Survival Analysis with Applications to Ga-Dikgale Children

Master of Science

MEW Makgaba

2014

Survival Analysis with Applications to Ga-Dikgale Children

by

Mokgoropo Enoch Walace Makgaba

Research dissertation submitted in fulfilment of the requirements for the degree of

Master of Science

in

Statistics

in the

Faculty of Science and Agriculture

(School of Mathematical and Computer Sciences)

at the

University of Limpopo

Supervisors: Prof. A. Tessera

Prof. M. Alberts

2014

DECLARATION

I declare that the dissertation hereby submitted to the University of Limpopo, for the degree of Master of Science in Statistics has not previously been submitted by me for a degree at this or any other university; that it is my work in design and in execution, and that all material contained herein has been duly acknowledged.

Makgaba, MEW **September 2014**

Surname, Initials **Date**

Dedication

I dedicate my dissertation work to my family.

Acknowledgements

I want to thank the following persons for their respective contributions to this dissertation:

- God for giving me strength and dedication to continue with this work.
- My wife, Makgano Mary, for her unconditional love, support and encouragement throughout this journey.
- My sons, Karabo Mogale and Ditheto Maribe, for their support and understanding. They are an inspiration.
- My parents, Mogale and Ramaite for encouraging me when I wanted to give up. I am grateful to have them in my life.
- My grandmother, Malebala “Marumo” Monama, for her support and encouragement though out my life.
- Not forgetting my little sister, brother and my nephew, Morongwa, Nkgesa and Marobo for being there for me.
- A special thank you to my supervisor, Prof. A. Tessera, for his guidance, support and encouragement. He was there to keep me on the race even during the times when I wanted to give up.
- My joint supervisor, Prof. M. Alberts, for her support and guidance.
- A special thank you to my friend and brother, Jack “Ace” Thosago for continuous support and encouragement.
- I also pass my gratitude to the staff working on the Dikgale HDSS, without you this would not have happened.

Abstract

The health and survival of children are important measures of the social wellbeing and health status of the community. The World Community made a commitment to reduce under-five mortality by two-thirds between 1990 and 2015.

The purpose of this study was to identify factors that have influence on child survival. The Dikgale Health and Demographic Surveillance System (HDSS) data for children born between 01 January 1996 and 31 December 2010 were analysed using cross-tabulation, logistic regression and survival analysis to determine factors that have influence on child survival.

The findings revealed that mother's survival status and child birth weight are significantly associated with child survival. The results showed that the odds that children born to mothers who are alive survive beyond five years are almost four times the odds that children born to mothers who are not alive survive beyond five years. The study also found that the odds that children born with birth weight 2.5kg or more survive beyond five years are almost two times that of children born with birth weight less than 2.5kg.

The results of this study may help in formulating strategies and interventions that improve the lifespan of children and assist in the reduction of child mortality.

KEY CONCEPTS

Child survival, Health Demographic Surveillance System, Cross-tabulation, Logistic regression, Survival analysis, Mother's survival status, Birth weight.

Contents

Dedication ... iii
Declaration ... ii
Acknowledgements ... iii
Abstract ... v
List of Figures ... ix
List of Tables ... x
1. CHAPTER 1: INTRODUCTION ... 1
1.1. Introduction ... 1
1.2. Research Problem ... 1
1.3. Aim of Study ... 2
1.4. Purpose of Study ... 2
1.5. Literature Review ... 3
1.6. Research Methodology ... 3
1.6.1. Data Analysis ... 3
1.7. Significance of the study ... 4
1.8. Conclusion ... 4
2. CHAPTER 2: LITERATURE REVIEW ... 5
2.1. Introduction ... 5
2.2. Child Mortality ... 5
2.3. Literature Reviewed ... 6
2.3.1. Mother's age at birth ... 6
2.3.2. Parental Education ... 7
2.3.3. Parental Death ... 8
2.3.4. Place of Delivery ... 9
2.3.5. Gender of the Child ... 10
2.3.6. Birth order and interval ... 11
2.3.7. Marital Status of the mother ... 12
2.3.8. Place of residence ... 12
2.3.9. Distance to the nearest health facility ... 12
2.3.10. Household Head ... 13

2.3.11. Parental Occupation 13
2.3.12. Other factors 14
2.4. Conclusion 14
3. CHAPTER 3: RESEARCH METHODOLOGY 16
3.1. Introduction 16
3.2. Study Site 16
3.3. Study Population 16
3.4. Data Collection 16
3.5. Logistic Regression 17
3.5.1. Fitting the multiple logistic regression model 17
3.5.2. Testing the significance of the coefficients 18
3.5.3. Confidence interval for logistic regression coefficients 21
3.5.4. Interpretation of the fitted logistic regression model 22
3.6. Survival analysis 23
3.6.1. Survival function 24
3.6.2. Hazard Function 25
3.6.3. Kaplan-Meier Method 26
3.6.4. Log Rank test 27
3.6.5. The Cox Proportional hazards Model 28
3.7. Variable Selection Method 32
4. CHAPTER 4: DATA ANALYSIS AND RESULTS 33
4.1. Introduction 33
4.2. Data and Data Analysis 33
4.2.1. Description of data 34
4.3. Under-Five Mortality 39
4.4. Application of survival analysis 48
4.4.1. Kaplan-Meier survival curves 48
4.4.2. Cox Proportional hazards model 57
4.5. Conclusion 59
5. Chapter 5: Discussion and Conclusion 60
5.1. Discussion of the results 60
5.1.1. Introduction 60

5.1.2. Factors that have effect on child survival and survival time 60
5.1.3. Factors that appear to have no effect on child mortality and survival time . 62
5.2. Conclusion 64
5.2.1. Summary of the results 64
5.2.2. Recommendations 65
5.2.3. Conclusion 65
References 66
Appendix 72

List of Figures

Figure 1: Survival probabilities for Dikgale HDSS children by gender 49
Figure 2: Survival probabilities for Dikgale HDSS children by mother's survival status. 51
Figure 3: Survival probabilities for Dikgale HDSS children by child birth weight 52
Figure 4: Survival probabilities for Dikgale HDSS children by Delivery attendant 54
Figure 5: Survival probabilities for Dikgale HDSS children by mother's educational level 55
Figure 6: Log-minus-log plot for proportional hazards checking.................................... 59

Figure A1: Hazard function against gender for children of Dikgale HDSS 72
Figure A2: Hazard function against the mother's survival status for children of Dikgale HDSS 73
Figure A3: Hazard function against by child birth weight 74
Figure A4: Survival probabilities for Dikgale HDSS children by breastfeeding status 75
Figure A5: Hazard function against by breastfeeding status 76
Figure A6: Survival probabilities for Dikgale HDSS children by the mother's attendance of ANC clinic 77
Figure A7: Hazard function for children by the mother's attendance of ANC clinic 78
Figure A8: Survival probabilities for Dikgale HDSS children by the place of delivery 79
Figure A9: Hazard function of children by the place of delivery 80
Figure A10: Hazard function of children by the delivery attendant 81
Figure A11: Hazard function for children by the mother's educational level 82
Figure A12: Survival probabilities for Dikgale HDSS children by the mother's marital status 84
Figure A13: Hazard function for children by the mother's marital status 84
Figure A14: Survival curve for children by the mother's age category 86
Figure A15: Hazard function for children by the mother's age category 86

List of Tables

Table 1: Classification of the Dikgale HDSS children by place of birth.......................... 34
Table 2: Classification of child survival status by gender, birth weight category and status of breast feeding 35
Table 3: Classification of child survival status by mother survival status, mother's age category, mother's marital status and the mother's educational level 36
Table 4: Classification of child survival status by antenatal clinic attendance, delivery place and delivery attendant 38
Table 5: Classification of children by status of survival up to the age of five years and gender, birth weight and breastfeeding status with p-values.. 40
Table 6: Classification of children by survival status up to the age of five years and mother's survival status, mother's age category, mother's marital status and mother's educational level with p-value for each factor... 42
Table 7: Classification of children by status of survival up to the age of five years and antenatal clinic attendance, delivery place and delivery attendant with p-value for each factor 44
Table 8: Odds ratio for child survival up to the age of five and child's birth weight 45
Table 9: Odds ratio for child survival up to the age of five and mother status (dead/alive) 45
Table 10: Odds ratio for child survival up to the age of five by delivery attendant 46
Table 11: Model Summary and Omnibus tests of mode coefficients 47
Table 12: Logistic regression analysis for the association between child survival and different factors 48
Table 13: Means for Survival Time by Gender and test to compare survival time and gender 50
Table 14: Means for Survival Time by mother's survival status and test to compare survival time and mother's survival status 51
Table 15: Means for Survival Time by birth weight and test to compare survival time and birth weight.. 53
Table 16: Means for Survival Time for children by delivery attendant and test to compare survival time by delivery attendant 54

Table 17: Means for Survival Time for children by the mother's educational level and the test to compare survival time for children by mother's educational level 56
Table 18: Test of Model Coefficient 57
Table 19: The Cox Proportional hazards model for child survival and different factors . 58

Table A: Overall summary of children in Dikgale HDSS by the mother's survival status 72
Table B: Overall summary of children in Dikgale HDSS by child birth weight 73
Table C: Overall summary of children in Dikgale HDSS by breast feeding 74
Table D: Means for Survival Time by child breast feeding 75
Table E: Test to compare survival time and child breast feeding 75
Table F: Overall summary of children by the mother's attendance of ANC clinic 76
Table G: Means for Survival Time of children by the mother's attendance of ANC clinic 77
Table H: Test to compare survival time of children by the mother's attendance of ANC clinic 77
Table I: Overall summary of children by the place of delivery 78
Table J: Means for survival time of children by the place of delivery 79
Table K: Test to compare survival time of children by the place of delivery 79
Table L: Overall summary of children by the delivery attendant 80
Table M: Overall summary of children in Dikgale HDSS by the mother's educational level 81
Table N: Means for Survival Time for children by the mother's educational level 82
Table O: Overall summary of children in Dikgale HDSS by mother's marital status 83
Table P: Test to compare survival time for children by the mother's marital status 83
Table Q: Overall summary of children HDSS by the mother's age category 85
Table R: Means for survival time for children by the mother's age category 85
Table S: Test to compare survival time for children by the mother's age category 85
Table T: Odds ratio for child gender 87

1. CHAPTER 1: INTRODUCTION

1.1. Introduction

The health and survival of children are important measures of the social well-being and health status of the community. It is because of the importance of child survival that the World Community promised to reduce under-five mortality by two-thirds between 1990 and 2015. This is commonly referred to as Millennium Development Goal 4. There has been a remarkable improvement in child survival and a significant reduction in under-five mortality between 1990 and 2012. In 1990, the under-five mortality was 12.6 million and the figure dropped by 48% to 6.6 million in 2012 (UNICEF, 2013). However, despite this remarkable improvement in child survival, if the current trends continue, the Millennium Development Goal can only be met in 2028 (UNICEF, 2013). One reason for this is the insufficient reduction in under-five mortality in developing countries, particularly in Sub-Saharan Africa.

To meet the Millennium Development Goal 4 before 2028 requires the acceleration of activities that help to improve child survival. One such activity is research on child survival and it is hoped that this research project will make some contribution towards this issue.

South Africa's under-five mortality rate dropped from 62 per 1,000 live births in 1990 to 47 per 1,000 live births in 2011 (UNICEF, 2012). According to Hall et al. (2012) the contributing factors to under-five mortality in South Africa include HIV transmitted from the mother to the child, newborn and childhood conditions such as diarrhoea and pneumonia that are commonly associated with poverty.

1.2. Research Problem

In order to be able to reduce child mortality, factors that contribute to mortality of children need to be identified so that the country can develop interventions that will assist in improving child survival thus reducing child mortality. This study aims to use the data from Dikgale Health and Demographic Surveillance System (HDSS) to determine the factors that have influence on child mortality in Dikgale HDSS. It is hoped that this study will assist the local government to plan ways to improve life expectancy of children in Dikgale HDSS in particular and the province in general.

The study uses logistic regression and survival analysis techniques to model the duration of child survival and identify factors that influence child survival. Logistic regression and cross tabulations were used to determine the factors that contribute towards child mortality. Data used in this study were on children born between 01 January 1996 and 31 December 2010. The study established factors that have influence on child mortality and child survival. The children born between 01 January 1996 and 31 December 2010 were extracted from the Dikgale HDSS database that contains information of all individuals who are residents in Dikgale HDSS.

1.3. Aim of Study

The aim of the study is to determine factors that have effect on child survival for children residing in Dikgale HDSS and who are born between 01 January 1996 and December 2010. The study will determine the association between child survival in Dikgale HDSS and various factors. The results of the study will be useful for decision makers to develop strategies that will accelerate the activities that will assist in reducing child mortality rate.

1.4. Purpose of Study

The purpose of the study is to identify factors and assess their effect on child survival for children residing in Dikgale HDSS. The study will model child survival probabilities and compare the survival and hazard probabilities in relation to various factors for children residing in Dikgale HDSS.

The study focuses on the following specific objectives:

- identify factors that have effect on child survival time.
- assess the effect of those factors on child survival for Dikgale HDSS children.

- estimate the effect of different factors on child survival.
- measure the effect of factors on child survival period.

1.5. Literature Review

Child survival is the responsibility of all the segments of the society (UNICEF, 2013). Improving the survival of children is one of the important goals set by the United Nations member states. Many researchers have identified the factors that affect child survival and child mortality over the years.

Mother's education has been identified to be a strong predictor of child mortality especially in developing countries. Hobcraft (1993), Van Den Broeck et al. (1996), Bennet (1999), Kabir et al. (2001), Heaton and Amoateng (2007), Rahman (2009), and Kumar and File (2010) are some of the researchers who determined using cross tabulation, logistic regression and Cox Proportional hazards model that mother's education is a strong predictor of child survival.

The death of the mother was found to have an impact on child survival. It is one of the factors that appeared to have a negative impact on child survival. Becher et al. (2004), Rahman (2009) and Ronsmans et al. (2010) are some of the researchers that found that mother's survival is a strong predictor of child survival.

Maternal age at birth was found to be another predictor that most of the researchers identified as predictor of child survival. Mahmood and Kiani (1994), Kabir et al. (2001), Mustafa and Odimegwu (2008) are some of the researchers who found that mother's age at birth is a predictor of child survival.

1.6. Research Methodology

1.6.1. Data Analysis

Cross tabulation, Logistic regression and survival analysis of data that included Kaplan Meier survival curves, log rank tests and Cox proportional hazards model were done using Statistical Programme for Social Sciences (SPSS) version 21. Test of independence between child survival and various predictor variables and calculation of odds ratios were done using cross tabulation. The predictors that were found to be significant were included in the multivariate logistic regression model which estimated the odds of different factors in the model. Logistic regression model was used to identify the factors which have effect on child survival in the presence of other factors.

Kaplan Meier survival curves were produced for children according to the groups of different predictors. The survival probabilities for the children were estimated and compared across different factors using log rank test. The factors found to be significant were included into the Cox proportional hazards model which was used to calculate the hazard ratios for children by the predictor variables in the model.

1.7. Significance of the study

The study seeks to identify factors that are associated with child survival and those that have effect on children survival time for the children in Dikgale HDSS. It will be useful for the local government to utilize the results of this study to develop strategies and interventions that will assist in reducing child mortality.

1.8. Conclusion

In conclusion, children are the most vulnerable group that needs constant attention to protect them from different factors that can impact negatively on their wellbeing. This study determines factors that are associated with child mortality and survival using data for children born between 01 January 1996 and 31 December 2010 extracted from Dikgale HDSS database.

2. CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

This chapter provides a review of studies that were conducted to determine the factors that have effect on child survival. The chapter also presents the statistics on child mortality in South Africa and globally.

2.2. Child Mortality

Investing in the health and wellbeing of the children of South Africa is an investment for the future development of the country (Bradshaw et al., 2003). The reduction of child mortality is a global public health priority (Atrash, 2011). According to Kyei (2011) the level of mortality, especially child mortality, reflects the state of public health and hygiene, the environmental sanitation, cultural mores about feeding and clothing, socio-economic development, and the people's attitude towards the dignity and value of life itself.

The Millennium Development Goal (MDG) 4 aims to reduce under five mortality by two-thirds between 1990 and 2015 (United Nations, 2011). According to Folasade (2000) substantial progress has been made towards the reduction of infant and child mortality, although improvement in child survival has been very poor in sub-Saharan Africa (SSA) (Fotso et al., 2007). The annual number of under five deaths has fallen from 12.6 million in 1990 to an estimated 6.6 million in 2012 (UNICEF 2013).

In South Africa, Statistics South Africa's report shows that registered deaths of children in South Africa under the age of 18 years have increased from 41,188 in 1997 to a peak of 78,566 in 2006 (Statistics South Africa, 2006). This increase in child mortality makes it difficult for the country to meet the MDG goal 4 by 2015.

Child mortality trends in South Africa have shown no sign of improvement over the past 15 years, which is a cause of great concern (Kibel et al., 2010). The Actuarial Society of South Africa (ASSA)'s 2003 model shows a steady increase in child mortality under 5 years rate from 56 deaths per 1000 live births in 1990 to 73 deaths per 1000 live births in 2000, then a decline to 67 deaths per 1000 live births in 2008 (Actuarial Society of South Africa, 2005).

2.3. Literature Reviewed

The studies reviewed have shown various factors that contribute towards child mortality in different parts of the world. Some of the factors that were found to be associated with child survival include water, sanitation and hygiene, child immunization, mother's age at birth, marital status of the mother, gender of the child, birth interval, parental education, place of residence, mother's working status, father's working status, parental death, household head, and health seeking behaviours. Some of the factors are discussed below.

2.3.1. Mother's age at birth

Syamala (2004) demonstrated the relationship between socio demographic factors and child survival. The data used for the study was from the survey "The levels of fertility and mortality in Goa" conducted by the International Institute for Population Sciences. Logistic regression was used to better understand the contribution of the variables on child survival status.

The results of the study showed that when mothers age at delivery increased from less than 19 to 20 -29, the survival chances increased by 9.21%. When the age of a mother increased to above 30 years the survival chances increased further by 5.26%.

A study was conducted to compare teenage mothers (below 20 years of age) and adults (20 – 34 years of age) to child survival from the Ethiopian Demographic Health Survey (DHS) by Taffa and Obare (2004). Cox Proportional hazards model was used to assess the socio-economic determinants of child survival. The study found that mother's age was strongly associated with child health. According to the researchers the results of this study agreed with the studies conducted in Kenya and Uganda using DHS data.

In their study Mahmood and Kiani (1994) used logistic regression and found that older maternal age unexpectedly enhances the survival of children, probably because of a cumulative increase in maternal experience with increasing age.

2.3.2. Parental Education

Education is the most influential factor in differentiating infant and child mortality levels within all the socioeconomic factors (Mondal et al. 2009). According to Mahmood and Kiani (1994) maternal educational level stands out as a significant factor affecting child survival. Hobcraft (1993) reviewed the evidence concerning the relative importance of maternal education on child survival and found that mother's education plays an important role in determining child survival even after controlling for other factors such as husband's level of education and occupational status.

Van den Broeck et al. (1996) studied maternal factors related to child survival in the rural area of Bwamanda, Northern Zaire (now known as Democratic Republic of Congo) and also found that maternal education is a contributing factor to child survival. The results from the studies by Kembo and Van Ginneken (2009) and Rahman (2009) agree with other researchers that maternal education has an impact on child survival.

According to Caldwell (1990), an educated woman may feel more deprived in a country where most other women are educated than in one where they are not, but her children stand a much greater chance of survival than those of uneducated women. It can be seen in the societies that educated mothers are providing better care to their children as compared to uneducated mothers. Education can delay women from having a child and marriage, which will have a positive effect on child survival (UNICEF, 2013).

Syamala (2004) also found that the association between mother's education and survival rate of children during infancy is strong and direct. Chowdhury et al. (2010) conducted a study to determine the socio-economic determinants of neonatal, post neonatal, infant and child mortality in Natore district of Bangladesh. Data were collected from all eligible reproductive aged women and logistic regression was used to analyse the data. The study found that the significant predictors of child mortality are mother's and father's education.

The effect of the father is also important in the child's wellbeing. Mondal et al. (2009) conducted a study in Bangladesh. Logistic regression model was used to analyse the data from the study. The study found that father's education has a significant influence on infant and child mortality.

Kumar and File (2010) used the Ethiopian Demographic and Health Survey 2005 data to investigate the predictors of child mortality in Ethiopia. The chi-square test was used to identify the relationship between child mortality and other explanatory variables. The study found that the highest number of deaths was observed for the illiterate mothers and lowest observed for the educated mothers. It was proven that mother's education played a significant role on child mortality.

According to UNICEF (2013) mother's educational level has been showing an increase in all regions in the world. It was found that mother's education has a positive effect on child survival. This includes being able to provide better health care, access to better housing and knowledge of the importance of child immunization.

2.3.3. Parental Death

Maternal mortality is a tragic loss and also has serious effects on the mother's surviving children (UNICEF, 2013). According to Razzaque et al. (2012) the effect of maternal death on the health and well-being of the child includes a child not getting day to day care, isolation of the child and no motivation for better life by the mother. Razzaque et al. (2012) studied the effect of maternal death on mortality of under five children in Bangladesh. The study examined survival probabilities of children born to mothers who died during pregnancy, children born from mothers who died from non-maternal causes and those children whose mothers are still alive. The study used life tables to examine the effect of different factors on the survival of under-five children. The results from the study showed that maternal death has effect on child survival. The study showed that the survival of children who were adopted immediately after their mother's death have a greater chance of survival compared to those staying in their own households without anyone to take care of them.

UNICEF (2013) reports that an infant who loses a mother in the first six weeks of life is at greater risk of death than an infant whose mother is alive. Anderson et al. (2007) conducted a study to determine the odds of child death when a mother dies from maternal or non-maternal causes in rural Haiti. The study used data for deaths among reproductive aged women between 1997 and 1999 in and around Jeremie, Haiti. The deaths were classified according to maternal or non-maternal causes of death and other variables including age of mother, village and parity. They found that maternal death has a significant effect on child survival.

The death of a parent, particularly of a mother, is likely to have a major effect on the health and survival of young children, especially in poor settings (Ronsmans et al., 2010). According to Becher et al. (2004) reduced child care, non-breast feeding of a child and improper bottle feeding are contributing reasons to child survival for children born to mothers who have died.

2.3.4. Place of Delivery

Place of delivery is an important predictor of child survival (Mondal et al., 2009). According to Ajaari et al. (2012) the place of delivery determines the quality of care that the mother and the child receive after the delivery. Health facilities provide a safer sanitary environment and medically correct birth assistance than what is available at home (Ajaari et al., 2012). The results from Hossain and Mondal (2009) agree with the results by Ajaari et al. (2012) that a mother's place of delivery is an important determinant of child survival. Mothers should be encouraged to deliver in the health facility under medical supervision of skilled medical professionals to decrease health risks of mothers and babies (Hossain and Mondal, 2009)

Ajaari et al. (2012) conducted a study to determine the impact of place of delivery on neonatal death in Rural Tanzania. The data from Rufiji Health and Demographic Surveillance System (RHDSS) in Tanzania were used to determine the impact of the place of delivery on neonatal death. Multivariate Logistic regression was used to analyse the data. The study found that the mother's place of delivery has a significant impact on neonatal mortality. According to Hossain and Mondal (2009) proper medical attention and hygienic conditions during their delivery time will reduce the risk of infections on the mother and the child, and also facilitate management of complications that might occur during delivery.

According to Rahman (2009) mother's place of delivery has a positive effect on the child survival and child death rate is 0.8 times lower among mothers whose place of delivery was at the hospital than mothers whose place of delivery was at their home. Health facility provides safer delivery facilities with skilled staff who will assist with the delivery of a child (Mondal et al., 2009) than at home where the child might be delivered by people who are not skilled in midwifery.

2.3.5. Gender of the Child

Mustafa and Odimegwu (2008) examined the importance of bio-social, demographic and economic factors associated with infant mortality using the 2003 Kenya Demographic Health Survey data for children. Logistic regression models were fitted to the factors affecting infant mortality and it was found that gender of the child was significantly associated with infant mortality. Mondal et al. (2009) also found that gender has a significant effect on child survival. They found that neonatal death and post natal death are lower for infant females compared to infant males but at the age of 1 year to below 5 years, female child death becomes higher compared to male child death.

According to Uddin et al., (2009) male children have more survival advantage than female children because male children are given more attention in terms of parental care, feeding patterns, intra family food distribution and treatment of illness.

Hirve and Ganatra (1997) conducted a prospective cohort study on the survival experience of under five children in rural Western India. The study used information from 45 villages in Shirur Development Block in Pune District in Maharashtra in India where 4129 children were followed up from birth until they reach the age of five. Information was obtained from their households on a three monthly basis. Survival analysis techniques were used to determine the role of different variables on child survival. The study found that girls had a better early survival period than boys but during late neonatal period, boys began to have better survival than girls.

Becher et al. (2004) found that their results contradict those by Mondal et al. (2009), Hirve and Ganatra (1997) and Mustafa and Odimegwu (2008). They found that the gender of a child does not have an impact on infant and child under five years survival. Chowdhury et al. (2013) also found that gender of a child does not have a significant effect on child survival.

2.3.6. Birth order and interval

Mondal et al. (2009) conducted a study on the effects of socio-economic and demographic variables on infant and child mortality. The study was conducted in Rajshahi District in Bangladesh. Logistic regression was used to analyse the data and the study found that birth interval was one of the most significant predictors of child mortality. The risk of dying for children born after the birth interval of 36 months or longer are 57.70% and 34.80% respectively lower compared to children born below the birth interval of 18 months.

In their study, Mustafa and Odimegwu (2008) found that birth order and birth interval were significantly associated with infant mortality for children from rural areas only. Syamala (2004) also found that birth order and birth interval affect significantly the probability of survival of the child. Kumar and File (2010) agree with these researchers that birth order has effect on child mortality.

According to Uddin et al. (2009) a shorter birth interval increases the risk of child death because this can result in low birth weight and may impair milk production for children born on close intervals. UNICEF (2013) agrees with Uddin et al. (2009) that short birth interval will increase the risk of prematurity and low birth weight thus increasing infant and child mortality.

2.3.7. Marital Status of the mother

Worku (2009) used the South Africa Demographic Health Survey (SADHS) of 2003 data to determine socio-economic and health related factors responsible for the child mortality in South Africa. The study used Cox Proportional hazards model to measure the effect of different factors on child survival in South Africa. It was found that mother's marital status is associated with child survival. The study showed that children of married mothers are 1.74 times more likely to survive compared to children of mothers who are not married. According to Worku (2009) married parents work together to ensure the wellbeing of the child while an unmarried mother has no one to assist her in ensuring the wellbeing of the child.

Mturi and Curtis (1995) investigated the determinants of infant and child mortality in Tanzania using the 1991/92 Tanzania Demographic and Health Survey (TDHS). The study used Cox Proportional hazards model to determine factors that have an influence on infant and child mortality in Tanzania. The results from the study show that marital status of the mother has effect on child survival.

2.3.8. Place of residence

In the study conducted by Taffa and Obare (2004) it was found that the place of residence in Ethiopia emerged as a strong determinant of under-five mortality. Kumar and File (2010) also found that place of residence has effect on child mortality rate. Their study was conducted using data from the Ethiopian Demographic and Health Survey (2005 EDHS) to determine the predictors of under-five mortality in Ethiopia. Cross tabulation technique was used to analyze the EDHS data.

2.3.9. Distance to the nearest health facility

Moisi et al. (2010) used the Epidemiological and Demographic Surveillance Systems data from Kilifi District in Kenya to evaluate the effect of travel time to the nearest health facility on childhood mortality. Proportional hazards models were developed and it was found that distance to the health facilities was not associated with childhood mortality.

2.3.10. Household Head

A family is the primary social unit for children in communities (Doctor, 2011). Doctor (2011) investigated whether household headship had an impact on child death. The study used the 2008 Nigeria Demographic Health Survey to illustrate whether household headship had an impact on child survival. Multivariate Logistic regression was used to assess the influence of household headship on child mortality. It was found that household headship is associated with child mortality.

According to Doctor (2011) a child born from a female headed household is 17% more likely to survive compared to a child born from a male headed household. The reason was that female headed households may understand maternal and child health better compared to male headed households. Furthermore women encourage each other to seek medical attention or advice if the child is sick (Doctor, 2011).

The effect of household head on child death in Nepal was studied by Adhikari and Podhisita (2010) using the Nepal Demographic and Health Survey (NDHS) data of 2006. Logistic regression was used to examine the effect of different variables including household head on child death. They found that female headed households were less likely to experience child death as compared to male headed households.

In a female headed household, the female head has the autonomy in decision making and is more likely to use contraceptives, which in turn reduces the risk of getting pregnant and thus minimizing the risk of child death (Adhikari and Podhisita 2010).

2.3.11. Parental Occupation

In their study Mondal et al. (2009) found that the mother’s occupation has a significant influence on child survival and is associated with nutritional status of the child. According to Mondal et al. (2009) a father’s occupational status determines the economic status, nutrition, housing conditions, access to health care facilities and clothing in the family.

Chowdhury et al. (2010) used data extracted from the Bangladesh Demographic and Health survey of 2007. Cox Proportional hazards model was used to identify factors that have an influence on child survival. The results showed that father's occupation has an influence on child survival. According to Chowdhury et al. (2010) this may be due to the fact that a father may be highly educated and providing better advantage in terms of food, nutrition and health services to a child better than the father who is not working.

2.3.12. Other factors

Hirve and Ganatra (1997) studied the role of birth weight, nutrition, immunization and other medical and social factors on child survival. The study used Kaplan Meier survival curves to analyse the prospective cohort data for children under five years in rural Western India. The study found that birth weight influences child survival and mortality.

Syamala (2004) found that child's survival probability was associated with the religion of the mother. The results showed that the children of Roman Catholics had higher degree of survival chances than those of Hindus.

Rahman (2009) used Proportional hazards model to show that mother's education, source of antenatal care, assistance during delivery, place of delivery and father's education have a significant effect on the child survival.

Chowdhury et al. (2013) conducted a study using data extracted from the Bangladesh Demographic and Health Survey (BDHS-2007). Cox proportional hazards model was used to analyse the data and the study found that antenatal care visits have a strong influence on the increase of child survival.

2.4. Conclusion

This chapter reviewed research studies conducted by various researchers on factors that influence child mortality and survival. Studies reviewed found that mother's age at birth, parental education, death of the mother, mother's place of delivery, child gender, mother's marital status, place of residence, distance to the nearest health facility, antenatal care, birth weight, nutrition and immunization are some of the factors that influence on child mortality and survival.

The present study is aimed at determining factors that influence child mortality and survival of Dikgale children born between 01 January 1996 and 31 December 2010. Chapter 3 describes the methodology that will be used to determine factors that influence child mortality in Dikgale HDSS.

3. CHAPTER 3: RESEARCH METHODOLOGY

3.1. Introduction

This chapter provides a review of the statistical methods used to analyse the Dikgale HDSS children data to determine factors that have effect on child survival and to estimate the effect of different factors on child survival. The chapter provides information on the study site, study population, data collection and data analysis that were used in this study. Logistic regression and survival analysis are the two statistical procedures that were used to analyse Dikgale HDSS data.

3.2. Study Site

Dikgale Health and Demographic Surveillance System (HDSS) centre is a member of International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) in developing countries. Dikgale is located in Capricorn District, Limpopo Province of South Africa and is situated approximately 40 km north-east of the city of Polokwane and 15 km from the University of Limpopo (Turfloop Campus). The centre was established in 1995 when demographic information of the population was collected. Yearly update documenting births, deaths and migrations are conducted.

3.3. Study Population

The study focused on all Dikgale children born between 01 January 1996 and 31 December 2010. This includes all children who were born in Dikgale HDSS and those who migrated into Dikgale HDSS.

3.4. Data Collection

Data were extracted from the Dikgale HDSS data base. The Dikgale HDSS collects information on births, deaths, migration, household relationships, health and socio-economic variables on a continuous basis. Households are visited once every year to collect the information to update the information collected the previous year.

3.5. Logistic Regression

Logistic regression is a statistical procedure that is used to describe the relationship between a dichotomous or polychotomous response variable and a set of explanatory variables. Here we are interested in the case where the response variable Y is dichotomous or binary. The possible values of Y are 0 (failure) and 1 (success) and let $E(Y)=\pi(\underline{x})=P(Y=1/\underline{X}=\underline{x})$.

The logit may be written as

$$g(\underline{x})=log\frac{\pi(\underline{x})}{1-\pi(\underline{x})}=\beta_1x_1+\beta_2x_2+...+\beta_px_p=\sum_1^p\beta_jx_j \qquad \text{where } x_1=1.$$

This implies that $\pi(\underline{x})=\dfrac{exp\left(\sum_1^p\beta_jx_j\right)}{1+exp\left(\sum_1^p\beta_jx_j\right)}$.

3.5.1. Fitting the multiple logistic regression model

The number of observations (n_i) and the number of successes (y_i) at $\underline{x}_i=(x_{i1},x_{i2},x_{i3},...,x_{ip})$ are sufficient for estimating $\underline{\beta}$. Therefore, we consider k pairs of observations $(y_1,\underline{x}_1),(y_2,\underline{x}_2),(y_3,\underline{x}_3),...,(y_k,\underline{x}_k)$ where y_i is the number of successes in n_i observations at $\underline{x}_i$ $\left(\sum_1^k n_i=n \text{ the sample size}\right)$. When we have just one observation at all $\underline{x}_i$ $n_i=1$ and $k=n$.

$$P(Y_i=y_i)=\begin{cases}\binom{n_i}{y_i}\left[\pi(\underline{x}_i)\right]^{y_i}\left[1-\pi(\underline{x}_i)\right]^{n_i-y_i} & y_i=0,1,...,n_i \\ 0 & otherwise\end{cases}.$$

The likelihood function is given by

$$L(\underline{\beta}; \underline{y}, \underline{x}) = \prod_{i=1}^{k} \binom{n_i}{y_i} \left[\pi(\underline{x}_i)\right]^{y_i} \left[1-\pi(\underline{x}_i)\right]^{n_i - y_i}$$

$$= \prod_{i=1}^{k} \binom{n_i}{y_i} \left[\frac{exp\left(\sum_1^p \beta_j x_{ij}\right)}{1+exp\left(\sum_1^p \beta_j x_{ij}\right)} \right]^{y_i} \left[1 - \frac{exp\left(\sum_1^p \beta_j x_{ij}\right)}{1+exp\left(\sum_1^p \beta_j x_{ij}\right)} \right]^{n_i - y_i}$$

$$ln\, L \propto \sum_{i=1}^{k} \sum_{j=1}^{p} \beta_j x_{ij} y_i - \sum_{i=1}^{k} n_i\, ln\left[1+exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)\right]$$

$$= \sum_{r=1}^{p} \left(\sum_{i=1}^{k} y_i x_{ir}\right) \beta_r - \sum_{i=1}^{k} n_i\, ln\left[1+exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)\right]$$

$$\frac{\partial\, ln\, L}{\partial \beta_r} = \sum_{i=1}^{k} y_i x_{ir} - \sum_{i=1}^{k} n_i x_{ir} \frac{exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)}{1+exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)} = \sum_{i=1}^{k} y_i x_{ir} - \sum_{i=1}^{k} n_i x_{ir} \pi(\underline{x}_i) \ .$$

The maximum likelihood estimate of $\underline{\beta}$ is the solution of the equations

$$\sum_{i=1}^{k} y_i x_{ir} - \sum_{i=1}^{k} n_i x_{ir}\, \hat{\pi}(\underline{x}_i) = 0 \qquad r = 1,2,...,p \ .$$

The equations are nonlinear and require iteration. SPSS, SAS, STATISTICA and BMDP are some of the statistical software packages that can be used to obtain the estimates.

3.5.2. Testing the significance of the coefficients

The significance of the coefficients are tested to determine if all the independent variables are of no value and should be excluded from the model. According to Hosmer and Lemeshow (2000) this involves testing the hypothesis that independent variables in the model are significantly related to the dependent variable.

To check if variables have a significant impact on the model, one model with all independent variables called the saturated model and the one without independent variables are compared. According to Hosmer and Lemeshow (2000), comparing the observed to predicted values in logistic regression are done using the likelihood function based on:

$$D = -2\,ln\left[\frac{likelihood\ of\ fitted\ mod\ el}{likelihood\ of\ saturated\ mod\ el}\right]$$

Where the content within the brackets in the equation above is called the likelihood ratio and the test is called the likelihood ratio test. Saturated model is whereby number of parameters is equal to the number of observations. D in the equation is called the deviance and it is used to determine how the model fits the data.

According to Hosmer and Lemeshow (2000), using the minus twice its log will assist in obtaining a quantity that has a known distribution and can also be used for hypothesis testing purpose. Abraham and Ledolter (2006) say that D is defined as twice the log-likelihood ratio of the saturated model and the model with all the parameters for the variables with their estimated probability of success given by $\hat{\pi}_i = \hat{\pi}(\underline{x}_i) = \dfrac{exp(\underline{x}_i'\hat{\underline{\beta}})}{1+exp(\underline{x}_i'\hat{\underline{\beta}})}$.

Therefore, substituting the log-likelihood function for full fitted model and the likelihood for saturated model in D above will give

$$D = -2\sum_{i=1}^{n}\left[y_i\,ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \qquad \text{where } \hat{\pi}_i = \hat{\pi}(\underline{x}_i).$$

According to Hosmer and Lemeshow (2000), D is very important to assess the goodness-of-fit. To assess the significance of the variables in the model, the deviance for the model with and the one without independent variables in the equation are compared.

If the value of D is close to zero it shows that the model fits the data. Under the null hypothesis $\underline{\beta} = \underline{0}$ the deviance follows a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters in the models. A well fitted model is significant at the 0.05 level of significance or better.

To assess the significance of the independent variables, the value of D for the model with variables is compared to the value of D for the model without variables. The difference between the deviances of the models is given by:

$$\begin{aligned} G &= D(model\ without\ variables) - D(model\ with\ variables) \\ &= 2\left[log\ likelihood(\ full\ model) - log\ likelihood(reduced\ model)\right] \end{aligned}$$

To test the null hypothesis that at least one of the variable coefficients in the full model is different from zero, the full model and the reduced model will be fitted and their respective D will be calculated and compared using G above. According to Quinn and Keogh (2002) statistic G can be termed a deviance when the likelihood ratio is the likelihood of a specific model divided by the likelihood of the saturated model.

According to Hosmer and Lemeshow (2000) the statistic G plays the same role as the numerator of the partial F test in linear regression. The statistic G can be written as follows:

$$G = 2\left\{ \sum_{i=1}^{n} \left[y_i \ln\left(\pi_i\right) + (1-y_i)\ln\left(1-\hat{\pi}_i\right)\right] - \left[n_1 \ln(n_1) + n_0 \ln(n_0) - n\ln(n)\right] \right\}$$

Where $n_0 = \sum y_i$ and $n_1 = \sum (1-y_i)$. With the null hypothesis set at $\beta_1 = 0$, then the G statistics follows a chi-square distribution with 1 degree of freedom. A well fitted model is significant at the 0.05 level of significance or better.

The Wald test is used to test the statistical significance of each coefficient β_i in the model. It tests the null hypothesis that says the coefficients of the independent variables is zero. The Wald statistics is obtained by dividing the maximum likelihood estimate of the slope parameter $\hat{\beta}$ by an estimate of its standard error. Wald test for logistic regression is provided by

$$Z = \frac{\hat{\beta}_i}{SE}$$

3.5.3. Confidence interval for logistic regression coefficients

The confidence interval for the coefficients and the intercept in the logistic regression model are based on the Wald test. The confidence interval for the coefficients is given by

$$\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_i)$$

where $Z_{1-\frac{\varepsilon}{2}}$ is the $\left(1-\alpha/2\right)100^{th}$ percentile of the $N\left(0,1\right)$ distribution and $SE(\hat{\beta}_i)$ is the estimate of the standard error of the respective parameter estimate. The linear part of the logistic regression model is called the logit. The logit is estimated using

$$g\left(\underline{x}\right) = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_p x_p = \sum_{1}^{p} \hat{\beta}_j x_j$$

The equation can be expressed as $\hat{g}(x) = x'\hat{\beta}$, where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)$ and the vector $x' = (x_1, x_2, ..., x_p)$ and $x_0 = 1$. The variance of the logit estimator is obtained by obtaining the variance of the sum. This can be done using

$$\hat{\text{Var}}[\hat{g}(\underset{\sim}{x})] = \sum_{i=1}^{p} x_i^2 \hat{Var}(\hat{\beta}_i) + \sum_{i=1}^{p} \sum_{k=i+1}^{p} 2 x_i x_k \hat{Cov}(\hat{\beta}_i, \hat{\beta}_k) .$$

The expression of the variance can also be written in a matrix expression of the estimator of the variance of the estimator of the coefficients in the model. Therefore the expression is as follows:

$$\hat{V}ar[g(\hat{\underset{\sim}{x}})] = \underset{\sim}{x}'\hat{V}ar(\hat{\beta})\underset{\sim}{x} = \underset{\sim}{x}'(X'VX)^{-1}\underset{\sim}{x} .$$

According to Hosmer and Lemeshow (2000) a $100(1-\alpha)\%$ Wald based confidence interval for the logit model is given by

$$g(\hat{\underset{\sim}{x}}) \pm Z_{1-\frac{\alpha}{2}} \hat{SE}[g(\hat{\underset{\sim}{x}})] .$$

3.5.4. Interpretation of the fitted logistic regression model

Interpretation of a fitted logistic regression model requires us to be able to make inferences from the estimated coefficients in the model. In logistic regression the slope measures the change in the logit function when independent variables changes. The interpretation informs the relationship between an independent variable and the dependent variable.

The concept of odds is used to interpret the coefficients in logistic regression model. For a binary variable, the odds that an event occurs is the ratio of the probability that the event occurs to the probability that it does not occur. If the probability that an event occurs is given by p , then the odds that the event occurs is given by $p/(1-p)$.

The odds ratios which are the measure of association between independent and dependent variables are used in logistic regression. Odds ratio is a relative measure of an event to occur in one group relative to another. According to Westergren et al. (2001) the odds ratio is a relative measure that shows how likely it is for an event to occur given an exposure to an independent variable. In Logistic regression the regression coefficients are used to estimate the log odds of an event per change in the independent variable. The odds ratio is calculated as an exponential function of the logistic regression coefficient (e^{β}) and is also associated with the change in the independent variable.

The odds ratio is given by

$$Odds\ Ratio = \frac{p_1 / (1-p_1)}{p_2 / (1-p_2)}.$$

When the odds ratio is equal to one it shows that the odds of an event to occur in the two groups are the same. If the odds ratio is greater than one, it shows that the odds for an event to occur in the first group is greater than the odds for the event to occur in the second group. When the odds ratio is less than 1 it shows that the odds of an event to occur in the first group is less than in the second group or the independent variable has an effect on the event to occur but with lower odds.

3.6. Survival analysis

Survival analysis is a collection of statistical techniques that are used to analyse data for which the outcome variable is the time until an event occurs (Kleinbaum and Klein, 2005). An event can be the death of a patient, failure of a component or remission of a disease, etc. In survival analysis, the dependent variable, which is usually called the survival time, it is the time from the entry into the study until the event occurs. Survival analysis is used to model the distribution of survival times and estimate the effect of different factors on survival.

Survival time T can be determined using three basic elements (Le, 1997). The elements are:

- time each individual enters the study
- occurrence of an event of interest
- measurement scale for the passage of time

Unlike other statistical procedures, survival analysis has the ability to take into account censored observations. Censoring occurs when a variable of interest is not completely observed. Censoring occurs when an individual has not experienced an event during the period of the study, when an individual is lost to follow up or when an individual withdraws from the study because of other factors. Censoring has an effect on the calculation of survival times for individuals in the study. The assumption of normality does not hold in survival analysis as the survival data is skewed.

The survival time can be modelled by using the survival function or the hazard function. These functions are reviewed in the following section.

3.6.1. Survival function

The survival function which is denoted by $S(t)$ is the probability that an individual survives longer than time t. According to Kleinbaum and Klein (2005) the survival function is important to the survival analysis because it assists in obtaining survival probabilities for different values of t which in turn provides important summary information from survival data. The survival function is defined as

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

where $F(t) = P(T \leq t)$ is the cumulative distribution function of T. The survival time T is a continuous variable and $0 \leq S(t) \leq 1$. The graph of the survival function is a smooth curve. However, in practice the graph follows a steps pattern going from 1 down to 0.

The survival distributions are skewed and they cannot be summarised using their mean and variance but they can be summarised using the medians and percentiles. According to Le (1997) the distribution of the survival time T is characterized by the probability density function f(t) and the cumulative distribution function $F(t)$ defined by

$$f(t) = \lim_{\delta t \to 0} \frac{P(t \leq T \leq t + \delta t)}{\delta t}$$

and

$$\mathrm{F}(t) = P(T \leq t) = \int_0^t f(x) d(x)$$

$$\mathrm{F}(t) = 1 - S(t).$$

3.6.2. Hazard Function

The hazard function h(t) is defined by

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t \mid T \ge t)}{\delta t} \right\}.$$

It is the limit of the conditional probability that an event occurs in the time interval $(t, t+dt)$ given that it has not occurred up to time t divided by the length of the time interval. It is an instantaneous rate. The function does not follow the probability rules as it does not lie between 0 and 1. Hazard function focuses on failure while survival function focuses on survival.

The hazard function is defined by

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t \mid T \ge t)}{\delta t} \right\} = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t) / \delta t}{P(T \ge t)} \right\}.$$

$$\lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t)}{\delta t} \right\} = F^{/}(t) = f(t)$$

Thus

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}.$$

It follows that

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

If we integrate from 0 to t and use the boundary conditions $S(0) = 1$ we get

$$S(t) = exp\left(-\int_0^t h(\tau) d\tau \right).$$

The hazard function can be written in terms of the survival function and conversely. The survival and hazard functions can be used alternatively to model the survival time.

3.6.3. Kaplan-Meier Method

The Kaplan-Meier method is a non-parametric technique within survival analysis which is useful for estimating survival function and provides graphical representation of the survival distributions (Lee and Wang, 2003). The method takes into account censored and uncensored observations when estimating the survival probabilities. The method was used to estimate the survival curves from the survival times without the assumptions of the probability distribution. The assumption of the Kaplan-Meier survival function is that the distribution of censored times is independent of the survival times.

Suppose that in N survival times there are m that are uncensored (event of interest occurred) and n are censored (event did not occur), which means that $N = m + n$. Suppose at time t_i there were m_i events occurring, r_i individuals at risk during the time interval t_{i-1} to t_i and c_i censored. According to Lee and Wang (2003) S (t) will be a discrete function with probabilities mass at each t_i and will be given as

$$S(t_i) = \prod_{j=1}^{i} (1 - \lambda_j) \qquad \text{where } \lambda_j = \frac{m_j}{r_{j-1}}.$$

The conditional probability that m_i of the N individuals at risk dies is given by

$$L_i = \frac{[f(t_i)]^{m_i} [S(t_i)]^{n_i - m_i}}{[S(t_i)]^{n_i}} = \lambda_i^{m_i} (1 - \lambda_i)^{n_i - m_i}$$

where t_i is the time just before t_{i-1}. According to Le (1997) the process is the same as the binomial distribution process with n_i trials, m_i successes and the probability of λ_i. In this study n_i is the i^{th} observation where the survival time was recorded. The record of the child is an observation in this study. This includes censored observations. The product of conditional probability will be given by $L = \prod_{i=1}^{k} \lambda_i^{m_i} (1 - \lambda_i)^{n_i - m_i}$.

The equation above can be seen as a likelihood estimator of the survivor function $S(t)$ and the maximum likelihood estimation of this survival function $\hat{S}(t)$ that maximises L results in

$$\hat{S}(t) = \prod_{t_i < t} (1 - \frac{m_i}{r_{i-1}}).$$

The equation is called the Kaplan-Meier estimator for $S(t)$. In this study the survivor function of the individuals will be calculated using this function. The standard error of $\hat{S}(t)$ can be calculated using the Greenwood's formula given by

$$\hat{S}(t) = \prod_{t_i < t}^{j} p_i \quad where\ p_i = \frac{n_i - m_i}{n_i} \qquad log\ \hat{S}(t) = \sum^{j} log\ p_i$$

$$var[log\ \hat{S}(t)] = \sum var(log\ p_i).$$

$$\therefore var[log\ \hat{S}(t)] = \sum var(log\ \frac{n_i - m_i}{n_i})]$$

Thus

$$s.e.[\hat{S}(t)] = \hat{S}(t) \left[\sum_i \frac{m_i}{n_i(n_i - d_i)} \right]^{1/2}.$$

The 95% confidence limit for the survival probability can be found using the formula

$$\hat{S}(t) \pm 1.96\, s.e[\hat{S}(t)].$$

3.6.4. Log Rank test

The Log rank test is a method that is used to compare two or more survival distributions. According to Lee and Wang (2003) the method is based on a set of scores that are assigned to the observations. The null hypothesis is that the risk of experiencing an event is the same in different groups. That is the survival distributions of two or more groups are the same. The alternative hypothesis is that the survival distributions of two or more groups are not the same or the risk of experiencing an event is different in different groups.

The log rank test arranges the survival times in a rank order using both censored and uncensored observations. The data is arranged into a 2x2 tables at each time an event is experienced.

The sum of all observations scores gives the log rank test of the groups. The variance of the sum of all observations scores is given by

$$Var(S) = \frac{n_1 n_2 \sum_{i=1}^{n_1+_2} w_i^2}{(n_1+n_2)(n_1+n_2-1)}.$$

where w_i is the score for the i^{th} observation and S is the sum of all the scores of the observations. The test statistic is given by

$$L = \frac{S}{Va(S)}.$$

The test statistic follows the chi-square distribution with 1 degree of freedom.

3.6.5. The Cox Proportional hazards Model

In 1972, Sir David Cox proposed the Cox Proportional hazards model which is a survival analysis regression model and is used to model the relationship between the hazard function and a set of covariates. It allows the difference between survival times to be tested while allowing for other factors. Cox Proportional hazards model makes no assumptions about the probability distribution of the baseline hazard (Bradburn et al, 2003). The response variable in the Cox Proportional hazards model is the hazard.

The Cox Proportional hazards model is represented mathematically as follows:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)$$

where the hazard function $h(t)$ is dependent on set independent variables and coefficients of the independent variables are given by $\beta_1, \beta_2, ..., \beta_p$ and $h_0(t)$ is the baseline hazard. According to Cox (1972) the independent variables may be the function of time. The coefficient β_i in the model represents the change in the hazard given a unit change in the independent variables. The function $\exp(\beta_i)$ is called the hazard ratio of the i^{th} variable in the model. If β_i is greater than 0, then the hazard ratio will be greater than 1, which will indicate that as the value of the i^{th} variable increases, then the hazard will increase and thus the length of survival decreases. The hazard ratio is used to measure the risk of an event to occur at time t for an individual in one group compared to an individual in another group.

The Cox Proportional hazards model works with the hazard function and it assumes that the hazard for one group is proportional to the hazard of another group. According to Lee and Wang (2003) if the equation above is divided by $h_0(t)$ in both sides and taking the logarithm of both sides will give

$$\log \frac{h(t)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p .$$

Then if different groups are compared the coefficients in the equation above will give the change in the logarithm of the hazard ratio per unit change of the variable and taking the exponents of the coefficients will give the hazard ratios of the groups while other variables are adjusted.

3.6.5.1. Fitting the Cox proportional hazards model

Like other regression models, fitting the Cox Proportional hazards model involves estimating the unknown coefficients of the independent variables in the model. Cox (1972) suggested the partial likelihood function method as an estimation method for the regression coefficients (β). The partial likelihood function is in the form of

$$PL(\beta)=\prod_{i=1}^{n}\prod_{t\geq 0}\left\{\frac{Y_i(t)r_i(\beta,t)}{\sum_j Y_j(t)r_j(\beta,t)}\right\}^{dN_i(t)}.$$

where $r_i(\beta,t)$ is the risk score for the i^{th} subject, $r_i(\beta,t)=\exp[\underline{X}_i'(t)\underline{\beta}]\equiv r_i(t)$, $dN_i(t)=\begin{cases}1 & \text{if the } i^{th} \text{ observation is an event}\\ 0 & \text{if the } i^{th} \text{ observation is censored}\end{cases}$ and $Y_i(t)=\begin{cases}1 & \text{if } t\leq t_i\\ 0 & \text{if } t>t_i\end{cases}$.

Taking the log of the partial likelihood function gives

$$l(\beta)=\sum_{i=0}^{n}\int_0^{\infty}[Y_i(t)r_i(t)-\log(\sum_j Y_j(t)r_j(t))]^{dN_i(t)}.$$

According to Therneau and Grambsch (2000), partial likelihood function is not a likelihood function as it is proportional to the probability of an observed dataset. Partial likelihood can be taken as likelihood for the purpose of asymptotic inferences.

Differentiating both sides of the $l(\beta)$ with respect to β gives the score vector $U(\beta)$ given by

$$U(\beta)=\sum_{i=1}^{n}\int_0^{\infty}[X_i(s)-\bar{x}(\beta,s)]dN_i(s).$$

where $\bar{x}(\beta,s)$ is a weighted mean of X, for those observations at risk at time s.

Therefore

$$\bar{x}(\beta,s)=\frac{\sum Y_i(s)r_i(s)X_i(s)}{\sum Y_i(s)r_i(s)}$$ where $Y_i(s)r_i(s)$ are the weights.

The estimation of β is found by solving the partial likelihood after equating it to zero. Newton-Raphson algorithm can be used to solve the partial likelihood equation starting with the initial $\hat{\beta}(0)$, and then iteratively computing the following:

$$\hat{\beta}^{(n+1)} = \hat{\beta}^{n} + \mathrm{I}^{-1}(\hat{\beta}^{n})U(\hat{\beta}^{n}).$$

where $\mathrm{I}^{-1}(\hat{\beta}^{n})$ is the negative second derivative which is a pxp information matrix given by

$$\mathrm{I}(\beta) = \sum_{i=1}^{n}\int_{0}^{\infty} V(\beta,s)dN_i(s).$$

where $V(\beta,s)$ is the weighted variance of X at time S and it is given by

$$V(\beta,s) = \frac{\sum_i Y_i(s)r_i(s)[X_i(s)-\bar{x}(\beta,s)]'[X_i(s)-\bar{x}(\beta,s)]}{\sum Y_i(s)r_i(s)}.$$

The estimated values of the parameters are consistent and asymptotically normally distributed with the mean of β and the variance of $\{\varepsilon\mathrm{I}(\beta)\}$, the inverse of the expected information matrix. The iteration will be done until the log partial likelihood of $l(\hat{\beta}^{(n+1)}) \approx l(\hat{\beta}^{n})$.

3.6.5.2. Testing the significance of the coefficients

The Wald, score and likelihood ratios tests are available for the Cox partial likelihood test to test hypothesis on the β's. SPSS and SAS output provides results in all three test statistics. The null hypothesis for the parameter β is given by $H_0 : \beta = 0$. The Wald test statistic is given by

$W = (\hat{\beta} - \hat{\beta}^{(0)})'\hat{\mathrm{I}}(\hat{\beta} - \beta^{0})$ where $\hat{\mathrm{I}} = \mathrm{I}(\hat{\beta})$ is the estimate of the information matrix.

The score test statistics $U'(\beta^{(0)})I(\beta^{(0)})^{-1}U(\beta^{(0)})$ is calculated iteratively using Newton-Raphson algorithm. According to Therneau and Grambsch (2000) the likelihood ratio test is considered the most reliable test and Wald test statistics as the least reliable test.

3.6.5.3. Confidence interval for Cox Proportional hazards model coefficients

The confidence interval for the estimated parameters are created based on the Wald statistic with the lower and upper confidence interval values given by the $\exp(\hat{\beta} \pm 1.96 se(\hat{\beta}))$.

3.7. Variable Selection Method

According to Bradburn et al. (2003) stepwise selection methods are used to select variables to build a model. SPSS has different variable selection approaches that are helpful in selecting and testing the variable's significance in order to increase the efficiency of the data analysis.

Survival analysis and logistic regression methods discussed in this chapter were used to determine factors that have an influence on child survival and the survival time for the children of Dikgale HDSS.

4. CHAPTER 4: DATA ANALYSIS AND RESULTS

4.1. Introduction

Different statistical techniques were used to determine the factors that are associated with child survival and estimate the effects of the factors on child survival. Logistic regression was used to determine factors associated with under-five mortality. Survival analysis was used to determine factors associated with child survival and estimate the effect on child survival.

4.2. Data and Data Analysis

Data used were extracted from the Dikgale HDSS database. The HDSS database was developed to capture the demographic surveillance records that are collected annually. The data analysis was conducted using the statistical package called Statistical Package for Social Science (SPSS). Data from the HDSS database was extracted and imported into SPSS for analysis.

The first part of this chapter provides the variables that were tested for significance for their impact on child survival. The variables were:

1) Child survival

2) Gender of the child

3) Mother alive or dead

4) Mother’s marital status

5) Mother’s age

6) Mother’s educational level

7) Delivery Place

8) Attendant of the delivery

9) Antenatal clinic visits by the mother when pregnant

10) Birth weight

11) Breastfeeding

Descriptive statistics was used to produce tables for the variables in the study. Cross tabulation was used to calculate the association between the variables and child survival. The survival time was measured in days and categorised into 365 days to represent the years of survival.

4.2.1. Description of data

Table 1: Classification of the Dikgale HDSS children by place of birth

		Emigrate		Total
		0	1	
Born in Dikgale HDSS	**Count**	1530	973	2503
	% within Inmigrate to Dikgale	61.1%	38.9%	100.0%
Immigrated to Dikgale HDSS	**Count**	884	99	983
	% within Inmigrate to Dikgale	89.9%	10.1%	100.0%
Total	**Count**	2414	1072	3486
	%	69.2%	30.8%	100.0%

Table 1 shows that the study focused on 3486 children who were born from 01 January 1996 to 31 December 2010. Out of the 3486 children, 2503 children were born in the Dikgale HDSS while 983 children were born outside Dikgale HDSS. Of those 983 children born outside Dikgale HDSS 99 emigrated out of Dikgale HDSS while 884 children remained in Dikgale HDSS during the study period.

Table 2: Classification of child survival status by gender, birth weight category and status of breast feeding

		Event					
		Alive/Censored		Died		Total	
		Count	Row N %	Count	Row N %	Count	Row N %
Gender	Female	1763	98.1%	34	1.9%	1797	100.0%
	Male	1651	97.8%	38	2.2%	1689	100.0%
	Total	3414	97.9%	72	2.1%	3486	100.0%
Birth weight category	Below 2.5 kg	612	95.3%	30	4.7%	642	100.0%
	Greater than or equal to 2.5 kg	1813	98.0%	37	2.0%	1850	100.0%
	Total	2425	97.3%	67	2.7%	2492	100.0%
Breastfed	No	129	94.9%	7	5.1%	136	100.0%
	Yes	2301	97.4%	61	2.6%	2362	100.0%
	Total	2430	97.3%	68	2.7%	2498	100.0%

Table 2 shows that out of the 3486 children in the study, 1797 were females while 1689 were males. Of the 1797 females in the study 1.9% died during the study period while 2.2% of the 1689 males in the study died during the study period. There were only 2492 records with birth weight recorded where 1850 children were of the birth weight 2.5 kg and above and 642 were born with weight below 2.5 kg. The study shows that 4.7% of 642 children who were born with weight below 2.5 kg died during the study period while the corresponding figures for children born with weight 2.5 kg and above were 2.0% of 1850.

The table above shows that there was information on breastfeeding for 2498 children. Out of the 2498 children whose information on breastfeeding was available 2362 children were breastfed during their time since they were delivered while 136 were never breastfed. Of the 2362 children who were breastfed 2.6% died during the study period while 5.1% of the 136 children who were never breastfed also died.

Table 3: Classification of child survival status by mother survival status, mother's age category, mother's marital status and the mother's educational level

		Event					
		Alive/Censored		Died		Total	
		Count	Row N %	Count	Row N %	Count	Row N %
Mother status	**Alive**	3283	98.2%	59	1.8%	3342	100.0%
	Died	131	91.0%	13	9.0%	144	100.0%
	Total	3414	97.9%	72	2.1%	3486	100.0%
Mothers age category	**11 to 19 years**	631	97.1%	19	2.9%	650	100.0%
	20 to 29 years	1709	98.2%	31	1.8%	1740	100.0%
	30 to 39 years	816	97.6%	20	2.4%	836	100.0%
	40+ years	258	99.2%	2	0.8%	260	100.0%
	Total	3414	97.9%	72	2.1%	3486	100.0%
Mother Marital status	**Never Married**	1655	97.0%	51	3.0%	1706	100.0%
	Married	363	98.6%	5	1.4%	368	100.0%
	Widow/Separated	162	98.8%	2	1.2%	164	100.0%
	Divorced	136	97.1%	4	2.9%	140	100.0%
	Total	2316	97.4%	62	2.6%	2378	100.0%
Mothers education level	**No Education**	21	91.3%	2	8.7%	23	100.0%
	Primary Education	168	94.9%	9	5.1%	177	100.0%
	Secondary Education	938	96.9%	30	3.1%	968	100.0%
	Matric	517	98.7%	7	1.3%	524	100.0%
	Higher Education	269	99.3%	2	0.7%	271	100.0%
	Total	1913	97.5%	50	2.5%	1963	100.0%

Table 3 shows the classification of child survival status by mother survival status, mother's age category, mother's marital status and mother's educational status. The table shows that mothers of 144 children died during the study period. The study shows that out of 72 children who died, 13 of them lost their mothers during the study period while the mothers of 59 children who died were still alive during the study period.

The highest number of children were born to mothers in the age group 20 to 29 years (1740), followed by children born to mothers in the age group 30 to 39 years (836). The lowest number of children was born to mothers in the age group is 40 years and older while 650 children were born to mothers in the age group 11 and 19 years of age.

Out of 72 children who died, 31 were born to mothers in the age group 20 to 29 years followed by children born to mothers whose age category is 30 to 39 years at 20. Two children who died were born to mothers in the age group above 40 years while 19 were born to mothers in the age group 11 to 19 years.

Out of the 650 children born to mothers in the age group 11 to 19 years 2.9% died. Out of children 1740 children born to mothers in the age group 20 to 29 years 1.8% died while 2.4% of the 836 children born to mothers in the age group 30 to 39 died. Out of the 260 children born to mothers in the age group 40 years and older 0.8% died.

Table 3 shows that information on mother's marital status was available for 2378 children. The highest number of children (1706) was born to mothers who were never married, followed by 368 children born to mothers who were married. There were 164 children born to widowed or separated mothers while 140 children were born to mothers who were divorced.

Out of the 1706 children born to mothers who were never married 3.0% died. Out of 368 children born to mothers who were married 1.4% died. 1.2% of 164 children born to widowed/separated mothers died while 2.9% of 140 children born to divorced mothers died.

Table 3 shows that information on mother's educational level was available for 1963 children. The highest number of children were born to mothers with secondary education (968) followed by children born to mothers with matric (524). The lowest number of children were born to mothers with no education (23) followed by those born to mothers with primary education (177).The study shows that 271 children were born to mothers with higher education. The table shows that 50 children with information on their mother's educational level died during the study period.

Table 4: Classification of child survival status by antenatal clinic attendance, delivery place and delivery attendant

		Event					
		Alive/Censored		Died		Total	
		Count	Row N %	Count	Row N %	Count	Row N %
ANC Clinic	**No**	59	93.7%	4	6.3%	63	100.0%
	Yes	2370	97.3%	66	2.7%	2436	100.0%
	Total	2429	97.2%	70	2.8%	2499	100.0%
Delivery place	**Home**	212	95.9%	9	4.1%	221	100.0%
	Health Facility	2187	97.3%	61	2.7%	2248	100.0%
	Total	2399	97.2%	70	2.8%	2469	100.0%
Delivery attendant	**Others**	1187	99.1%	11	0.9%	1198	100.0%
	Health Professional	2227	97.3%	61	2.7%	2288	100.0%
	Total	3414	97.9%	72	2.1%	3486	100.0%

Table 4 shows that information on mother's antenatal care (ANC) clinic attendance during their pregnancy was available for 2499 children. The table shows that 63 children were born to mothers who have never attended ANC clinic during their pregnancy while 2436 children were born to mothers who attended ANC clinic during pregnancy.

Out of 63 children born to mothers who did not attend ANC clinic during their pregnancy 6.3% died while 2.7% of the children born to mothers who attended ANC clinic died.

Table 4 shows that information on delivery place of the child was available for 2469 children. The table shows that 221 children were delivered at home while 2248 children were delivered in health facilities. Out of the 221 children who were delivered at home 4.1% died while 2.7% of the children who were delivered in health facilities died.

Table 4 shows that 1198 children were not delivered by a health professional while 2288 children were delivered by a health professional. Out of 1198 children who were not delivered by a health professional 0.9% died while 2.7% of the children were delivered by a health professional died.

4.3. Under-Five Mortality

4.3.1. Introduction

As mentioned in the previous chapters one of the objectives of this study was to determine the factors that have effect on under-five mortality. The study focused on the data for children of Dikgale HDSS born between 01 January 1996 to 31 December 2010 who died before reaching the age of five while still residing in Dikgale HDSS and those who survived until the age of five. The data excludes those children who emigrated before they reached the age of five because their event status is not known. Children who died after the age of five who emigrated are not included into the dataset.

4.3.2. Descriptive analysis

Tables 5–7 give results of tests of association between child survival up to the age of 5 years and selected factors. Table 5 shows the association between children survival up to the age of five years and child birth weight, child gender and breastfeeding of the child. Table 6 and table 7 show the association between children survival up to the age of five years and the mother's survival status, the mother's age category, the mother's educational level, the mother's marital status, ANC clinic attendance by the mother, the delivery attendant and the place of delivery. The test of independence was used to measure the statistical significance of variables at 5% level of significance.

Table 5: Classification of children by status of survival up to the age of five years and gender, birth weight and breastfeeding status with p-values

		Survival status						
		Alive up to age five		Died before age five		Total		p-value
		Count	Row N %	Count	Row N %	Count	Row N %	
Gender	Female	1023	97.0%	32	3.0%	1055	100.0%	0.535
	Male	959	96.5%	35	3.5%	994	100.0%	
	Total	1982	96.7%	67	3.3%	2049	100.0%	
Birth weight category	Below 2.5 kg	351	92.6%	28	7.4%	379	100.0%	0.013
	Greater than or equal to 2.5 kg	811	96.0%	34	4.0%	845	100.0%	
	Total	1162	94.9%	62	5.1%	1224	100.0%	
Breast fed	No	69	90.8%	7	9.2%	76	100.0%	0.095
	Yes	1098	95.1%	56	4.9%	1154	100.0%	
	Total	1167	94.9%	63	5.1%	1230	100.0%	

Table 5 shows the association between children survival up to the age of five and gender of the child, birth weight category of the child and whether the child was breastfed. The table shows that out of the 2049 children, 1055 were females while 994 were males. Out of the 1055 female children 3.0% died while 3.5% of male children died.

Since the p-value (0.535) is more than the significance level (0.05), we can conclude that there is no association between gender and child survival.

Table 5 shows that the birth weight information was available for 1224 children. The table shows that 379 out of the 1224 children with information on birth weight were born with birth weight below 2.5 kg while 845 were born with birth weight of 2.5 kg or more. Table 5 indicates that 7.4% of the 379 children born with birth weight below 2.5 kg died while 4.0% of children born with birth weight equal to 2.5 kg or more died.

Since the p-value (0.013) is less than the significance level (0.05), we can conclude that there is an association between birth weight and child survival beyond the age of five years.

Table 5 shows that information on breastfeeding status was available for 1230 children. Out of the 1230 children, 76 were never breastfed before while 1154 children were breastfed. Out of the 76 children who were never breastfed 9.2% died while 4.9% of those children who were breastfed died.

Since the p-value (0.095) is more than the significance level (0.05), there is no association between breastfeeding and child survival beyond the age of five years.

Table 6 shows that 1917 children were born to mothers who are still alive while mothers of 132 children died. Out of the 1917 children whose mothers are alive, 2.8% died while 9.8% of the children whose mothers have died also died.

Since the p-value (0.000) is less than the significance level (0.05), then it can be concluded that there is a significant relationship between mothers status and child survival beyond the age of five years.

Table 6 shows that the highest number of children (920) was born to mothers of age category 20 to 29 years followed by children (513) born to mothers of age category was 30 to 39 years. The lowest number of children (203) was born to mothers of age category above 40 years and older followed by 413 children born to mothers of age category 11 – 19 years.

Since the p-value (0.116) is more than the significance level (0.05), there is no significant relationship between mother's age and child survival beyond the age of five years.

Table 6: Classification of children by survival status up to the age of five years and mother's survival status, mother's age category, mother's marital status and mother's educational level with p-value for each factor

		Survival status						
		Alive up to age five		Died before age five		Total		p-value
		Count	Row N %	Count	Row N %	Count	Row N %	
Mother status	Alive	1863	97.2%	54	2.8%	1917	100.0%	0.000
	Died	119	90.2%	13	9.8%	132	100.0%	
	Total	1982	96.7%	67	3.3%	2049	100.0%	
Mothers age category	11 to 19 years	394	95.4%	19	4.6%	413	100.0%	0.116
	20 to 29 years	892	97.0%	28	3.0%	920	100.0%	
	30 to 39 years	495	96.5%	18	3.5%	513	100.0%	
	40+ years	201	99.0%	2	1.0%	203	100.0%	
	Total	1982	96.7%	67	3.3%	2049	100.0%	
Mothers Marital status	Never Married	820	94.7%	46	5.3%	866	100.0%	0.037
	Married	263	98.1%	5	1.9%	268	100.0%	
	Widow/Separated	121	98.4%	2	1.6%	123	100.0%	
	Divorced	106	96.4%	4	3.6%	110	100.0%	
	Total	1310	95.8%	57	4.2%	1367	100.0%	
Mothers education level	No Education	13	86.7%	2	13.3%	15	100.0%	.038
	Primary Education	80	90.9%	8	9.1%	88	100.0%	
	Secondary Education	415	93.3%	30	6.7%	445	100.0%	
	Matric	208	97.2%	6	2.8%	214	100.0%	
	Higher Education	92	97.9%	2	2.1%	94	100.0%	
	Total	808	94.4%	48	5.6%	856	100.0%	

Table 6 shows that information on the mother's marital status was available for 1367 children. The highest number of children (866) was born to mothers who were never married followed by 268 children born to married mothers. The lowest number of children (110) was born to mothers who are divorced while 123 children were born to mothers who were widowed/ separated.

Since the p-value (0.037) is less than the significance level (0.05), it can be concluded that there is a significant relationship between mothers marital status and child survival beyond the age of five years.

Table 6 shows that information on the mother’s educational level was available for 856 children. The highest number of children (445) was born to mothers with secondary education followed by 214 children born to mothers with matric. The lowest number of children (15) was born to mothers with no education followed by 88 children born from mothers with primary education. The table indicates that 94 children were born to mothers with higher education.

Since the p-value (0.038) is less than the significance level (0.05), it can be concluded that there is a significant relationship between mother’s educational and child survival beyond the age of five years.

Table 7 shows that the highest number of children (1190) was born to mothers who attended ANC clinics during their pregnancy while 41 children were born to mothers who did not attend ANC clinic during their pregnancy. Out of the 41 children who were born to mothers who did not attend ANC clinic during pregnancy 9.8% died while 5.1% of children born to mothers who attended ANC clinic died.

Since the p-value (0.192) is more than the significance level (0.05), there is no significant relationship between mother’s attendance of ANC clinic during pregnancy and child survival beyond the age of five years.

Table 7 shows that the information on mother’s delivery place was available for 1216 children. The study reveals that 149 children were delivered at home while 1067 were delivered in health facilities. Out of the 149 who were delivered at home 5.4% died while 5.3% of those who were delivered at the health facility also died.

Since the p-value (0.989) is more than the significance level (0.05), there no significant relationship between mother's place of delivery and child survival beyond the age of five years.

Table 7: Classification of children by status of survival up to the age of five years and antenatal clinic attendance, delivery place and delivery attendant with p-value for each factor

		Status of survival						
		Alive up to age five		Died		Total		p-value
		Count	Row N %	Count	Row N %	Count	Row N %	
ANC Clinic	No	37	90.2%	4	9.8%	41	100.0%	0.192
	Yes	1129	94.9%	61	5.1%	1190	100.0%	
	Total	1166	94.7%	65	5.3%	1231	100.0%	
Delivery place	Home	141	94.6%	8	5.4%	149	100.0%	0.989
	Health Facility	1010	94.7%	57	5.3%	1067	100.0%	
	Total	1151	94.7%	65	5.3%	1216	100.0%	
Delivery attendant	Others	949	99.0%	10	1.0%	959	100.0%	0.000
	Health Professional	1033	94.8%	57	5.2%	1090	100.0%	
	Total	1982	96.7%	67	3.3%	2049	100.0%	

Table 7 shows that 1090 children out of the 2049 children were delivered by a health professional while 959 children were not delivered by a health professional. Out of the 1090 who were delivered by a health professional 5.2% died while 1.0% of those who were not delivered by a health professional also died.

Since the p-value (0.000) is less than the significance level (0.05), it can be concluded that there is a significant relationship between child survival beyond the age of five years and whether the child was delivered by a health professional or not.

Table 5, table 6 and table 7 show that birth weight, mother's survival status, mother's marital status, mother's educational level and the delivery attendant are the factors related to child survival.

Table 8: Odds ratio for child survival up to the age of five and child's birth weight

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Birth weight category (Below 2.5 kg / Greater than or equal to 2.5 kg)	.526	.314	.880
For cohort Event = Alive up to age five	.965	.935	.996
For cohort Event = Died	1.836	1.130	2.983
N of Valid Cases	1224		

Table 8 shows that the odds that children born with birth weight 2.5kg or more survive beyond five years are almost two times the odds that children born with birth weight less than 2.5kg survive beyond five years.

Table 9: Odds ratio for child survival up to the age of five and mother status (dead/alive)

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Mother status (Alive / Died)	3.769	2.001	7.099
For cohort Event = Alive up to age five	1.078	1.018	1.141
For cohort Event = Died	.286	.160	.510
N of Valid Cases	2049		

Table 9 shows that the odds that children born to mothers who are alive survive beyond five years are almost four times the odds that children born to mothers who are not alive survive beyond five years.

Table 10: Odds ratio for child survival up to the age of five by delivery attendant

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Delivery attendant (Others / Health Professional)	5.236	2.659	10.313
For cohort Event = Alive up to age five	1.044	1.028	1.060
For cohort Event = Died	.199	.102	.388
N of Valid Cases	2049		

Table 10 shows that the odds that children delivered by non-health professionals survive beyond five years are almost 5.2 times that for children delivered by health professionals.

4.3.3. Application of Logistic Regression Model

In this section logistic regression model is fitted with status of under-five survival as the response variable and the significant factors from the previous section as explanatory variables. Based on the results of table 5, table 6 and table 7 child's birth weight, mother's survival status (dead/alive), mother's marital status, mother's educational level and delivery attendant are associated with child survival. All the explanatory variables were used to check the variables that are significant in the Logistic regression model. The results were further verified using the backwards and forward selection methods.

Table 11 shows the model summary and Hosmer and Lemeshow test for the model in the study. The Cox & Snell R-square for the model is 0.0280, indicating a relationship of 2.8% between the predictors and predictions. The p-value for the Hosmer and Lemeshow Test is 0.822 and it is greater than the significance level of 0.05. Therefore, it can be concluded that the model is a good fit.

Table 8: Model Summary and Omnibus tests of mode coefficients

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	456.441[a]	.028	.083

Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	2.898	6	.822

Omnibus Tests of Model Coefficients

		Chi square	Df	Sig.
Step 1	Step	34.225	8	.000
	Block	34.225	8	.000
	Model	34.225	8	.000

Table 11 shows that the chi square value for the model is 34.225 with 8 degrees of freedom and the p-value (0.000) less than the significance level of 0.05. This shows that the model with all variables is better than the model with only an intercept.

Table 12 shows the logistic regression results which show the factors that are associated with child survival up to the age of 5 years when controlling the effect of other variables. Since the p-values for mother's status and birth weight are less than the significance level of 0.05, it can be concluded that the under-five survival is influenced by the child's birth weight and the mother's status of survival.

The table shows that the odds of under-five survival for a child whose mother is alive is 4.902 times that of a child whose mother is dead. The odds ratio was found by taking an inverse of the $\exp(\beta)$ for mother status. The odds of under-five survival for a child born with weight that is equal to 2.5 kg or more is 2.103 times that of a child born with weight below 2.5 kg.

Table 9: Logistic regression analysis for the association between child survival and different factors

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Birthweight_cat(1)	.744	.295	6.342	1	.012	2.103	1.179	3.752
	Mother_status(1)	-1.589	.361	19.394	1	.000	.204	.101	.414
	Delivery_attendant(1)	-.422	.461	.840	1	.359	.655	.266	1.618
	Secondary	.298	.316	.891	1	.345	1.348	.725	2.505
	Primary	.778	.479	2.633	1	.105	2.177	.851	5.569
	Married	-.880	.488	3.253	1	.071	.415	.159	1.079
	Widowed	-1.054	1.042	1.023	1	.312	.349	.045	2.687
	Divorced	.199	.565	.124	1	.725	1.220	.403	3.690
	Constant	-1.869	.405	21.253	1	.000	.154		

4.4. Application of survival analysis

In this section, the application of survival analysis is described as it was used for Dikgale HDSS data. The data used in this section consists of all the 3486 records. Unlike logistic regression, survival analysis uses all the information from censored data as well as uncensored data.

Kaplan-Meier estimates and Cox Proportional hazards model were fitted to the Dikgale HDSS data for children born from 01 January 1996 to 31 December 2010 using SPSS. The first section shows survival plots produced using the Kaplan-Meier survival function and the second section shows the Cox Proportional hazards model fitted to the data.

4.4.1. Kaplan-Meier survival curves

Kaplan-Meier survival curves were plotted to determine the influence of each factor on children survival probabilities in the study.

Figure 1 shows the survival probabilities for the children by gender. The figure shows that female children have a higher survival probabilities compared to male children from birth.

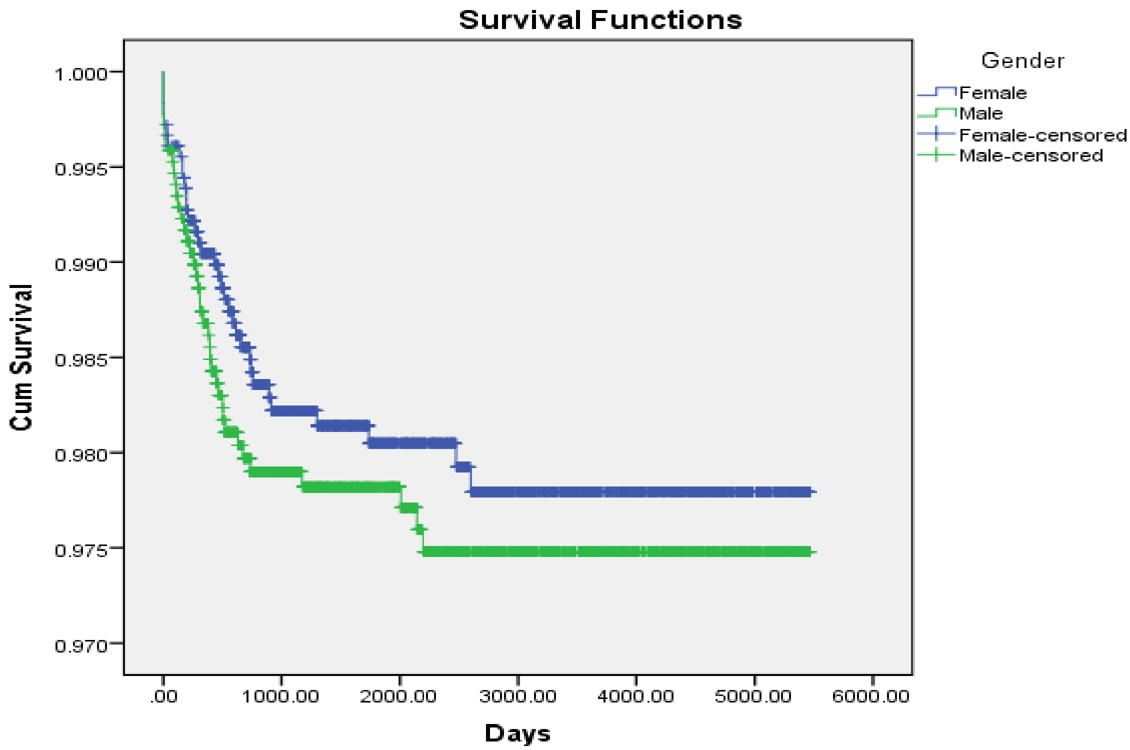

Figure 1: Survival probabilities for Dikgale HDSS children by gender

Table 13 below shows that p-value (0.448) is higher than the significance level of 0.05; therefore we can conclude that gender does not have significant effect on child survival.

Table 13 shows the means for survival time by mother's survival status and the test to compare the survival time by mother's survival status. The mean survival time for a female child was higher at 5366.011 days compared to a male child whose mean survival time is 5346.116 days. The table also shows that the 95% confidence intervals for male and female children are overlapping which again shows that gender does not have significant effect on child survival.

Table 10: Means for Survival Time by Gender and test to compare survival time and gender

Gender	Mean[a]				
	Estimate	Std. Error	95% Confidence Interval		
			Lower Bound	Upper Bound	
Female	5366.011	17.860	5331.006	5401.017	
Male	5346.116	19.888	5307.135	5385.096	
Overall	5356.825	13.331	5330.695	5382.955	
Overall Comparisons					
			Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)			.575	1	.448
Breslow (Generalized Wilcoxon)			.771	1	.380
Tarone-Ware			.682	1	.409

Figure 2 shows the survival probabilities of children by their mother's survival status. The survival probabilities for the children whose mothers have died dropped faster compared to the survival probabilities for children whose mothers are still alive. The figure shows that children whose mothers are alive have a higher survival probabilities compared to children whose mothers died.

Table 14 below shows that the p-value (0.00) is less than the significance level of 0.05. Therefore, it can be concluded that mother's survival status has a significant effect on child survival.

Table 14 shows the means for survival time by mother's survival status and the test to compare the survival time by mother's survival status. The mean survival time for a child whose mothers are still alive was higher at 5372.651 days compared to a child whose mothers have died which is 5001.599 days. The table also shows that the 95% confidence intervals for children whose mothers are alive and children whose mothers have died are not overlapping which implies that mother's survival status has an effect on child survival.

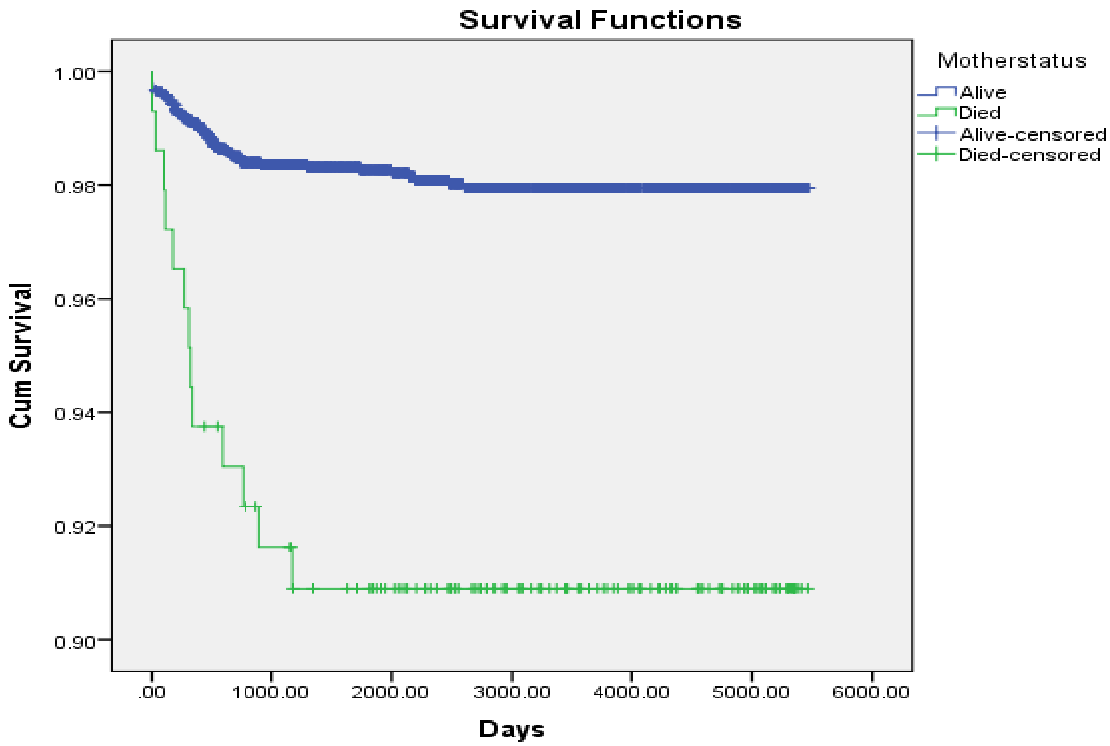

Figure 2: Survival probabilities for Dikgale HDSS children by mother's survival status

Table 11: Means for Survival Time by mother's survival status and test to compare survival time and mother's survival status

Mother status	**Mean[a]**				
	Estimate	**Std. Error**	**95% Confidence Interval**		
			Lower Bound	**Upper Bound**	
Alive	5372.651	12.716	5347.728	5397.574	
Died	5001.599	122.259	4761.972	5241.226	
Overall	5356.825	13.331	5330.695	5382.955	
Overall comparisons					
			Chi-Square	**Df**	**Sig.**
Log Rank (Mantel-Cox)			31.925	1	.000
Breslow (Generalized Wilcoxon)			34.920	1	.000
Tarone-Ware			33.937	1	.000

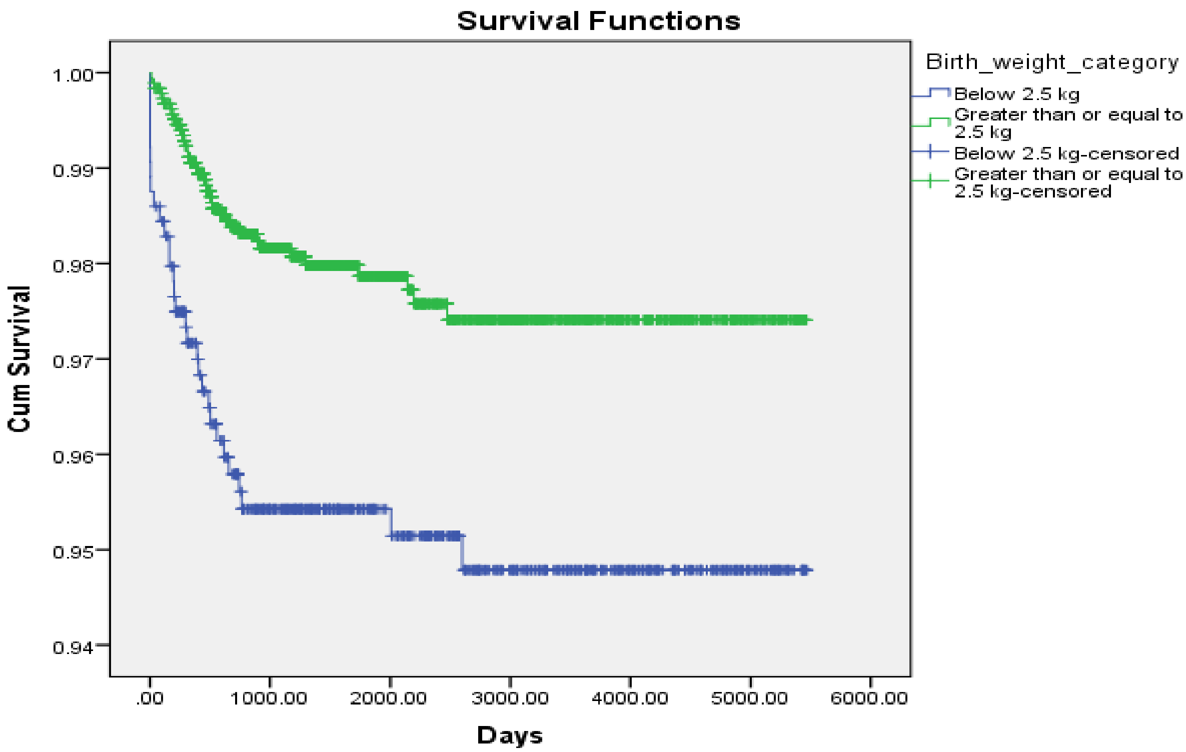

Figure 3: Survival probabilities for Dikgale HDSS children by child birth weight

Figure 3 shows the survival probabilities of children by child birth weight. The survival probabilities for the children in both birth weight categories dropped over number of days. Children born with weight below 2.5 kg have lower survival probabilities than children born with weight equal to 2.5 kg or more from birth. The survival probabilities for children whose birth weight is below 2.5 kg decreases faster than the one for children whose birth weight is greater than or equal to 2.5 kg.

Table 15 below shows that the p-value is less than the significance level of 0.05. Therefore, it can be concluded that the birth weight has an effect on child survival.

The table also shows the means for survival time by birth weight and the test to compare the survival time by birth weight. The mean survival time for the children whose birth weight is 2.5 kg or more is higher at 5347.149 days than the mean survival time for children with birth weight below 2.5 kg. The table also shows that the 95% confidence intervals for children with birth weight greater than or equal to 2.5 kg and children with body weight below 2.5 kg are not overlapping, and again it can be concluded that the birth weight has an effect on child survival.

Table 12: Means for Survival Time by birth weight and test to compare survival time and birth weight

Birth weight category	Mean[a]				
	Estimate	Std. Error	95% Confidence Interval		
			Lower Bound	Upper Bound	
Below 2.5 kg	5212.282	45.957	5122.207	5302.357	
Greater than or equal to 2.5 kg	5347.149	19.780	5308.380	5385.918	
Overall	5314.385	18.898	5277.345	5351.425	
Overall Comparisons					
			Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)			11.616	1	.001
Breslow (Generalized Wilcoxon)			13.988	1	.000
Tarone-Ware			12.992	1	.000

Figure 4 shows the survival probabilities for Dikgale HDSS children by delivery attendant. The survival probabilities for all children dropped over the time span. Children who were delivered by a health professional had a lower survival probability than those children who were not delivered by a health professional. Survival probabilities for children who were delivered by a health professional decreased faster than the survival probabilities for children who are delivered by others.

Table 16 below shows that the p-value is less than the significance level of 0.05. Therefore, it can be concluded that the delivery attendant has an effect on child survival.

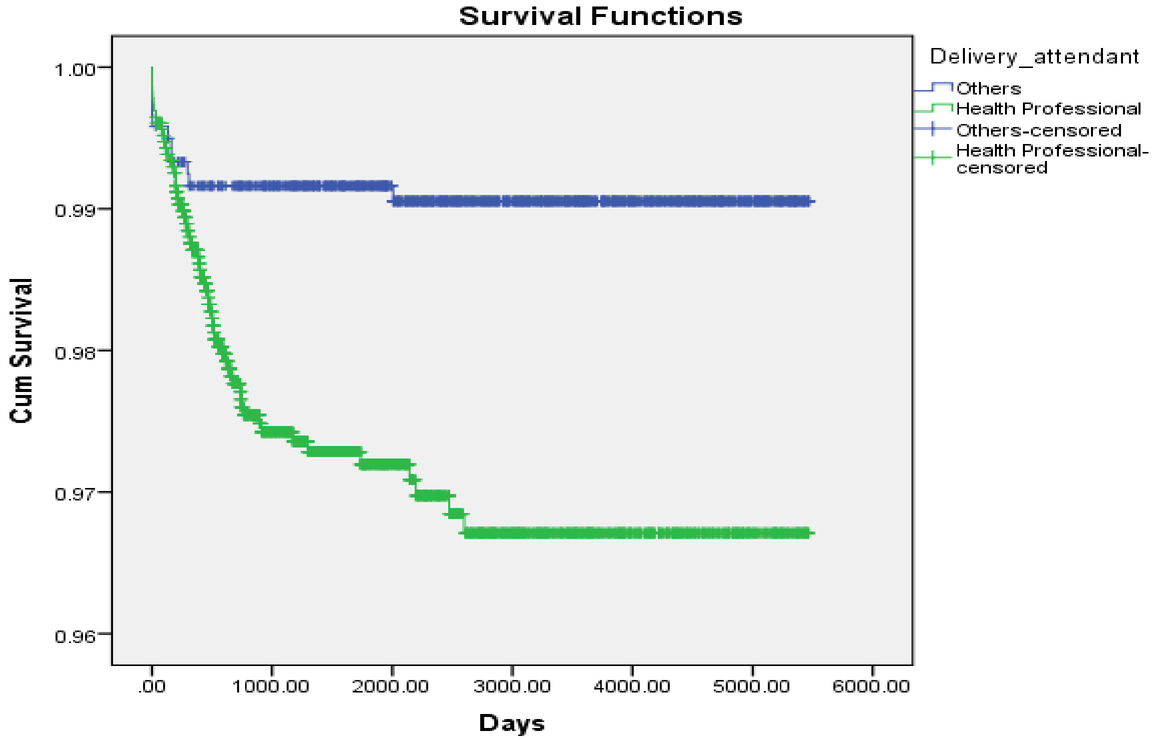

Figure 4: Survival probabilities for Dikgale HDSS children by Delivery attendant

Table 13: Means for Survival Time for children by delivery attendant and test to compare survival time by delivery attendant

Delivery attendant	Mean[a]				
	Estimate	Std. Error	95% Confidence Interval		
			Lower Bound	Upper Bound	
Others	5422.344	14.621	5393.687	5451.000	
Health Professional	5312.364	19.842	5273.473	5351.255	
Overall	5356.825	13.331	5330.695	5382.955	

Overall Comparisons			
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	14.910	1	.000
Breslow (Generalized Wilcoxon)	12.821	1	.000
Tarone-Ware	13.857	1	.000

Table 16 shows the means for survival time for the children by delivery attendant and the test to compare survival time by delivery attendant. The mean survival time for children who were not delivered by a health professional was higher at 5422.344 days than the mean survival time for children delivered by a health professional at 5312.364 days. The table also shows that the 95% confidence intervals for children delivered by health professional and children delivered by others are not overlapping which implies that delivery attendant has an effect on child survival.

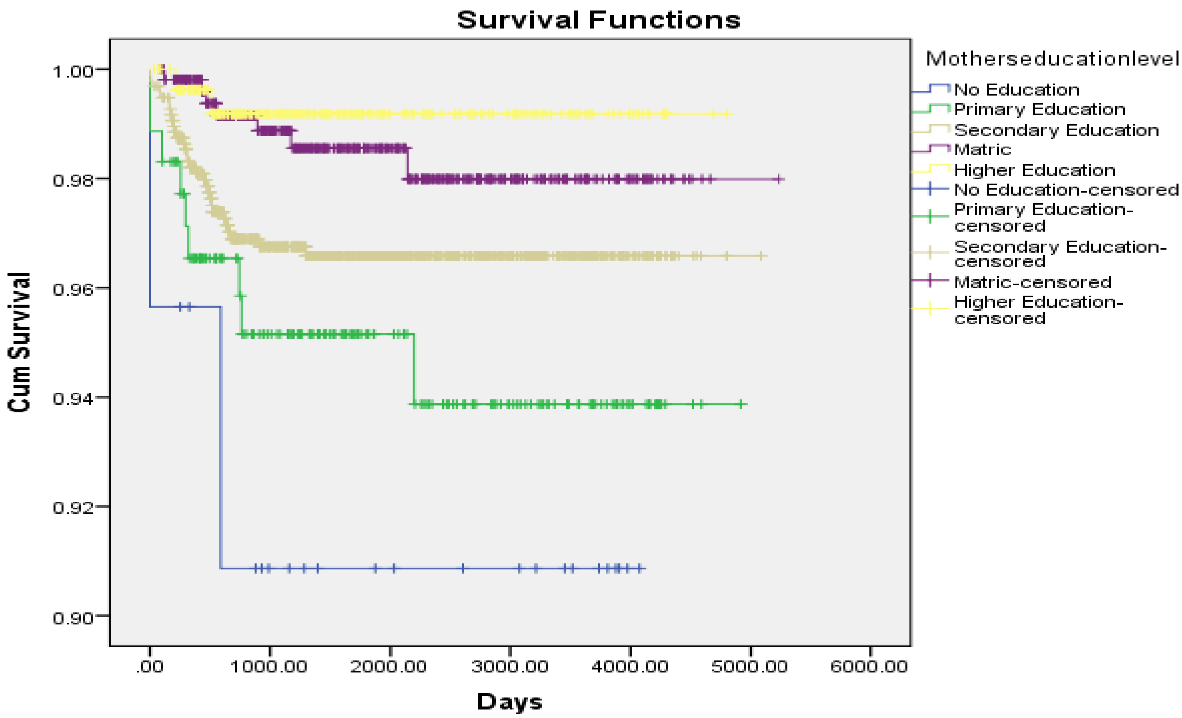

Figure 5: Survival probabilities for Dikgale HDSS children by mother's educational level

Figure 5 shows the survival probabilities for Dikgale HDSS children by their mother's educational level. Children whose mother had no education had lower survival probability followed by those from mothers with primary education. Children whose mothers had higher education had higher survival probability followed by those whose mothers had secondary education.

Table 17 below shows that the p-value is less than the significance level of 0.05. Therefore, it can be concluded that the mother's education has effect on child survival.

Table 14: Means for Survival Time for children by the mother's educational level and the test to compare survival time for children by mother's educational level

Mother's educational level	Mean[a]				
	Estimate	Std. Error	95% Confidence Interval		
			Lower Bound	Upper Bound	
No education	3731.057	232.235	3275.877	4186.236	
Primary education	4657.984	83.619	4494.092	4821.877	
Secondary education	5097.820	22.261	5054.188	5141.451	
Higher education	4765.878	25.467	4715.963	4815.793	
Overall	5092.888	19.878	5053.928	5131.849	
Overall Comparisons					
			Chi-Square	**Df**	**Sig.**
Log Rank (Mantel-Cox)			10.932	3	.012
Breslow (Generalized Wilcoxon)			10.768	3	.013
Tarone-Ware			10.862	3	.012

Table 17 shows the mean survival time for the children by their mother's educational level and the overall test to compare the survival time by the mother's educational level. The highest mean survival time was for children whose mothers had secondary education at 5097.820 days followed by children whose mothers had higher education at 4765.878 days. Children born from mothers with no education had the lowest mean survival time at 4657.984 days followed by children born from mothers with primary education with mean survival time of 4657.984 days.

The 95% confidence intervals for the mother's educational level are not overlapping. This supports the claim that mother's educational level has an effect on child survival.

Kaplan-Meier survival curves above show that the probability of survival of 0.5 has not been reached. It can be concluded that median survival time for children of Dikgale HDSS has not been reached during the study period. Based on the results of the Kaplan-Meier survival curves there is insufficient statistical evidence to support the claim that gender of the child, child breast feeding, antenatal clinic visit by the mother, mothers place of delivery, mother's marital status and mother's age have effect on child survival. The survival curves do not show any statistical significance at 5% level of significance.

Birth weight, mother survival status (alive/dead), delivery attendant and mother's educational status were found to have effect on child survival probability.

4.4.2. Cox Proportional hazards model

In this section, the Cox Proportional hazards model will be fitted to describe the relationship between survival time and some selected independent variables. The model will include all significant variables from Kaplan-Meier survival probability analysis in section 4.4.1 above.

Table 15: Test of Model Coefficient

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	Df	Sig.	Chi-square	Df	Sig.	Chi-square	Df	Sig.
982.893	75.965	6	.000	40.088	6	.000	40.088	6	.000

Table 18 shows that the model with the selected four variables is significant at 5 % level of significance. This shows that the model can be used to determine the relationship between survival time and the four variables.

Table 19 shows the p-values and the hazard ratios of factors that were found to have an effect on survival time. The factors were tested at 5% level of significance. In the presence of all variables that were significant in the Kaplan-Meier survival curves, mother status (dead or alive) and birth weight category are the only variables with the p-value less than 0.05. Therefore it can be concluded that mother's status and birth weight are the only variables which have a significant relationship with child survival.

Table 16: The Cox Proportional hazards model for child survival and different factors

	B	SE	Wald	Df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Mother status	-1.924	.318	36.619	1	.000	.146	.078	.272
Birth weight category	.815	.271	9.045	1	.003	2.258	1.328	3.840
Delivery attendant	-.319	.413	.598	1	.439	.727	.324	1.632
Primary	-.373	.406	.845	1	.358	.688	.311	1.526
Secondary and above	.129	.288	.202	1	.653	1.138	.647	2.001

The model shows that in the presence of all variables having influence on each other in the model, mother survival status and birth weight are the only variables that have effect on child survival. Therefore the survival of the children is influenced by mother's survival status and child birth weight. Table 19 show that the death hazard for children whose mother has died is 6.85 (the inverse of Exp (B) is 1/0.146) times the hazard of a child whose mother is alive. The death hazard for children born with birth weight that is 2.5kg and above is 2.258 times the hazard for children born with birth weight below 2.5kg.

Figure 6 shows the log-minus-log plot for proportional hazards checking. The curve for the children whose mothers have died is higher compared with the one for children whose mothers are still alive. Therefore, the proportional hazards assumption is satisfied as the two curves are not overlapping.

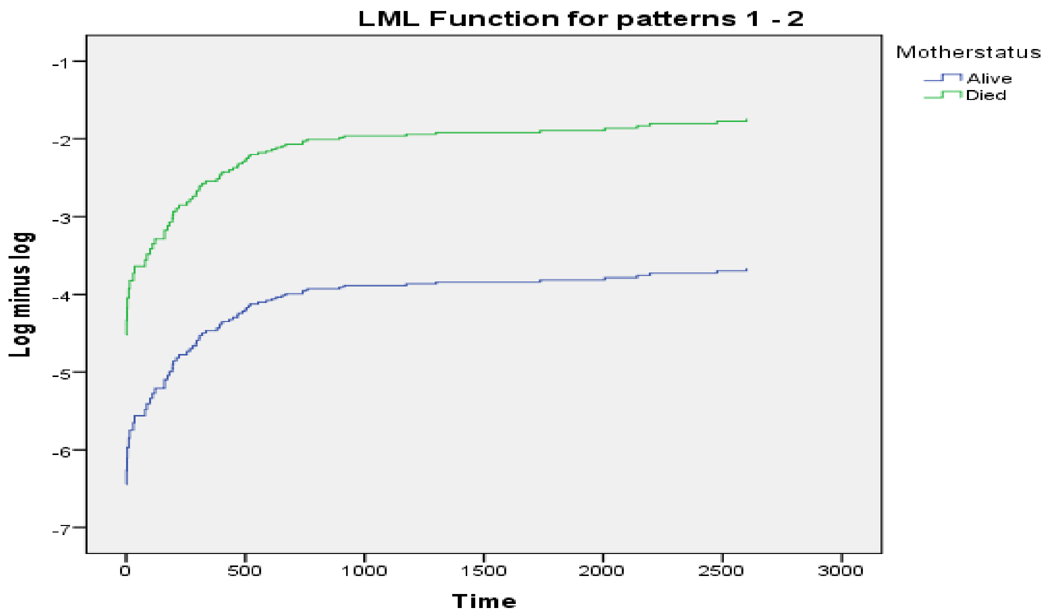

Figure 6: Log-minus-log plot for proportional hazards checking

4.5. Conclusion

In this chapter cross tabulation, logistic regression and survival analysis were used to determine the effect of various factors on child survival and child survival time. Logistic regression model was fitted to the variables that were found to have effect on child survival. Logistic regression model has shown that mother status (alive/ dead) and child birth weight have significant effect on under-five survival.

Kaplan-Meier plots were fitted to the Dikgale HDSS data for all children and the factors that were found to have effect on survival time were included into the Cox Proportional hazards model to determine the relationship between survival time and various factors. The model showed how each factor influences child survival while being influenced by other factors. It was found that the mother's status (alive/ dead) and birth weight have effect on child survival time.

5. Chapter 5: Discussion and Conclusion

In this chapter, the results found in chapter four will be discussed. The results are compared with other studies that have been reviewed in chapter 2. The chapter will give the summary of the results, the recommendation and the concluding remarks.

5.1. Discussion of the results

5.1.1. Introduction

The objective of the study was to determine the effect of various factors on under-five mortality and child survival time for children born between 01 January 1996 and 31 December 2010. The data used in the study was extracted from the Dikgale HDSS (a longitudinal demographic surveillance system) database. The study used cross tabulation, logistic regression and survival analysis methods to assess and determine factors which have effect on child mortality and child survival time.

5.1.2. Factors that have effect on child survival and survival time

Based on the results from cross tabulation, it was found that under-five survival is affected by child birth weight, the mother's survival status (alive/dead), the mother's marital status and presence of a delivery attendant.

The results show that the presence of a delivery attendant has effect on child survival. The results show that children who were not delivered by a health professional survive longer than those who are delivered by a health professional. Some of the children that are delivered by health professionals might be due to complications experienced by the mothers. Figure 4 also shows that children who were not delivered by a health professional have a higher survival probability than those children delivered by a health professional. Mothers should still be encouraged to get a health professional to assist them during delivery. The results of this study contradict the findings by Ajaari et al. (2012), Rahman (2009) and Hossain and Mondal (2009) who found that the place of delivery is an important determinant of child survival. The reason for lower survival probabilities for children delivered by a health professional might be that mothers visit a health facility if there are serious complications when giving birth.

Logistic regression model was used to determine the relationship between under-five survival and various independent variables. Cox Proportional hazards model was used to determine the relationship between child survival time and different independent variables. The models found that the mother's survival status (alive/dead) and birth weight have effect on child survival and child survival time.

The results by Chowdhury et al. (2013) agree with this study that the survival status of the mother and child birth weight have effect on under-five mortality and child survival. The odds that children born to mothers who are alive survive beyond five years are almost four times the odds that children born to mothers who are not alive. This shows that children whose mothers are still alive in Dikgale HDSS are likely to survive longer than children whose mothers have died.

The results agree with the results of Razzaque et al. (2012) which showed that the survival status of the mothers is associated with the under-five survival status of children from Matlab HDSS in Bangladesh. The results are also consistent with the results of Becher et al. (2004) and Anderson et al. (2007) who also found that the death of the mother is strongly related to child survival.

The death of a child when the mother is no longer alive can be attributed to the reduction of child care, no proper support system for the child when necessary and no breastfeeding for younger ones (Razzaque et al., 2012).

Children with birth weight equal to 2.5 kg or more had a higher chance of surviving beyond five years (OR: 1.90, 95% CI: 1.14 – 3.18) compared to children whose birth weight is below 2.5 kg. The results showed that the odds that children born with birth weight 2.5kg or more survive beyond five years are almost two times the odds that children born with birth weight less than 2.5kg survive beyond five years. These results are consistent with the results from the study that was conducted by Dutt and Srinivasa (1997) which showed that child birth weight is associated with child survival. Low birth weight is defined as weight at birth of less than 2.5 kg and children with low birth weight are approximately 20 times more likely to die than those born with weight greater than 2.5 kg (WHO, 2004).

The results agree with the study by Hirve and Ganatra (1997) conducted in rural Western India that birth weight exerted influence on child survival both in early child life and at around the age of five. Mothers need to maintain good nutrition while pregnant to reduce the risk of low birth weight.

Logistic regression and survival analysis results in this study have shown that gender of the child, breastfeeding, whether the mother attended antenatal clinic while pregnant, the place where the child was delivered, the presence of delivery attendant, the mother's age, education of the mother and marital status of the mother have no effect on child survival and child survival time.

The results are not consistent with the results by Worku (2009) which showed that duration of breastfeeding, marital status of the mother, the mother's educational level and delivery place have a significant impact on child mortality and survival.

The results are also not consistent with the findings of other researchers such as Mondal et al. (2009) and Mustafa and Odimegwu (2008) who found that breast feeding of a child have an influence on child survival.

Mustafa and Odimegwu (2008)'s results on place of delivery and mother's highest educational level is consistent with the results found in this study. Kembo and Van Ginneken (2009) also found that the mother's age, child gender and maternal education do not have a significant impact on child survival.

5.1.3. Factors that appear to have no effect on child mortality and survival time

Child gender does not appear to have an influence on child survival when using both cross tabulation and Kaplan Meier survival curves. The study by Hirve and Ganatra (1997) found that the girl child had a higher survival time compared to a boy child until the late neonatal period where boys appeared to survive longer than girls. The results are contradictory to the study by Mustafa and Odimegwu (2008) who found gender to be associated with infant mortality.

Kaplan Meier survival curves show that children born to mothers who are 20 years and older have a higher survival probability than children that are born to mothers with age less than 20 years. The log rank test indicates that there is no significant statistical association between mother's age and child survival. This is not consistent with the results by Taffa and Obare (2004) and Syamala (2004) who found a strong relationship between child survival and mother's age.

This study found that the mother's delivery place is not associated with child survival. The result is in contradiction with that by Rahman (2009) which showed that mother's place of delivery is highly associated with child survival with children born in a health facility survive longer than those who are born elsewhere.

The studies by Caldwell (1990), Van den Broeck et al. (1996), Syamala (2004), Chowdhury et al. (2010), Rahman (2009) and Kembo and Van Ginneken (2009) found that maternal education has a relationship with child survival. The results of those studies contradict the result of this study that mother's educational level plays no role in child survival. This might be because of the maternal education which mothers are getting when they visit the health facilities for maternal care.

The result is consistent with the result by Worku (2009) who found that the mother's education is not related to child mortality using logistic regression and survival analysis techniques to analyze data for South African children under the age of five years.

In South Africa women are encouraged to breastfeed their children because it is believed that breastfeeding plays an important role in child survival. According to Abada et al. (2001) breastfeeding contributes to the immunologic defense system of the child and increases their resistance to diseases. The Department of Health in South Africa is promoting breastfeeding of a child to influence the growth of the child. Abada et al. (2001) found that breastfeeding does not significantly influence child survival at 5% level of significance, but at 10% level of significance cross tabulation showed that breastfeeding has a significant influence. The results of this study showed that breastfeeding does not have an influence on child survival.

Women are encouraged to visit the health facilities during their pregnancies to assist in reducing the maternal death rates in South Africa. The Department of Health is offering free antenatal services to pregnant women since 1994 (Department of Health, 2007). The results showed that antenatal care visits to the health facility by a pregnant woman have no influence on child survival. The results contradict what Uddin et al. (2009) found which showed that maternal health care services have a strong relationship with the child survival.

Literature shows that mothers who utilized antenatal services during their pregnancy are aware of what the facilities can offer for their child health. The results of the study might be because all mothers are informed of the necessary effect of child health services on the child growth and wellbeing. According to Pervin et al. (2012) the antenatal services include the communication of health-related information, the prevention and management of complications, screening of risk factors and preparation of delivery in a safe place and by skilled attendants.

5.2. Conclusion

5.2.1. Summary of the results

The results in the study found that the mother’s survival status and child birth weight are important predictors of child survival. All the methods used have identified the mother’s survival status and child birth weight as important predictors of child survival. A child whose mother is alive has a higher chance of survival than a child whose mother is dead. Also, a child with birth weight 2.5 kg or more has a higher probability of survival than a child with birth weight less than 2.5 kg.

Cox Proportional hazards model and Logistic regression have found that gender of the child, mother’s marital status, mother’s age, mother’s educational level, delivery place, who attendant to the delivery, antenatal care clinic visits by the mother, whether the child was ever breastfed were found not to be significant in the model.

5.2.2. Recommendations

It is recommended that the results of this study be used to formulate strategies and interventions that will improve the lifespan of children and assist in reducing child mortality. The results of this study will assist the country to move towards improving child survival thus moving towards the MDGs target of reducing child mortality by 2015. The study can be replicated in other HDSS in the country and other areas especially rural areas to assist in the improvement of national policies on child health.

Further studies to determine the social and environmental factors that have effect on birth weight are needed so that policies can be drafted to be able to deal with low birth weight and assist in increasing child birth weight. WHO report on low birth weight has shown that birth weight is a good indicator of public health challenges including long-term maternal nutrition, ill health and poor pregnancy health care.

It is also important that further studies be conducted to assess the factors that have effect on survival of children whose mothers are dead. This will assist in the development of policies that will strengthen the overall improvement in child survival.

5.2.3. Conclusion

The main purpose of this study was to determine the factors that have effect on child survival and child survival time. Based on the results from this study the child birth weight and mothers survival (alive / dead) were the factors that have significant effect on child survival and child survival time. The results of the study with regard to the gender of the child, whether the child was breastfed, mother’s educational level, antenatal clinic visits by the mother during pregnancy, delivery place of the child and the mother’s age have been found not to have effect on child survival and child survival time.

References

Abada, T. Travato, F. & Lalu, N. 2001. Determinants of breastfeeding in the Philippines: A survival analysis. *Social Science and Medicine*, 52:71-81.

Abraham, B. & Ledolter, J. 2006. Introduction to Regression Modeling. Thomson USA: Brooks/Cole.

Actuarial Society of South Africa. 2005. ASSA2003. Cape Town. South Africa (accessed 12 April 2011).

Adhikari, R. & Podhisita, C. 2010. Household headship and child death: Evidence from Nepal.*BMC International Health and Human Rights*, 10(13), 1 – 8.

Ajaari, J. Masanja, H. Weiner, R. Abokyi, S. & Owusu-Agyei, S. 2012. Impact of Place of Delivery on Neonatal Mortality in Rural Tanzania. *International Journal of MCH and AIDS*, 1(1): 49 -59.

Anderson, F. Morton, S. Naik, A. & Gebrian, B. 2007. Maternal Mortality and the Consequences on Infant and Child Survival in Rural Haiti. *Maternal Child Health Journal*, 11: 395 – 401.

Atrash, H. 2011. Parents' death and its implications for child survival. *Journal of Human Growth and Development*, 21(3): 759 -770.

Becher, H. Muller, O. Jahn, A. Gbangou, A. Kynast-Wolf, G. & Kouyate, B. 2004. Risk factors of infant and child mortality in rural Burkina Faso. *Bulletin of the World Health Organisation*, 82 (4): 265 -273

Bennet, J. 1999.Correlates of child mortality in Pakistan: A hazard model analysis. *The Pakistan Development Review*, 30(1): 85-118.

Bradburn, M. Clark, T. Love, S. & Altman, D. 2003. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*, 89: 431 - 436.

Bradshaw, D. Bourne, D. & Nannan, N. 2003. What are the leading causes of death among South African children? *MRC Policy Brief*, 3.

Caldwell, J. 1990. Cultural and Social Factors Influencing Mortality Levels in Developing Countries. *Annals of the American Academy of Political and Social Science*, 510: 44-59.

Chowdhury, A. Hossain, M. Khan, M. & Hoq, M. 2013. Cox's Proportional hazards Model analysis of Child survival in Bangladesh. *Asian Journal of Social Sciences and Humanities*, 2(2): 608 – 618.

Chowdhury, Q. Islam, R. & Hossain, K. 2010. Socio-economic determinants of neonatal, post neonatal, infant and child mortality. *International Journal of Sociology and Anthropology*, 2(6): 118 – 125.

Cox, D. 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society,* Series B (34): 187—220.

Department of Health. 2007. Guidelines for Maternity Care in South Africa, Pretoria, South Africa.

Doctor, H. 2011. Does living in a female-headed household lower child mortality? The case of rural Nigeria. *Journal of Rural and Remote Health Research*, 11:163.

Dorrington, R. Bradshaw, D. Johnson, L. & Daniel, L. 2006. The demographic impact of HIV/AIDS in South Africa: National and provincial indicators 2006. Centre for Actuarial Research, Medical Research Council and Actuarial Society of South Africa, Cape Town, South Africa.

Dutt, D. & Srinivasa, D. 1997. Impact of Maternal and Child Health Strategy on Child Survival in a Rural Community of Pondicherry. *Indian Pediatrics*, 34: 785-791.

Folasade, I. 2000. Environmental factors, situation of women and child mortality in South Western Nigeria. *Social Science and Medicine*, 51: 1473 – 1489.

Fotso, J. Ezeh, A. Madise, N. & Ciera, J. 2007. Progress towards the child mortality millennium development goal in urban sub-Saharan Africa: the dynamic of population growth, immunization, and access to clean water. *BioMed Central Public Health*, 7:218.

Hall, k. Woolard, I. Lake, L. & Smith, C. 2012. South African Child Gauge. Children's Institute, University of Cape Town, South Africa.

Heaton, B. & Amoateng, A. 2007. The family context of racial difference in child mortality in South Africa: Human Science Research Council, Pretoria, South Africa.

Hirve, S. & Ganatra, B. 1997. A Prospective Cohort Study on the Survival Experience of Under Five Children in Rural Western India. *Indian Pediatrics*, 34: 995 -1001.

Hobcraft, J. 1993. Women's education, child welfare and child survival: a review of the evidence. *Health Transition Review*, 3(2): 169-173.

Hossain, K. & Mondal, N. 2009. A Study on the Health Factors of Infant and Child Mortality in Rajshahi, Bangladesh. *Asian Journal of Medical Sciences*, 1(3): 82 – 87.

Hosmer, D. & Lemeshow, S. 2000. Applied Logistic Regression. (2nd Ed.). John Wiley & Sons, Inc, United States of America.

Hypher, N. 2011. The role of social protection in achieving equitable reduction in child mortality. Centre for Social Protection, IDS.

Kabir, A. Islam, M. Ahmed, M. & Khalique, B. 2001. Factors influencing Infant and child mortality in Bangladesh. *The Sciences*, 5: 292-295.

Kembo, J. & Van Ginneken, J. 2009. Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis. *Demographic Research*, 21(13): 367-384.

Kibel, M. Lake, L. Pendlebury, S. & Smith, C. 2010. South African Child Gauge 2009/2010, Cape Town: Children's Institute, University of Cape Town, South Africa.

Kleinbaum, D. & Klein, M. 2005. Survival Analysis: A Self-Learning Text. Springer, United States of America.

Kumar, P. & File, G. 2010. Infant and child mortality in Ethiopia: A statistical Analysis Approach. *Ethiopian Journal of Education and Science*, 5(2): 51 – 57.

Kyei, K. 2011. Socio-economic factors affecting under five mortality in South Africa – An investigative study. *Journal of emerging trends in economic and management sciences*, 2 (2): 104 – 110.

Le, C. 1997. Applied Survival Analysis, New York: John Wiley & Sons, INC.

Lee, T. & Wang, J. 2003. Statistical Methods for Survival Data Analysis, John Wiley & Sons, INC, New York

Mahmood, N. & Kiani, M. 1994. Health Care Determinants of Child Survival in Pakistan. The Pakistan Development Review, 33(4): 759-771.

Moisi, J. Gatakaa, H. Noor, A. Williams, T. Bauni, E. Tsofa, B. Levine, O. & Scott, A. 2010. Geographical access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. *BioMed Central Public Health*, 10:142.

Mondal, N. Hossain, K. & Ali, K. 2009. Factors influencing infant and child mortality: A case study of Rajshahi District, Bangladesh. *Journal of Human Ecology*, 26(1): 31-39.

Mturi, A. & Curtis, S. 1995. The determinants of infant and child mortality in Tanzania, *Health Policy and Planning*, 10(4): 384 – 394.

Mustafa, H. & Odimegwu, C. 2008. Socioeconomic determinants of Infant Mortality in Kenya: Analysis of Kenya DHS 2003. *Journal of Humanities and Social Sciences*, 2(2): 1-16.

Peng, C. Lee, K. & Ingersoll, G. 2002. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1): 3 -14.

Pervin, J. Moran, A. Rahman, M. Razzaque, A. Sibley, L. Streatfield, P. Reichnbach, L. Koblinsky, M. Hruschaka, D. & Rahman, A. 2012. Association of Antenatal Care with facility delivery and prenatal survival – A population based study in Bangladesh. *Biomed Central Pregnancy and child birth*, 12:111.

Quinn, G. & Keogh, M. 2002. Experimental Design and Data Analysis for Biologists, United Kingdom: Cambridge University Press.

Rahman, M. 2009. Factors Affecting on Child Survival in Bangladesh: Cox Proportional hazards Model Analysis. *Journal of Tropical Medicine*, 6(1): 1 – 7.

Razzaque, A. Hossain, A. DaVanzo, J. Hoque, M. Alam, N. Bhuiya, A. & Streatfield, P. 2012. Effect of Maternal Mortality on Survival of Under-five Children: Evidence from Matlab, Bangladesh. *Asian Population Studies*, 9(3): 1-15.

Ronsmans, C. Chowdhury, M. Dasgupta, S. Ahmed, A & Koblinsky, M. 2010. Effect of parent's death on child survival in rural Bangladesh: a cohort study. *The Lancet*, 375: 2024-2031.

Statistics South Africa, 2006. Mortality and Causes of death in South Africa. 2007. Findings from Death Notifications, Pretoria, South Africa.

Sartorius, B. Sartorius, K. Chirwa, T. & Fonn, S. 2011. Infant mortality in South Africa – distribution, associations and policy implications, 2007: an ecological spatial analysis. *International Journal of Health Geographics*, 10:61.

Syamala, T. 2004. Relationship between socio demographic factors and child survival: Evidence from Goa, India. *Journal of Human Ecology*, 16(2): 141 – 145.

Taffa, N. & Obare, F. 2004. Pregnancy and child health outcomes among adolescents in Ethiopia. *Ethiopian Journal of Health Development*, 18(2): 91-95.

Therneau, T. & Grambsch, P. 2000. Modeling Survival data: Extending the Cox Model. New York: Springer-Verlag.INC.

Uddin, J. Hossain, Z. & Ullah, M. 2009. Child Mortality In a Developing Country: A Statistical Analysis. *Journal of Applied Quantitative Methods*, 4(3): 270-283.

United Nations. 2011. Millennium Development Goals Report. New York, USA.

United Nations Children's Fund. 2004. Low Birth weight. New York, USA.

United Nations Children's Fund. 2008. The State of the World's children. New York, USA.

United Nations Children's Fund. 2009. Maternal and Newborn Health: A Global Challenge. New York, USA.

United Nations Children's Fund. 2012. Levels and Trends in Child Mortality Report. New York, USA.

United Nations Children's Fund. 2013. Committing to Child Survival: A Promise Renewed: Progress Report 2013. New York, USA.

United Nations Children's Fund and World Health Organization. 2004. Low Birth weight: Country, regional and global estimates, UNICEF, New York, USA.

Van Den Broeck, J. Eeckles, R. & Massa, G. 1996. Maternal Determinants of Child Survival in a Rural African Community. *International Journal of Epidemiology 1996*, 25: 998 - 1004

Westergren, A. Karlsson, S. Andersson, P. Ohlsson, O. & Hallberg, I. 2001. Eating difficulties, need for assisted eating, nutritional status and pressure ulcers in patients admitted for stroke rehabilitation. *Journal of Clinical Nursing*; 10: 257-269.

Worku, Z. 2009. Factors that affect under-five mortality among South African children: Analysis of the South African Demographic and Health Survey Dataset, Proceedings of the World Congress on Engineering and Computer Science, WCECS, San Francisco, USA.

Appendix

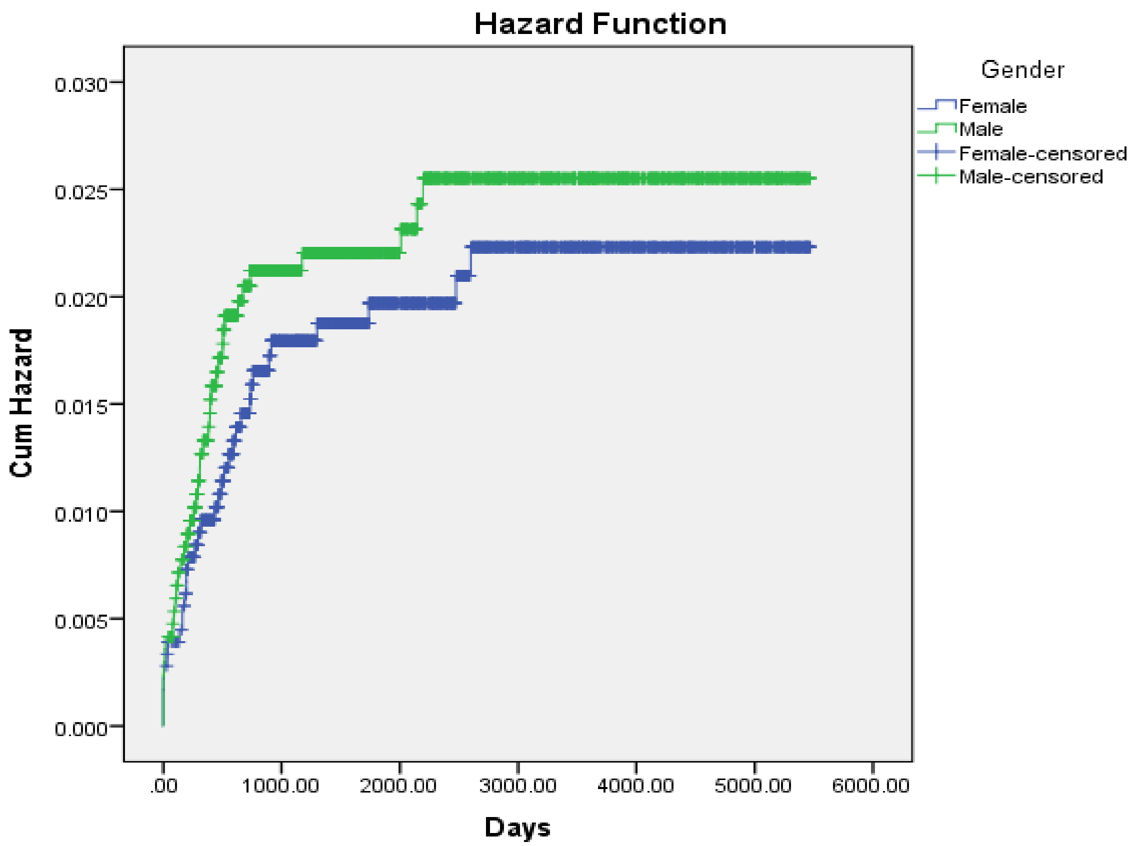

Figure A1: Hazard function against gender for children of Dikgale HDSS

Table A: Overall summary of children in Dikgale HDSS by the mother's survival status

Gender	Total N	N of Events	Censored	
			N	Percent
Female	3342	59	3283	98.2%
Male	144	13	131	91.0%
Overall	3486	72	3414	97.9%

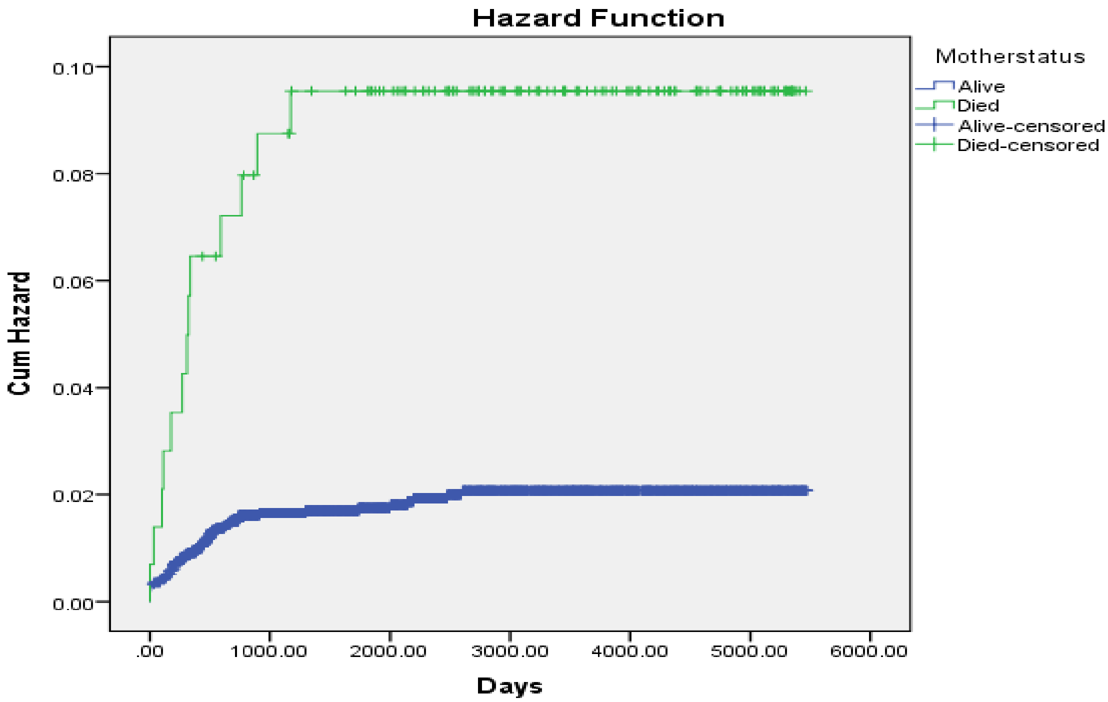

Figure A2: Hazard function against the mother's survival status for children of Dikgale HDSS

Table B: Overall summary of children in Dikgale HDSS by child birth weight

Birth weight category	Total N	N of Events	Censored	
			N	Percent
Below 2.5 kg	642	30	612	95.3%
Greater than or equal to 2.5 kg	1850	37	1813	98.0%
Overall	2492	67	2425	97.3%

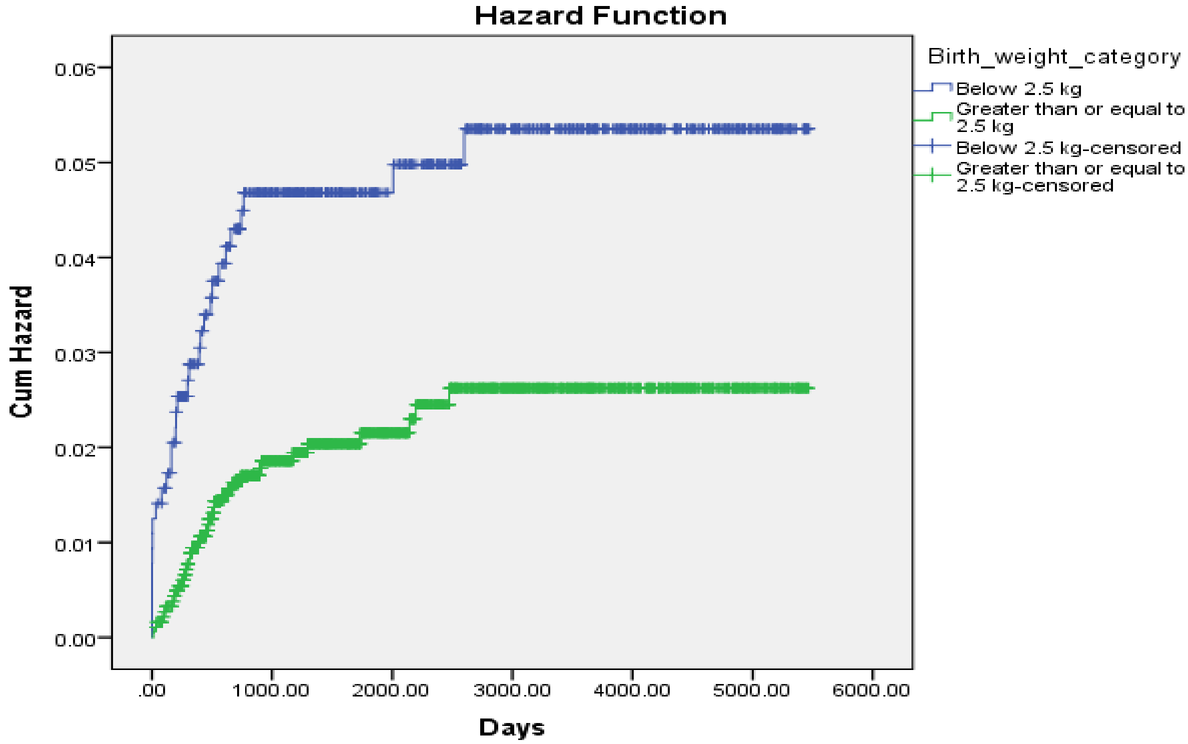

Figure A3: Hazard function against by child birth weight

Table C: Overall summary of children in Dikgale HDSS by breast feeding

Breast fed	Total N	N of Events	Censored	
			N	Percent
No	136	7	129	94.9%
Yes	2362	61	2301	97.4%
Overall	2498	68	2430	97.3%

Table D: Means for Survival Time by child breast feeding

Breast fed	Mean[a]			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No	5146.324	106.936	4936.730	5355.918
Yes	5320.200	19.081	5282.801	5357.598
Overall	5312.547	18.971	5275.363	5349.731

Table E: Test to compare survival time and child breast feeding

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	3.285	1	.070
Breslow (Generalized Wilcoxon)	3.928	1	.047
Tarone-Ware	3.666	1	.056

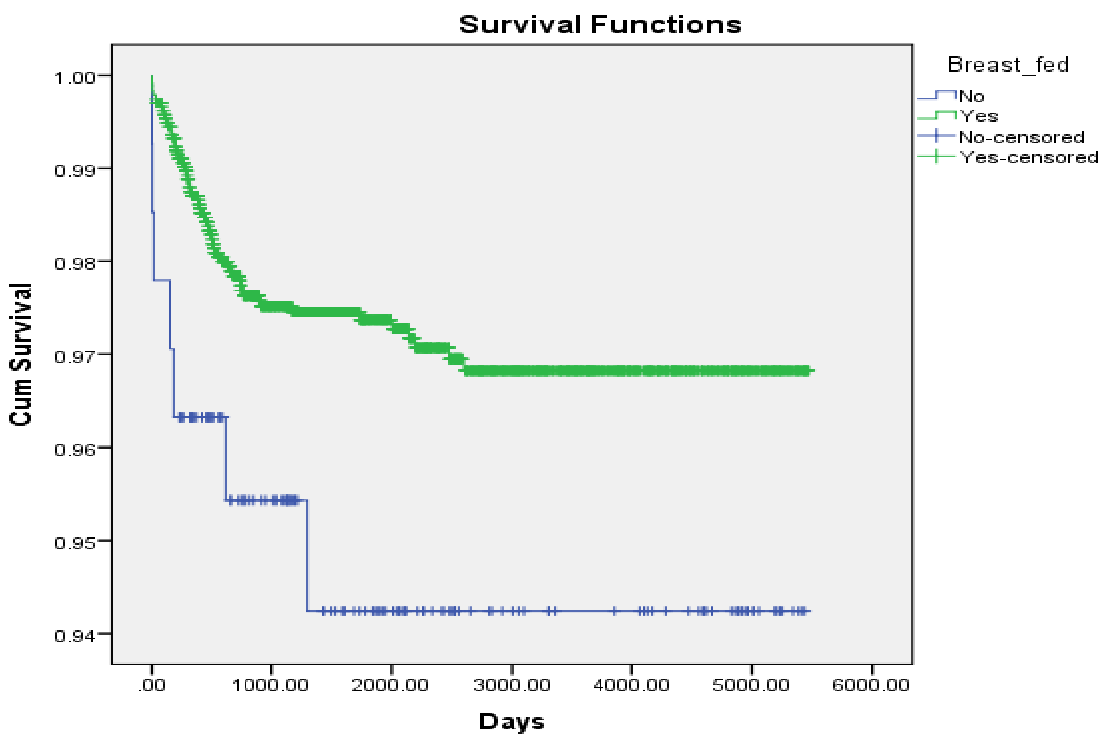

Figure A4: Survival probabilities for Dikgale HDSS children by breastfeeding status

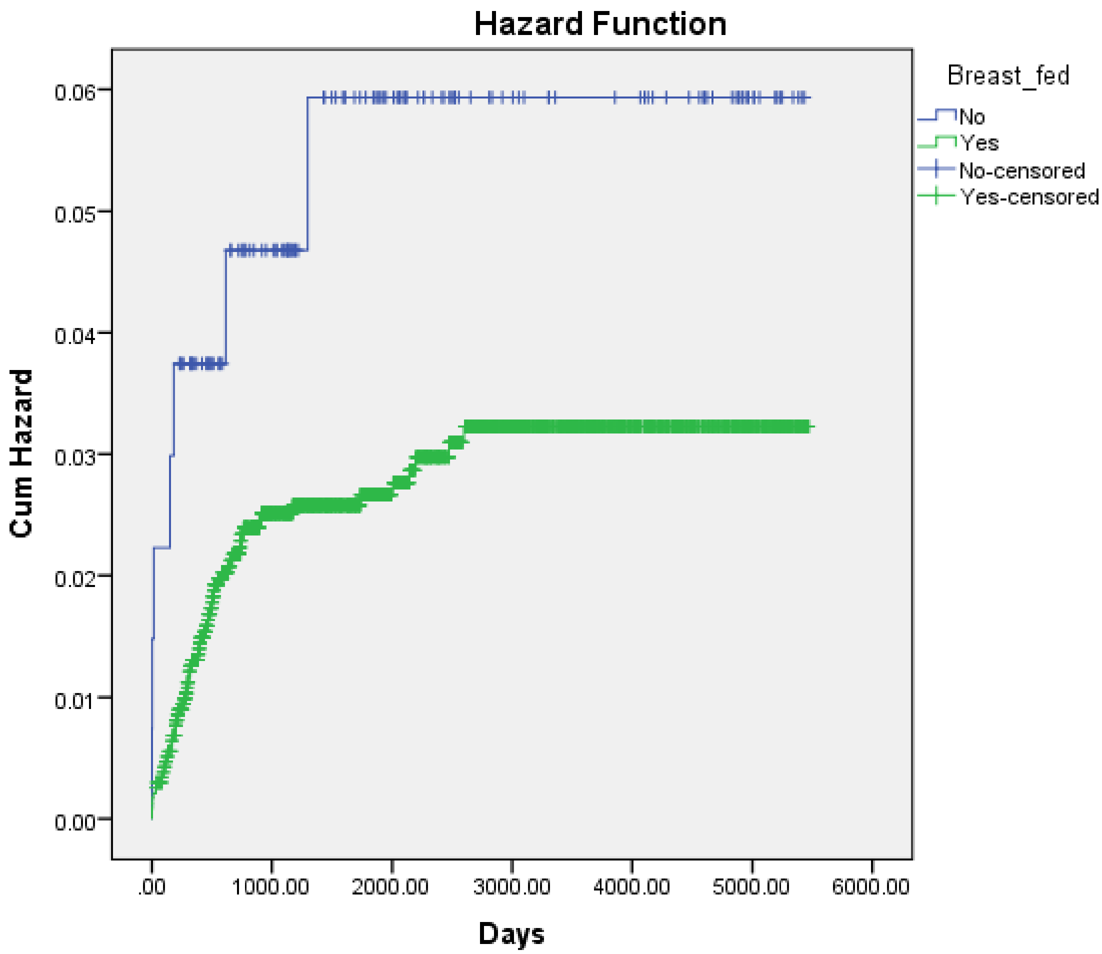

Figure A5: Hazard function against by breastfeeding status

Table F: Overall summary of children by the mother's attendance of ANC clinic

ANC Clinic	Total N	N of Events	Censored	
			N	Percent
No	63	4	59	93.7%
Yes	2436	66	2370	97.3%
Overall	2499	70	2429	97.2%

Table G: Means for Survival Time of children by the mother's attendance of ANC clinic

ANC Clinic	Mean[a]			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No	5112.526	165.989	4787.186	5437.865
Yes	5312.742	19.244	5275.025	5350.460
Overall	5308.216	19.202	5270.579	5345.853

Table H: Test to compare survival time of children by the mother's attendance of ANC clinic

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	2.743	1	.098
Breslow (Generalized Wilcoxon)	3.627	1	.057
Tarone-Ware	3.265	1	.071

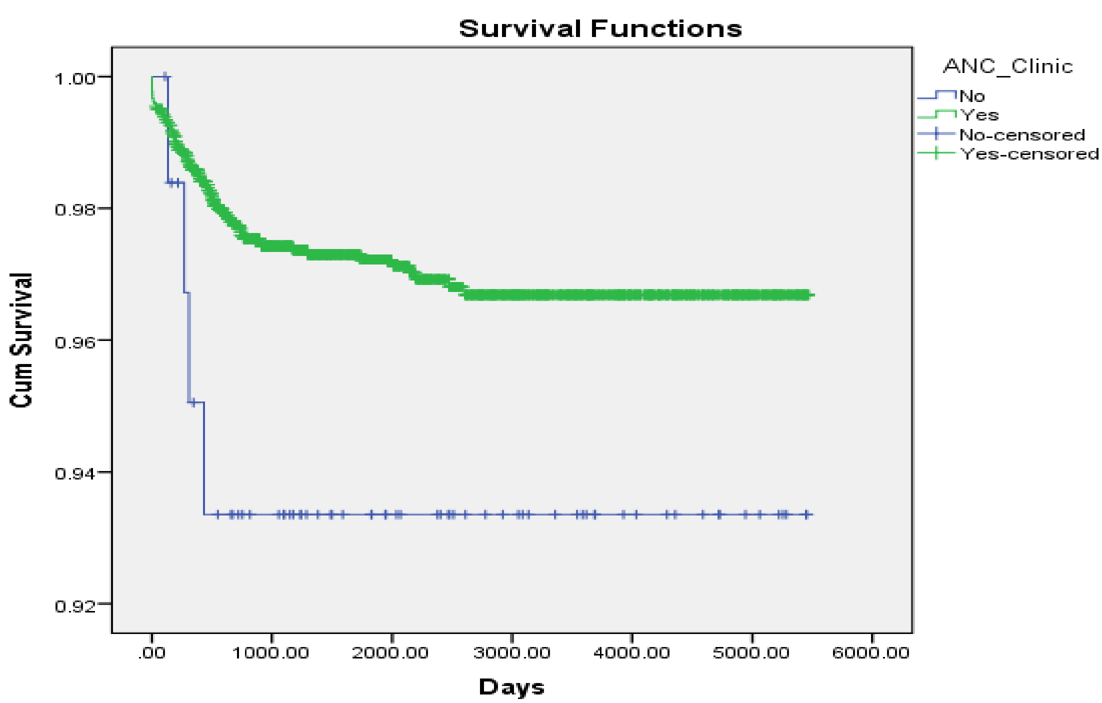

Figure A6: Survival probabilities for Dikgale HDSS children by the mother's attendance of ANC clinic

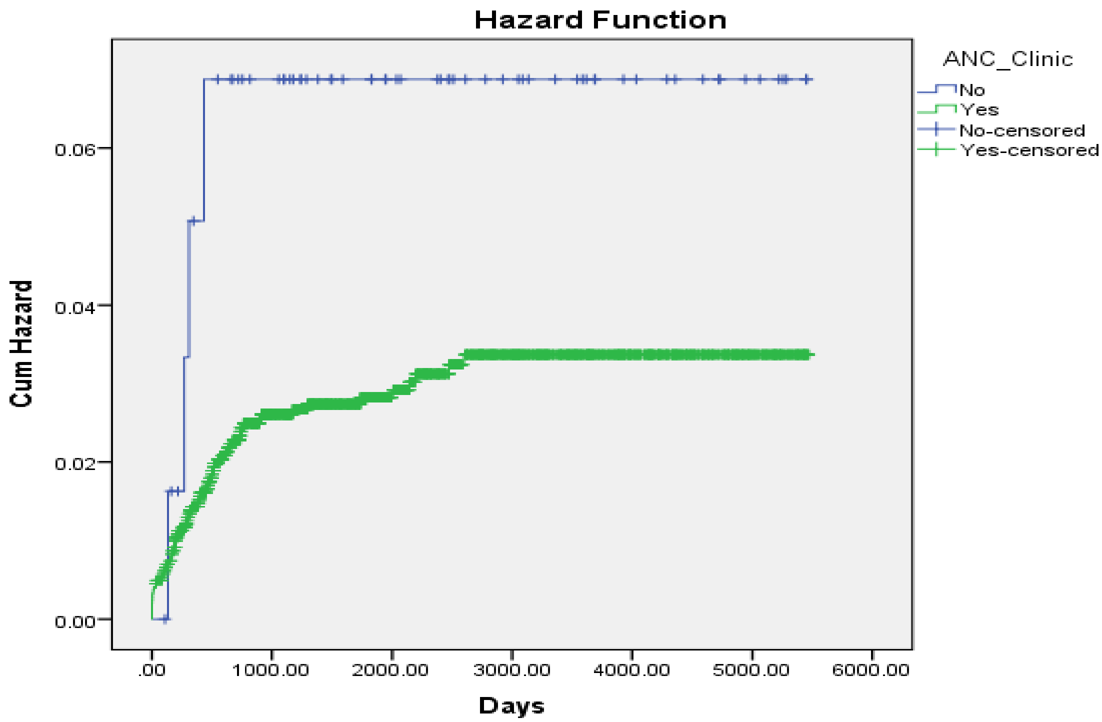

Figure A7: Hazard function for children by the mother's attendance of ANC clinic

Table I: Overall summary of children by the place of delivery

Delivery place	Total N	N of Events	Censored	
			N	Percent
Home	221	9	212	95.9%
Health Facility	2248	61	2187	97.3%
Overall	2469	70	2399	97.2%

Table J: Means for survival time of children by the place of delivery

Delivery place	Mean[a]			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Home	5250.416	71.675	5109.933	5390.899
Health Facility	5309.346	20.227	5269.702	5348.991
Overall	5306.162	19.443	5268.054	5344.269

Table K: Test to compare survival time of children by the place of delivery

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.913	1	.339
Breslow (Generalized Wilcoxon)	1.498	1	.221
Tarone-Ware	1.211	1	.271

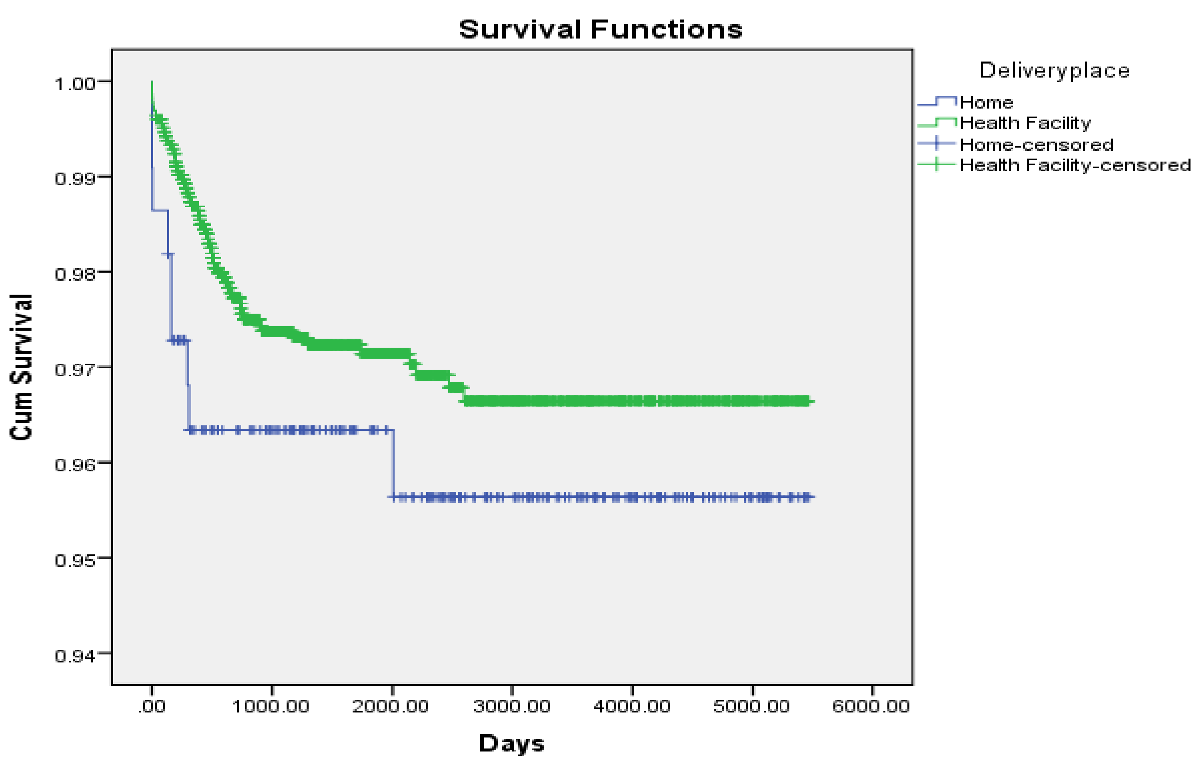

Figure A8: Survival probabilities for Dikgale HDSS children by the place of delivery

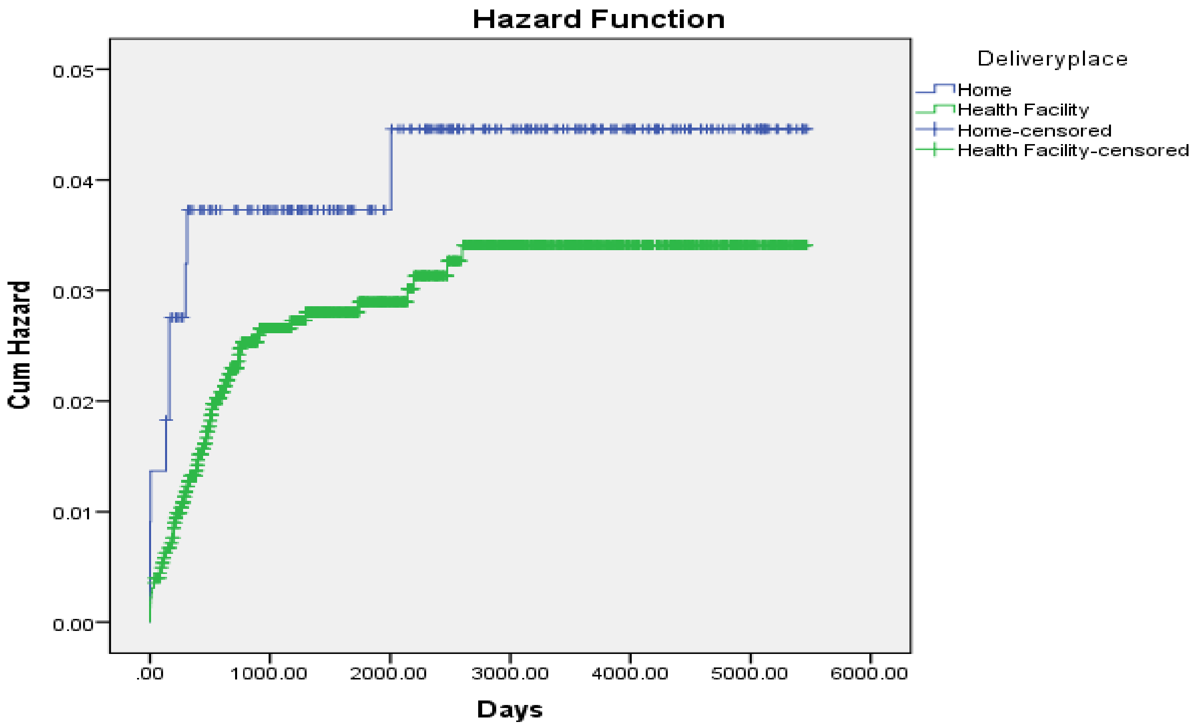

Figure A9: Hazard function of children by the place of delivery

Table L: Overall summary of children by the delivery attendant

Delivery attendant	Total N	N of Events	Censored	
			N	Percent
Others	1198	11	1187	99.1%
Health Professional	2288	61	2227	97.3%
Overall	3486	72	3414	97.9%

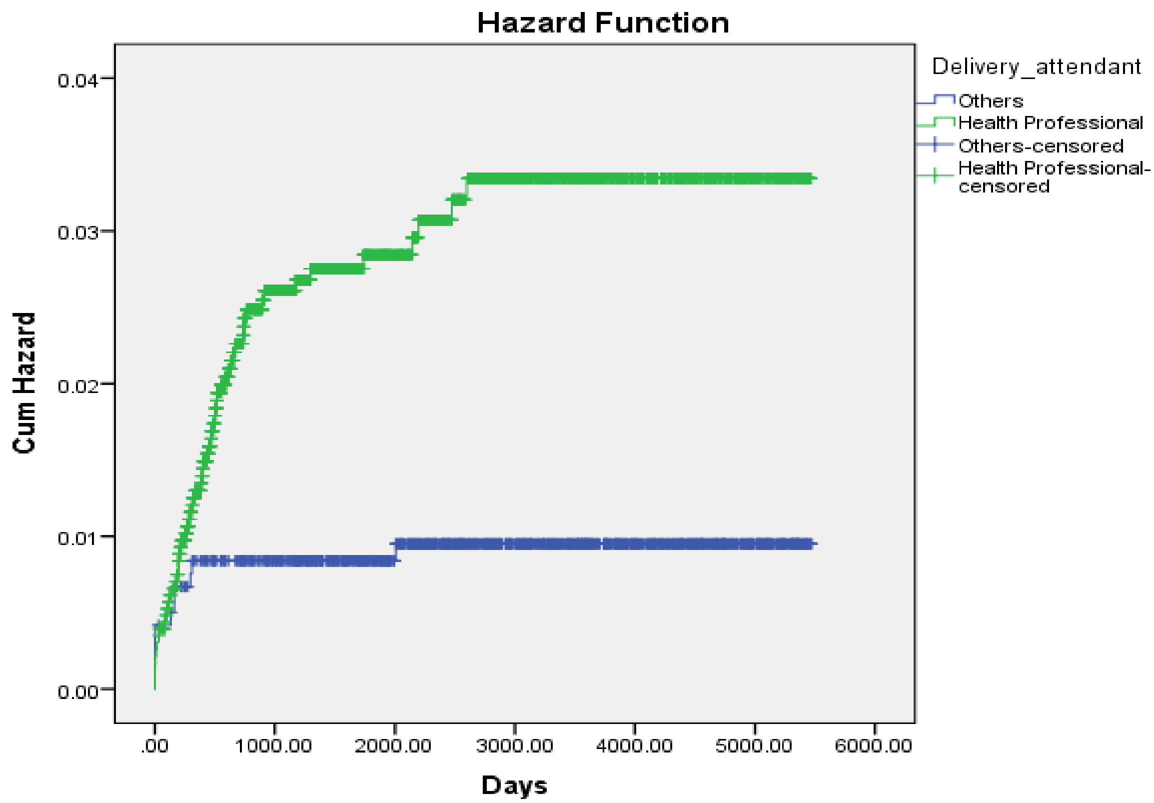

Figure A10: Hazard function of children by the delivery attendant

Table M: Overall summary of children in Dikgale HDSS by the mother's educational level

Mother education	Total N	N of Events	Censored	
			N	Percent
No education	23	2	21	91.3%
Primary education	177	9	168	94.9%
Secondary education	1492	37	1455	97.5%
Higher education	271	2	269	99.3%
Overall	1963	50	1913	97.5%

Table N: Means for Survival Time for children by the mother's educational level

Mother's educational level	Mean[a]			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Never Married	5295.216	24.015	5248.147	5342.285
Married	5394.956	31.998	5332.240	5457.672
Widow/Separated	5398.154	45.571	5308.836	5487.472
Divorced	5310.692	73.604	5166.427	5454.957
Overall	5323.412	18.151	5287.837	5358.988

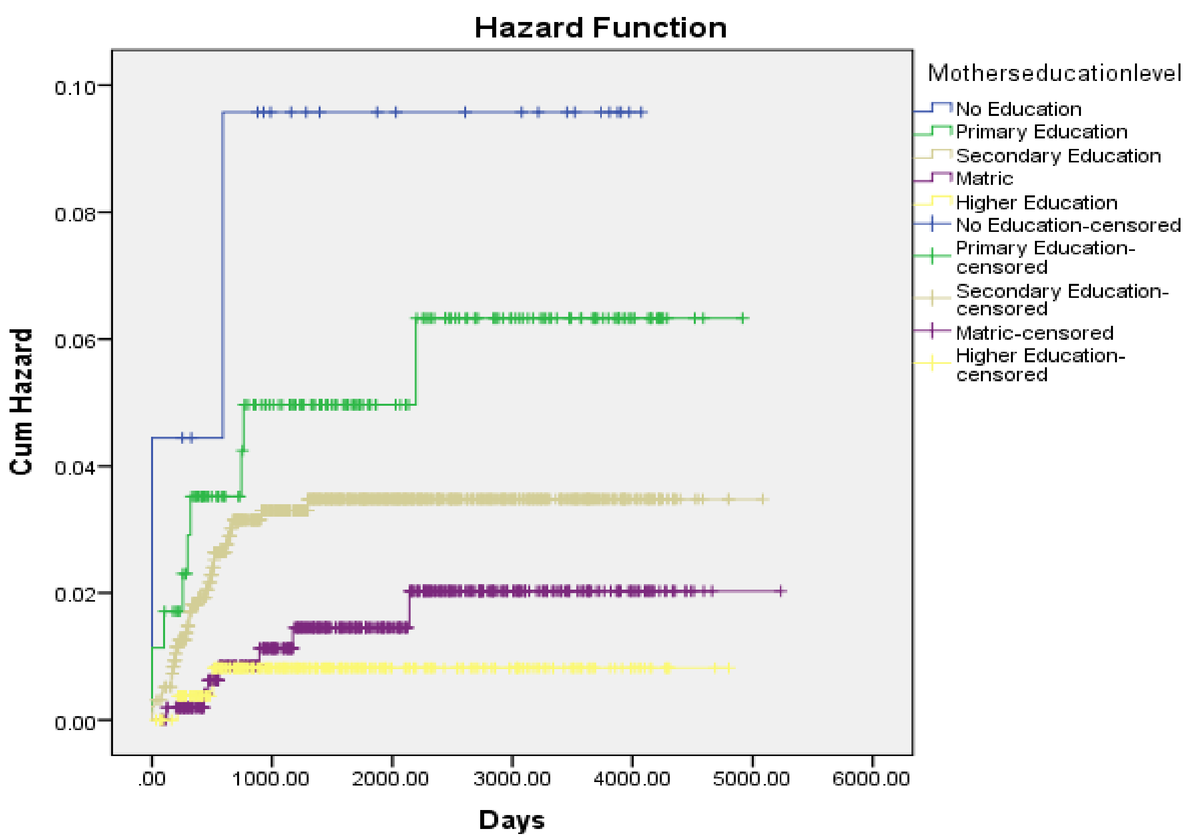

Figure A11: Hazard function for children by the mother's educational level

Table O: Overall summary of children in Dikgale HDSS by mother's marital status

Mother's Marital status	Total N	N of Events	Censored	
			N	Percent
Never Married	1706	51	1655	97.0%
Married	368	5	363	98.6%
Widow/Separated	164	2	162	98.8%
Divorced	140	4	136	97.1%
Overall	2378	62	2316	97.4%

Table P: Test to compare survival time for children by the mother's marital status

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	5.437	3	.142
Breslow (Generalized Wilcoxon)	4.327	3	.228
Tarone-Ware	4.829	3	.185

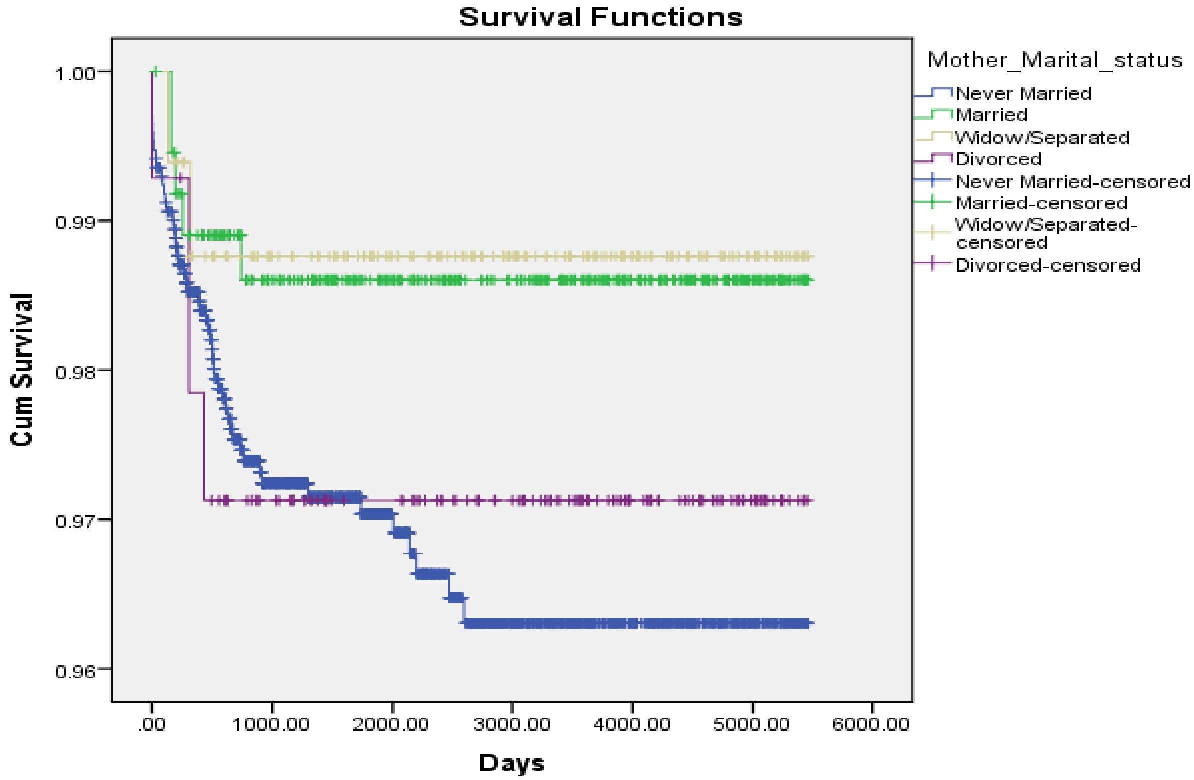

Figure A12: Survival probabilities for Dikgale HDSS children by the mother's marital status

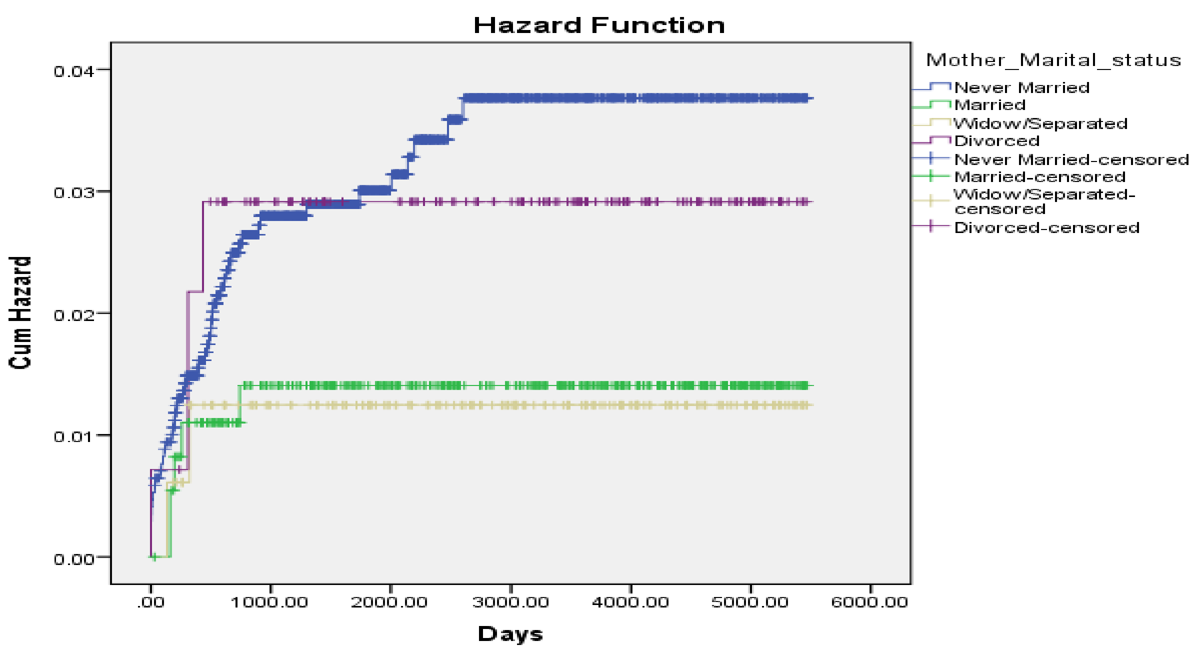

Figure A13: Hazard function for children by the mother's marital status

Table Q: Overall summary of children HDSS by the mother's age category

Mother's age category	Total N	N of Events	Censored	
			N	Percent
Below 20 years	650	19	631	97.1%
20 years and above	2690	53	2637	98.0%
Overall	3340	72	3268	97.8%

Table R: Means for survival time for children by the mother's age category

Mother's age category	Mean[a]			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Below 20 years	5278.100	35.018	5209.466	5346.734
20 years and above	5360.241	15.101	5330.643	5389.838
Overall	5351.227	13.985	5323.816	5378.638

Table S: Test to compare survival time for children by the mother's age category

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	1.870	1	.172
Breslow (Generalized Wilcoxon)	2.235	1	.135
Tarone-Ware	2.099	1	.147

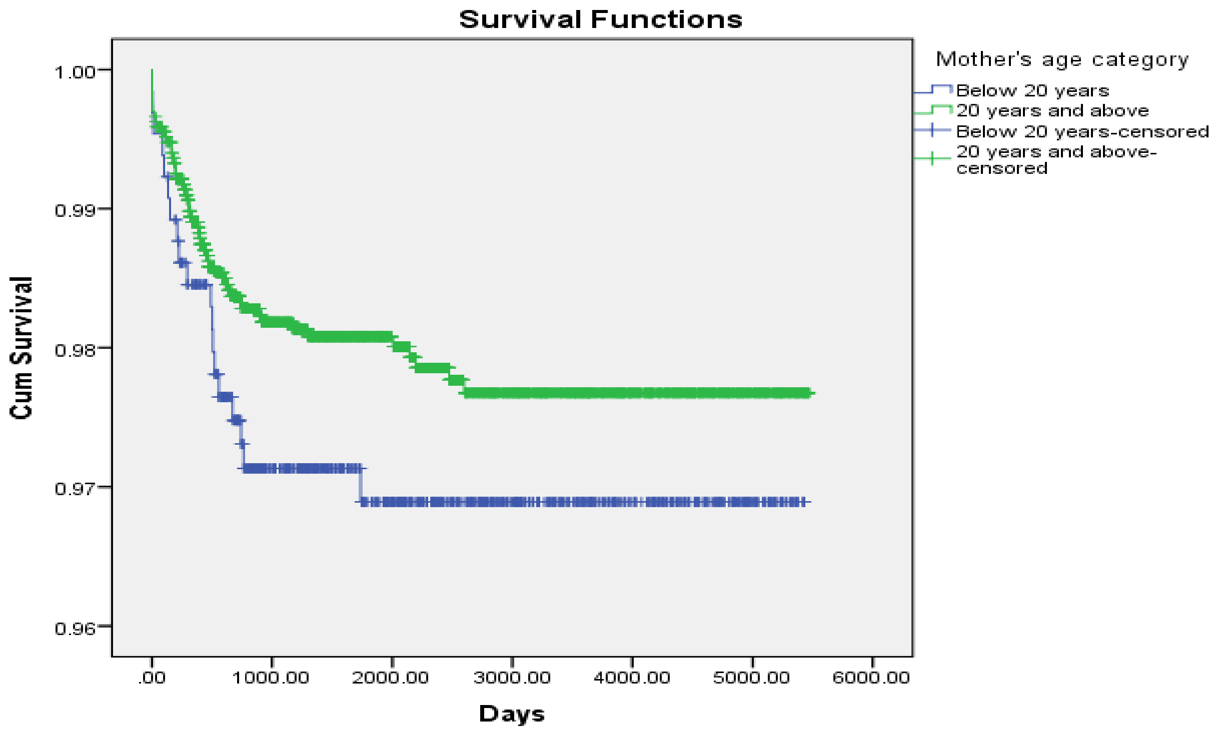

Figure A14: Survival curve for children by the mother's age category

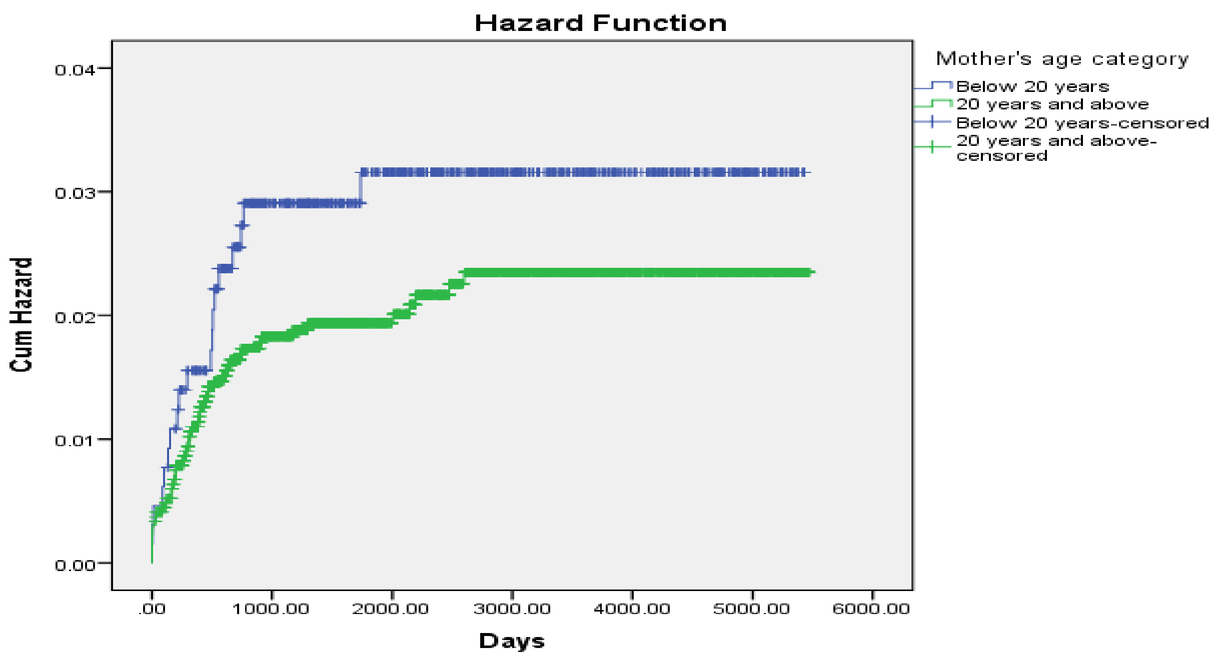

Figure A15: Hazard function for children by the mother's age category

Table T: Odds ratio for child gender

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Event (Alive up to age five / Died)	1.167	.717	1.900
For cohort Gender = Female	1.081	.838	1.393
For cohort Gender = Male	.926	.733	1.170
N of Valid Cases	2049		