unacceptable housing. A further 1.4 billion lack access to safe water, while another 2.9 billion people have no access to adequate sanitation [56], all of which are essential for good hygiene. Unable to afford clean fuels, the poor largely rely on biomass fuels for cooking and heating. Inside the smoky dwellings of developing countries, air pollution is often higher than outdoors in the world's most congested cities.

As already mentioned, the health indicators for Lesotho are generally not improving as there are some formidable challenges to be addressed. The infant mortality rate has remained largely unchanged from 2005 to 2012 (84/1000 and 74/1000, respectively); under-five mortality has gone from 108 to 100/1000 while the maternal mortality ratio was estimated to have increased from 2005 to 2012, a trend which is opposite to the expectations of the Millennium Development Goal 5 (Lesotho Demographic and Health Survey 2009). Reducing child mortality is the fourth Millennium Development Goal, whose target is to reduce the under-five mortality rate by two-thirds between 1990 and 2015. Despite numerous interventions and action plans, very little evidence exists on why the infant and child mortality rates are increasing in Lesotho. If Lesotho is committed to achieving the MDG on child mortality, it is prudent to understand clearly the factors that are contributing to the high levels of mortality. This study therefore explores the household environmental and socio-economic characteristics and their effect on child and infant mortality in Lesotho.

1.4 Objectives of the Study

The general aim of the study is to explore the relationship between households' environmental, demographic factors and socio-economic characteristics on child mortality. The specific objectives are:

- To use survival analysis or time to event models to understand factors associated with child mortality in Lesotho.
- To identify the environmental determinants of child mortality, controlling for other covariates.

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

- Households' access to safe water has no effect on child mortality.

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

The rest of the thesis is structured as follows: The literature on child mortality in Chapter 2. This is where the findings by other studies are presented. Chapter 3 discusses survival analysis methods. Regression models in survival analysis such as Cox PH hazard model, residuals for Cox regression model, are discussed in chapter 4. Chapter 5 represents parametric models such as Exponential PH model, Weibull PH model, etc. Frailty models are represented in Chapter 6, while Chapter 7 gives results from exploratory. Chapter 7 is about the application to the DHS under five mortality using 2009 LDHS data, and how the dependent variables were derived from the data sets. Chapter 8 the results from the models. Finally, Chapter 9 gives the discussion, and conclusions

Chapter 2

Literature Review

There is a relatively large literature that focuses on the determinants of child mortality (for a survey, see [82]). Theoretical frameworks are often presented as health production functions, which capture the structural relation between health outcomes, and the household behavioral variables, like nutrition, breastfeeding, child spacing, etc. (see [65]). In the framework of a health production function, child mortality risks depend on both observed health inputs and unobserved biological endowment or frailty. Not properly taking account of these unobserved characteristics or the outcomes within a family or cluster may lead to inconsistent and inefficient estimators (for example, see [61]).

There are a number of different analytical frameworks through which to view the effects of different determinants on childhood mortality. Demographic research by Mosley and Chen [51] and by Schultz [65] made the distinction between variables considered to be exogenous or socio-economic (i.e. cultural, social, economic, community, and regional factors) and endogenous or biomedical factors (i.e. breastfeeding patterns, hygiene, sanitary measures, and nutrition). The effects of the exogenous variables are considered indirect because they operate through the endogenous biomedical factors. Likewise, the bio-medical factors are called intermediate variables or proximate determinants because they constitute the middle step between the exogenous variables and child mortality ([34];[51];[65].

Mosely [51] were among the first to study the intermediate biomedical factors affecting child mortality, labeled 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.

Several studies have been carried out on infant and child mortality using census and survey data. In Lesotho, all of these studies have used indirect methods, mostly the Trussel technique, Preston method and Coale-Demeny model life

table to estimate child mortality [71].

For instance in other countries, [33] and [53] combine the Trussell technique for estimating child mortality based on the Coale-Demeny model life table with multivariate linear regression; [79] employs the Trussell-Preston methods and multivariate regression analysis to calculate mortality indices for each woman; [55] utilizes the Coale and Trussel technique as well as multiple regression analysis using census data to estimate mortality; [20] and [58] also employ the Trussell technique while [38] uses cross-tabulation and regression analysis. All these studies, which use either the DHS or census data to measure the effect of socio-economic, environmental or demographic covariates on child mortality, find demographic, socio-economic and environmental factors (type of toilet facility, type of bathing facility, source of drinking water) to be significantly related to infant and child mortality.

In other countries, [5] and [21] use indirect methods to estimate levels and trends of mortality in that country. Although the results from former studies indicate that owning a pit latrine does not have a significant effect on child mortality (which is explained by the argument that just because a household has sanitation facilities does not mean that it will be used hygienically or by all members of the household), the latter results indicate that the source of drinking water and sanitation facilities are strong predictors of infant mortality.

[81] employs a logistic regression to examine the effect of some environmental and socio-economic factors that determine childhood diarrhoea in Eritrea, using data from the 1995 Eritrea Demographic and Health Survey (EDHS). The results show that the type of floor material, household economic status and place of residence are significant predictors of diarrhoea. Similarly [72], in a comparative study of rural areas of Ghana, Egypt, Brazil and Thailand, find out that children's health is affected by environmental conditions and the economic status of the household.

Time to event modeling is applied by [26] to assess the impacts of water and sanitation on child mortality in Egypt. 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 hazard rate framework is elegantly utilized by [78], in which a flexible parametric framework for analyzing infant and child mortality is developed. Their model predicts that a significant number of deaths of children under five years can be averted by providing access to electricity, improving the education of women, providing sanitation facilities and reducing indoor air pollution. In particular, reducing indoor air pollution and increasing the educational level of women might have substantial impacts on child mortality. In a related study, [78] examine the linkages between child mortality and morbidity, and the quality of the household and community environment in rural China using a compet-

ing risks approach. The key findings are that: (1) the use of unclean cooking fuels (wood and coal) significantly reduces the neonatal survival probability in rural areas; (2) access to safe water or sanitation reduces child mortality risks by about 34 per cent in rural areas; (3) a higher maternal education level reduces child mortality and female education has strong health externalities; (4) access to safe water/sanitation and immunization reduces diarrhoea incidence in rural areas, while access to modern sanitation facilities (flush toilets) reduces diarrhoea prevalence in rural areas; (5) significant linkages between Acute Respiratory Infections (ARI) incidence and use of unclean cooking fuels are found using the city level data constructed from the survey.

[78], using the results from the 2000 Ethiopia DHS, examines the environmental determinants of child mortality by constructing three hazard models (the Weibull, the piecewise Weibull and the Cox model) to examine three age-specific mortality rates: neonatal, infant, and under-five mortality by location (urban/rural), female education attainment, religion affiliation, income quintile, and access to basic environmental services (water, sanitation and electricity). The estimation results show a strong statistical association between child mortality rates and poor environmental conditions.

There is general consensus in the literature that a household's socio-economic and environmental characteristics do have significant effects on child and infant mortality. This is true for studies which employ both direct and indirect techniques to estimate infant and child mortality.

As observed in most studies, a household's income has a significant effect on the survival prospects of children. Higher mortality rates are experienced in low income households as opposed to their affluent counterparts.

The mother's level of education is strongly linked to child survival. Higher levels of educational attainment are generally associated with lower mortality rates, since education exposes mothers to information about better nutrition, use of contraceptives to space births, and knowledge about childhood illnesses and treatment. Larger differences have been found to exist between the mortality of children of women who have attained secondary education and above and those with primary level of education or less.

On the households environmental characteristics, safe source of drinking water supply has a significantly negative effect on child mortality. The same holds true for those with sanitation, which in most cases is taken to be access to a flush toilet or a ventilated improved pit latrine.

Differentials by urban/rural residence have commonly been observed, with urban areas having more advantages and therefore better child survival prospects.

As concerns the demographic variables, the patterns of mortality by mater-

nal age and birth order are typically U-shaped [44]. Children born to both relatively old and young women have higher mortality rates than others; the interpretation of the effect of maternal age at birth on infant mortality must be biological, i.e., it depends on reproductive maturity. Moreover, first and higher order births also have higher mortality rates since the birth order reflects the components of the child biological endowment. As for the child's gender, it is widely believed that male mortality is higher due to biological disadvantages. Twins face a higher mortality risk.

Chapter 3

Survival Analysis

Survival analysis is a statistical method that deals with time to event data [80]. According to [31] the name survival data arose because originally events were most often deaths. This term is now used for all kind of time to event data. In all cases the event can be viewed as a transition from one state to another [43].

The main concept in survival analysis is about modelling the time to event, also called the failure time. Survival time is defined as the length of time that is measured from a specified time origin to the time the event of interest occurred [27]. When determining the survival time it is important to know the time origin (starting point), a scale for measuring the passage of time must be agreed upon, and the definition of the event (often called failure) must be entirely clear [43].

Difficulty in survival analysis arises when some individuals have experienced the event while others have not had the event by the end of study and thus their actual survival times are unknown, which lead to the concept of censoring.

Censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly [66]. According to [43] there are three types of censoring; the first one is right censoring, which occurs if the event occurs after the observed survival time. It follows that right censored survival time is less than the actual survival time. Secondly, left censoring, which occurs if we observe the presence of a condition but do not know where it began; the actual survival time here is less than the observed censoring time. The last one is the interval censoring which occurs when the individual is known to have experienced an event within an interval of time but the actual survival time is not known.

An important assumption for methods presented here for the analysis of censored survival data is that the individuals who are censored are at the same risk of subsequent failure as those who are still alive and uncensored. This implies

that a subject whose survival time is censored at time C must be representative of all other individuals who have survived to that time. If this is the case, the censoring process is called non-informative [80]. Statistically, if the censoring process is independent of the survival time, if

$$P(X \geq x, C \geq x) = P(X \geq x)P(C \geq x)$$

where X is the actual time to event. Thus independent censoring is a special case of non-informative censoring [80].

3.1 The Survivor Function and the Hazard Function

Suppose we have a group of individuals with survival times $t_1, t_2, \ldots, t_N$ some of which may be censored. These values can be regarded as the values of continuous variable T, which has probability density function $f(t)$, and cumulative distribution function $F(t)$ [80], where $F(t)$ is given by

$$F(t) = P(T < t) = \int_0^t f(u)du.$$

This represents the probability that the survival time is less than some value t [9]. The survivor function, which represents the probability that an individual will survival beyond t, is given by

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

Since the survival distributions are usually skewed and there may be censored observations, the mean and the variance are not used to summarize the distribution of T, but rather the median and quantiles are used. These can be estimated from the survival function [43]. For example the median survival time is that value t_m of T satisfying $S(t_m) = 0.5$

For a continuous random variable T [80], the probability density function(pdf) is given by

$$f(t) = F'(t) = -S'(t), t \geq 0.$$

and the hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t which is given by

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t};\ t \geq 0. \tag{3.1}$$

By using conditional probability laws, and the mathematical definition of derivatives, the above equation can be rewritten in the following manner.

$$\begin{aligned} h(t) &= \lim_{\delta t \to 0} \frac{P(t \leq T \leq t + \delta t)}{\delta t P(T \geq t)} \\ &= \lim_{\delta t \to 0} \left[\frac{F(t + \delta t) - F(t)}{\delta t} \right] \frac{1}{P(T \geq t)} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

The relationship of $S(t)$ and $h(t)$ is given below

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}.$$

Likewise,

$$S(t) = \exp\left[-\int_0^t h(u)du \right] = \exp(-H(t)),\ t \geq 0,$$

and $H(t) = \int_0^t h(u)du$ is called cumulative hazard function, which can be obtained from the survival function since $H(t) = -\log S(t)$ while the probability density function can be given by

$$f(t) = h(t) \exp\left[-\int_0^t h(u)du \right],\ t \geq 0.$$

All these functions give a mathematical equivalent specification of the distributions of the survival time T. If one of them is known, then the other two can be determined. The survival function is most useful for comparing the survival progress of two or more groups. The hazard function gives a more useful description of the risk of failure at any time point. It is the instantaneous probability of failure at a given time given the individual survived up to that time.

The methods of estimation can be broadly grouped into parametric and non-parametric methods. Other methods such as the semi-parametric approach due to Cox [10] (namely the Cox proportional hazards model) have also been developed. These methods will be briefly explained in this project.

The survival function is always a decreasing function while the hazard function is always an increasing function.

3.2 Types of Survival Distributions: Parametric Distribution

Survival data are usually right skewed or skewed to the right, thus symmetric distributions such as the Normal are not useful in modelling such data [43].

Typically asymmetric distributions are the exponential, Weibull and log-logistic distributions [80]. Only the exponential and the Weibull model will be briefly discussed in this section. The aim is to derive some basic relationships when specific survival distributions are assumed.

3.2.1 Exponential Distribution

The exponential distribution is characterized by the following probability density function (p.d.f)

$$f(t;\lambda) = \lambda e^{-\lambda t}, \quad t > 0$$

The cumulative distribution function (c.d.f) is given by

$$F(t) = 1 - e^{-\lambda t}$$

and the survivor function is

$$S(t) = 1 - F(t) = e^{-\lambda t}.$$

while hazard function is given by

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{\lambda t}} = \lambda \text{ (constant)}.$$

From this it can be seen that the exponential distribution has a constant hazard, which means that the risk of death is independent of time, which is quite an unrealistic assumption, because intuitively the risk of death may increase or decrease as an individual ages, for example.

According to [45] an important property of the exponential distribution is the lack of memory property. Suppose that the random variable T is associated with survival time, and is exponentially distributed with parameter λ. Consider the probability that an individual survives for a time greater than t_1 given that he or she has survived up until time t_0. Then

$$\begin{aligned} P(T > t_1 | T > t_0) &= \frac{P(T > t_1 \text{ and } T > t_0)}{P(T > t_0)} \\ &= \frac{P(T > t_1)}{P(T > t_0)} \\ &= \frac{S(t_1)}{S(t_0)} \\ &= \frac{e^{-\lambda t_1}}{e^{-\lambda t_0}} \\ &= e^{\lambda(t_1 - t_0)}. \end{aligned}$$

This can be interpreted in the following manner. Given survival to time t_0, the excess life beyond t_0 still has the exponential distribution with parameter λ. This

result also explains why the exponential distribution may not be such a realistic distribution for modeling time-to-event data [37]. However, since the equation is simple and the calculations relatively easy, this model can be appealing in certain circumstances and for also explaining basic properties of time to event data.

The following is the hazard plot for exponentially distributed time-to-event data.

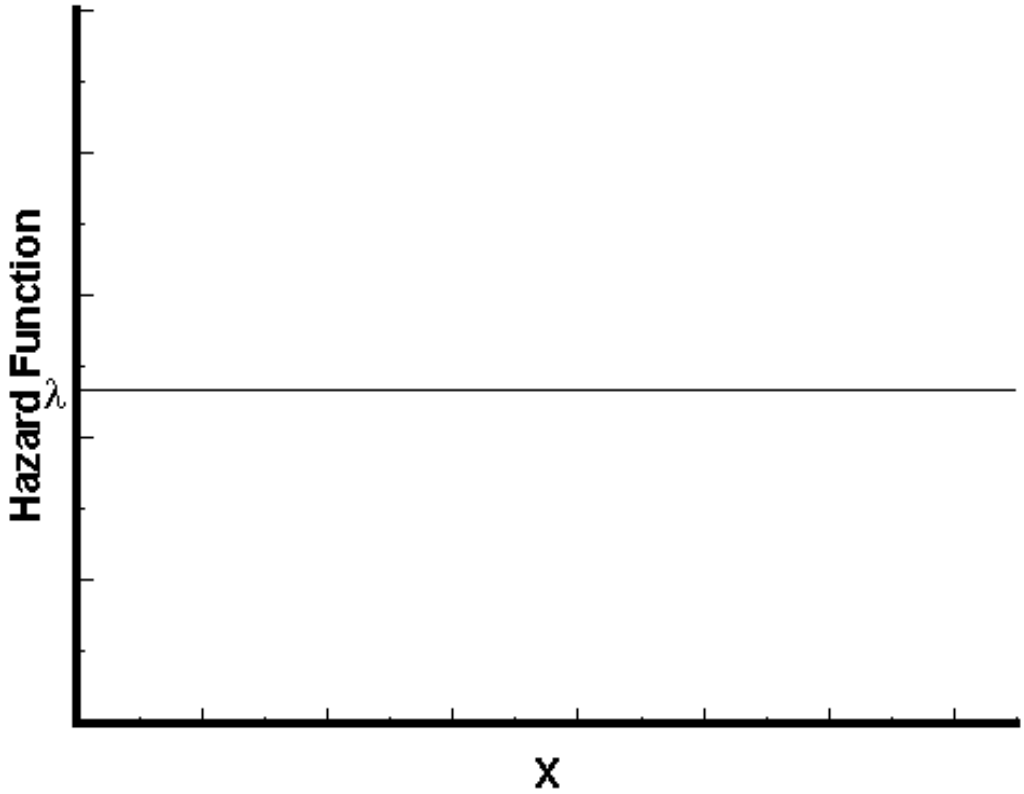


Figure 3.1: The hazard plot assuming exponentially distributed time to event variable

Figure 3.1 shows that exponential hazard function is always constant with no flexibility at all. Thus such a distribution ought to be used with caution because in reality it is an almost unattainable assumption.

3.2.2 Weibull Distribution

The two parameter p.d.f. for the Weibull distribution is given by

$$f(t;\gamma,\lambda) = \lambda\gamma t^{\gamma-1}e^{-\lambda t^{\gamma}}, \quad t > 0.$$

Here γ is known as the shape parameter, and δ is the scale parameter [80]. Note that when $\gamma = 1$ the Weibull distribution reduces to the exponential distribution with parameter δ. The c.d.f. of the Weibull distribution is given by

$$F(t) = 1 - e^{-\lambda t^{\gamma}}, \quad t > 0$$

and thus the corresponding survivor function is

$$S(t) = e^{-\lambda t^{\gamma}}$$

and hence the hazard function is

$$h(t) = \frac{f(t)}{S(t)} = \lambda \gamma t^{\gamma - 1}.$$

Clearly for $\gamma \neq 1$, the hazard is not constant, in contrast to the exponential distribution. The hazard function takes a different shape depending on the shape parameter γ, as summarized in the following table:

Table 3.1: Hazard function for different values of γ

Values of γ	Shape of $h(t)$
$0 < \gamma < 1$	Exponential decay
$\gamma = 1$	Constant ($h(t) = \delta$)
$\gamma = 2$	Straight line
$\gamma > 2$	Exponential growth

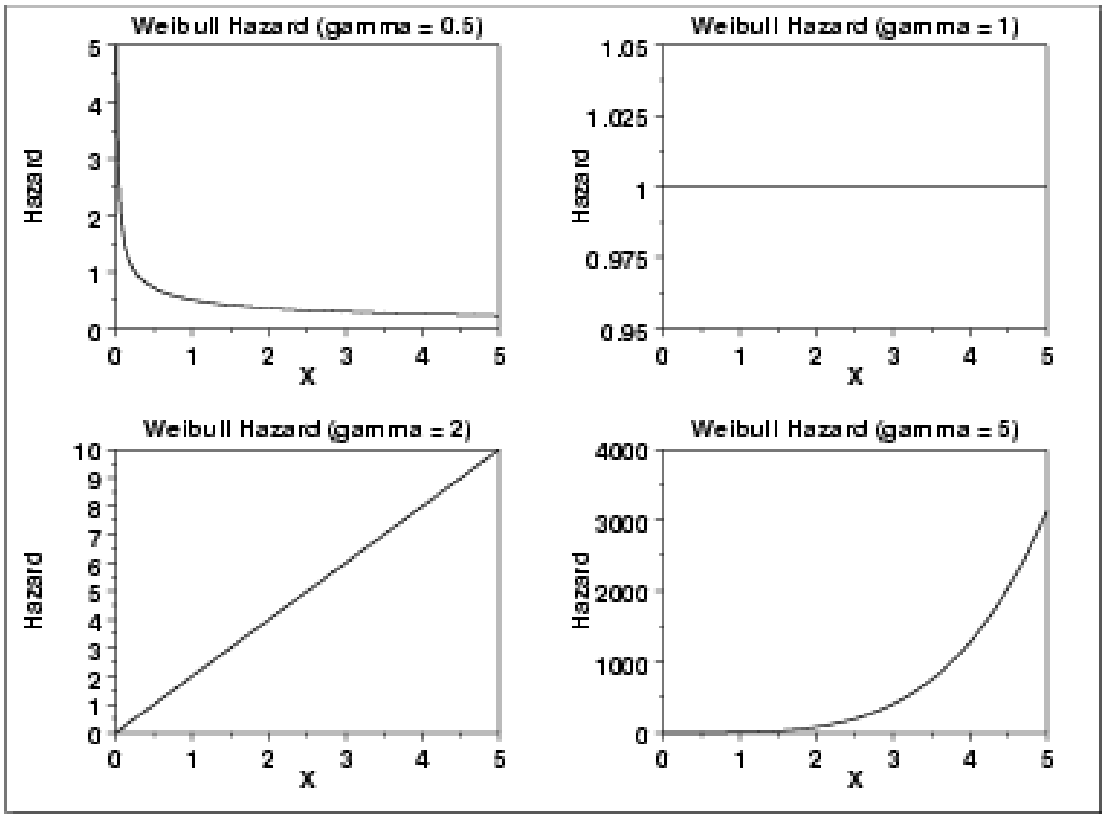


Figure 3.2: The hazard plot for parametric data, representing the Weibull function

Figure 3.2 shows that the Weibull hazard function can be constant, increasing or decreasing. Other continuous distributions that can be used are log-normal, log-logistic, gamma and other non-measure distributions for continuous distribution.

3.3 Non-Parametric Methods

In this subsection we briefly describe two common non-parametric methods that are used to explore and describe time to event data. These are the Kaplan-Meier estimate of the survival function and the log-rank test [48] used to compare two groups with time to event outcomes or observations.

3.3.1 The Kaplan-Meier Estimate for the Survivor Function

Let

$$t_{(1)} < t_{(2)} < \ldots < t_{(m)},$$

denote the distinct ordered actual times of death of m individuals out of a sample of N individuals (not counting censoring times). Let d_i be the number of deaths at $t_{(i)}$, and let n_i be the number alive just before $t_{(i)}$. This is the number exposed to risk at time $t_{(i)}$. By [43] the Kaplan-Meier or product limit estimate of the survivor function is

$$\hat{S}(t) = \prod_{i:t_{(i)}<t} \left(1 - \frac{d_i}{n_i}\right). \tag{3.2}$$

Likewise, to survive to time t an individual must first survive to $t_{(1)}$. The individual must then survive to $t_{(2)}$ given that he/she had already survived to $t_{(1)}$ and so on. Note we assume there are no deaths between $t_{(i-1)}$ and $t_{(i)}$, therefore the probability of dying between these times is zero. The conditional probability of surviving at time $t_{(i)}$ is the complement of $\frac{d_i}{n_i}$, that is $\left(1 - \frac{d_i}{n_i}\right)$. The overall unconditional probability of surviving to t is obtained by multiplying the conditional probabilities for all relevant time intervals up to t.

Non-Parametric Maximum Likelihood

Consider the likelihood contribution of a case that experiences an event or censored at time t_i. Taking c_i to represent the number of cases censored between $t_{(i-1)}$ and $t_{(i)}$, and taking d_i be the number of cases which die or experience the event at $t_{(i)}$ [27]. Then the likelihood function takes the following form which is,

$$L = \prod_{i=1}^{m} \left[S(t_{(i-1)}) - S(t_{(i)})\right]^{d_i} \left[S(t_{(i)})\right]^{c_i} \tag{3.3}$$

where the product is over the m distinct, and taking $t_{(0)} = 0$ with $S(t_{(0)}) = 1$. According to [43] to estimate m parameters representing the values of the survival function at the death times $t_{(1)}, t_{(2)}, \ldots, t_{(m)}$, the conditional probability of surviving from $S(t_{(i-1)})$ to $S(t_{(i)})$ is denoted by $\pi_i = \frac{S(t_{(i)})}{S(t_{(i-1)})}$. Thus $S(t_{(i)})$ can be written as

$$S(t_{(i)}) = \pi_1 \pi_2 \ldots \pi_i$$

and the likelihood becomes

$$L = \prod_{i=1}^{m} (1 - \pi_i)^{d_i} \pi_i^{c_i} \left(pi_1 \pi_2 \ldots \pi_{i-1}\right)^{d_i + c_i}$$

taking into account that all cases who die at $t_{(i)}$ or are censored between $t_{(i)}$ and $t_{(i+1)}$ contribute a term π_j to each of the previous times of death from $t_{(1)}$ to $t_{(i-1)}$. In addition, those who die at $t_{(i)}$ contribute $1-\pi_i$, and the censored cases contribute an additional π_i. Let $n_i = \sum_{j>i} (d_j + c_j)$ denote the total number exposed to risk at $t_{(i)}$. Thus, collecting the terms on each π_i, the likelihood becomes

$$L = \prod_{i=1}^{m} (1 - \pi)^{d_i} \pi_i^{n_i - d_i} \tag{3.4}$$

a binomial likelihood. The maximum likelihood estimator of π_i is then given by

$$\hat{\pi}_i = \frac{n_i - d_i}{n_i} = 1 - \frac{d_i}{n_i}.$$

The K-M estimator follows from multiplying these conditional probabilities.

Greenwood's formula

As shown from the likelihood obtained above, it follows that the large sample variance of $\hat{p}_i$ conditional on the data n_i and d_i is given by the usual binomial formula, as

$$Var(\hat{\pi}_i) = \frac{\pi_i\,(1-\pi_i)}{n_i}. \tag{3.5}$$

assuming the $cov(\hat{\pi}_i, \hat{\pi}_j) = 0$ for $i \neq j$, that is the covariances of the contributions from different times of death are all zero. This can be verified by taking logs and then first and second derivatives of the log-likelihood function. To obtain the large sample variance of $\hat{S}(t)$, the K-M estimate of the survival function, we apply the delta method twice [43], starting by taking logs so that instead of the variance of a product we can find the variance of a sum, working with

$$K_i = \log \hat{S}(t_{(i)}) = \sum_{j=1}^{i} \log \hat{\pi}_j.$$

To find the variance of the log of $\hat{\pi}_i$, we apply the delta method for the first time, so the large-sample variance of a function f of a random variable X is

$$Vax(f(x)) = (f'(X))^2 var(X).$$

Thus, for log function, the variance becomes

$$Var(log\hat{\pi}_i) = \left(\frac{1}{\pi_i}\right)^2 var(\pi_i) = \frac{1-\pi_i}{n_i\pi_i}.$$

For the reason that K_i is a sum and the covariances of the $\pi_i's$ and hence of the $\log \pi_i's$ are zero, we find that

$$Var(\log \hat{S}(t_{(i)})) = \sum_{j=1}^{i} \frac{1-\pi_j}{n_j\pi_j} = \sum \frac{d_j}{n_j(n_j-d_j)}. \tag{3.6}$$

Using delta method again, this time to get the variance of the survivor function from the variance of its log, we get

$$Var(\hat{S}(t_{(i)})) = \left[\hat{S}(t_{(i)})\right]^2 \sum_{j=1}^{i} \frac{1-\hat{\pi}_j}{n_{j\hat{\pi}_j}}. \tag{3.7}$$

This result is known as Greenwood's formula.

3.3.2 Mantel-Haenszel

Consider the problem of comparing two or more survivor functions, for example urban versus rural in Lesotho. Let

$$t_{(1)} < t_{(2)} < \ldots < t_{(m)}$$

denote the distinct times of death observed in the total sample, obtained by combining all groups of interest. Let

d_{ij} = death at time $t_{(i)}$ in group j, and,

n_{ij} = number at risk at time $t_{(i)}$ in group j,

d_i = total number of deaths, and

n_i = child at risk at time $t_{(i)}$.

If the survival probabilities are the same in all groups, then the d_i deaths at time $t_{(i)}$ should be distributed among the k groups in proportion to the number at risk. Thus, conditional on d_i and n_{ij},

$$E(d_{ij}) = d_i \frac{n_{ij}}{n_i}$$

where the last term shows that we can also view this calculation as applying an overall failure rate $\frac{d_i}{n_i}$ to the n_{ij} subject in group j.

The Mantel-Haenszel statistic (log-rank test) tests the null hypothesis that the risk or hazard of death is the same in the two groups or more groups. Without loss of generality we consider the case of two groups [43]. In other words, if the study was about comparing the hazard of death (child mortality as in the current study) in the rural and the urban areas of Lesotho, the null hypothesis would be that there is no different in the risk of death of a child between the two groups. The test is described in detail below.

Suppose the two groups are denoted by 1 and 2 the urban and the rural areas respectively, and that there are k distinct times, $t_1 < t_2 < \ldots < t_k$, across the two groups. The test uses a conditional argument based on the number at risk of failing just prior to each observed failure time. Suppose that at time t_i there are d_i deaths and n_i at risk in total, with d_{i1} and d_{i2} deaths and n_{i1} and n_{i2} at risk in group 1 and 2 respectively such that $d_{i1} + d_{i2} = d_i$ and $n_{i1} + n_{i2} = n_i$. At each death time t_i such data can be summarized in a 2×2 table below.

Table 3.2: Number of deaths at time t_i

Group	No.of deaths	No. survived	Total
1	d_{i1}	$n_{i1} - d_{i1}$	n_{i1}
2	d_{i2}	$n_{i2} - d_{i2}$	n_{i2}
Total	d_i	$n_i - d_i$	n_i

Except for the tied survival times, $d_i = 1$, and hence either d_{i1} or d_{i2} have the value 0 or 1. If a child is censored at time t_i then that child is considered to be at risk at that time and is included in n_i. In other words the assumption is that censoring occurs after the event [43]. If the null hypothesis is true, then the number of deaths at any time is expected to follow the hypergeometric distribution, and therefore

$$E(d_{i1}) = e_{i1} = \frac{n_{i1}d_i}{n_i}$$

and the variance is given by

$$Var(d_{i1}) = \frac{d_i(n_i - d_i)n_{i1}n_{i2}}{n_i^2(n_i - 1)}.$$

The difference between d_{i1} and e_{i1} is the basis for the test statistics for testing the null hypothesis. The log-rank test is the combination of these differences over all death times [43]. Summing the various measures over the death times gives

$$\begin{aligned} O_1 &= \sum_i d_{i1}, \\ E_1 &= \sum_i e_{i1}, \text{ and} \\ V_1 &= \sum_i Var(d_{i1}) \end{aligned}$$

where E_1 can be seen as the expected number of deaths occurring in group 1 over the entire period while O_1 are the total observed deaths in the group. The variance of the difference $O_1 - E_1$ assuming independent event times is given by V_1 . The test statistic is then given by

$$\chi_1^2 = \frac{(O_1 - E_1)^2}{V_1}$$

which, under H_0, is χ^2 distributed with 1 degree of freedom [48]. If the calculated value is larger than the value corresponding to the χ^2 distribution at a significance level of α, then the null hypothesis of no differences is rejected and one can conclude that the risk of death is different in the two groups [43].

Alternatively, assuming the deviations $d_{i1} - e_{i1}, i = 1, 2, \ldots, k$, are independent,

$$Z = \frac{O_1 - E_1}{\sqrt{V_1}}$$

should have an approximately standard normal distribution, and the null hypothesis is rejected for large values of Z. In particular at 5% level of significance the null hypothesis is reject if the observed Z is greater than 1.96. The ratios

$\frac{O_1}{E_1}$ and $\frac{O_2}{E_2}$ are referred to as the relative death rates and they measure the ratio of the death rate in each group to the death rate among both groups combined. The ratio of these two relative rates estimates the death rate in group 1 relative to the death rate in group 2.

The log rank test can be generalized to test equality of death rates in $s > 2$ groups. The test statistic, with $(s-1)$ degrees of freedom, would then be given by

$$\chi^2_{s-1} = \frac{(O_1 - E_1)^2}{V_1} + \frac{(O_2 - E_2)^2}{V_2} + \frac{(O_3 - E_3)^2}{V_3} + \dots$$

If the calculated value exceeds the table value at α significant level, we reject the null hypothesis of no group differences survivor or hazard functions.

Some important remarks in the derivation of the log-rank statistic are stated below. First the vector of observed-minus-expected failures does not in fact have independent components and the central limit theorem usually applied to prove asymptotic normality fails. Further still, differences between observed and expected failures are given equal weight regardless of the risk set (namely number of cases still under observation) at observed failure times. Such weighting will have implications on the overall test statistic. These more delicate aspects of significance tests are studied in the book on counting processes and survival analysis by Fleming and Harrington [23].

Chapter 4

Regression Models in Survival Analysis

4.1 Introduction

The non-parametric methods described in the previous chapter, namely the log-rank and the Kaplan-Meier curves, cannot control for covariates therefore extensions to include covariates is necessary. These non-parametric methods do not control for covariates and they require categorical predators. When we have several prognostic variables, we must use multivariate approaches, but we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations [12]. We need another method to model survival data with the presence of censoring. One very popular model in survival data is the Cox proportional hazard (PH) model, proposed by Cox [10].

4.1.1 Cox Proportional Hazard (PH) Model

The proportional hazard model introduced by Cox [10] is a regression model with event time as dependent variable. However the regression model is formulated through the dependence of the hazard function on covriates. It allows the inclusion of information about known (observed) covariates in models of time-to-event data in an easy way.

Cox proportional hazard model is given by

$$h(t|X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) = h_0(t)\exp(\beta' X)$$

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, $X = (x_1, x_2 \ldots, x_p)$ is the values of the vector of exploratory variables for a particular individual, and $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of regression coefficients.

The corresponding survival function given the covariate X is

$$S(t|X) = [S_0(t)] \exp(\beta' X)$$

where $S_0(t) = \int_0^t h_0(u)du$ denotes the baseline survival function and the component of the vector β are unknown regression parameters. Thus the survival function of an individual with covariate vector X is a power of the baseline survival function.

This model is also called the semi-parametric model, because it takes no assumptions about the form of $h_0(t)$ (non-parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model) [27]. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients β, hazard ratio and adjusted hazard curves [12].

The measure of effect is called hazard ratio. The hazard ratio of two individuals with different baseline or time-invariant covariates X and X^* is

$$\hat{HR} = \frac{h_0(t) \exp(\hat{\beta}' X)}{h_0(t) \exp(\hat{\beta}' X^*)} = \exp(\sum \hat{\beta}'(X - X^*)) \tag{4.1}$$

This hazard ratio is time-independent, which is why this is called the proportional hazard model. In other words from (4.1) we see that the hazard of the individual with covariates X is proportional to the hazard of the individual with covariates X^*.

4.2 Partial Likelihood Estimate for Cox PH Model

When fitting the Cox proportional hazard model, the main focus is the estimation of the regression parameters $(\beta_1, ..., \beta_p)$ where individual parameter β_j is interpreted as the log-hazard ratio corresponding to covariate or predictor variable X_j [11]. If X_j is a two group classifier then $\exp(\beta_j)$ is the hazard ratio of one group to the other group used as the reference, while if X_j is continuous the $\exp(\beta_j)$ is the hazard ratio for a unit increase in X_j. One approach is to attempt to maximize the likelihood function for the observed data simultaneously with respect to β. A more popular approach was proposed by Cox [11] in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. The partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameter ($h_0(t)$ in the Cox PH model).

Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals. Let the ordered death or event times of r individuals who experience the event of interest be $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ 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

$$\frac{\exp(\beta' X_i(t_j))}{\sum_{k \epsilon R(t_{(j)})} \exp(\beta' X_k(t_{(j)}))}$$

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

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' X_i(t_j))}{\sum_{k \epsilon R(t_{(j)})} \exp(\beta' X_k(t_{(j)}))} \tag{4.2}$$

in which $X_i(t_{(j)})$ is the vector of covariate values for individual i who dies at $t_{(j)}$. The general method of partial likelihood was discussed by Cox [11]. This likelihood function is only for the uncensored individuals. Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals and δ_i be the event indicator, which is zero if the i^{th} survival time is censored, and unity otherwise. The likelihood function in the above equation can be expressed by

$$L(\beta) = \prod_{j=1}^{n} \left[\frac{\exp(\beta' X_i(t_j))}{\sum_{k \epsilon R(t_{(j)})} \exp(\beta' X_k(t_{(j)}))} \right]^{\delta_i} \tag{4.3}$$

where $R(t_i)$ is the risk set at time t_i.

The partial likelihood is valid when there are no ties in the dataset; that means there are no two subjects who have the same event time.

4.3 Cox PH Assumption Checking

The main assumption of the Cox PH model is proportional hazards [43]. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e, the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality, and among them is the graphical method, adding time-dependent covariates as well as a formal test based on the Schoenfeld residuals [43].

4.3.1 Graphical Method

We can obtain Cox PH survival function by the relationship between hazard function and survival function, that is

$$S(t, X) = [S_0(t)]^{\exp\left(\sum_{i=1}^{p} \beta_i x_i\right)}$$

where $X = (x_1, x_2, \ldots, x_p)$ is the values of the vector of explanatory variables for a particular individual. When we take the logarithm twice, we can easily

show that

$$\ln\left[-\ln S(t,X)\right] = \sum_{i=1}^{p} \beta_i x_i + \ln\left[-\ln S_0(t)\right].$$

Then the difference in log-log curves corresponding to two different individual with variable $X_1 = (x_{11}, x_{12}, \ldots, x_{1p})$ and $X_2 = (x_{21}, x_{22}, \ldots, x_{2p})$ is given by

$$\ln\left[-\ln S(t,X_1)\right] - \ln\left[-\ln S(t,X_2)\right] = \sum_{i=1}^{n} \beta_i \left(x_{1i} - x_{2i}\right)$$

which does not depend on t provided the two covariate vectors X_1 and X_2 are not time dependent. This relationship is very helpful to help us identify situations where we may have proportional hazards. By plotting estimated log(-log(survival)) versus survival time for two groups we would see parallel curves if the hazards are proportional [43]. This method does not work well for continuous predictors or categorical predictors that have many levels because the group becomes cluttered. Furthermore, the curves are sparse when there are few time points and it may be difficult to tell how close to parallel is close enough [43].

However, looking at the K-M curves and log(-log(survival)) is not enough to be certain of proportionality since they are univariate analysis that do not necessarily show whether hazards will still be proportional when a model includes many other predictors, but they support our argument for proportionality. There are, however, other statistical methods for checking the proportionality [43].

4.3.2 Adding Time-dependent Term in the Cox Model

We create time-dependent term by creating an interaction between the predictor or covariate of interest and a function of survival time and including it in the model. For example, if the predictor of interest is X_j, then we create a time-dependent term or covariate given by $X_j(t), X_j(t) = X_j \times g(t)$, where $g(t)$ is a function of time, for instance, $t, \log t$ or heaviside function of t. The model assessing PH assumption for X_j adjusted for other covariates is

$$h(t, X(t)) = h_0(t) \exp\left[\beta_1 x_1 + \beta_2 x_2 + \ldots \beta_j x_j + \ldots \beta_p x_p + \delta x_j \times g(t)\right]$$

where $X(t) = (x_1, x_2, \ldots, x_p, x_j(t))'$ is the values of the vector of explanatory variables for a particular individual. The null hypothesis to check the PH assumption for X_j is: $\delta = 0$ and it's alternative hypothesis is: $\delta \neq 0$ ($H_0 : \delta = 0$ virsus $H_a : \delta \neq 0$). The test procedure can be carried out using either a Wald test or likelihood ratio test. In the Wald test, the test statistic post-estimation

$$W = \left(\frac{\hat{\delta}}{se(\hat{\delta})}\right)^2$$

which is asymptotically distributed as chi-square with one degree of freedom. The likelihood ratio (LR) test statistic that compares the likelihood under the

null hypothesis H_0 and the likelihood under the alternative hypothesis H_a. It requires that the model under H_0 be nested within the model under H_a. Thus the LR test statistic is given by

$$LR = -2 \ln \left(\frac{L_0}{L_a} \right) = -2 \left(\ell_0 - \ell_a \right),$$

where ℓ_0, ℓ_a are the log likelihoods under the two hypotheses respectively. The statistic follows a chi-square distribution with one degree of freedom under the null hypothesis. If a time-dependent covariate is significant, i.e, the null hypothesis is rejected, then the predictor is declared not to satisfy the PH assumption. In the same way, we can also assess the PH assumption for several predictors simultaneously.

4.4 Residuals for Cox Regression Model

There are methods for calculating residuals in survival analysis, especially in Cox regression model by which each method has specific use, such as goodness-of-fit, to identify possible outliers and influential observations or in general to check necessary assumptions. In this thesis, we study six methods of residuals, namely Cox-Snell, Modified Cox-Snell, Schoenfeld, Martingale, Deviance and Score residuals.

4.4.1 Cox-Snell Residuals

These residuals are commonly used in the analysis of survival data, and were named after [13]. The Cox-Snell residual for the i^{th} individual, $i = 1, 2, \ldots, n$ is given by

$$r_{Ci} = \exp \left(\beta' x_i \right) \hat{H}_0(t_i) \tag{4.4}$$

where $\hat{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i, the observed survival time of that individual. This residual can be derived from a general result in mathematical statistics on the distribution of a function of a random variable. According to this result, if T is the random variable associated with the survival time of an individual, and $S(t)$ is the corresponding survivor function, then the random variable $Y = -\log S(t)$ has an exponential distribution with unit mean, irrespective of the form of $S(t)$ [16]. The proof of this result is obtained in the following paragraph, which can be omitted without loss of continuity.

According to a general result, if $f_X(x)$ is the probability density function of the random variable X, the density of the random variable $Y = g(X)$ is given by

$$f_Y(y) = f_X \frac{\left\{ g^{-1}(y) \right\}}{|\frac{dy}{dx}|}$$

where $f_X\left\{g^{-1}(y)\right\}$ is the density of X expressed in terms of y. Using this result, the probability density function of the random variable $Y = -\log S(T)$, is given by

$$f_Y(y) = \frac{f_T\left\{g^{-1}(y)\right\}}{|\frac{dy}{dx}|} \tag{4.5}$$

where $f_T(t)$ is the probability density function of T. Now,

$$\frac{dy}{dx} = \frac{\{-\log S(t)\}}{dt} = \frac{f_T(T)}{S(t)}$$

and when the absolute value of this function is expressed in terms of y, the derivative becomes

$$\frac{f_T\left\{S^{-1}(e^{-y})\right\}}{S^{-1(e^{-y})}} = \frac{f_T\left\{S^{-1}(e^{-y})\right\}}{e^{-y}}$$

Finally, on substituting for the derivative in equation (4.5), we find that

$$f_Y(y) = e^{-y}$$

is the probability of the exponential random variable with unit mean. The crucial step in the argument is, if the model fitted to the data is satisfactory, then a model-based estimate of the survivor function for the i^{th} individual at t_i, the survival time of that individual will be closed to the corresponding true $S_i(t_i)$ [16]. This suggests that if the correct model has been fitted, the values $S_i(t_i)$ will have properties similar to those of $S_i(t_i)$. Then, the negative logarithms of the estimated survivor functions, $-\log \hat{S}_i(t_i)$, $i = 1, 2, \ldots, n$ will behave as n observations from a unit exponential distribution. These estimates are the Cox-Snell residuals [14].

If the observed survival time for an individual is right-censored, then the corresponding value of the residual is also right-censored. The residual will therefore be a censored sample from the unity exponential distribution.

The Cox-Snell residuals, r_{Ci} have properties that are dissimilar to those of residuals used in linear regression analysis, for example. In particular, they will not be symmetrically distributed about zero, and in fact they cannot be negative. Furthermore, since the Cox-Snell residuals are assumed to have an exponential distribution when an appropriate model has been fitted they have a highly skewed distribution and the mean and variance of the i^{th} residual will both be unity [15].

4.4.2 Modified Cox-Snell Residuals

Suppose that the i^{th} survival time is a censored observation, t_i^*, and let t_i be the actual, but unknown survival time, so that $t_i > t_i^*$. The Cox-Snell residual for this individual evaluated at the censored survival time, is then given by

$$r_{Ci} = \hat{H}_i(t_i^*) = -\log \hat{S}_i(t_i^*)$$

where $\hat{H}_i(t_i^*)$ and $\hat{S}_i(t_i^*)$ are the estimated cumulative hazard and survivor functions, respectively, for the i^{th} individual at the censored survival time [16].

If the fitted model is correct, the values r_{Ci} can be taken to have a unity exponential distribution. The cumulative hazard function of the individual increases linearly with time, and so the greater the value of the time t_i for the i^{th} individual the greater the value of the Cox-Snell for that individual [9]. It then follows that the residual for the i^{th} individual at the actual (unknown) failure time, $\hat{H}_i(t_i)$, will be greater than the residual evaluated at the observed censored survival time. To take account of this, Cox-Snell residual can be modified by the addition of a positive constant $\triangle$, which can be called the excess residual. Modified Cox-Snell residuals are therefore of the form

$$r'_{Ci} = \begin{cases} r_{Ci}, & \text{for uncensored observations} \\ \\ r_{Ci} + \triangle, & \text{for censored observations} \end{cases}$$

where r_{Ci} is the Cox-Snell residual for the i^{th} observation, defined in equation (4.4). To identify the suitable value for $\triangle$, we use the lack of memory property by [15], suppose that the random variable T has an exponential distribution with mean λ^{-1}, and consider the probability that T exceeds $t_0+t_1, t_{1\geq 0}$ conditional on T being at least equal to t_0. From the standard result for conditional probability given in the previous chapter this probability is

$$P\left(T \geq t_0 + t_1 | T \geq t_0\right) = \frac{P\left(T \geq t_0 + t_1 \text{ and } T \geq t_0\right)}{P\left(T \geq t_0\right)}.$$

The numerator of this expression is simply $P\left(T \geq t_0 + t_1\right)$ after simplifying it and so the required probability is the ratio of the probability of survival beyond $t_0 + t_1$ to the probability of survival beyond t_0, that is $\frac{S(t_0+t_1)}{S(t_0)}$. The survivor function for the exponential distribution is given by the $S(t) = e^{-\lambda t}$. Hence

$$P\left(T \geq t_0 + t_1 | T \geq t_0\right) = \frac{\exp\left\{-\lambda(t_0 + t_1)\right\}}{\exp(-\lambda t_0)} = e^{-\lambda t}$$

which is the survivor function of an exponential random variable at time t_1, that is $P(T \geq t_1)$. This result means that, conditional on survival to time t_0, the excess survival time beyond t_0 also has an exponential distribution with mean λ^{-1}. In other words, the probability of survival beyond time t_0 is not affected by the knowledge that the individual has already survived to time t_0.

From this result, since r_{Ci} has a unity exponential distribution, the excess residual $\triangle$, will also have a unity exponential distribution by [14]. The expected value of $\triangle$ is therefore unity, suggesting that $\triangle$ may be unity, and this leads to modified Cox-Snell residuals, given by

$$r'_{Ci} = \begin{cases} r_{Ci}, & \text{for uncensored observations} \\ r_{Ci} + 1, & \text{for censored observations} \end{cases} \tag{4.6}$$

The i^{th} modified Cox-Snell residual can be expressed in an alternative form by introducing an event indicator, δ_i which takes the value zero if the observed survival time of the i^{th} individual is censored and unity if it is uncensored. Then the modified Cox-Snell residual is given by

$$r'_{Ci} = 1 - \delta + r_{Ci} \tag{4.7}$$

Notice that from the definition of the type of residual, r'_{Ci} must be greater than unity for a censored observation. Also as for the unmodified residuals the r'_{Ci} can take any value between zero and infinity, and they will have a skew distribution.

On the basis of empirical evidence, [17] found that the addition of unity to a Cox-Snell residual for a censored observation inflated the residual t_0 by too great an extent. They suggest that the median value of the excess residual be used rather than the mean. For the unity exponential distribution the survivor function is $S(t) = e^{-t}$, and so the median $t(50)$ is such that $e^{-t(50)} = 0.5$ where $t(50) = \log 2 = 0.693$. Thus, a second version of the modified Cox-Snell residual has

$$r''_{Ci} = \begin{cases} r_{Ci}, & \text{for uncensored observations} \\ r_{Ci} + 0.693, & \text{for censored observations} \end{cases} \tag{4.8}$$

However, if the proportion of censored observations is not too great the set of residuals obtained from each of these two forms of modification will not appear too different.

4.4.3 Martingale Residuals

This residual is used to examine overall test of goodness-of-fit of a Cox model [9]. The modified residuals r'_{Ci} defined in equation (4.7) have mean of unity for uncensored observations. Accordingly, these residuals might be further refined by relocating the r'_{Ci} so that they have a mean of zero when an observation is uncensored. If in addition the resulting values are multiplied by -1, we obtain the residuals

$$r_{Mi} = \delta_i - r_{Ci} \tag{4.9}$$

These residuals are known as Martingale residuals, since they can also be derived using what are known as Martingale methods. In this derivation the r_{Ci} are based on the Nelson-Aalen estimate of the cumulative hazard function.

A comprehensive account of the Martingale approach to the analysis of survival data has been presented by a number of authors, including ([4],[23] and [70]).
Martingale residuals takes values between $-\infty$ and unity, with the residuals for censored observations, where $\delta_i = 0$ being negative. It can be shown that these residuals sum to zero and in large samples the Martingale residuals are uncorrelated with one another and have an expected value of zero.

In this respect, they have properties similar to those possessed by residuals uncounted in linear regression analysis. Another way of looking at the Martingale residuals is to note that the quantity r_{Mi} in equation (4.9) is the difference between the observed number of deaths for the i^{th} individual in the interval $(0, t_i)$ and the corresponding estimated expected number of the basis of the fitted model. To see this note that the observed number of deaths is unity if the survival time t_i is uncensored, and zero if censored, that is δ_i. The second term in equation (4.9) is an estimate of $H_i(t_i)$, the cumulative hazard of death for the i^{th} individual over the interval $(0, t_i)$. Since we are dealing with just one individual, this can be viewed as the expected number of deaths in that interval. This shows another similarity between the Martingale residuals and residual from other areas of data analysis.

4.4.4 Deviance Residuals

Although Martingale residuals share many of the properties possessed by residuals encountered in other situations, such as in linear regression analysis, they are not symmetrically distributed about zero, even when the fitted model is correct. This skewness makes plots based on the residuals difficult to interpret. The deviance residuals, which were introduced by [25] are much more symmetrically distributed about zero. They are defined by

$$r_{Di} = sgn(r_{Mi}) \left[-2 \left\{r_{Mi} + \delta_i \log \left(\delta_i - r_{Mi}\right)\right\}\right]^{\frac{1}{2}} \tag{4.10}$$

where r_{Mi} is the Martingale residual for the i^{th} individual, and the function $sgn(.)$ is the sign function. This is the function that takes the value $+1$ if its argument is positive and -1 if negative. Thus $sgn(r_{Mi})$ ensures that the deviance residuals have the same sign as the Martingale residuals.

The original motivation for these residuals is that they are components of the deviance. According to [19] the deviance is a statistic that is used to summarize the extent to which the fit of a model of current interest deviates from that of a model which is a perfect fit to the data. This latter model is called the saturated or full model in which $\beta-$coefficients are allowed to be different for each individual. The statistic is given by

$$D = -2 \left\{\log \hat{L}_c - \log \hat{L}_f \right\}$$

where $\hat{L}_c$ is the maximized partial likelihood under the current model and $\hat{L}_f$ is the maximized partial likelihood under the full model. The smaller the value of the deviance, the better the model.

The deviance can be regarded as a generalization of the residual sum of squares used in modeling normal data to the analysis of non-normal data, and features prominently in generalized linear modeling by [19]. Note that differences in deviance between two alternative models are the same as differences in the value

of the statistic $-2\log\hat{L}$. The deviance residuals are $D = \sum r_{Di}^2$, so that observations that correspond to relatively large deviance residuals are those that are not well fitted by the model.

Another way of viewing the deviance residuals is that they are Martingale residuals that have been transformed to produce values that are symmetric about zero when the fitted model is appropriate [70]. To see this, first recall that the Martingale residuals r_{Mi} can take any values in the interval $(-\infty, 1)$. For large negative values of (r_{Mi}), the term in the square brackets in equation (4.10) is dominated by r_{Mi}. Taking the square root of this quantity has the effect of bringing the residual closer to zero. Thus Martingale residuals in the range $(-\infty, 0)$ are shrunk toward zero. Now consider the Martingale residuals in the interval $(0, 1)$. The term $\delta_i \log(\delta_i - r_{Mi})$ in equation (4.10) will only be non-zero for uncensored observations, and will then have the value $\log(1 - r_{Mi})$. As r_{Mi} gets closer to unity, $1 - r_{Mi}$ gets closer to zero and $\log(1 - r_{Mi})$ takes large negative values. The quantity in square brackets in equation (4.10) is then dominated by this logarithmic term, and so the deviance residuals are expanded toward $+\infty$ as the Martingale residual reaches its upper limit of unity.

One final point to note is that although these residuals can be expected to be symmetrically distributed about zero when an appropriate model has been fitted they do not necessarily sum to zero.

4.4.5 Schoenfeld Residuals

Two advantages of the residuals described in previous sections are that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. Both of these advantages are overcome in a residual proposed by [64]. These residuals were originally termed partial residuals, for the reason given in the sequel, but are now commonly known as Schoenfeld residuals. These residuals differ from those considered previously in one other important respect. This is that there is no single value of the residual for each individual, but a set of values, one for each explanatory variable included in the fitted Cox regression model. The i^{th} partial or Schoenfeld residual for X_j, the j^{th} explanatory variable in the model, is given by

$$r_{Pji} = \delta_i \{x_{ji} - \hat{a}_{ji}\} \tag{4.11}$$

where x_{ji} is the value of the j^{th} explanatory variable, $j = 1, 2, \ldots, p$ for the i^{th} individual in the study,

$$\hat{a}_{ji} = \frac{\sum_{\iota \epsilon R(t_i)} x_{j\iota} \exp\left(\hat{\beta' x_\iota}\right)}{\sum_{\iota \epsilon R(t_i)} \exp\left(\hat{\beta' x_i}\right)} \tag{4.12}$$

and $R(t_i)$ is the set of all individuals at risk at time t_i. Note that non-zero values of these residuals only arise for uncensored observations. Moreover, if the

largest observation in a sample of survival times is uncensored, the value of $\hat{a}_{ji}$ for that observation, from equation (4.12), will be equal to x_{ji} and so $r_{Pji} = 0$.

The i^{th} Schoenfeld residual, for the explanatory variable X_j, is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j, which is given by

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \delta_i \left\{ x_{ji} - a_{ji} \right\} \tag{4.13}$$

where

$$a_{ji} = \frac{\sum_\iota x_{j\iota} \exp(\beta' x_\iota)}{\sum_\iota \exp(\beta' x_\iota)} \tag{4.14}$$

The i^{th} term in this summation, evaluated at $\hat{\beta}$, is then the Schoenfeld residual for X_j, given in equation (4.11). Since the estimates of the $\beta' s$ are such that

$$\frac{\partial \log L(\beta)}{\partial \beta_j}|_{\hat{\beta}} = 0$$

the Schoenfeld residuals must sum to zero. These residuals also have the property that, in a large sample, the expected values of r_{Pji} is zero and they are uncorrelated with one other.

It turns out that a scaled version of the Schoenfeld residuals, proposed by [25], is more effective in detecting departures from the assumed model. Let the vector of Schoenfeld residuals for the i^{th} individual be denoted by $r_P = (r_{P1i}, r_{P2i}, \ldots, r_{Ppi})'$. The scaled or weighted Schoenfeld residuals, r^*_{Pji}, are then the components of the vector

$$r^*_{Pi} = rvar(\hat{\beta}) r_{Pi}$$

where r is the number of deaths among the n individuals, and $var\hat{\beta}$ is the variance-covariance matrix of the parameter estimates in the fitted Cox regression model. These scaled Schoenfeld residuals are therefore quite straightforward to compute.

4.4.6 Score Residuals

Other types of residuals that are useful in some aspect of model checking, and which, like the Schoenfeld residual, are obtained from the first derivative of the logarithm of the partial likelihood function with respect to the parameter $\beta_j, j = 1, 2, \ldots, p$ [69]. However, the derivative in equation (4.13) is now expressed in a quite different form, which is

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left\{ \delta_i \left(x_{ji} - a_{ji} \right) + \exp\left(\beta' x_i \right) \sum_{t_r < t_{t_i}} \frac{\left(a_{ji} - x_{ji} \right) \delta_r}{\sum_{\iota \epsilon R(t)}} \exp\left(\beta' x_\iota \right) \right\} \tag{4.15}$$

x_{ji} is the i^{th} value of the j^{th} explanatory variable. δ_i is the event indicator which is zero for censored observations and unity otherwise, a_{ji} is given in equation (4.14), and $R(t_r)$ is the risk set at time t_r. In this formulation, the contribution of the i^{th} observation to the derivative only depends on information up to time t_i. In other words, if the study was actually concluded at time t_i, the i^{th} component of the derivative would be unaffected. Residuals are then obtained as the estimated value of the n component of the derivative

From equation (4.15) the first derivative of the logarithm of the partial likelihood function, with respect to β_j, is the efficient score for β score for β_j, and so these residuals are known as score residuals.

From equation (4.15) the i^{th} score residual $i = 1, 2, \ldots, n$ for the j^{th} explanatory variable in the model, X_j, is given by

$$r_{Sji} = \delta_i \left(x_{ji} - \hat{a}_{ji}\right) + \exp\left(\hat{\beta' x_i}\right) \sum_{t_r < t_{t_i}} \frac{\left(a_{ji} - x_{ji}\right) \delta_r}{\sum_{\iota \epsilon R(t)}} \exp\left(\beta' x_\iota\right)$$

using equation(4.11), this may be written in the form

$$r_{Sji} = r_{Pji} + \exp(\hat{\beta' x_i}) \sum_{t_r < t_{t_i}} \frac{\left(a_{ji} - x_{ji}\right) \delta_r}{\sum_{\iota \epsilon R(t)}} \exp\left(\beta' x_\iota\right) \tag{4.16}$$

which shows that the score residuals are modifications of the Schoenfeld residuals. As for the Schoenfeld residual the score residuals sum to zero, but will not necessarily be zero when an observation is censored [69].

Chapter 5

Parametric Model

The Cox PH model outlined in the previous chapter is the most common approach used to model survival data particularly in health research such as clinical trials. This is probably due to the fact that this model allows one to make inferences about the model parameters without assuming any distribution for the survival time. However, when the PH assumption is not tenable, these models will not be suitable. On the other the PH assumption may be tenable but a suitable distribution for the time to event variable. In this case a fully parametric model can be used. In this section, we introduce parametric models, in which specific probability distributions are assumed for survival times. In the first section we introduce the parametric PH model. In the second section we will present the accelerated failure time (AFT) model and more detailed discussion of the exponential, Weibull, log-logistic, log-Normal and gamma AFT models.

5.1 Parametric PH Model

The parametric PH model is the parametric version of the Cox PH model. It is given with the similar form to the Cox PH model. The hazard function at time t for the particular patient with a set of p covariate $(x_1, x_2 \ldots, x_p)$ is given as follows:

$$h(t|X) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) = h_0(t) \exp(\beta' X).$$

The key difference between the two kinds of model is that the baseline hazard function is assumed to come from a specified distribution leading to a fully parametric PH model, where as the Cox PH model has no such constraint. The coefficients are estimated by partial likelihood in Cox model but by a fully maximum likelihood estimation in the parametric PH model. Other than this, the two types of models are equivalent. Hazard ratios have the same interpretation and PHs still hold in both. A number of different parametric PH models may be derived by choosing different hazard functions. The commonly applied models are Exponential, Weibull, or Gompertz models.

5.1.1 Exponential PH Model

Exponential PH model is the special case of the Weibull model when $\gamma = 1$. The hazard function for exponential PH model is assumed to be constant over time. The baseline survival and hazard function for this model are given by

$$S(t) = \exp(-\lambda_0 t) \tag{5.1}$$

and the hazard function is given by

$$h(t) = \lambda_0. \tag{5.2}$$

The hazard model for a particular patient with covariates $x_{i1}, x_{i2}, ..., x_{ip}$ is given by

$$h(t|X) = \lambda_0 \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) = \lambda_0 \exp(\beta' X).$$

The extension of the exponential PH model to the piecewise constant exponential model is possible [6]. For the piecewise exponential model, the follow up period say $[0, T]$ is divided into K intervals $(t_j, t_{j+1}]$ for $j = 1, 2, \ldots, K$, with $t_1 = 0$ for simplicity. The approach assumes that the baseline hazard is constant within each interval but can vary across intervals, so that $h_0(t) = \exp(\alpha_j) = \lambda_j$ for $t_j < t < t_{j+1}$ that is the baseline hazard function is approximated by a step function. The piecewise exponential model is given by

$$\lambda_{ij} = \lambda_j \exp(\beta' x_i) \tag{5.3}$$

where λ_{ij} is the hazard corresponding to individual i in interval j and $\exp(\beta' x_i)$ is the relative risk for an individual with covariate value x_i compared to the baseline at any given time.

For piecewise exponential model approach, a log-linear model is used to estimate both the effects of the covariates and the underlying hazard function. Estimates of the underlying hazard function and the regression parameters can be obtained using maximum likelihood. The maximum likelihood estimate of the baseline hazard function in interval j for given regression coefficients β is given by

$$\hat{\lambda}_j = \frac{d_j}{\sum_{i \epsilon R_j} \exp(\hat{\beta}' x_i) t_{ij}}$$

where d_j denotes the number of events in interval j, R_j denotes the risk set entering j, and t_{ij} is the observed survival time for individual i in interval j ([28];[29]; [54]). The great challenge related to the use of the piecewise exponential model is to find an adequate grid of time-points needed in its construction. The advantage of this method is the ability to incorporate time-dependent covariates. If there were any time-dependent covariates, their values at the beginning of each interval could be assigned to the records for that time interval.

5.1.2 Weibull PH Model

The survival time is said to follow the Weibull distribution with scale parameter λ and shape parameter γ, so the survival and hazard function of a $W(\lambda, \gamma)$ distributions are given by

$$S(t) = \exp(-\lambda t^{\gamma}) \tag{5.4}$$

and the baseline hazard function is

$$h(t) = \lambda_0 \gamma t^{\gamma-1} \tag{5.5}$$

with $\lambda, \gamma > 0$. The hazard rate increases where $\gamma > 1$ and decreases when $\gamma < 1$ as time goes on. For Weibull PH model, the hazard function of a particular patient with covariates $x_{i1}, x_{i2}, \ldots, x_{ip}$ where the i denotes the individual, is given by

$$h(t|X) = \lambda_0 \gamma t^{\gamma-1} \exp(\beta' X)$$

where $\lambda_0 \exp(\beta' X)$ denotes the scale parameter and γ the shape parameter. Therefore the Weibull family with fixed γ possesses the PH property. This indicates that the effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains constant. The corresponding survival function is given by

$$S(t|X) = \exp\left\{-\exp(\beta' X)\lambda_0 t^{\lambda}\right\}.$$

The baseline survival function (with no covariates) for Weibull distribution can be transformed to obtain the following equation

$$\log(-\log(S(t))) = \log \lambda_0 + \gamma \log t \tag{5.6}$$

where a plot of $\log(-\log(S(t)))$ versus $\log(t)$ should give approximately a straight line if the Weibull distribution assumption is reasonable. The intercept and the slope of the line will be rough estimates of $\log \lambda_0$ and γ respectively. If the two lines for two groups in this plot are essentially parallel, this means that the proportional hazards model is valid. Furthermore, if the straight line has a slope of nearly one, the simpler exponential distribution is reasonable. In other words, for an exponential distribution, there is $\log S(t) = -\lambda_0 t$. Thus we can consider the graph of $-\log S(t)$ versus $\log t$. This should be a line that goes through the origin if exponential distribution is appropriate.

5.1.3 Gompertz PH Model

Under the Gompertz PH model the baseline survival and hazard distribution is given by

$$S(t) = \exp\left(\frac{\lambda_0}{\theta}\left(1 - e^{\theta t}\right)\right) \tag{5.7}$$

and the hazard is

$$h(t) = \lambda_0 \exp(\theta t) \tag{5.8}$$

where t is between $[0, \infty)$ and $\lambda_0 > 0$. The parameter θ denotes the shape of the hazard function. If $\theta = 0$, the hazard function reduces to that of an exponential distribution. That is, 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 patient is given by

$$h(t|x) = \lambda_0 \exp(\theta t) \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) = \lambda_0 \exp(\beta' X) \exp(\theta t).$$

It can easily be shown that the Gompertz model has PH property [31]. Although the Gompertz distribution is rarely used in practice it has some appealing properties such as the PH property.

5.2 Accelerated Failure Time (AFT) Model

One alternative parametric model used to analyze time-to-event data is the accelerated failure time model. These models assume a given probability distribution for the survival time. The AFT model is an alternative to the PH model for the analysis of time-to-event data. In the AFT model we measure the direct effect of the explanatory variables on the survival time instead of hazard, as we do in PH model. This formulation allows for an easier interpretation of the results because the parameters measure the effect of a given covariate on the survival time directly. 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 the PH model, the AFT model describes the relationship between survival probabilities and a set of covariates.

The AFT assumption can be expressed as $S_2(t) = S_1(\eta t)$ for $t \geq 0$, where $S_1(t)$ and $S_2(t)$ are the survival functions for group one and group two, respectively and η is a constant called the acceleration factor comparing group one and group two. In regression framework the acceleration factor η could be parameterized as $\exp(\alpha)$ where α is a parameter to be estimated from the data. With this parameterization, the AFT assumption can be expressed as $S_2(t) = S_1(\exp(\alpha)t)$ or equivalently, $S_2(\exp(-\alpha)t) = S_1(t)$ for $t \geq 0$. The AFT assumption can also be expressed in terms of random variables for survival time rather than the survival function. If T_2 is a random variable (following some distribution) representing the survival time for group two and T_1 is a random variable representing the survival time for group one, then the AFT assumption can be expressed as $T_1 = \eta T_2$.

The acceleration factor is the key measure of association obtained in an AFT model. It allows us to evaluate the effect of predictor variables on survival time

just as the hazard ratio allows the evaluation of predictor variables on the hazard.

The acceleration factor describes the "stretching out" or contraction of survival functions when comparing one group to another. More precisely, the acceleration factor is a ratio of survival times corresponding to any fixed value of $S(t)$; more examples will be shown in the following subsections.

In this section we will be estimating the Exponential AFT model, Weibull AFT model and the Log-logistic AFT model.

5.2.1 Exponential AFT Model

The exponential survival and hazard functions are given by $S(t) = \exp(-\lambda t)$ and $h(t) = \lambda$ respectively. In this subsection we show how $S(t)$ can be reparameterized as an AFT model. The AFT assumption, for comparing two levels of covariates, is that the ratio of times to any fixed value of $S(t) = q$ is constant for any probability q. We develop the model with the survival function and solve for t in terms of $S(t)$. We then scale t in terms of the predictor. For example, from exponential survival function

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

solving for t by first taking the natural log, multiplying by -1 on both sides, and then multiplying by the reciprocal of λ, yields

$$t = [-\ln S(t)] \times \frac{1}{\lambda},$$

letting $\frac{1}{\lambda} = \exp(\alpha_0 + \alpha_1 \text{GROUP})$ where GROUP is 1 for group two and is 0 for group one. Then t becomes

$$t = [-\ln S(t)] \times \exp(\alpha_0 + \alpha_1 \text{GROUP})$$

it can be seen how the predictor variable GROUP is used to scale the time to any fixed value of $S(t)$. Suppose $S(t) = q$, therefore t is going to be

$$t = [-\ln q] \times \exp(\alpha_0 + \alpha_1 \text{GROUP}).$$

The acceleration factor η is found by taking the ratio of the times to $S(t) = q$ for GROUP = 1 and GROUP = 0, that is

$$\eta = \frac{[-\ln(q)] \times \exp(\alpha_0 + \alpha_1)}{[-\ln(q)] \times \exp(\alpha_0)} = \exp(\alpha_1)$$

After canceling, η reduces to $\exp(\alpha_1)$.
The parameter estimates can be used to estimate time $\hat{t}$ to any value of q; for example they can estimate time in years for the first ($q = 0.25$), second

(median ($q = 0.5$)), and third quartiles ($q = 0.75$). The main property of the exponential model is that he corresponding acceleration factor and hazard ratio (for example. GROUP= 1 and GROUP= 0) are reciprocals of each other, as shown on the table below:

Table 5.1: Acceleration factor and hazard ratio

AFT		HR	
$\eta > 1 \Rightarrow$	Exposure benefits survival	$HR > 1 \Rightarrow$	Exposure harmful to survival
$\eta < 1 \Rightarrow$	Exposure harmful to survival	$HR < 1 \Rightarrow$	Exposure benefits survival
$\eta = 1 \Rightarrow$	No effect from exposure	$HR = 1 \Rightarrow$	No effect from exposure

Although the exponential PH and AFT models focus on different underlying assumptions, they are in fact the same model. The only difference is in their parameterization. The resulting estimates for the survival function, hazard function, and median survival do not differ between these models.

5.2.2 Weibull AFT Model

An AFT can also be formulated with the Weibull distribution. We derive the AFT parameterization similarly to that done with the exponential model, by solving for t in terms of a fixed $S(t)$. The Weibull survival function is given by

$$S(t) = \exp\left(-\lambda t^{\gamma}\right).$$

Solving for t by first taking the natural log, multiplying by negative 1 on both sides, raising to the power of $\frac{1}{\gamma}$, and then multiplying by the reciprocal of $\lambda^{\frac{1}{\gamma}}$, yields the expression for t as follows.

$$t = \left[-\ln S(t)\right]^{\frac{1}{\gamma}} \times \frac{1}{\lambda^{\frac{1}{\gamma}}}$$

letting

$$\frac{1}{\lambda^{\frac{1}{\gamma}}} = \exp\left(\alpha_0 + \alpha_1 \text{GROUP}\right)$$

then t becomes

$$t = \left[-\ln S(t)\right]^{\frac{1}{\gamma}} \times \exp\left(\alpha_0 + \alpha_1 \text{GROUP}\right).$$

By reparameterizing $\frac{1}{\lambda^{\frac{1}{\gamma}}} = \exp\left(\alpha_0 + \alpha_1 \text{GROUP}\right)$, it can be seen that the predictor variable GROUP is used to scale the time to any fixed value of $S(t)$. For any fixed probability $S(t) = q$. For example to find an expression for the median survival time t_m, substitute $q = 0.5$, then t, becomes

$$t = \left[-\ln q\right]^{\frac{1}{\gamma}} \times \exp\left(\alpha_0 + \alpha_1 \text{GROUP}\right)$$

but the median survival time, $q = 0.5$ that implies that

$$t = \left[-\ln(0.5)\right]^{\frac{1}{\gamma}} \times \exp\left(\alpha_0 + \alpha_1 \text{GROUP}\right)$$

likewise the acceleration factor, η(GROUP = 1versus GROUP = 0) is given by

$$\eta = \frac{[-\ln(q)]^{\frac{1}{\gamma}} \exp(\alpha_0 + \alpha_1)}{[-\ln(q)]^{\frac{1}{\gamma}} \exp(\alpha_0)} = \exp(\alpha_1)$$

acceleration factor η is obtained as the ratio of the times to $S(t) = q$ for (GROUP = 1versus GROUP = 0). After canceling η reduces to $\exp(\alpha_1)$. As with the PH form of the model, this result depends on γ not the varying by groups status; otherwise η would depend on q.

Relating Weibull AFT and PH Coefficients

Corresponding coefficients obtained from the PH and AFT forms of the Weibull models are related as $\beta_j = -\alpha_j$ for the j^{th} covariate. This can most easily be seen by formulating the parameterization equivalently in terms of $\ln(\lambda)$ for both the PH and AFT form of the model as shown below, For AFT:

$$\lambda^{\frac{1}{\gamma}} = \exp[-(\alpha_0 + \alpha_1 \text{GROUP})]$$

taking natural log on both sides yields

$$\frac{1}{\gamma} \ln \lambda = -(\alpha_0 + \alpha_1 \text{GROUP})$$

solving for $\ln \lambda$ that is,

$$\ln \lambda = -\gamma (\alpha_0 + \alpha_1 \text{GROUP})$$

and for proportional hazard

$$\lambda = (\beta_0 + \beta_1 \text{GROUP})$$

solving for $\ln \lambda$ yields

$$\ln \lambda = \beta_0 + \beta_1 \text{GROUP}.$$

So the relationship of coefficients is

$$\beta_j = -\alpha_j \gamma$$

so that

$$\beta = -\alpha$$

for exponential ($\gamma = 1$).

5.2.3 Log-logistic AFT Model

The log-logistic distribution accommodates an AFT model but not a PH model. It's hazard function is shown below

$$h(t) = \frac{\lambda \gamma t^{\gamma - 1}}{1 + \lambda t^{\gamma}}, \; p > 0, \; \lambda > 0.$$

Unlike the Weibull model, a log-logistic AFT model is not a PH model. However, the log-logistic AFT model is a proportional odds (PO) model. A proportional odds survival model is a model in which the odds ratio is assumed to remain constant over time. This is analogous to a proportional hazard model where the hazard ratio is assumed constant over time.

We develop the AFT parameterization by solving for t in terms of a fixed $S(t)$, where

$$S(t) = \frac{1}{1+\lambda t^{\gamma}} = \frac{1}{1+(\lambda^{\frac{1}{\gamma}})^{\gamma}}.$$

Solving for t from the expression for $S(t)$, by first taking the reciprocals, subtracting 1, raising to the power $\frac{1}{\gamma}$, and then multiplying by the reciprocal of $\lambda^{\frac{1}{\gamma}}$, yields the expression for t as follows:

$$t = \left[\frac{1}{S(t)} - 1\right]^{\frac{1}{\gamma}} \times \frac{1}{\lambda^{\frac{1}{\gamma}}}$$

letting

$$\frac{1}{\lambda^{\frac{1}{\gamma}}} = \exp\left[\alpha_0 + \alpha_1 \text{GROUP}\right]$$

likewise, t becomes

$$t = \left[\frac{1}{S(t)} - 1\right]^{\frac{1}{\gamma}} \times \exp\left[\alpha_0 + \alpha_1 \text{GROUP}\right].$$

By reparameterizing $\frac{1}{\lambda^{\frac{1}{\gamma}}} = \exp\left[\alpha_0 + \alpha_1 \text{GROUP}\right]$; we allow the predictor variable GROUP to be used for the multiplicative scaling of time to any fixed value of $S(t)$. The expression for t started letting $S(t) = q$ substituting this on t yields

$$t = \left[\frac{1}{q} - 1\right]^{\frac{1}{\gamma}} \times \exp\left[\alpha_0 + \alpha_1 \text{GROUP}\right].$$

The acceleration factor η is found by taking the ratio of the time to $S(t) = q$ for GROUP = 1 and GROUP = 0. After canceling η reduces to $\exp(\alpha_1)$, that is, $\eta(\text{GROUP} = 1 \text{GROUP} = 0)$ is

$$\eta = \frac{\left[q^{-1} - 1\right]^{\frac{1}{\gamma}} \exp(\alpha_0 + \alpha_1)}{\left[q^{-1} - 1\right]^{\frac{1}{\gamma}} \exp(\alpha_0)} = \exp(\alpha_1).$$

Chapter 6

Frailty Models

6.1 Introduction

There may be times when the proportional hazard is a plausible model for the time to event data but there exists individual to individual heterogeneity or when we are dealing with clustered data of some kind. Such heterogeneity may sometimes be explained through inclusion of random effects, otherwise known as frailty [40]. A possible cause of this is when covariates that are important in describing the survival of an individual are omitted. If standard methods are applied to this data (such as Cox PH model) the resulting estimate will be biased. To account for frailty in model, an unmeasured random effect is incorporated in the hazard function, under the assumption that frailty is independent of any censoring that may take place. The frailty term acts multiplicatively on the hazard function. The following introduction to frailty is based on the 1979 paper by [77].

Let $h_i(t, x, z)$ be the hazard function for an individual in population cluster i or group (or individual i as a cluster) with a vector of covariate x, at some time t, and with a frailty of z. The definition of frailty, as defined by Vaupel et al, [77], states that the ratio of the hazards for two different individuals in population group i is equal to the ratio of their frailties. Mathematically this is expressed as

$$\frac{h_i(t, x, z)}{h_i(t, x', z')} = \frac{z}{z'}$$

or

$$z' h_i(t, x, z) = z h_i(t, x, 1) \tag{6.1}$$

where an individual with a frailty of 1 might be viewed as a standard individual. If an individual has a frailty of 2, and the fixed effect is age, then in frailty terms that person is twice as likely to die at any particular age, at any particular time, than a standard individual. On the other hand, a person with a frailty of 0.5 is only half as likely to die. In other words, if $z > 1$ then an individual is more

frail than a standard individual, if $z < 1$ the subject is less frail than an average individual. Thus the frailties can be interpreted as relative risks.

The above definition of frailty assumes that each individual maintains a constant level of frailty, from birth to death. However, it does not imply that individuals with the same frailty are identical. Also, it is more convenient to define frailty in terms of the hazard, rather than the age-specific probability of death, $q(x)$ for the following reasons:

- $q(x)$ is bounded above by one (because it is a probability) and thus the range of the frailty values would also be bounded above.
- $q(x)$ is a nonlinear function of the size of the age interval used.

For simplicity, let $h_i(t, x, z)$ and $h_i(t, x, 1)$ be denoted by $h(z)$ and h respectively. Then the equation (6.1)can be rewritten as

$$h(z) = zh.$$

The following relationships for the cumulative hazard and hence the survivor function clearly follow

$$H(z) = zH$$

likewise

$$S = e^{-H}$$

this implies that

$$S(z) = S^z$$

where $S = S(t, x, 1)$ for some vector x and time t.

6.2 Univariate and Multivariate Frailty in Survival Data

In univariate survival data, each cluster has only one individual with only one survival outcome. Korosok [45] stated that univariate survival models can only have univariate frailty under certain conditions. Multivariate survival data consist of clusters of more than one individual. The cluster may be multiple survival outcomes for a single individual or one or more survival comes out for multiple correlated individuals, such as relatives.

Multivariate survival data can be modeled using univariate or multivariate frailties. The frailties across different clusters are assumed to have a distribution; they account for unexplained heterogeneity at the cluster level. The frailty term z_{ij} being a scalar and the same across all different $j's$ with the same cluster. Individuals within a cluster share a common frailty (also termed shared frailty). In a multivariate frailty model, each cluster has two or more frailties, and can come from a multivariate distribution (in the case of correlated frailty).

6.3 The Distribution of Frailty

Hougaard [30] and [32] discussed the choice of frailty distributions, mainly for the case of shared frailty [30] and [32] and also described how the theoretical properties of the various models and the distributions impacting the fit of the model are used in selection of the frailty distribution. The gamma distribution has typically been used to fit the frailty random effect. According to Hougaard [32], the gamma distribution was chosen for mathematical reasons but there are no known biological reason motivating the choice of the gamma distribution. If one chooses a gamma distribution for the frailty, the advantages are that they have simple densities for which parameters are easily obtained through likelihood estimation. For simplicity the gamma frailty distribution is chosen with the same shape parameter and different scale parameter for the survivors at a given age or time [32].

Other than the gamma distribution, Hougaard [32] discussed how the choice of frailty distribution was extended to the natural exponential family, where the gamma distributions are the simplest family. Examples of these distributions include the inverse Gaussian and the positive stable distributions.

In univariate frailty models (generally parametric), because of the identifiability problem, the distribution specified for the frailty is often given a pre-specified fixed mean of 1 for multiplicative (or proportional) hazard model to identify the frailty distribution.

Identifiability refers to being able to uniquely estimate both the parameters of the hazard function as well as of the frailty distribution in univariate data [8].

6.4 Estimation in the Frailty

The baseline hazard $h_0(t)$ can be specified explicitly or left unspecified. Under a parametric assumption for $h_0(t)$, parameters in the resulting model can be estimated using maximum likelihood estimation (MLE) procedures. However, if $h_0(t)$ is left unspecified, then the unknown parameters in the shared frailty model have to be estimated by various approaches or methods such as

- Expectation Maximization (EM) algorithm [42],
- Penalized Partial Likelihood (PPL) approach [70],
- Markov Chain Monte Carlo (MCMC) methods [75],
- Monte Carlo EM (MCEM) approach [62], and
- Different method using Laplace approximation [63].

The choice of estimation method depends largely on the choice of frailty distribution. When a gamma frailty is assumed the EM algorithm can be used. However, when a log-normal frailty is used, the estimation procedures are based on numerical integration methods such as the Laplace approximation methods. This thesis will focus on the maximum likelihood estimation (M.L.E.) procedures.

6.5 Univariate Frailty Model

At the observation level, frailty is introduced as an unobservable multiplicative effect denoted by z on the hazard function such that

$$h(t|z) = zh(t) \tag{6.2}$$

where $h(t)$ is a non frailty hazard function, such as hazard function of any of the parametric models seen in the previous chapters. The frailty z is a random positive quantity and, for model identifiability, is assumed to have mean one and variance θ. [80]. Using the relationship between the cumulative hazard function and survivor function gives the expression for the survivor function given the frailty to be as follows

$$S\left(t|z\right) = \exp\left(-\int_0^t h\left(u|z\right)du\right) = \exp\left(-z\int_0^t \frac{f(u)}{S(u)}du\right) = \left(S(t)\right)^z \tag{6.3}$$

where $S(t)$ is the survivor function that corresponds to $h(t)$ because z is unobservable, it must be integrated out of $S(t|z)$ to obtain the unconditional survivor function. Let $g(z)$ be the probability density function of z, in which case an estimable form of our frailty model is achieved as

$$S_\theta(t) = \int_0^\infty S\left(t|z\right)g(z)dz = \int_0^\infty \left(S(t)\right)^z g(z)dz. \tag{6.4}$$

Given the unconditional survivor function, we can obtain the unconditional hazard and density in the usual way, that is

$$f_\theta = -\frac{d}{dt}S'_\theta(t),\ h_\theta = \frac{f_\theta(t)}{S_\theta(t)}.$$

Hence, a univariate frailty model is merely a typical parametric survival model, with the additional estimation of an over-dispersion parameter θ. In a standard survival regression, the likelihood calculations are based on $S(t),\ h(t)$ and $f(t)$. In a univariate frailty model, the likelihood is based analogously on $S_\theta(t),\ h_\theta(t)$ and $f_\theta(t)$. Any continuous distribution supported on the positive numbers that has expectation one and finite variance θ is allowed here. For mathematical tractability, however, we limit the choice to either the gamma$(\frac{1}{\theta}, \theta)$ distribution or the inverse-Gaussian (IG) distribution with parameters one and $\frac{1}{\theta}$, denoted as

$IG(1, \frac{1}{\theta})$. The gamma (α, β) distribution has probability density function given by

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}} \tag{6.5}$$

and the IG(α, β) distribution has density function given by

$$f(x) = \left(\frac{\beta}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{\beta}{2\alpha}\left\{\frac{x}{\alpha} - 2 + \frac{\alpha}{x}\right\}\right). \tag{6.6}$$

Therefore, by integrating equation (6.4) it can be shown that the gamma frailty will result in the frailty survival function (in terms of the non frailty survivor function, $S(t)$), given by

$$S_\theta(t) = (1 - \theta \log\{S(t)\})^{-\frac{1}{\theta}} \tag{6.7}$$

and the inverse Gaussian frailty will result in the frailty survival model given by

$$S_\theta(t) = \exp\left\{\frac{1}{\theta}\left(1 - [1 - 2\theta \log\{S(t)\}]^{\frac{1}{2}}\right)\right\}. \tag{6.8}$$

Regardless of the choice of frailty distribution, $\lim_{\theta\to 0} S_\theta(t) = S(t)$ and thus the frailty function reduces to $S(t)$ when there is no heterogeneity present.

6.5.1 Likelihood for Univariate Frailty Model

The relationship between the survival function and hazard function still holds unconditional on z, and thus we can obtain the population hazard function using

$$h_\theta = -\frac{d}{dt} S_\theta(t) \left[S_\theta(t)\right]^{-1}. \tag{6.9}$$

It can be shown equivalently that the hazard function is given by

$$h_\theta(t) = h(t)E(z|T > t).$$

That is, the unconditional hazard is the average hazard at any given time [32]. As with the standard survival model the response is given as (t_{0i}, t_i, δ_i) where $i = 1, 2, \ldots, n$ is the i^{th} observation corresponding to the time $(t_{0i}, t_i]$, with either failure occurring at time t_i (or $\delta_i = 1$) as a combination of the failure and the censored observation in the usual representation. Thus the log-likelihood is given by

$$\begin{aligned} \ln(L) &= \ln \prod_{i=1}^{n} \left[S_{\theta_i}(t_i)\right]^{1-\delta_i} \left[f_{\theta_i}(t_i)\right]^{\delta_i} \\ &= \sum_{i=1}^{n} \left[\ln\{S_{\theta_i}(t_i)\} - \ln\{S_{\theta_i}(t_{0i})\} + \delta_i h_{\theta_i}(t_i)\right] \end{aligned}$$

where δ_i is the event indicator variable, and the subscript i is used such that, for example $h_{\theta_i}(t) = h_\theta(t|x_i)$.

6.6 Shared Frailty Model

An extension of the univariate (unshared) frailty model is where individuals are allowed to share the same frailty value. Sharing a frailty value also generates dependence between those individuals who share frailties, whereas conditional on the frailty those individuals are independent.

If the data consist of n groups with the i^{th} group comprised of n_i individuals for $i = 1, 2, \ldots, n$ we write the hazard of the frailty model as

$$h_{ij}(t|z_i) = z_i h_{ij}(t), \tag{6.10}$$

where z_i denote the random variable, $j = 1, 2, \ldots, n_i$ with $h_{ij}(t) = h(t|x_{ij})$ denote hazard without frailty. That is, for any member of the i^{th} group, the standard hazard function is now multiplied by the shared frailty z_i. For instance, in the case of Weibull PH regression, the conditional hazard for an individual is given by

$$h_{ij}(t|z_i) = z_i h_{ij}(t) = z_i \exp(X_{ij}\beta)\gamma t^{\gamma-1} \tag{6.11}$$

with the conditional survival function given by

$$S_{ij}(t|z_i) = \{S_{ij}(t)\}^{z_i} = \exp\left(-z_i \exp\left\{X_{ij}\beta\right\} t^{\gamma}\right). \tag{6.12}$$

6.6.1 Likelihood Estimation

By [80] likelihood of the data can be obtained by computing the group-level conditional likelihoods and integrating out the frailty. For data having n groups with n_i observation per group consisting of the three response $(t_{0ij}, t_{ij}, d_{ij})$ where $i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n_i$ which indicate the start time, end time, and failure or censoring indicator to the j^{th} individual from the i^{th} group .

For frailty z_i, the likelihood of the ij^{th} individual is given by

$$\begin{aligned} L_{ij}(z_i) &= \frac{S_{ij}(t_{ij|z_i})}{S_{ij}(t_{0ij}|z_i)} \left\{h_{ij}(t_{ij}|z_i)\right\}^{d_{ij}} \\ &= \left[\frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})}\right]^{z_i} \left[z_i h_{ij}(t_{ij})\right]^{d_{ij}} \end{aligned}$$

where $S_{ij}(t|z_i) = [S_{ij}(t)]^{z_i}$ and $\left\{h_{ij}(t_{ij}|z_i)\right\}^{d_{ij}} = z_i \left\{h_{ij}(t_{ij})\right\}^{d_{ij}}$, by defining $D_i = \sum_{j=1}^{n_i} d_{ij}$, the likelihood of the i^{th} group is given by

$$L_i(z_i) = (z_i)^{D_i} \prod_{j=1}^{n_i} \left[\frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})}\right]^{z_i} [h_{ij}]^{d_{ij}}. \tag{6.13}$$

Unconditionally, we can get the marginal likelihood of the i^{th} group by integrating out z_i, that is

$$L_i = \int_0^{\infty} L_i(z_i) g(z_i) dz_i$$

where $g(z_i)$ is the probability density function of the frailty in the case of the gamma distribution with mean 1 and variance θ, the pdf is given by

$$g(z_i) = \frac{z_i^{\frac{1}{\theta}-1} \exp\left(-\frac{z_i}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) \theta^{\frac{1}{\theta}}}.$$

For gamma frailty distribution, the marginal likelihood contribution becomes

$$L_i = \left(\prod_{j=1}^{n_i} [h_{ij}(t_{ij})]^{d_{ij}} \right) \frac{\Gamma(\frac{1}{\theta} + D_i)}{\Gamma(\frac{1}{\theta})} \theta^{D_i} \left\{ 1 - \theta \sum_{j=i}^{n_i} \ln \frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})} \right\}^{-\frac{1}{\theta} - D_i}. \quad (6.14)$$

So given the conditional group likelihoods, we can estimate the regression parameters and frailty variance θ by maximizing the overall marginal log-likelihood, that is

$$\ln L = \sum_{i=1}^{n} \ln L_i.$$

The details of estimation and inference procedures of the models are well explained by Gutierrez [46]. Model selection was ascertained by comparing log likelihood values and on the basis of Akaike's information criterion (AIC) [3]. Lower AIC value gives the best model fit.

Chapter 7

Application to the Lesotho DHS Under Five Mortality

7.1 Exploratory Data Analysis

This chapter demonstrates the application of the models discussed in the previous chapters using data on maternal, socioeconomic and environmental concomitant variables associated with infant and child mortality in Lesotho using DHS data for 2009. In particular the chapter examines the extent to which the survival outcomes of siblings are associated with the above class of variables.

The extra variability of infant and child survival across households and communities even after accounting for different known determinants of mortality will be taken into account by means of frailty terms.

Frailty models were fitted using the STATA software version 11.

7.2 Measurement of the Family and Community Frailty Effect

For the frailty model, we suppose that conditional on the frailty, z_i the hazard function $h_{ik}(t)$ for the failure time of the k^{th} children in the family ($k = 1, 2.3 \ldots, k; i = 1, 2, 3 \ldots, n$) follows the usual proportional hazards form and is given by

$$h_{ik}(t) = h_0(t) z_i \exp\left(\beta' X_{ik}\right), t > 0 \tag{7.1}$$

where z_i is group-level (frailty). These frailties are unobservable, assumed to be independent and identically distributed with unity mean and unknown variance θ. Each family could have different values of random effects and the variability in the $z_i s$ reflect heterogeneities of risks between families.

The frailty is often assumed to follow gamma distribution for the sake of computational convenience and convergence [42]; [59]; and [70] and this model is expected to yield correct $z-$ratios, on which researchers rely heavily for their conclusion.

We used the STCOX command in STATA to compute the coefficients for the family and community frailty effects for infant and child mortality. We fitted the following two models to the data:

- Model I: Single random effect to allow for clustering by family;
- Model II: Single random effect to allow for clustering by community.

The estimated coefficients from the two models are interpreted just as in a standard hazard model, while the estimated parameters describing the distributions of the frailty effects are interpreted as variances of the frailty distribution. If the variance is zero, observations from the same family or community are independent. A larger variance implies greater heterogeneity in frailty across families or communities and greater correlation among individuals belonging to the same family or community.

7.3 Data

The data used in this study was the 2009 Lesotho Demographic and Health Survey (LDHS). The sample was selected using a two-stage stratified random sampling design that relied on a sampling frame. Fieldwork was conducted between April and September 2009 and achieved an overall response rate of 97% of households and 96% of women aged $15-49$ who were eligible for an individual interview. The interview includes a retrospective maternity history that collects data on date of birth, survival status, and age at death for all children each woman had given birth to.

7.4 Variables in the Study

The table below gives a summary distribution of the socio-economic, demographic, health and environmental characteristics that are considered as the most important determinants of child survival status.

Table 7.1: Urban and Rural areas percentage distribution of live births by some of the selected explanatory variables

Variables	**Category**	**Urban**	**percent**	**Rural**	**percent**
Child is alive	No	55	8.2	338	10.2
	Yes	617	91.8	2989	89.8
	Total	672	100	3327	100
Preceding birth	less $<$ 18 months	21	6.1	115	5.4
	$18-35$ months	61	17.8	737	34.3
	$>$ 36 months	260	76.0	1297	60.4
	Total	342	100	2149	100
Highest education	No education	7	1.1	83	5.5
	Primary incomplete	95	14.1	1188	35.7
	Primary complete	125	18.6	1010	30.4
	Secondary+	445	66.2	1046	31.4
	Total	672	100	3327	100
Succeeding birth	$<$ 19 years	16	20.5	107	14.5
	$19-35$ years	33	42.3	422	57.2
	36 or more	29	37.2	209	28.3
	Total	78	100	738	100
Type of toilets	Pit Latrine	39	6.2	5	0.2
	Water closet	537	84.5	1209	38.5
	No facility	57	9.0	1930	61.4
	Total	633	100	3144	100
Water source	Piped water	516	86.0	1677	53.2
	Well water	48	8.0	911	28.9
	River, streams, rain water	36	6.0	564	17.9
	Total	600	100	3152	100
Breastfeeding status	Never	55	8.2	234	7.1
	Ever	613	91.8	3076	92.9
	Total	668	100	3310	100
Breastfeeding duration	$<$ 6 months	174	25.9	635	19.1
	$\geq$ 6 months	498	74.1	2692	80.9
	Total	672	100	3327	100
Birth size	Small	66	9.9	507	15.4
	Average	488	72.8	2138	64.9
	Large	116	17.3	648	19.7
	Total	670	100	3293	100
Birth order	1st birth	328	48.8	1166	35.1
	2-4th	311	46.3	1621	48.7
	$>$ 4th birth	33	4.9	540	16.2
	Total	672	100	3327	100
Child is twin	Single birth	658	97.9	3222	96.8
	Multiple birth	14	2.1	105	3.2
	Total	672	100	3327	100

Continued on next page

Table 7.1 – *Continued from previous page*

Variables	**Category**	**Urban**	**percent**	**Rural**	**percent**
Place of delivery	Home	86	13.1	1616	49.5
	Public	528	80.4	1583	48.5
	Private	43	6.5	65	2.0
	Total	657	100	3264	100
Mother's age at first birth	<20yrs	290	43.2	1819	54.7
	≥20	382	56.8	1508	45.3
	Total	672	100	3327	100
Religion	Roman	274	40.9	1419	42.7
	Christian	239	36.7	1594	47.9
	Other Christian	12	1.8	78	2.3
	Islam	55	8.2	236	7.1
	Total	670	100	3327	100
Wealth index	Poorest	1	0.2	1174	35.3
	Poorer	89	13.2	1552	46.6
	Rich	582	86.6	601	18.1
	Total	672	100	3327	100
Sex of child:	Male	316	47.0	1695	51.0
	Female	356	53.0	1632	49.1
	Total	672	100	3327	100

Table 7.1 shows the distribution of the urban and rural areas in Lesotho in $2005 - 2009$. As shown in the table about 91.8% of births in urban areas were live births compared to 89.8% in the rural areas. In urban areas, most of the births took place at the public hospitals (80.4% of all births)while in rural areas almost an equal percentage of births took place in public hospitals and at homes (48.5% in public hospitals and 49.5% at home). The percentage of female births seem to be higher than male births in urban areas (53% females and 47% males) and lower in rural areas (49% females and 51% males). Most of the respondents in both urban and rural areas reported their babies' sizes to be of average body size. The percentage of teenage deliveries is about 54.7% in rural areas while it was 43.2% in urban areas. Likewise, those who gave birth at age of 20 and above is about 56.8% in urban areas while it was 45.3% in the rural areas.

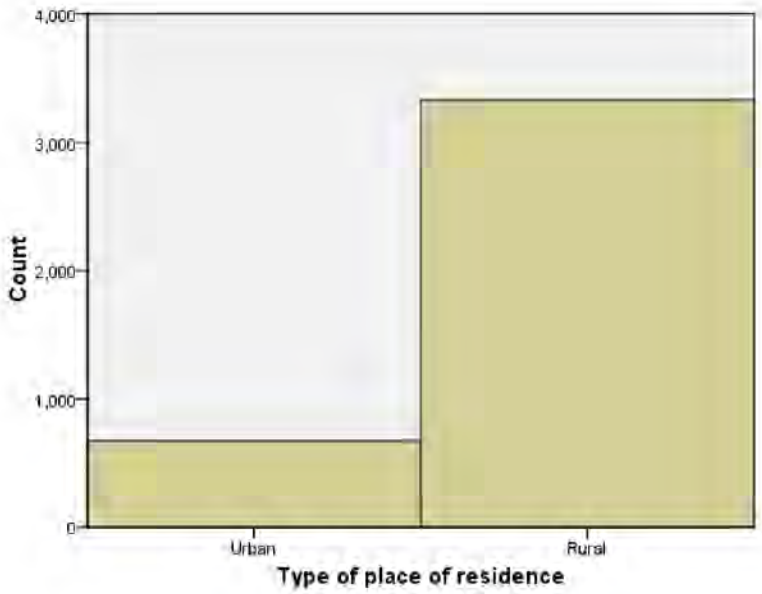


Figure 7.1: Proportion of the residence during the survey period 2005 – 2009 for Lesotho.

Figure 7.2: Proportion of those who survived, and died reported during the survey period $2005 - 2009$ for Lesotho.

Figure 7.1 and 7.2 represent the number of children survived and died during the study as well as the number of rural and urban areas present during the study in Lesotho in $2005 - 2009$. The above figure shows a larger number of people from rural areas than urban areas that participated in the study. It also shows that few children died during the study.

7.5 Dependent Variable and Explanatory Variables

In our application the age at a child's death or survival is used as the time to event where the event whether a child is dead or is alive at the time of the survey.

The explanatory variables that determine infant and child survival status are socioeconomic, demographic, and biological variables[39]. These variables include the gender of the child, age of the child, preceding birth interval, birth order of the child, mother's age at child birth, place of residence, mother's education level, mother's work status, household economic status, etc. as shown in the following table.

Table 7.2: Determinants of infant and child survival in Lesotho.

Variable	**Definition**
Socioeconomic	
Mother's highest educational	None (0), Primary (1), Secondary (2) Higher (3)
Mother's occupation	Not working (0), Agriculture (1), Sales (2), Other (3)
Wealth index	Poorest (1), Poorer (2), Rich (3)
Type of residence	Urban (1), Rural (2)
Demographic	
Age at birth	Age of the mother at time of child birth
Age at first birth	Age of the mother at her first birth
Sex of the child	Male (1), Female (2)
Type of place of residence	Urban (1), Rural(2)
Biological variables	
Birth order	1st birth(0), 2-4th birth(1), >4th birth(2)
Birth size	Average (1), Large/ very large (2), small/ very small (3)
Breast feeding	Never (0), Ever (1)
Breast feeding duration	< 6 months(0), ≥ 6 months (1)
Previous birth interval	Time space in months between this and preceding child birth
Place of delivery	Home (1), Public Health Sector (2), Private Health Sector (3)

7.6 Data Processing

Some continuous variables were categorized before starting the analysis. Age of mother at birth was categorized into six age groups which are, $15-19, 20-24, 25-34, 30-34, 35-39$ and $40-44$. Breast-feeding status was categorized into ever and never while birth order was categorized into three groups: first order, $2-4$ birth order and 5+ birth order. The preceding birth interval was grouped into two groups: less than two years and two or more years. Because some first birth orders do not have a preceding birth interval, preceding birth interval and birth order were combined in one variable of five categories: first order, 2-4 birth order with less than 2 years of preceding spacing, $2-4$ birth order with 2 years or more of preceding spacing, 5th or more birth order with less than 2 years of preceding spacing and 5th or more birth order with two years or more of preceding spacing. This categorization method of birth spacing variables had been used by [52] in their study, and is used here because of its appropriateness.

Categorized variables were further edited by combining some of their groups in one or two groups either because of the small number of observations in those categories or to make the analysis and the interpretation more meaningful. For instance, in the variable mother's occupation, occupations like professional, technical, managerial, clerical, services, domestic and manual were all combined with sales occupation in one group because of the small number of observations in those occupational categories. At the end of fitting each model, the Wald test was used to test the overall significance of the variables selected by the model. The p-value produced by the test was measured using the rank-order of the explanatory factors in terms of their importance in determining the outcome, since the overall P-value for a factor is a measure for the relative need of that variable in explaining the variability in the outcome [49]. However, several authors have identified problems with the use of the Wald statistics [50];[1]. This technique enables adjusting for many explanatory factors and controlling for many confounders at the same time as it enables easy detection of interactions between explanatory factors. It is flexible, easy to use and usually gives meaningful interpretation by giving the magnitude and the direction of the association between explanatory and outcome variables [41].

Chapter 8

Results

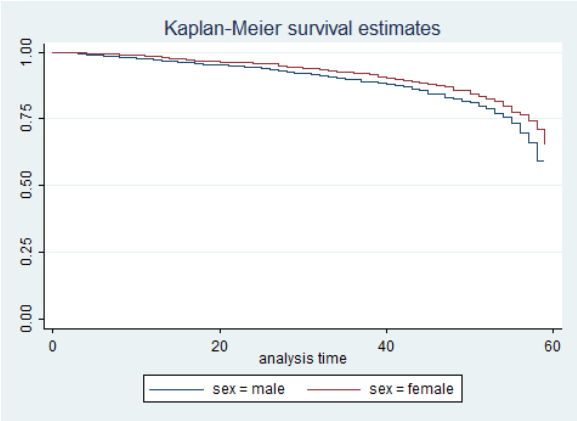
8.1 Selected Explanatory Variables

As a preliminary exploratory analysis was conducted to study the association of predictor variables for child survival.

Table 8.1: Chi-square test of association

Variable	χ^2	df	p-value
Maternal highest educational level	5.881	3	0.1180
Wealth index	9.300	4	0.0540
Type of place of residence	2.460	1	0.1175
Age at birth	23.099	6	0.0010
Sex of the child	9.085	1	0.0030
Birth order	11.258	2	0.0040
Birth type	83.940	1	0.0000
Birth size	40.939	2	0.0000
Breast feeding	483.433	1	0.0000
Breast feeding duration	143.204	1	0.0000
Previous birth interval	12.625	2	0.0020
Place of delivery	10.301	2	0.0060

Table 8.1 shown above represents variables found to be more significant than other variables that were in the data. The software used to check whether the variable is significant or not was SPSS. These variables will be used in different models (Cox models) to check which model is preferable. The results indicate that there is association between child survival and the following predictor variables: Age at birth, Sex of the child, Birth order, Birth type, Birth size, Breast feeding, Breast feeding duration and Previous birth interval; the p-values are all less than the 5 percent level of significant.



(a) Kaplan-Meire survival estimate by gender

Figure 8.1: Figures of the Kaplan-Meire by gender in Lesotho

Figure 8.1 shows the plot of the K-M curves for females and males shown on the same graph. Notice that the K-M curve for females is consistently higher than the K-M curve for males. This figure indicate that females have a better survival curves than males.

Table 8.2: The results of Cox proportion hazard model

Variable	β	SE	Wald	df	P-value	Exp(β)	95.0% CI for Exp(β) Lower	Upper
Birth size: Small			17.0363	2	0.0002			
Average	-0.5321	0.1306	16.5986	1	<0.001	0.5874	0.4547	0.7587
Large	-0.2893	0.1599	3.2730	1	0.0704	0.7488	0.5473	1.0244
Birth order: 1st birth			2.7900	2	0.2490			
2nd-4th birth	0.0246	0.1431	0.0300	1	0.8634	1.0249	0.7743	1.3567
>4th birth	0.3523	0.2443	2.0803	1	0.1492	1.4224	0.8812	2.2958
Child is twin	0.7272	0.1815	16.0565	1	<0.001	2.0693	1.4499	2.9533
Place of birth: Home			23.15427	2	9.38E-06			
Public	-0.5396	0.1133	22.6696	1	<0.001	0.5830	0.4669	0.7280
Private	-0.5985	0.3710	2.6016	1	0.1068	0.5497	0.2656	1.1374
Age at 1st birth	0.1697	0.1164	2.1255	1	0.1449	1.1850	0.9432	1.4887
Wealth index: Poorest			2.6039	2	0.2720			
Poorer	0.2027	0.1289	2.4730	1	0.1158	1.2247	0.9513	1.5768
Rich	0.0893	0.1514	0.3480	1	0.5553	1.0934	0.8127	1.4712
Age group: 45-49			422.834	6	<0.0010			
15-19	1.7270	0.1450	141.7460	1	<0.0010	5.6250	4.233	7.4750
20-24	0.7680	0.1350	32.4200	1	<0.0010	2.1560	1.655	2.8090
25-34	0.5120	0.1360	14.2260	1	<0.0010	1.6680	1.655	2.1760
30-34	0.5420	0.1380	15.4740	1	<0.0010	1.7190	1.2790	2.2510
35-39	0.4740	0.142	11.1900	1	<0.0010	1.6060	1.217	2.1210
40-44	0.2120	0.1500	1.990	1	<0.0010	1.2360	0.9210	1.6610

Table 8.2 shows the results obtained after applying the Cox PH models to the LDHS 2009 data using age at death as the time to event. The exponent of the parameter estimates gives the hazard ratio associated with a particular variable in the first column.

The hazard of death for a child whose mother's place of birth is public hospital is 0.5830 ($95\%CI : 0.4669, 0.7280$) times the hazard of death for a child whose mother's place of birth is at home (reference category). This indicates a reduced hazard of death of about 41.7% for a child born at the public hospital than a child born at home, and is statistically significant ($p < 0.05$).

A child born with average body size has a hazard of death of 0.5874 ($95\%CI : 0.4547, 0.7587$), times the hazard of a child born with a small body size. This shows a reduced hazard of death of an average size baby of about 41.3% than the small size baby. The effect is statistically significant ($p < 0.001$).

The hazard of death for a child whose mother's age is between $15 - 19$ years and $20 - 24$ years is 5.6250 ($95\%CI : 4.233, 0.7.4750$) and 2.1560 ($95\%CI : 1.655, 2.8090$) respectively. This shows a higher mortality hazard for mother's aged $15 - 19$ with wide confidence interval than that of mothers aged $45 - 49$ years (the reference category), meaning that mothers aged $15 - 19$ years do not take good care of their babies. The reason might be because they are still teenagers or they are full time students, so they do not have enough time for babies. Likewise the mortality hazard for mothers aged $40 - 44$ years is 1.2360 ($95\%CI : 0.9210, 1.6610$) times higher than that of mothers aged $45 - 49$ years the reference category. This indicates that mothers aged $40 - 44$ years have not enough time to take care of their babies compared with mothers aged $45 - 49$ years. The reason here might be because they are working, they wake up in the morning, go to work and come back in the evening [47].

As shown in the table above some variables such as wealth index, large size baby, birth order are not statistically significant.

8.1.1 Results Using Household Frailty

The following results were obtained using different softwares namely the R and STATA software. In R we used Cox-PH formula and in STATA we used STCOX command. These softwares are used to compute coefficients and relative risks (hazard ratio) for household (family) and community effects for under five child mortality using Lesotho 2009 data. With this data we fitted the single random effect to allow for clustering by household. STATA was found to be more suitable software than R because STATA yields smaller output values for each variable than R.

The following tables provide the Cox PH and the Weibull results which consist of two models per table. The left-hand side model is the model without household or alternatively family frailty and the right-hand side model is the model with household frailty. For the same effects the results consists of the Hazard ratio (HR) or relative risks, Standard error (SE), Z test statistic values, p-value of Z test statistics and 95% confidence interval ($95\%CI$)). It will be

noted that the hazard ratios for both models (with and without frailty) are very close to each other.

The estimated parameters describing the distributions of the frailty effects are interpreted as variances of the frailty distribution. If the variance is zero, it means that household to household heterogeneity is negligible. A large variance implies greater heterogeneity in frailty across households and a greater correlation among individuals belonging to the same household.

Some variables shown on these tables have no results when the frailty effect is taken into account; that indicates that the variable gives an error or keeps on running without giving results when the frailty effect is included in the model. This indicates that the frailty effect has an effect on the variable.

Table 8.3: Cox PH model results with and without household frailty

Cox PH without household Frailty model							Cox PH with household Frailty Model					
Variables	HR	SE	z	P>z	95% CI.		HR	SE	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.3450	0.0816	-4.5	<0.001	0.2170	0.5486	0.3358	0.0803	-4.5600	<0.001	0.2101	0.5368
25-29	0.2558	0.0686	-5.08	<0.001	0.1511	0.4330	0.2539	0.0690	-5.0500	<0.001	0.1492	0.4322
30-34	0.1877	0.0568	-5.53	<0.001	0.1037	0.3398	0.1871	0.0571	-5.4900	<0.001	0.1029	0.3404
35-39	0.1969	0.0655	-4.89	<0.001	0.1026	0.3780	0.1977	0.0664	-4.8300	<0.001	0.1024	0.3818
40-44	0.1924	0.0707	-4.48	<0.001	0.0937	0.3956	0.1909	0.0711	-4.4500	<0.001	0.0920	0.3961
45-49	0.1675	0.0736	-4.06	<0.001	0.0708	0.3965	0.1640	0.0732	-4.0500	<0.001	0.0684	0.3932
Birth order: 1*st* birth	-	-	-	-	-	-	-	-	-	-	-	
$2-4$th order	1.3335	0.1983	1.9300	0.0530	0.9963	1.7848	1.3274	0.1987	1.8900	0.0590	0.9897	1.7800
>4 birth	2.0723	0.5300	2.8500	0.0040	1.2555	3.4202	2.0338	0.5262	2.7400	0.0060	1.2248	3.3771
Birth size: Small												
Average	0.4803	0.0607	-5.8100	<0.001	0.3750	0.6153	0.4770	0.0609	-5.8000	<0.001	0.3714	0.6126
Large	0.5871	0.0916	-3.4100	0.0010	0.4324	0.7972	0.5809	0.0916	-3.4400	0.0010	0.4265	0.7914
Place of birth: Home												
Public	0.6194	0.0703	-4.2200	<0.001	0.4960	0.7736	0.6158	0.0706	-4.2300	<0.001	0.4919	0.7709
Private	0.6755	0.2491	-1.0600	0.2870	0.3279	1.3915	0.6849	0.2544	-1.0200	0.3080	0.3308	1.4183
Wealth: Poorest	-	-	-	-	-	-	-	-	-	-	-	
Poorer	1.2297	0.1574	1.6200	0.1060	0.9570	1.5803	1.2184	0.1577	1.5300	0.1270	0.9455	1.5702
Rich	1.3568	0.2037	2.0300	0.0420	1.0108	1.8211	1.3511	0.2051	1.9800	0.0470	1.0034	1.8192
Breast feeding: Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever:	0.0947	0.0113	-19.7600	<0.001	0.0749	0.1196						
Breast feeding duration:<6 months	-	-	-	-	-	-	-	-	-	-		
≥ 6 months	0.1312	0.0145	-18.4200	<0.001	0.1057	0.1628	-	-	-	-	-	
θ	-	-	-	-	-	-	0.0958	0.0692	-	-	-	

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 2.44$, the value of $p = 0.059$ at 5 percent level of significant

Table 8.3 shows the results of a Cox model for the survival times namely, age at death or survival of a child. The left-hand side results represent the Cox model without household frailty while the right-hand side model results represent the Cox model with household frailty. The results obtained on the left-hand side of the table are almost the same as the results obtained on the right-hand side of the table. Likewise, variables that are significant on the left-hand side are also significant on the right-hand side of the table, vice versa.

For example, on the right-hand side of the table, the hazard of death for a child whose mother's age is $20-24$ years is reduced about 65.5% from the hazard of death for a child whose mother's aged is between $15-19$, years and it is statistically significant $p < 0.05$ level of significance. This indicates that when Cox PH is used the mortality hazard of babies being born to mothers aged $20-24$ years is 0.3450 $(95\%CI : 0.2170, 0.5486)$; when the unobserved household effect is taken into account, the mortality hazard is reduced about 66.42% from the mortality hazard for a child whose mother's age is between $15-19$. In the Cox PH with frailty model, the mortality hazard for children born to mothers aged $25-29$ years is 0.2539 and it's confidence interval is as little as 0.1492 and as much as 0.4322. Clearly here, when the unobserved household effect is taken into account, the hazard of death of a child decreases from that of the Cox PH without frailty. Likewise, for mothers aged $30-34, 35-39, 40-44$ and $45-49$ years, their mortality rate is lower than the mortality rate for mothers aged $15-19$ years which is the reference category.

The average size babies at birth have 47.70% excess hazard than the small size babies at birth when the unobserved household effect is taken into account. This means that small child babies (reference category) are more likely to die than average size babies, and it's p-value is $p < 0.001$, indicating highly significant.

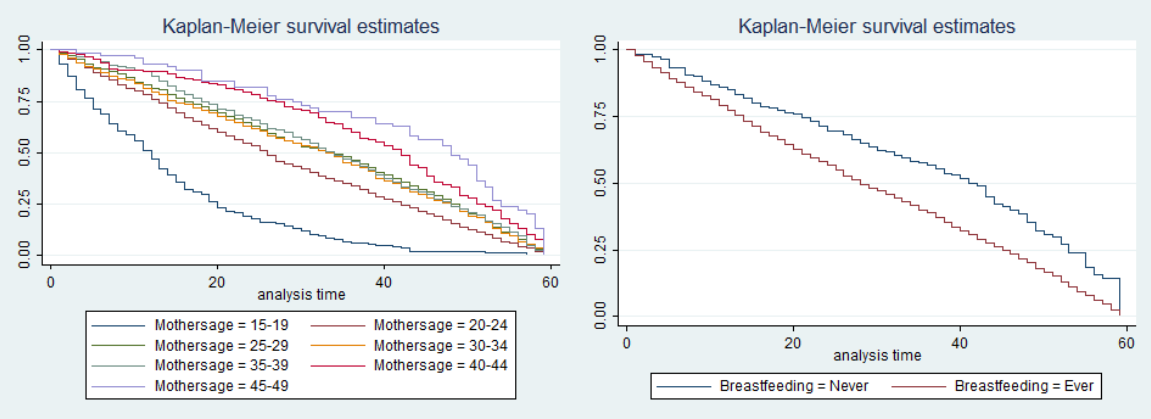
The mortality rate (hazard of death) for a child born at public hospital is 0.6158 $(0.419, 07709)$ when the unobserve household effect is taken into account. This indicates a hazard of death of about 61% times lower than the hazard of death of a child born at home, and it is said to be statistically significant with a p-value which is less than 5% level of significant. Likewise, the hazard of death for a child born at private hospital is 0.6849 $(95\%CI : 0.3308, 1.4183)$. This indicates a reduced hazard of death for a child born at private hospital of about 31.51% than a hazard of death for a child born at home. It is statistically insignificant $(p = 0.3080)$.

Infants who are breastfed have significantly lower hazard of death HR $= 0.0947$ $(95\%CI : 0.0749, 0.1196)$ than those who have never been breastfed $(p < 0.001)$.

Some variables are statistically insignificant $(p > 0.05)$ such as $2-4th$ birth order, and the wealth index (poorer).

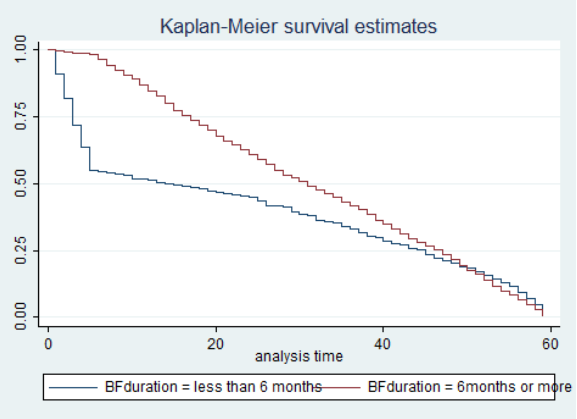
When the unobserved frailty component (household) effect is taken into account as a gamma frailty, the variance of frailty θ is estimated to be 0.0958 with non-significant likelihood ratio test for the presence of heterogeneity ($p > 0.05$). This indicates that the variance is not significantly different from 0, therefore there is not much evidence of household to household heterogeneity given the fixed effects variables that have been included in the model.

As shown on the table above, Cox PH with household frailty is considered the best model as it has the lowest mortality hazard for each variable than Cox PH without frailty.



(a) The KM plot for mother's age

(b) The KM plot for breast feeding status



(c) The KM plot for breast feeding duration

Figure 8.2: Figures of the Cox PH assumption by mother's age, breastfed duration and breastfed status in Lesotho

Figure 8.2 shows the graphs of KM (Kaplan Meier). The first panel of the graphs represents mother's age and the breastfeeding status, and likewise the bottom represents the breastfeeding duration in Lesotho.

Parallel lines for mother's age and breastfeeding status, implies that the proportional hazard assumptions for mother's age and breastfed status is not violated. Likewise for the bottom panel, crossing lines for breastfeeding duration indicate that the PH assumption is violated.

Table 8.4: Test of proportional-hazards assumption using Schoenfeld residual

variable	ρ	χ^2	df	P-value
Breastfeeding status	0.04249	6.57	1	0.0440
Mother's age	0.13086	66.78	1	0.0104
breastfed duration	-0.00176	0.01	1	0.9170
global test		69.87	3	<0.0001

Table 8.4 shown above shows that for both mother's age and the breast-feeding-status, there is evidence that the proportional hazard assumption has been violated ($p < 0.05$), while for the breastfed duration, there is no evidence that the proportional hazard assumption has been violated ($p > 0.05$).

Table 8.5: Hazard ratios of child mortality associated with various socio-demographic variables

	Weibull model without household frailty						Weibull model with household frailty					
Variable	HR	SE.	z	P>z	95% CI.		HR.	SE.	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.3675	0.0853	-4.3200	<0.001	0.2332	0.5791	0.3611	0.0845	-4.3600	<0.001	0.2283	0.5711
25-29	0.3120	0.0776	-4.6800	<0.001	0.1916	0.5081	0.3120	0.0779	-4.6600	<0.001	0.1912	0.5090
30-34	0.2363	0.0662	-5.1500	<0.001	0.1365	0.4092	0.2372	0.0667	-5.1200	<0.001	0.1367	0.4114
35-39	0.2613	0.0792	-4.4300	<0.001	0.1442	0.4734	0.2642	0.0804	-4.3700	<0.001	0.1454	0.4798
40-44	0.2630	0.0889	-3.9500	<0.001	0.1356	0.5101	0.2651	0.0902	-3.9000	<0.001	0.1361	0.5163
45-49	0.2566	0.1057	-3.3000	0.0010	0.1145	0.5753	0.2571	0.1067	-3.2700	0.0010	0.1139	0.5801
Birth order: 1st birth	-	-	-	-	-	-	-	-	-	-	-	-
2-4th birth	1.2693	0.1850	1.6400	0.1020	0.9539	1.6890	1.2588	0.1843	1.5700	0.1160	0.9448	1.6772
>4th birth	1.7424	0.4262	2.2700	0.0230	1.0789	2.8142	1.7020	0.4198	2.1600	0.0310	1.0496	2.7599
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.4681	0.0590	-6.0200	<0.001	0.3655	0.5993	0.4658	0.0592	-6.0100	<0.001	0.3631	0.5976
Large	0.5711	0.0891	-3.5900	<0.001	0.4207	0.7753	0.5679	0.0892	-3.6000	<0.001	0.4174	0.7727
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.6488	0.0735	-3.8200	<0.001	0.5196	0.8102	0.6471	0.0738	-3.8200	<0.001	0.5174	0.8092
Private	0.6714	0.2475	-1.0800	0.2800	0.3260	1.3828	0.6793	0.2519	-1.0400	0.2970	0.3285	1.4050
Wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	1.2295	0.1571	1.6200	0.1060	0.9572	1.5793	1.2244	0.1577	1.5700	0.1160	0.9512	1.5761
Rich	1.3338	0.2002	1.9200	0.0550	0.9938	1.7901	1.3296	0.2012	1.8800	0.0600	0.9883	1.7886
Breast feeding:Never												
Ever	0.0977	0.0116	-19.6600	<0.001	0.0774	0.1232						
Breast feeding duration: <6 months	-	-	-	-	-	-	-	-	-	-	-	-
≥ 6 months	0.1327	0.0145	-18.4600	<0.001	0.1071	0.1645						
$\ln\gamma$	0.5512	0.0429	12.8400	0.0000	0.4670	0.6353	0.5529	0.0428	12.9300	<0.001	0.4691	0.6368
γ	1.7353	0.0745	-	-	1.5952	1.8876	1.7383	0.0744	-	-	1.5985	1.8903
$1/\gamma$	0.5763	0.0247	-	-	0.5298	0.6269	0.5753	0.0246	-	-	0.5290	0.6256
θ	-	-	-	-	-	-	0.0745	0.0661	-	-	0.0131	0.4238
$\ln\theta$	-	-	-	-	-	-	-2.5969	0.8870	-2.9300	0.0030	-4.3353	-0.8585

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 1.55$, the value of $p = 0.107$ at 5 percent level of significant

Table 8.5 shows the results of a parametric regression model assuming a Weibull distribution for the survival times namely age at death or survival of a child. The left-hand side results are for a Weibull without family frailty model and the right-hand side results are for Weibull with family frailty model. The results obtained on the left-hand side model are almost the same as on the right-hand side model. Variables that are significant on the left-hand side are also significant on the right-hand side and those that are not significant on the left-hand side are also insignificant on the right-hand side.

In both models (left-hand side and right-hand side), there is a decreasing mortality hazard for children born to mothers aged $20-24$ years at birth compared to children born to mothers aged $15-19$ years, that is 0.3675 ($95\%CI$: $0.2332, 0.5791$) and 0.3611 ($95\%CI$: $0.2283, 0.5791$). This means that the mortality hazard for children born to mothers aged $20-24$ is lower than the mortality hazard for children born to mothers aged $15-19$ years. Even though the mortality hazard for babies born to mothers aged $20-24$ years is lower when the Weibull without frailty is taken into account, it is even smaller when the unobserved frailty effect is taken into account. In the Weibull with frailty model, mothers aged $25-29$ years, the mortality hazard for babies being born by these mothers remains the same when the unobserved household effect is taken into account, that is (HR=0.3120; 95% CI: 0.1916, 0.5081) for Weibull without frailty model while for Weibull with frailty model it is ($HR = 0.3120; 95\%CI : 0.1912, 0.5090$). All HR are less than 1. It kis indicating the higher mortality hazard for children born to mothers aged $15-19$. When the unobserved household effect is taken into account the mortality hazard for children born to mothers aged $30-34, 35-39, 40-44$ and $45-49$ years increases, but it does not exceed the reference category ($15-19$ years). This means the reference category still has higher mortality hazard compared to all these mothers aged mentioned on the previous sentence. The reason behind this might be because these mothers are still teenagers and they need their mothers to take care of them too. Another reason might be that she might be scared to tell her parents that she is pregnant, so she decides to abort the baby. The table shows that the variable age of the mother is statistically significant because its p-value is less than 5% level of significant in both models.

On the right-hand side output, higher than the fourth birth order has a mortality of 1.7020 ($95\%CI : 1.0496, 2.7599$) times higher than that of the 1st birth order, this implies about 70.2% increase in the mortality hazard than the mortality hazard for the first birth order, and when the unobserved household effect is not taken into account the mortality hazard is higher than the fourth birth order is 1.7424 ($95\%CI : 1.0789, 2.8142$). On the left-hand side the risk of death times higher than the hazard for first birth order; this clearly shows that the presence of unobserved frailty decreases the risk of death by 4% on the right-hand side model, it is statistically significant because the confidence interval does not include one, with the p-value of < 0.005.

On the right hand side, the relative risk for average birth size babies is about 0.4658 ($95\%CI : 0.3631, 0.5976$) times that of small size babies at birth, while the larger size babies at birth have a relative risk of about 0.5679 ($95\%CI : 0.4174, 0.7727$) times the relative risk of small size babies (the reference category) at birth, and these are statistically significant. The reason behind this or the cause of death, might be that a child being born before time, for example born in seventh month instead of ninth month, or the mother had a miscarriage.

Some categorical variables on the table are significant and non-significant variables, such as the birth order and the place of birth of a child.

When the unobserved frailty component (household) effect is taken into account as a gamma frailty, the variance of frailty θ is estimated to be 0.0745 and an insignificant likelihood ratio test for the presence of heterogeneity ($p > 0.05$). This indicates that the variance is not significantly different from 0, therefore we conclude that observations from the same household (family) are independent.

The estimate for the shape parameter γ in Weibull without and with frailty models is 1.7353 and 1.7383 respectively, suggesting a slightly decreasing hazard over time. The Weibull with frailty model is considered the best fit, because it consists of lower mortality hazard than the Weibull without frailty.

8.1.2 Results Using Community Frailty

In the following tables we will be using community frailty as an unobserved frailty effect, while in the previous tables we used household frailty as an unobserved frailty effect. The following tables provide the Cox PH and Weibull results which consist of two models per table. The left-hand side model is the model without community frailty and the right-hand side model is the model with community frailty. For the same effects the results consist of the Hazard ratio(HR) or relative risk, Standard error(SE), Z test statistics value, p-value of Z test statistics and 95% confidence interval ($95\%CI$)). It will be noted that the hazard ratios for both models (with and without frailty) are almost the same.

The estimated parameter describing the distributions of the frailty effects are interpreted as variances of the frailty distribution. If the variance is zero, it means that community to community heterogeneity is negligible. A large variance implies greater heterogeneity in frailty across community and a greater correlation among individuals belonging to the same community.

Some variables shown on these tables have no results when the frailty effect is taken into account; that indicates that the variable did not converge when the frailty effect is included in the model.

Table 8.6: Cox PH model results with and without community frailty

Cox PH model without community frailty model							Cox PH model with community frailty model					
Variables	HR	SE.	z	P>z	95% CI.		HR	SE.	z	p>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.3450	0.0816	-4.5	<0.001	0.2170	0.5486	0.3289	0.0808	-4.5200	<0.001	0.2032	0.5325
25-29	0.2558	0.0686	-5.08	<0.001	0.1511	0.4330	0.3664	0.0897	-4.1000	<0.001	0.2268	0.5920
30-34	0.1877	0.0568	-5.53	<0.001	0.1037	0.3398	0.2827	0.0747	-4.7800	<0.001	0.1684	0.4744
35-39	0.1969	0.0655	-4.89	<0.001	0.1026	0.3780	0.3209	0.0871	-4.1900	<0.001	0.1885	0.5463
40-44	0.1924	0.0707	-4.48	<0.001	0.0937	0.3956	0.3160	0.0932	-3.9100	<0.001	0.1773	0.5632
45-49	0.1675	0.0736	-4.06	<0.001	0.0708	0.3965	0.2790	0.1106	-3.2200	<0.001	0.1283	0.6066
Birth order: 1st birth	-	-	-	-	-	-	-	-	-	-	-	-
$2-4th$ birth	1.3335	0.1983	1.9300	0.0530	0.9963	1.7848	-	-	-	-	-	
$>$ 4birth	2.0723	0.5300	2.8500	0.0040	1.2555	3.4202	-	-	-	-	-	
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.4803	0.0607	-5.8100	<0.001	0.3750	0.6153	0.7327	0.1046	-2.1800	0.0290	0.5539	0.9693
Large	0.5871	0.0916	-3.4100	0.0010	0.4324	0.7972	0.5809	0.0916	-3.4400	0.0010	0.4265	0.7914
Place of birth: Home	-	-	-	-	-		-	-	-	-	-	-
Public	0.6194	0.0703	-4.2200	<0.001	0.4960	0.7736	0.5909	0.0688	-4.5200	<0.001	0.4704	0.7423
Private	0.6755	0.2491	-1.0600	0.2870	0.3279	1.3915	0.5425	0.2056	-1.6100	0.1070	0.2581	1.1403
Wealth: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	1.2297	0.1574	1.6200	0.1060	0.9570	1.5803	-	-	-	-	-	-
Rich	1.3568	0.2037	2.0300	0.0420	1.0108	1.8211	-	-	-	-	-	-
Breast feeding: Never	-	-	-	-	-		-	-	-	-	-	-
Ever:	0.0947	0.0113	-19.7600	<0.001	0.0749	0.1196	0.0779	0.0105	-18.9000	<0.001	0.0598	0.1015
Breast feeding duration:< 6 months	-	-	-		-	-	-	-	-	-	-	-
$\geq$6 months	0.1312	0.0145	-18.4200	<0.001	0.1057	0.1628	0.1126	0.0137	-17.9700	<0.001	0.0887	0.1429
θ	-	-	-	-	-	-	0.2148	0.0883	-	-	-	

Likelihood-ratio test of $\theta = 0 : \chi_1^2 = 10.1500$, the value of $p = 0.0010$ at 5 percent level of significant

Likewise, Table 8.6 shows that, on the right-hand side of the table, the hazard of death for a child whose mother has ever breastfed is 0.0779 ($95\%CI$: $0.0598, 0.1015$) times the hazard of death for a child whose mother has never breastfed. This indicates a hazard of death of about 7.8% times lower than the hazard of death for a child whose mother has never breastfed, and it is statistically significant since its p-value is less than 5% level of significance and the confidence interval does not include one. The reason behind this might be that breastfed babies are better nourished than those babies who drink milk, for example Nana, (powdered milk for small babies or infants, to drink after mixing it with boiled water) when they are young; breastfed is more preferable than other milk.

When the unobserved community effect is taken into account, the mortality rate for a child born at public hospital is 0.5909 ($95\%CI : 0.4704, 0.7423$). This shows that the hazard of death for a child born at public hospital is 59.1% times lower than the hazard of death for a child born at home, and is statistically significant ($p < 0.05$). The reason behind this might be because at the hospital there are qualified nurses and doctors who know how to assist a women when she is in labor whereas for home births there is usually no one qualified to help when a women is in labor.

On the right hand-side mortality rate for a child whose mother breastfed for 6 months or more is 0.1126 ($95\%CI : 0.0887, 0.1429$). Again the hazard of death is reduced by 88.7% from the hazard of death for a child whose mother breastfed for less than 6 months, likewise, it is statistically significant ($p < 0.05$).

The mortality hazard for average size child has increased from 48.03% to 73.27% when unobserved community effect is taken into account and is statistically significant ($p < 0.05$).

The mortality hazard for a child whose mother is aged between $20 - 24$ years is reduced by $67, 11\%$ from the hazard of death for a child whose mother is aged between $15-19$ years and it is statistically significant ($p < 0.05$), on the right hand side. The inclusion of the unobserved community effect has caused the increase in mortality for children whose mothers are aged $25-29, 30-34, 35-39, 40-44$ and $45-49$ years: that is 25.58% to 36.64%, 18.77% to 28.27%, 19.69% to 32.09, 19.24% to 31.60% and 16.75% to 27.90 respectively, compared to the mother aged $15-19$ years. The reasons behind this are that mothers aged $15-19$ years are still too young to have a baby; they do not have enough time to take care of their babies; and their parents are still taking care of them.

Other estimates shown on the table are statistically insignificant ($p > 0.05$), such as the place of birth (private hospital), wealth index, birth order ($2 - 4th$ birth) and birth size (large).

When the unobserved frailty component (community) effect is taken into ac-

count as a gamma frailty, the variance of frailty θ is estimated to be 0.2148 and an insignificant likelihood ratio test for the presence of heterogeneity ($p < 0.05$). This indicates that the variance is significantly different from 0, therefore we conclude that observations from the same community are not independent.

As shown in the table above the Cox PH without frailty is considered the best model than Cox with frailty, this is because Cox PH without frailty has lower mortality hazard for each variable.

Table 8.7: Hazard ratios of child mortality associated with various socio-demographic variables

	Weibull without community frailty						Weibull with community frailty					
Variable	HR	SE.	z	P>z	95% CI.		HR	SE.	z	P>z	95% CI.	
Breast feeding: Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever	0.0977	0.0116	-19.6600	0.0000	0.0774	0.1232	0.0852	0.0120	-17.5000	0.0000	0.0647	0.1123
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.6488	0.0735	-3.8200	0.0000	0.5196	0.8102	0.6015	0.0689	-4.4400	0.0000	0.4805	0.7528
Private	0.6714	0.2475	-1.0800	0.2800	0.3260	1.3828	0.5717	0.2144	-1.4900	0.1360	0.2742	1.1922
Breast feeding duration:$<$ 6 months	-	-	-	-	-	-	-	-	-	-	-	-
$\geq$ 6 months	0.1327	0.0145	-18.4600	0.0000	0.1071	0.1645	0.1178	0.0149	-16.8800	0.0000	0.0919	0.1510
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.4681	0.0590	-6.0200	0.0000	0.3655	0.5993	0.7125	0.1001	-2.4100	0.0160	0.5411	0.9383
Large	0.5711	0.0891	-3.5900	0.0000	0.4207	0.7753	0.9671	0.1645	-0.2000	0.8440	0.6930	1.3496
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.3675	0.0853	-4.3200	0.0000	0.2332	0.5791	0.3268	0.0794	-4.6000	0.0000	0.2030	0.5263
25-29	0.3120	0.0776	-4.6800	0.0000	0.1916	0.5081	0.3683	0.0889	-4.1400	0.0000	0.2295	0.5910
30-34	0.2363	0.0662	-5.1500	0.0000	0.1365	0.4092	0.2896	0.0755	-4.7500	0.0000	0.1738	0.4827
35-39	0.2613	0.0792	-4.4300	0.0000	0.1442	0.4734	0.3241	0.0869	-4.2000	0.0000	0.1916	0.5483
40-44	0.2630	0.0889	-3.9500	0.0000	0.1356	0.5101	0.3298	0.0960	-3.8100	0.0000	0.1865	0.5834
45-49	0.2566	0.1057	-3.3000	0.0010	0.1145	0.5753	0.3534	0.1345	-2.7300	0.0060	0.1676	0.7453
wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	1.2295	0.1571	1.6200	0.1060	0.9572	1.5793	-	-	-	-	-	-
Rich	1.3338	0.2002	1.9200	0.0550	0.9938	1.7901	-	-	-	-	-	-
Birth order: 1st birth	-	-	-	-	-	-	-	-	-	-	-	-
2-4th birth	1.2693	0.1850	1.6400	0.1020	0.9539	1.6890	-	-	-	-	-	-
$>$4th birth	1.7424	0.4262	2.2700	0.0230	1.0789	2.8142	-	-	-	-	-	-
$\ln \gamma$	0.5634	0.0424	13.2800	0.0000	0.4802	0.6465	0.5803	0.0427	13.5900	0.0000	0.4966	0.6639
γ	1.7566	0.0745	-	-	1.6165	1.9089	1.7865	0.0763	-	-	1.6431	1.9424
$1/\gamma$	0.5693	0.0242	-	-	0.5239	0.6186	0.5598	0.0239	-	-	0.5148	0.6086
θ	-	-	-	-	-	-	0.1580	0.0835	-	-	0.0561	0.4449
$\ln \theta$	-	-	-	-	-	-	-1.8454	0.5284	-3.4900	0.0000	-2.8811	-0.8098

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 5.17$, the value of $p = 0.012$ at 5 percent level of significant

Table 8.7 shows that in model one (left-hand side model) the hazard ratio for ever breastfed children is about 9.8% and it is as little as 0.0774 or as much as 0.1232 compared to hazard ratio of never breastfed children, while in model two (right-hand side model) it is about 8.5% and it is as little as 0.0647 or as much as 0.1123 compared to hazard ratio of never breastfed children. Ever breastfed children is significant in both models ($p < 0.001$).

The mortality hazard for children born to mother's age $20-24$ years is reduced by 67.32% when community frailty is taken into account than the mortality hazard for children born to mothers aged $15-19$, and is significant in both models, but the most notable finding is the change in the p-value of mothers aged $45-49$. It is significant ($p < 0.05$)in model one but is not significant ($p > 0.05$) in model two. As in Cox PH with and without frailty above, the inclusion of the unobserved community effect has caused the increase in mortality for children whose mothers are aged $25-29, 30-34, 35-39, 40-44$ and $45-49$, years that is 31.20% to 36.83%, 23.63% to 28.96%, 26.13% to 32.41, 26.30% to 32.98% and 25.66% to 35.34 respectively; compared to the mothers aged $15-19$ years, they have a lower mortality hazard.

Other categorical variables have a variable which is significant and not significant, such as the place of birth (public) has a hazard rate of 0.6488 ($95\%CI: 0.5196, 0.8102$) times the hazard rate of children born at home, in model one (left-hand side model); in model two (right-hand side model) has the hazard rate of 0.6015 ($95\%CI: 0.4805, 0.7528$) times the hazard rate of children born at home and it is significant in both models. Other significant and insignificant categories are the birth size: average size at birth is significant ($p < 0.05$) while large size at birth is insignificant ($p > 0.05$). The hazard for ever breastfed babies in model one (left-hand side model) and two (right-hand side model) respectively are 0.0977 ($95\%CI: 0.0774, 0.1232$) and 0.0852 ($95\%CI: 0.00647, 0.1123$) times the hazard of never breastfed children, They are statistically significant ($p < 0.05$).

On the right hand side, infants who are breastfed for ≥ 6 months have significantly lower hazard of death ($HR = 0.1178; 95\%CI: 0.0919, 0.1510$) than those who are breastfed to less than six months.

When the unobserved frailty component (community) effect is taken into account as a gamma frailty, the variance of frailty θ is estimated to be 0.1580 ($95\%CI: 0.056, 0.4449$) and a significant likelihood ratio test for the presence of heterogeneity ($p < 0.05$). This indicates that the variance is significantly different from 0, therefore we conclude that there is greater heterogeneity in frailty across communities and a greater correlation among individuals belonging to the same community.

The estimates for the shape parameter γ in Weibull with and without frailty models are 1.7353 and 1.7865 respectively, suggesting a slightly decreasing haz-

ard over time. The Weibull with frailty model is considered the best fit, because it consists of lower mortality hazard than the Weibull without frailty.

Table 8.8: The AIC and BIC for each model used

Model without household frailty			Model with household frailty	
Model	AIC	BIC	AIC	BIC
Exponential	2633.717	2727.578	2634.873	2734.991
Lognormal	2580.877	2680.996	2581.588	2687.964
Weibull	2503.627	2603.745	2504.079	2610.454
Log-logistic	2518.436	2618.554	2518.978	2625.353
Cox PH	5359.543	5453.404	5357.106	5450.967

The Weibull model with household frailty shown on the above table has a smaller Akaike information criterion (AIC) value (2504.079) compared to other models, so it is a best fit model for household frailty.

Table 8.9: The Akaike information criterion (AIC) and Baysian information criterion (BIC) for each model used

Model without community frailty			Model with community frailty	
Model	AIC	BIC	AIC	BIC
Exponential	2096.708	2178.713	2098.56	2186.114
Lognormal	1977.533	2065.078	1979.533	2073.331
Weibull	1955.798	2043.343	1952.6331	2046.43
Log-logistic	1957.283	2044.828	1959.22	2053.018
Cox PH	4739.029	4814.068	4728.879	4803.918

The Weibull model with community frailty seems to have a smaller Akaike information criterion (AIC) value (1952.6331), and indicates the best fit model for community frailty. This value is also the lowest AIC value when comparing models with household frailty and community frailty. This suggests that the Weibull model is the best fit model for DHS data with frailty model in Lesotho.

Chapter 9

Discussion And Conclusion

9.1 Discussion of Results

The primary goal of the study was to asses the determinants of under-five mortality by applying appropriate models to account for sibling-level correlation and thus provide valid estimates and correct statistical inference needed for policy decision making.

The results will be discussed according to factors such as demographic factors (wealth index, place of residence), socioeconomic factors (age of a mother, sex of a child) as well as biological factors (birth order, type of birth, breast feeding, place of delivery).

9.1.1 Biological Factors

The results from Cox model and Cox regression models with and without frailty for child mortality that analyses the age of a mother at birth, indicate that from all the biological factors analyzed in this study only a few factors are found to be significant. Those factors include the birth order, place of birth, birth size, breast feeding and breast feeding duration which appeared to be significant factors impacting deaths of children under this study.

Although other biological factors are not statistically significant they may still have an impact on childhood mortality. For example, children born by an older woman (age $30-49$) have a reduced mortality compared to children born by teenage mothers. These findings are, however, consistent with [36]who found that children of older women had lower risk for infant mortality when compared to children of younger mothers although their results were significant. Test for difference in mortality among the age categories between mothers aged 15-19 and other age group such as age $45-49$ shows that there is a significant different between neonatal deaths of children born to women aged $40-49$. Although all

mothers' age are significant it is observed that children born by older women aged 35 and above have a lower mortality relative to children born by a teenager.

9.1.2 Environmental Factors

Most of the environmental factors typically associated with childhood mortality do not have a significant impact on child mortality in Lesotho. However we did not include them in the results. Type of floor material was expected to be important for children older than one month, as research has indicated that when the floor is dirty children are not likely to be affected because they have started crawling or walking and are easily expose to the dirt [22]. Environmental contamination is one of the five groups of proximate determinants identified by [51], which assumes the direct influence of the risk of morbidity and mortality among children. According to [51], the transmission of infectious disease comes through different paths and that unsafe drinking water is one which could lead to diarrhoea and other intestinal disease.

9.2 Conclusion

In conclusion, the main aim of the study was to discover out the best statistical method that can be used when investigating factors associated with child mortality in Lesotho. Cox model with frailty is recommended for providing statistically valid estimates of the effects of proximate determinants after adjusting for the background variables and unobserved random effect.

One objective of this study was to ascertain the risk factors associated with child mortality for under five year of age and specifically the factors that affect neonatal and child-age mortality. The factors identified as risk factors for neonatal mortality are birth order, mother's age, place of birth, age at first birth, breastfeeding and breastfeeding duration.

The other objective of this study was to find different effects of risk factors on three ages of mortality. The age of a mother, and breastfeeding were found to be more pronounced for neonatal age, which is consistent with other findings from other countries that use the DHS data [18]. Some of the factors analyzed in this study that are identified in the literature as important to child mortality in other countries, are not important to child mortality in Lesotho. Those factors include place of residence, sex of a child and the source of drinking water. These variables should be investigated further by interacting these variables with each other. This approach would allow for testing, for example, whether the effect of source of drinking water is different in the rural areas than in the urban areas. The effect of HIV status of the mother should be investigated further as well.

Appendix A

Parametric Models

A.1 Results Using Household as Frailty

The following tables represent the results obtained using different models, such as Exponential model, Log-normal model and Log-logistic model. Each table consist of a model with and without frailty model.

Table A.1: Hazard ratios of child mortality associated with various socio-demographic variables

	Exponential model without household frailty						Exponential model with household frailty					
Variable	HR.	SE.	z	P>z	95% CI.		HR.	SE.	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-		-	-	-	-	-
20-24	0.5315	0.1222	-2.7500	0.0060	0.3386	0.8341	0.5262	0.1216	-2.7800	0.0050	0.3345	0.8277
25-29	0.5033	0.1235	-2.8000	0.0050	0.3111	0.8143	0.5054	0.1245	-2.7700	0.0060	0.3119	0.8190
30-34	0.3869	0.1072	-3.4300	0.0010	0.2248	0.6658	0.3889	0.1080	-3.4000	0.0010	0.2257	0.67028
35-39	0.4387	0.1314	-2.7500	0.0060	0.2439	0.7891	0.4432	0.1332	-2.7100	0.0070	0.2459	0.7989
40-44	0.4756	0.1587	-2.2300	0.0260	0.2473	0.9146	0.4810	0.1613	-2.1800	0.0290	0.2493	0.9279
45-49	0.4941	0.2009	-1.7300	0.0830	0.2227	1.0962	0.4966	0.2031	-1.7100	0.0870	0.2228	1.1070
Birth order: 1st birth	-	-	-	-	-	-		-	-	-	-	-
2-4th birth	1.2136	0.1764	1.3300	0.1830	0.9127	1.6137	1.2061	0.1758	1.2900	0.1980	0.9065	1.6049
>4th birth	1.5610	0.3789	1.8300	0.0670	0.9700	2.5120	1.5348	0.3747	1.7500	0.0790	0.9511	2.4767
Place of birth: Home	-	-	-	-	-	-		-	-	-	-	-
Public	0.4707	0.0593	-5.9800	<0.001	0.3677	0.6027	0.4698	0.0595	-5.9700	<0.001	0.3666	0.6021
Private	0.5815	0.0906	-3.4800	0.0010	0.4285	0.7893	0.5805	0.0908	-3.4800	0.0010	0.4272	0.7889
Birth size: Small	-	-	-	-	-	-		-	-	-	-	-
Average	0.6864	0.0781	-3.3100	0.0010	0.5491	0.8579	0.6863	0.0784	-3.2900	0.0010	0.5486	0.8586
Large	0.6709	0.2474	-1.0800	0.2790	0.3257	1.3820	0.6759	0.2500	-1.0600	0.2900	0.3273	1.3956
Wealth index: Poorest	-	-	-	-	-	-		-	-	-	-	-
Poorer	1.2262	0.1565	1.6000	0.1100	0.9548	1.5746	1.2235	0.1570	1.57	0.1160	0.9514	1.5733
Rich	1.3077	0.1969	1.7800	0.0750	0.9736	1.7565	1.3055	0.1975	1.76	0.078	0.9705	1.7560
Breast feeding: Never	-	-	-	-	-	-		-	-	-	-	-
Ever	0.1007	0.0120	-19.4000	<0.001	0.0799	0.1270		-	-	-	-	-
Breast feeding duration: $<$ 6 months	-	-	-	-	-	-	-	-	-	-	-	-
$\geq$ 6 months	0.1379	0.0151	-18.0700	<0.001	0.111229	0.1709		-	-	-	-	-
$\ln\theta$	-	-	-	-	-	-	-2.9313	1.1721	-2.5	0.0120	-5.2284	-0.6341
θ	-	-	-	-	-	-	0.0533	0.0625	-	-	0.0054	0.5304

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 0.84$, the value of $p = 0.0179$ at 5 percent level of significant

Table A.1 shows the results of the parametric regression model assuming a Exponential distribution for the survival times. The left-hand side results are for a Exponential without family frailty model and the right-hand side results are for Exponential with family frailty model.

Table A.2: Hazard ratios of child mortality associated with various socio-demographic variables

	Lognormal model without household						Lognormal model with household frailty					
Variable	Coef.	SE.	z	P>z	95% CI.		Coef.	SE.	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.7017	0.1665	4.2100	<0.001	0.3754	1.0280	0.7107	0.1674	4.2500	<0.001	0.3826	1.0388
25-29	0.7661	0.1786	4.2900	<0.001	0.4160	1.1161	0.7649	0.1793	4.2700	<0.001	0.4135	1.1162
30-34	1.0084	0.2009	5.0200	<0.001	0.6147	1.4020	1.0056	0.2014	4.9900	<0.001	0.6108	1.4004
35-39	0.8855	0.2190	4.0400	<0.001	0.4562	1.3147	0.8788	0.2198	4.000	<0.001	0.4480	1.3097
40-44	0.9523	0.2489	3.8300	<0.001	0.4645	1.4400	0.9484	0.2499	3.7900	<0.001	0.4586	1.4383
45-49	0.9529	0.3197	2.9800	0.0030	0.3263	1.5796	0.9520	0.3211	2.9600	0.0030	0.3227	1.5813
Birth order: 1st birth	-	-	-	-	-	-	-	-	-	-	-	-
2-4th birth	-0.1696	0.1053	-1.6100	0.1070	-0.3761	0.0368	-0.1653	0.1055	-1.5700	0.1170	-0.3721	0.0415
>4th birth	-0.3859	0.1792	-2.1500	0.0310	-0.7371	-0.0347	-0.3727	0.1800	-2.0700	0.0380	-0.7254	-0.0200
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.5559	0.1020	5.4500	<0.001	0.3560	0.7558	0.5567	0.1023	5.4400	<0.001	0.3561	0.7572
Private	0.3654	0.1217	3	0.0030	0.1268	0.6040	0.3657	0.1222	2.9900	0.0030	0.1263	0.6051
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.3479	0.0851	4.0900	<0.001	0.1810	0.5147	0.3485	0.0854	4.0800	<0.001	0.1811	0.515898
Large	0.3855	0.2625	1.4700	0.1420	-0.1291	0.9000	0.3798	0.2633	1.4400	0.1490	-0.1363	0.8960
Wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	-0.14007	0.094787	-1.48	0.139	-0.32585	0.045707	-0.13876	0.095213	-1.46	0.145	-0.32537	0.0479
Rich	-0.2575	0.1106	-2.3300	0.0200	-0.4742	-0.0408	-0.2571	0.1111	-2.3100	0.0210	-0.4747	-0.0394
Breast feeding duration: $<$ 6 months	-	-	-	-	-	-	-	-	-	-	-	-
$\geq$ 6 months	1.4990	0.0914	16.4000	<0.001	1.3199	1.6782						
Breast feeding: Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever	1.8641	0.1062	17.5500	<0.001	1.6558	2.0723						
$\ln \sigma$	0.2678	0.0385	6.9600	<0.001	0.1924	0.3432	0.2670	0.0386	6.9200	<0.001	0.1914	0.3426
σ	1.3071	0.0503	-	-	1.2122	1.4094	1.3060	0.0504	-	-	1.2109	1.408568
θ	-	-	-	-	-	-	0.0670	0.0646	-	-	0.0101	0.4435
$\ln \theta$	-	-	-	-	-	-	-2.7031	0.9643	-2.800	0.0050	-4.5931	-0.8131

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 1.29$, the value of $p = 0.128$ at 5 percent level of significant

Table A.2 shows the results of the parametric regression model assuming a log-normal distribution for the survival times namely age at death or survival of a child. The left-hand side results are for a log-normal model without household frailty and the right-hand side results are for a log-normal results with household frailty.

Table A.3: Hazard ratios of child mortality associated with various socio-demographic variables

	Log-logistic model without household frailty						Log-logistic model with household frailty model					
_Variable	Coef.	SE.	z	P>z	95% CI.		Coef.	SE.	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.5990	0.1415	4.2300	<0.001	0.3217	0.8763	0.6075	0.1425	4.2600	<0.001	0.3283	0.8867
25-29	0.6890	0.1510	4.5600	<0.001	0.3931	0.9850	0.6881	0.1517	4.5400	<0.001	0.3908	0.9853
30-34	0.8635	0.1697	5.0900	<0.001	0.5309	1.1960	0.8600	0.1703	5.0500	<0.001	0.5264	1.1937
35-39	0.7950	0.1844	4.3100	<0.001	0.4336	1.1564	0.7880	0.1852	4.2600	<0.001	0.4251	1.1509
40-44	0.8092	0.2041	3.9600	<0.001	0.4091	1.2092	0.8043	0.2052	3.9200	<0.001	0.4021	1.2065
45-49	0.8237	0.2543	3.2400	<0.001	0.3253	1.3221	0.8214	0.2559	3.2100	<0.001	0.3198	1.3231
Birth order:1st birth		-	-	-	-	-	-	-	-	-	-	-
2-4th birth	-0.1460	0.0881	-1.6600	0.0980	-0.3188	0.0267	-0.1413	0.0884	-1.6000	0.1100	-0.3145	0.0319
>4th birth	-0.3314	0.1480	-2.2400	0.0250	-0.6215	-0.0413	-0.3181	0.1488	-2.1400	0.0330	-0.6098	-0.0263
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.4588	0.0810	5.6600	<0.001	0.3000	0.6175	0.4602	0.0814	5.6500	<0.001	0.3007	0.6197
Large	0.3342	0.0978	3.4200	0.0010	0.1426	0.5258	0.3360	0.0982	3.4200	0.0010	0.1435	0.5285
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.2641	0.0697	3.7900	<0.001	0.1275	0.4007	0.2651	0.0700	3.7900	<0.001	0.1279	0.4022
Private	0.2376	0.2202	1.0800	0.2810	-0.1941	0.6693	0.2308	0.2210	1.0400	0.2960	-0.2024	0.6640
Wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	-0.1227	0.0781	-1.5700	0.1160	-0.2757	0.0303	-0.1207	0.0785	-1.5400	0.1240	-0.2745	0.0332
Rich	-0.1837	0.0917	-2.0000	0.0450	-0.3634	-0.0039	-0.1822	0.0922	-1.9800	0.0480	-0.3629	-0.0015
Breast feeding: Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever	1.5735	0.1019	15.4400	<0.001	1.3737	1.7732	-	-	-	-	-	-
Breast feeding duration:< 6 months	-	-	-	-	-	-	-	-	-	-	-	-
≥ 6 months	1.3032	0.0890	14.6500	<0.001	1.1288	1.4776	-	-	-	-	-	-
$\ln\gamma$	-0.5855	0.0430	-13.6200	<0.001	-0.6698	-0.5013	-0.5871	0.0430	-13.6500	<0.001	-0.6714	-0.5028
γ	0.5568	0.0239	-	-	0.5118	0.6057	0.5559	0.0239	-	-	0.5110	0.6048
θ	-	-	-	-	-	-	0.0719	0.0656	-	-	0.0121	0.4293
$\ln\theta$	-	-	-	-	-	-	-2.6320	0.9115	-2.8900	0.0040	-4.4185	-0.8455

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 1.46$, the value of $p = 0.114$ at 5 percent level of significant

Table A.3 shows the results of the parametric regression model assuming a Log-logistic distribution for the survival times. The left-hand side results are for a Log-logistic without family (household) frailty model and the right-hand side results are for Log-logistic with family frailty model.

A.2 Results Using Community as Frailty

Table A.4: Hazard ratios of child mortality associated with various socio-demographic variables

	Exponential model without community frailty						Exponential model with community frailty					
Variable	HR.	SE.	z	P>z	95% CI.		HR.	SE.	z	P>z	95% CI.	
Age of a mother: 15-19	-	-	-	-	-	-		-	-	-	-	-
20-24	0.5315	0.1222	-2.7500	0.0060	0.3386	0.8341	0.4725	0.1093	-3.2400	0.0010	0.3003	0.7434
25-29	0.5033	0.1235	-2.8000	0.0050	0.3111	0.8143	0.5830	0.1334	-2.3600	0.0180	0.3724	0.9129
30-34	0.3869	0.1072	-3.4300	0.0010	0.2248	0.6658	0.4703	0.1169	-3.0300	0.0020	0.2889	0.7656
35-39	0.4387	0.1314	-2.7500	0.0060	0.2439	0.7891	0.4990	0.1287	-2.6900	0.0070	0.3010	0.8273
40-44	0.4756	0.1587	-2.2300	0.0260	0.2473	0.9146	0.6051	0.1660	-1.8300	0.0670	0.3535	1.0359
45-49	0.4941	0.2009	-1.7300	0.0830	0.2227	1.0962	0.6597	0.2370	-1.16	0.2470	0.3263	1.3340
Birth order: 1st birth	-	-	-	-	-	-		-	-	-	-	-
2-4th birth	1.2136	0.1764	1.3300	0.1830	0.9127	1.6137		-	-	-	-	-
>4th birth	1.5610	0.3789	1.8300	0.0670	0.9700	2.5120		-	-	-	-	-
Place of birth: Home	-	-	-	-	-	-		-	-	-	-	-
Public	0.4707	0.0593	-5.9800	<0.001	0.3677	0.6027	0.6646	0.0718	-3.7800	<0.0001	0.5378	0.8214
Private	0.5815	0.0906	-3.4800	0.0010	0.4285	0.7893	0.5820	0.2130	-1.4800	0.1390	0.2841	1.1924
Birth size: Small	-	-	-	-	-	-		-	-	-	-	-
Average	0.6864	0.0781	-3.3100	0.0010	0.5491	0.8579	0.7199	0.0956	-2.4700	0.0130	0.5548	0.9339
Large	0.6709	0.2474	-1.0800	0.2790	0.3257	1.3820	0.9670	0.1574	-0.2100	0.8370	0.7028	1.3306
Wealth index: Poorest	-	-	-	-	-	-		-	-	-	-	-
Poorer	1.2262	0.1565	1.6000	0.1100	0.9548	1.5746		-	-	-	-	-
Rich	1.3077	0.1969	1.7800	0.0750	0.9736	1.7565		-	-	-	-	-
Breast feeding:Never	-	-	-	-	-	-		-	-	-	-	-
Ever	0.1007	0.0120	-19.4000	<0.001	0.0799	0.1270	0.0988	0.0128	-17.8300	<0.001	0.0767	0.1275
Breast feeding duration: $<$ 6 months	-	-	-	-	-	-		-	-	-	-	-
$\geq$ 6 months	0.1379	0.0151	-18.0700	<0.001	0.111229	0.1709	0.1356	0.0162	-16.7300	<0.001	0.1073	0.1714
$\ln\theta$	-	-	-	-	-	-	-3.7357	2.7718	-1.3500	0.1780	-9.1683	1.6970
θ	-	-	-	-	-	-	0.0239	0.0661	-	-	0.0001	5.4577

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 0.14$, the value of $p = 0.355$ at 5 percent level of significant

Table A.4 shows the results of the parametric regression model assuming a Exponential distribution for the survival times. The left-hand side results are for the Exponential without community frailty model and the right-hand side results are for the Exponential with community frailty model.

Table A.5: Hazard ratios of child mortality associated with various socio-demographic variables

	Log-normal model without community frailty						Log-normal model with community frailty					
Variable	Coef.	SE.	z	P>z	95% CI.		Coef.	SE.	z	P>z	95% CI.	
Breast feeding:Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever	1.8641	0.1062	17.5500	<0.001	1.6558	2.0723	1.8640	0.1062	17.5400	<0.001	1.6557	2.0722
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.5559	0.1020	5.4500	<0.001	0.3560	0.7558	0.3374	0.0747	4.5200	<0.001	0.1911	0.4838
Private	0.3654	0.1217	3	0.0030	0.1268	0.6040	0.5741	0.2500	2.3000	0.0220	0.0841	1.0641
Breast feeding duration: $<$ 6 months	-	-	-	-	-	-	-	-	-	-	-	-
$\geq$ 6 months	1.4990	0.0914	16.4000	<0.001	1.3199	1.6782	1.4990	0.0914	16.4000	<0.001	1.3198	1.6781
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.3479	0.0851	4.0900	<0.001	0.1810	0.5147	0.2046	0.0975	2.1000	0.0360	0.0134	0.3958
Large	0.3855	0.2625	1.4700	0.1420	-0.1291	0.9000	0.0121	0.1161	0.1000	0.9170	-0.2155	0.2398
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.7017	0.1665	4.2100	<0.001	0.3754	1.0280	0.6556	0.1569	4.1800	<0.001	0.3481	0.9630
25-29	0.7661	0.1786	4.2900	<0.001	0.4160	1.1161	0.5573	0.1562	3.5700	<0.001	0.2511	0.8636
30-34	1.0084	0.2009	5.0200	<0.001	0.6147	1.4020	0.7979	0.1691	4.7200	<0.001	0.4665	1.1293
35-39	0.8855	0.2190	4.0400	<0.001	0.4562	1.3147	0.6338	0.1757	3.6100	<0.001	0.2894	0.9782
40-44	0.9523	0.2489	3.8300	<0.001	0.4645	1.4400	0.7508	0.1953	3.8500	<0.001	0.3681	1.1335
45-49	0.9529	0.3197	2.9800	0.0030	0.3263	1.5796	0.6167	0.2630	2.3400	0.0190	0.1013	1.1322
Wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	-0.14007	0.094787	-1.48	0.139	-0.32585	0.045707						
Rich	-0.2575	0.1106	-2.3300	0.0200	-0.4742	-0.0408						
Birth order: 1st birth	-	-	-	-	-	-	-	-	-	-	-	-
2-4th birth	-0.1696	0.1053	-1.6100	0.1070	-0.3761	0.0368	-0.1653	0.1055	-1.5700	0.1170	-0.3721	0.0415
>4th birth	-0.3859	0.1792	-2.1500	0.0310	-0.7371	-0.0347	-0.3727	0.1800	-2.0700	0.0380	-0.7254	-0.0200
$\ln \sigma$	0.0803	0.0370	2.1700	0.0300	0.0077	0.1529	0.0803	0.0370	2.1700	0.0300	0.0077	0.1529
σ	1.0836	0.0401	-	-	1.0077	1.1652	1.0836	0.0401	-	-	1.0077	1.1652
θ	-	-	-	-	-	-	<0.001	0.0006	-	-	<0.001	.
$\ln \theta$	-	-	-	-	-	-	-13.6693	531.8832	-0.0300	0.9790	-1056.1410	1028.8030

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 0.00$, the value of $p = 1.00$ at 5 percent level of significant

Table A.5 shows the results of the parametric regression model assuming a Log-normal distribution for the survival times. The left-hand side results are for the Log-normal without community frailty model and the right-hand side results are for the Log-normal with community frailty model.

Table A.6: Hazard ratios of child mortality associated with various socio-demographic variables

	Log-logistic model without community frailty						Log-logistic model with community frailty					
Variable	Coef.	SE.	z	P>z	95% CI.		Coef.	SE.	z	P>z	95% CI.	
Breast feeding:Never	-	-	-	-	-	-	-	-	-	-	-	-
Ever	1.5735	0.1019	15.4400	<0.001	1.3737	1.7732	1.5755	0.1025	15.3800	<0.001	1.3747	1.7763
Place of birth: Home	-	-	-	-	-	-	-	-	-	-	-	-
Public	0.2641	0.0697	3.7900	<0.001	0.1275	0.4007	0.3169	0.0694	4.5700	<0.001	0.1810	0.4528
Private	0.2376	0.2202	1.0800	0.2810	-0.1941	0.6693	0.4086	0.2143	1.9100	0.0570	-0.0115	0.8287
Breast feeding duration:< 6 months	-	-	-	-	-	-	-	-	-	-	-	-
≥ 6 months	1.3032	0.0890	14.6500	<0.001	1.1288	1.4776	1.3049	0.0893	14.6100	<0.001	1.1298	1.4799
Birth size: Small	-	-	-	-	-	-	-	-	-	-	-	-
Average	0.4588	0.0810	5.6600	<0.001	0.3000	0.6175	0.2003	0.0878	2.2800	0.0230	0.0281	0.3724
Large	0.3342	0.0978	3.4200	0.0010	0.1426	0.5258	0.0201	0.1055	0.1900	0.8490	-0.1867	0.2270
Age of a mother: 15-19	-	-	-	-	-	-	-	-	-	-	-	-
20-24	0.5990	0.1415	4.2300	<0.001	0.3217	0.8763	0.6343	0.1486	4.2700	<0.001	0.3432	0.9255
25-29	0.6890	0.1510	4.5600	<0.001	0.3931	0.9850	0.5969	0.1476	4.0400	<0.001	0.3077	0.8862
30-34	0.8635	0.1697	5.0900	<0.001	0.5309	1.1960	0.7773	0.1584	4.9100	<0.001	0.4669	1.0878
35-39	0.7950	0.1844	4.3100	<0.001	0.4336	1.1564	0.6525	0.1656	3.9400	<0.001	0.3279	0.9771
40-44	0.8092	0.2041	3.9600	<0.001	0.4091	1.2092	0.7154	0.1785	4.0100	<0.001	0.3655	1.0653
45-49	0.8237	0.2543	3.2400	<0.001	0.3253	1.3221	0.6222	0.2419	2.5700	0.0100	0.1481	1.0964
Birth order: 1st birth		-	-	-	-	-	-	-	-	-	-	-
2-4th birth	-0.1460	0.0881	-1.6600	0.0980	-0.3188	0.0267	-	-	-	-	-	
>4th birth	-0.3314	0.1480	-2.2400	0.0250	-0.6215	-0.0413	-	-	-	-	-	
Wealth index: Poorest	-	-	-	-	-	-	-	-	-	-	-	-
Poorer	-0.1227	0.0781	-1.5700	0.1160	-0.2757	0.0303	-	-	-	-	-	-
Rich	-0.1837	0.0917	-2.0000	0.0450	-0.3634	-0.0039	-	-	-	-	-	-
$\ln\gamma$	-0.6610	0.0423	-15.6400	<0.001	-0.7438	-0.5782	-0.6626	0.0428	-15.5000	<0.001	-0.7464	-0.5788
γ	0.5163	0.0218	-	-	0.4753	0.5609	0.5155	0.0220	-	-	0.4741	0.5606
θ	-	-	-	-	-	-	0.0174	0.0712	-	-	<0.001	52.9959
$\ln\theta$	-	-	-	-	-	-	-4.0518	4.0929	-0.9900	0.3220	-12.0738	3.9702

Likelihood-ratio test of $\theta = 0 : \chi^2_1 = 0.06$, the value of $p = 0.401$ at 5 percent level of significant

Table A.6 shows the results of the parametric regression model assuming a Log-logistic distribution for the survival times. The left-hand side results are for a Log-logistic without community frailty model and the right-hand side results are for Log-logistic with community frailty model.

Appendix B

Stata Codes

B.1 Cox Model and Parametric Regression Models

B.1.1 Cox Model without Frailty

stcox i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.breastfeedduration i.wealthin

B.1.2 Cox Model with Frailty

stcox i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.breastfeedduration i.wealthin, shared(household)

stcox i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.breastfeedduration i.wealthin, shared(community)

B.1.3 Parametric Regression Models without Frailty

**********Model 1: exponential distribution gamma shared frailty model**********
streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.breastfeedduration i.wealthin, distribution(exponential)

**********Model 2: lognormal distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.breastfeedduration i.wealthin, distribution(lognormal)

**********Model 3: weibull distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(weibull)

**********Model 4: loglogistic distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(loglogistic)

B.1.4 Parametric Regression Models with Household Frailty

**********Model 1: exponential distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(exponential) frailty(gamma) shared(household)

**********Model 2: lognormal distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(lognormal) frailty(gamma) shared(household)

**********Model 3: Weibull distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(weibull) frailty(gamma) shared(household)

**********Model 4: loglogistic distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, distribution(loglogistic) frailty(gamma) shared(household)

B.1.5 Parametric Regression Models with Community Frailty

**********Model 1: exponential distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, dist(exponential) frailty(gamma) shared(community)

**********Model 2: Weibull distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, dist(weibull) frailty(gamma) shared(community)

**********Model 3: lognormal distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, dist(lognormal) frailty(gamma) shared(community)

**********Model 4: loglogistic distribution gamma shared frailty model**********

streg i.mothersage i.Birthorder i.Birthsize i.Placeofdelivery i.Breastfeeding i.
breastfeedduration i.wealthin, dist(loglogistic) frailty(gamma) shared(community)

B.1.6 KM Plots And Testing PH Assumptions

.sts graph, by(mothersage)

This code works for every variables, just need to interchange the variables.
For example instead of putting mothersage you can put breastfeedduration or
Placeofdelivery and so on.

**********Testing the PH assumption using schoenfeld residuals**********

After cox results, then run this code

.estat phtest, rank detail

Appendix C

R codes

C.1 Cox without Frailty

```
ab = coxph(Surv(Time,censored)  relevel(as.factor(mothersage), "1"), data =
Data)
summary(ab)
ac = coxph(Surv(Time,censored) relevel(as.factor(Birthorder),"2"), data = Data)
summary(ac)
ad = coxph(Surv(Time,censored) relevel(as.factor(Breastfeeding), "1"), data =
Data)
summary(ad)
ad = coxph(Surv(Time,censored) relevel(as.factor(Birthsize),"3"), data = Data)
summary(ad)
ae = coxph(Surv(Time,censored)  relevel(as.factor(Placeofdelivery),"1"), data
= Data)
summary(ae)
af = coxph(Surv(Time,censored) relevel(as.factor(wealthin),"1"), data = Data)
summary(af)
```

C.2 Cox with Frailty(Household)

```
ba = coxph(Surv(Time,censored) relevel(as.factor(mothersage), "1") + frailty(household),
data = Data)
summary(ba)
bc = coxph(Surv(Time,censored) relevel(as.factor(Birthorder),"2")+ frailty(household),
data = Data)
summary(bc)
bd = coxph(Surv(Time,censored) relevel(as.factor(Breastfeeding), "1")+ frailty(household),
data = Data)
summary(bd)
```

```
ad = coxph(Surv(Time,censored)  relevel(as.factor(Placeofdelivery),"3")+ frailty(household),
data = Data)
summary(ad)
be = coxph(Surv(Time,censored)  relevel(as.factor(Birthsize),"1")+ frailty(household),
data = Data)
summary(be)
bf = coxph(Surv(Time,censored)  relevel(as.factor(wealthin),"1")+ frailty(household),
data = Data)
summary(bf)
```

C.3 Cox with Frailty (Community)

```
ba = coxph(Surv(Time,censored)  relevel(as.factor(Breastfeeding), "1") + frailty(community),
data = Data)
summary(ba)
bc = coxph(Surv(Time,censored)  relevel(as.factor(Placeofdelivery),"2")+ frailty(community),
data = Data)
summary(bc)
bd = coxph(Surv(Time,censored)  relevel(as.factor(breastfeedduration ), "1")+
frailty(community), data = Data)
summary(bd)
ad = coxph(Surv(Time,censored)  relevel(as.factor(birthsize),"3")+ frailty(community),
data = Data)
summary(ad)
be = coxph(Surv(Time,censored)  relevel(as.factor(Mother'sage),"1")+ frailty(community),
data = Data)
summary(be)
```

Bibliography

[1] Alan Agresti. *An introduction to categorical data analysis*, volume 135. Wiley New York, 1996.

[2] Omar B Ahmad, Alan D Lopez, and Mie Inoue. The decline in child mortality: a reappraisal. *Bulletin of the World Health Organization*, 78:1175–1191, 2000.

[3] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

[4] Theodore Wilbur Anderson, Theodore Wilbur Anderson, Theodore Wilbur Anderson, and Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*, volume 2. Wiley New York, 1958.

[5] James A Baker. Radiofrequency medical instrument and methods for vessel welding, October 3 2000. US Patent 6,126,658.

[6] Norman Breslow. Covariance analysis of censored survival data. *Biometrics*, 30(1):89, 1974.

[7] UNICEF. International Child Development Centre. *Women in transition*. Number 6. UNICEF, International Child Development Centre, 1999.

[8] David Clayton, Michael Hills, and A Pickles. *Statistical models in epidemiology*, volume 161. IEA, 1993.

[9] David Collett. Modelling survival data. In *Modelling survival data in medical research*, pages 53–106. Springer, 1994.

[10] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

[11] David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

[12] David R Cox et al. Regression models and life tables. *JR stat soc B*, 34(2):187–220, 1972.

[13] David R Cox and EJ Snell. On test statistics calculated from residuals. *Biometrika*, 58(3):589–594, 1968.

[14] David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1982.

[15] David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

[16] David Roxbee Cox and E Joyce Snell. *Analysis of binary data*, volume 32. CRC Press, 1989.

[17] John Crowley and Marie Hu. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72(357):27–36, 1977.

[18] Siân L Curtis. Assessment of the quality of data used for direct estimation of infant and child mortality in dhs-ii surveys. 1995.

[19] AC Davison and A Gigli. Deviance residuals and normal scores plots. *Biometrika*, 76(2):211–221, 1989.

[20] Barthélémy Kuate Defo. Areal and socioeconomic differentials in infant and child mortality in cameroon. *Social science and medicine*, 42(3):399–420, 1996.

[21] Merimaaria Espo. *Infant mortality and its underlying determinants in rural Malawi*. Tampere University Press, 2002.

[22] Olufunke A Fayehun. Household environmental health hazards and child survival in sub-saharan africa. *N/a*, 2010.

[23] TR Fleming and DP Harrington. Counting processes and survival analysiswiley. *New York*, 1991.

[24] Michel Garenne and Enéas Gakusi. Health transitions in sub-saharan africa: overview of mortality trends in children under 5 years old (1950-2000). *Bulletin of the World Health Organization*, 84(6):470–478, 2006.

[25] Patricia M Grambsch and Terry M Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

[26] Eduard Hála, Ivan Wichterle, and Jiři Polák. *Vapour–Liquid Equilibrium Data at Normal Pressures*. Elsevier, 2002.

[27] David D Hanagal. *Modeling survival data using frailty models*. CRC Press, 2011.

[28] Nicholas HG Holford and Lewis B Sheiner. Kinetics of pharmacologic response. *Pharmacology and therapeutics*, 16(2):143–166, 1976.

[29] Nicholas HG Holford and Lewis B Sheiner. Understanding the dose-effect relationship. *Clinical pharmacokinetics*, 6(6):429–453, 1980.

[30] Philip Hougaard. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71(1):75–83, 1984.

[31] Philip Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

[32] Philip Hougaard. *Analysis of multivariate survival data*. Springer Science and Business Media, 1995.

[33] Lawrence Faustino Jada. *Determinants of levels and differentials of early childhood mortality in Kenya:" based on Kenya demograi'iiic and health survey (kdiis)-1989".* PhD thesis, N/a, 1992.

[34] Anrudh K Jain and Pravin Visaria. Infant mortality in india: differentials and determinants. *N/a*, 1988.

[35] C David Jenkins. *Building better health: a handbook of behavioral change*, volume 590. Pan American Health Org, 2003.

[36] Ahmad Kabir, Mohammad Shahidul Islam, Muhammad Shibbir Ahmed, and K Barbhuiya. Factors influencing infant and child mortality in bangladesh. *The Sciences*, 1(5):292–295, 2001.

[37] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley and Sons, 2011.

[38] Luna Kamau, Tovi Lehmann, William A Hawley, AS Orago, and Frank H Collins. Microgeographic genetic differentiation of anopheles gambiae mosquitoes from asembo bay, western kenya: a comparison with kilifi in coastal kenya. *The American journal of tropical medicine and hygiene*, 58(1):64–69, 1998.

[39] Luna Kamau, Tovi Lehmann, William A Hawley, AS Orago, and Frank H Collins. Microgeographic genetic differentiation of anopheles gambiae mosquitoes from asembo bay, western kenya: a comparison with kilifi in coastal kenya. *The American journal of tropical medicine and hygiene*, 58(1):64–69, 1998.

[40] Niels Keiding. Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 371–412, 1997.

[41] BR Kirkwood and JAC Sterne. Essential medical statistics. malden ma: Blackwell science, 2005.

[42] Joshua D Klein and Susan Lurie. Heat treatments for improved postharvest quality of horticultural crops. *HortTechnology*, 2(3):316–320, 1992.

[43] David G Kleinbaum and Mitchel Klein. *Survival analysis: a self-learning text*. Springer Science and Business Media, 2006.

[44] John Knodel. The design and analysis of focus group studies: A practical approach. *Successful focus groups: Advancing the state of the art*, 1:35–50, 2002.

[45] Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Science and Business Media, 2004.

[46] Stephen S Lim, Theo Vos, Abraham D Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A AlMazroa, Markus Amann, H Ross Anderson, Kathryn G Andrews, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The lancet*, 380(9859):2224–2260, 2013.

[47] Monica Akinyi Magadi, Nyovani Janet Madise, and Roberto Nascimento Rodrigues. Frequency and timing of antenatal care in kenya: explaining the variations between women of different communities. *Social science and medicine*, 51(4):551–561, 2001.

[48] Nathan Mantel. Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure. *Journal of the American Statistical Association*, 58(303):690–700, 1966.

[49] Ravai Marindo, Steve Pearson, and John B Casterline. Condom use and abstinence among unmarried young people in zimbabwe: which strategy whose agenda? *N/a*, 2007.

[50] Scott Menard. Applied logistic regression analysis: Sage university series on quantitative applications in the social sciences, 1995.

[51] Wiley Henry Mosley, Lincoln C Chen, et al. Child survival: strategies for research continued. *Population and development review*, 10(Supplement), 1984.

[52] Akim J Mturi and Sian L Curtis. The determinants of infant and child mortality in tanzania. *Health Policy and Planning*, 10(4):384–394, 1995.

[53] Gaudencia Mukolwe Okumbe. *Demographic and socio-economic correlates of neonatal mortality in Kenya*. PhD thesis, N/a, 1996.

[54] P Oliver. Poem sacred to the memory of j. *Wil. lard. Boston*, 4:B650, 1981.

[55] Daniel WR Omariba. *Socio-economic determinants of child survival in Upper Matasia sub-location, Kajiado district, Kenya*. PhD thesis, University of Nairobi, 1993.

[56] World Health Organization et al. Global status report on alcohol 2002. *N/a*, 2002.

[57] World Health Organization et al. Who traditional medicine strategy 2002-2005. *N/a*, 2002.

[58] FO Ouma. Environmental risk and socio-economic factors influencing infant and child mortality in siaya district: A case study of jera sub-location (ma thesis). *Population Studies and Research Institute, University of Nairobi*, 1991.

[59] Erik Parner et al. Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*, 26(1):183–214, 1998.

[60] Daniel D Reidpath and Pascale Allotey. Infant mortality rate as an indicator of population health. *Journal of epidemiology and community health*, 57(5):344–346, 2003.

[61] Geert Ridder and İnsan Tunalı. Stratified partial likelihood estimation. *Journal of Econometrics*, 92(2):193–232, 1999.

[62] Samuli Ripatti, Klaus Larsen, and Juni Palmgren. Maximum likelihood inference for multivariate frailty models using an automated monte carlo em algorithm. *Lifetime Data Analysis*, 8(4):349–360, 2002.

[63] Samuli Ripatti and Juni Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022, 2000.

[64] David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

[65] T Paul Schultz. Studying the impact of household economic and community variables on child mortality. *population and Development Review*, pages 215–235, 1984.

[66] Steve Selvin et al. *Statistical analysis of epidemiologic data.* Number Ed. 3. Oxford University Press, 2004.

[67] Priya Shyamsundar. *Poverty–environment Indicators*. Environment Department, World Bank, 2002.

[68] Joseph E Stiglitz and Ha-Joon Chang. *Joseph Stiglitz and the World Bank: the rebel within*. Anthem Press, 2001.

[69] Terry M Therneau, Patricia M Grambsch, and Thomas R Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, 1990.

[70] TM Therneau and PM Grambsch. Modeling survival data: extending the cox modelspringer-verlag. *New York*, 2000.

[71] Duncan Thomas, John Strauss, and Maria-Helena Henriques. Child survival, height for age and household characteristics in brazil. *Journal of Development Economics*, 33(2):197–234, 1990.

[72] Ian M Timæus and Louisiana Lush. Intra-urban differentials in child health. *Health transition review*, pages 163–190, 1995.

[73] James Trussell and German Rodriguez. Heterogeneity in demographic research. 1990.

[74] UNICEF. *The state of the world's children 2008: Child survival*, volume 8. Unicef, 2007.

[75] Florin Vaida and Ronghui Xu. Proportional hazards model with random effects. *Statistics in medicine*, 19(24):3309–3324, 2000.

[76] Gerard J Van den Berg. Duration models: specification, identification and multiple durations. *Handbook of econometrics*, 5:3381–3460, 2001.

[77] James W Vaupel, Kenneth G Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.

[78] Limin Wang. Determinants of child mortality in ldcs: empirical findings from demographic and health surveys. *Health policy*, 65(3):277–299, 2003.

[79] KP Wanjohi. The influence of environmental factors on infant and child mortality: A study of six districts in kenya (ma thesis). *Population Studies and Research Institute, University of Nairobi*, 1996.

[80] Andreas Wienke. *Frailty models in survival analysis*. CRC Press, 2010.

[81] Gebremariam Woldemicael. Demographic research volume 18, article 2, pages 27-58 published 07 march 1988. *N/a*, 1988.

[82] Kenneth I Wolpin. Determinants and consequences of the mortality and health of infants and children. *N/a*, 1997.

[83] UNDP WRI. Unep, and world bank, 2000. *World resources*, page 389, 1999.