Modeling Determinants of Time-To-Death in Premature Infants Admitted to Neonatal Intensive Care Unit in Jimma University Specialized Hospital.

By:

Million Wesenu

A Thesis Submitted to the Department of Statistics, College of Natural Sciences, Jimma University in the Partial Fulfillment of the Requirements for the Degree of Master of Science (M.Sc.) in Biostatistics.

June.2016

Jimma, Ethiopia

Modeling Determinants of Time-To-Death in Premature Infants Admitted to Neonatal Intensive Care Unit in Jimma University Specialized Hospital.

MSc Thesis

By:

Million Wesenu

Advisor: K. Sudhir (PhD.)

Co-advisor: Tafere Tilahun (M.Sc.)

June, 2016

Jimma, Ethiopia

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Million Wesenu
Date: _____
Signature: _____

JIMMA UNIVERSITY, JIMMA, ETHIOPIA

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES
JIMMA UNIVERSITY

As thesis research advisors, we her by certify that we have read the thesis prepared by Million Wesenu under our guidance, which is entitled "**Modeling Determinants of Time-to Death in Premature Infants Admitted to Neonatal Intensive Care Unit in Jimma University Specialized Hospital**", in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Dr. K. Sudhir (PhD)	_____	_____
Advisor	Signature	Date
Tafere Tilahun (M.Sc.)	_____	_____
Co-Advisor	Signature	Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

ACKNOWLEDGMENT

First and foremost, I am indebted to the Almighty God because of whose full mercy and grace I could complete my study.

During my graduate studies in Jimma University several persons and institutions collaborated directly and indirectly to my research. That is why I wish to dedicate this section to recognize their support.

My special gratitude goes to my advisor Dr. K. Sudhir, for his immense and invaluable advice and guidance that contributed to the successful realization of this study. I would like to express my sincere appreciation to my co-advisor Mr.Tafere Tilahun for his precious suggestions and comments during the entire time of my thesis work. It is a great pleasure working with him. I can only hope that his cooperation will keep on going in the future. I am so pleased to say thank you to my instructors Mr. Belay Birlie (PhD candidate) and Mr. Geremaw Mulata, Mr. Gashu Gadisa and my brother Tegene Wesenu. I got inspiration and good support from them during all my studies. I would like to thank you my beloved classmates for their huge support and true love in my two years of study.

My sincere thanks also go to all staff member of department of statistics of Jimma University for their unreserved knowledge sharing and cooperation. I would like to thank the host Jimma University and my sponsor Haromaya University for providing me to attend my training and Jimma University Specialized Hospital and staff members Jimma Hospital to undertake this study with their cooperation and permission in using the data with special thanks for Dr. Diriba Fufa, w/ro Yeshi Muluneh and Ato Mohammed Abbaraya for their willingness to help me.

Last but not least, I would like to thank my parents who have invested all their life to hold up me. My parents, I would like to thank you vastly for your sweet words at every time that gave me a long power and anticipate.

DEDICATION

I dedicate this work to my dear parents Mr. Wesenu Demissie and Mis. Zewditu Demissie for making me who I am today, for their support and for teaching me the value of education. To all my sisters and brothers for their daily encouragement and inspiration!!!

ABSTRACT

The purpose of this study was to identify factors leading to mortality and statistically modeling the survival of premature infants. A sample of 490 preterm was taken from a hospital record at Jimma university specialized hospital from January 2013 to December 2015. To estimate, compare and model the survival time as well as examine the association between the survival time with different demographic, health and risk behavior variables the log rank and generalized Wilcoxon test, Cox proportional hazard model and the parametric regression model were applied. The result from log rank and generalized Wilcoxon test revealed that the survival probability of premature infants is statistically Significance difference in experiencing the death event among groups classified by prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, gestational age, respiratory distress syndrome and initial temperature. The mean survival time of premature infants' was 21.23 days. The cox proportional hazad and parametric log-logistic regression(with odds ratio) model showed that prenatal Asphyxia (OR=2.479,P-value=0.01), hyaline membrane disease (OR=2.636,p-value=0.0001), Sepsis (OR=2.072,p-value=0.005), Jaundice (OR=2.737,p-value=0.000), temperature (OR=0.811,p-value=0.018), respiratory distress syndrome (OR=3.287,p-value=0.000), Gestational age of [30-32) (OR=0.336,p-value=0.017) and gestational age of (32-34] (OR=0.241,p-value=0.002) when gestational age of (26-28] as a references significantly contribute to a shorter survival time of premature infants'.In conclusion, the findings of this study shows that prenatal Asphyxia, hyaline membrane disease, sepsis, Jaundice, temperature, gestational age and respiratory distress syndrome as the most determinant and statistically associated with time to death of premature infants admitted to NICU. It is therefore recommended that people ought to be cognizant on the burden of these risk factors and well informed about the prematurity.

***Key Words**: premature infant, time to death, Cox proportional hazards model, log-logistic regression model.*

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AIC	Akaike Information Criterion
ANC	Antenatal Care Visit
BIC	Bayesian Information Criterion
CI	Confidence interval
CSA	Central Statistical Agency
DHS	Demographic and Health Survey
EDHS	Ethiopian Demographic and Health Survey
FMOH	Federal Ministry of Health
HIR	High Infant Risk
HR	Hazard Ratio
JUSH	Jimma University Specialized Hospital
LBW	Low Birth Weight
LRT	Likelihood Ratio Test
MDG	Millennium Development Goal
NICU	Neonatal Intensive Care Unit
NMR	Neonatal Mortality Rate
OR	Odds Ratio
PDF	Probability Density Function
PH	Proportional Hazard
PO	Proportional Odds
SVD	Spontaneous Vertex Delivery
UNICEF	United Nation Children Fund
WHO	World Health Organization

1. INTRODUCTION

1.1. **Background of the study**

Preterm birth is the term used to define births that occur before 37 completed weeks or 259 days of gestation. Premature infants come early into the world. They are born fragile, small and weighing less than full term infants. Many of the babies who survive face greater risks of significant health problems and disability throughout their lives (i.e, learning disabilities, visual and hearing problems, chronic lung disease and other long-term diseases) which translate into significant increased costs to healthcare, the economy and the broader society (WHO, 2012). It can be further sub categorized as late preterm delivery from 34 to 36, moderately preterm from 32 to 34, very preterm less than 32, and extremely preterm less than 28 weeks of gestation (Offiah *et al*., 2012). Premature is one of the major causes of infants' death which is not an acute disease and compared to term infants experience more difficulty with feeding, blood glucose control, jaundice, temperature instability, respiratory distress and sepsis either singly or in combination (Engle *et al*., 2007).

Premature birth is the major cause of prenatal morbidity and mortality all over the world (Rehana P., 2006). Globally, an estimated 13 million infants are born before 37 completed weeks of gestation annually. Rates are generally highest in low and middle income countries and increasing in some middle and high income countries (Lawn *et al*., 2010). More than 1 in 10 of the world's babies born in 2010 were born prematurely, making an estimated 15 million preterm births, of which more than 1 million died as a result of their prematurity (Howson *et al.*, 2013) . Preterm are now the second leading cause of death in children less than 5 years and the single most important cause of death in the critical first month of life. Preterm birth accounts for 3.1% of all Disability Adjusted Life Years (DALYs) in the Global Burden of Disease, more than for HIV and malaria (WHO and Offiah *et al.*, 2012). Also, deaths constitute 28% of the 4 million annual new born deaths with 99% of these deaths occurring in developing countries (FMOH, 2011). Morbidity, mortality and prolonged hospital stay of preterm babies result in significant cost to the health sector, parents and the society (Mohammad *et al*., 2009).

In Ethiopia, According to report of United Nations of children fund (UNICEF, 2012), one of the main causes of neonatal death is preterm birth accounts for 23% of all other causes of neonatal death. Given the frequency of preterm birth worldwide, it is likely that most people will experience the tragedy of preterm birth at some point in their lives, either in family members or indirectly through friends. Also, Ethiopian Demographic and Health Survey in 2011, high rate of neonatal mortality (37 deaths per 1,000 live births) is reported and preterm birth is believed to be a major and direct cause of neonatal mortality. Infant and under-five mortality rates remain very high in Ethiopia. One in every 13 babies born in Ethiopia did not survive to celebrate its first birth day and one in every eight children died before its fifth birth day by Belaynew *et al* . (2015).

Preterm birth has multiple factors whose solutions will not come through a single discovery but rather from an array of discoveries addressing multiple biological, clinical, and social behavioral risk factors. Causal factors linked to preterm birth include medical conditions of the mother or fetus, genetic influences, environmental exposure, infertility treatments, behavioral and socio-economic factors as well as iatrogenic prematurity (Pennell *et al.*, 2007). Approximately 45% to 50% of preterm births are idiopathic, 30% are related to preterm rupture of membranes and another 15% to 20% result from medically indicated or elective preterm deliveries (Goldenberg *et al*., 2008). The evolution of neonatal intensive care is one of the recent advances that ensure survival of the preterm neonate but sadly this is not readily available in most developing countries (Aalen, 1994). This is not surprising as neonatal intensive care is expensive because of the cost of sophisticated equipment, need for constant power supply, constant use of laboratory facilities and high staff to patient ratio (Lancaster,1979).

It is currently known that the study of risk factors for infant mortality is very important, as, particularly in the newborn, it can be considered one of the best quality indicators for health care, as well as an indicator for population social and economic welfare (Risso *et al*., 2010). Neonatal death rate was overall 27.4%, which was significantly higher in gestational age subgroup of less than 28 weeks compared with other gestational age subgroups. The most prevalent etiologies of neonatal death were respiratory distress syndrome (73.8%), congenital abnormalities (13.8%) and sepsis (5.4%), respectively. Preeclampsia and history of multiple

pregnancies were more prevalent in non-survived neonates. According to multivariable regression modeling, low gestational age, low birth weight, low Apgar scores, need for intensive supports, history of disease in mother, occurrence of pneumothorax, multiple gestation and preeclampsia could all strongly predict occurrence of death in premature infants (Behzan *et al*. ,2015). Over 80% of all cases of death in the world are result of neonatology premature birth and the complications which are result of preterm birth are significant socioeconomic problem and the hyaline membrane disease (HMD) is in the leading place, in the field of prenatal medicine and neonatology are applied preventive treatments in order to reduce the risk of death of preterm newborns (Anna *et al.*, 2016).

A cross sectional study conducted in north-west Ethiopia showed that 11.6% from the total 422 mothers gave a preterm birth. Presence of chronic illness (AOR=4.5; 95% CI (2, 10.2)), problem in current pregnancy (AOR=2.9; 95% CI: (1.3, 6.7)), premature rupture of membrane (AOR=6.2; 95% CI: (2.7, 14)) and has antenatal follow up (AOR=0.24; 95% CI: (0.09, 0.6)) were found to be significantly associated with preterm birth on the multivariate logistic regression carried out by (Tigist *et al.,* 2013). The retrospective study on the association between dependent (preterm death) and independent variables were assessed by using binary logistic regression and variables having significant association with binary logistic regression were entered to multiple logistic regressions for checking statistically significant association of the variables and showed Ante Natal Care visit, Tetanus vaccination of mother, mode of delivery are significant(Belyanew *et al.*,2015).

Here more of the model do not deal with the factors associated to time to event data, since time to event data has censored observation. Also the modeling is biased when many variables are included in the model to identify the significant factors upon which the hazard function depends. The key solution for such data to analysis is using different variable selection method in survival model which is appropriate for modeling determinant of time to death in premature infants.

This study was intended to model the survival time of premature death of infants admitted to neonatal intensive care units (NICU) in Jimma University specialized hospital (JUSH) using survival Analysis framework. The primary variable in survival analysis is survival time, time to death of infants

admitted to NICU for this study. The term "survival time" is used loosely for the time period from a starting time point to the occurrence of a certain event. despite, survival models have a long history in the biostatistician and medical literature (Cox *et al*., 1984), there are very few literatures regarding the use of survival analysis in modeling the survival time of premature death compared to other statistical models, such as, Logistic regression and multivariate logistic regression.

For this study, a non-parametric such as Kaplan- Meier analysis, log-rank and generalized Wilcoxon tests, cox proportional hazard and parametric survival model were used to estimate, compared and modeling the determinants of time to death in premature infants admitted to NICU. The non-parametric methods work well for homogeneous samples and test whether or not survival probability differences between groups of covariates.

Cox regression model also used to estimate and test the significance of the parameters when the baseline distribution is unspecified and exponential, Weibull and log-logistic baseline distribution used in parametric cox regression to fit an appropriate model and to investigate the relationship between different potential covariates with response of time to event. The method of variable selection techniques in cox regression analysis, such as the step wise selection or stepwise deletion which is popular was used. Finally, Parameter estimates in the model are obtained by maximizing the partial likelihood and maximum likelihood and model were compared based on BIC, AIC and R^2.

1.2. **Statement of the problem**

The birth of a preterm infant results in significant health consequences to the infant and emotional and economic costs for families and communities. Even if premature birth is not an acute disease, it is one of the major causes of infants' death and it continues to be significant public health burden. The average cost of medical care for a premature and low birth-weight baby for the first year of life is high in developing country like Ethiopia. These high medical expenses could burden the parents and family. In other case insurance coverage may provide additional benefits in the health plan in order to cover the medical cost incurred during the first year of preemies' life and hence it helps to reduce the burden of the family.

The first 4 weeks of life carries one of the highest risks of death of any 4 weeks period in the human lifespan. Of the 130 million babies born every year, about 4 million die in the first 4 weeks of life in the wide world. Reducing Neonatal morbidity and mortality are now a major focus of child health strategies (lawn *et al.*, 2001). The study conducted in Fawzy Moaz Hospital in Egypt, were reported that 48% death of preterm infants admitted to NICU (Fakher *et al*., 2005). According to Ethiopian Demographic and Health Survey in 2011, in Ethiopia, high rate of neonatal mortality (37 deaths per 1,000 live births)is reported and preterm birth is believe to be a major and direct cause of neonatal mortality (EDHS, 2011) . A total of 225 neonates were admitted during the study period of Jan 2012 to Dec 2012 in JUSH. One of The main causes of admissions were found to be prematurity (30.7%) and the hospital neonatal mortality rate was 15.9%.The average length of hospital stay was 9.5 days with SD of 8.2 days using binary logistic regression analysis (Habtamu *et al*.,2013).

Parents' awareness of premature birth and its consequences is still at a low level. Survival rates pattern or trend over the years is very important to give a sign to the parents whether their premature baby's life is at a risky level or otherwise (WHO, 2012). Even if this problem is a serious in developing country like Ethiopia, many studies done using multivariate logistic regression and binary logistic regression which is not appropriate for time to event data to identify risk factors when the data have censored/incomplete information, though the researcher wants to aspires an appropriate survival model for time-to-death in preterm of infants in neonatal intensive care units regarding the risk factor that aggravate the death of premature infants.

Even though, several studies on death of premature infants' used different statistical models to explore its determinant factor, its time to death still needs to be studied. This case does not hold in other situations, which are not common for all death events. This study focuses on modeling the determinant of time-to-death of preterm using various non-parametric, cox regression and parametric models. It was investigated the major risk factors of preterm death which will help to guide health professionals and health policy makers to identify indicators for monitoring preterm birth strategy and applying necessary preventive and appropriate measures to decrease preterm birth. May ultimately it will helps to reduce infant mortality rate;

also it will help to fill the research gaps in the study area and as base line information for other areas of the country.

The survival time for premature infants depends on different factors, such as socio-demographic, health conditions and laboratory factors. Therefore, this study is motivated to identify the major risk factors associated with survival of infants with prematurity which is not acute disease. The crucial questions that the study answered were:

i. What are the significant factors to determine the risk factors for preterm death in Jimma University specialized hospital?
ii. Which fitted model is statistically plausible\good?
iii. Which group of gestational age has more effect on preterm death?

1.3. Objectives of the Study

1.3.1. General objectives

The main objective of this study is modeling the determinants of time-to-death in premature infants admitted to neonatal intensive care unit (NICU) in Jimma University Specialized Hospital.

1.3.2. Specific objectives

i. Identify significant factors that are associated with time-to- death in premature infants.
ii. Fit an appropriate statistical model for survival time to death in premature infants.
iii. Predict the survival time and compare the survival curves of time-to-death among different levels of factors.

1.4. Significance of the study

The result of this study will provide information on time- to- death of premature infants admitted to neonatal intensive care unit in Jimma University Specialized Hospital and its determinant factors. Specifically;

i. To provide information about the covariates or risk factors of time- to- death preterm infants' .
ii. Provides information to government and concerned bodies in setting policies and strategies.
iii. Use as a stepping stone for further studies related to time-to- death of premature infants.

2. LITERATURE REVIEW

2.1.Description of the preterm birth

Premature babies who are known as 'preemies' comes into the world earlier than full term babies. Full term babies born ranges 37 to 42 weeks after the mother's last menstrual period while premature babies born before 37 weeks of pregnancy. Babies who are born very early in pregnancy are extremely small and fragile (Azizah, 2009). A collaborative effort to assess factors affecting newborn survival at NICU results suggest a need for greater efforts to identify and reduce risk factors associated with premature death, and to adequately evaluate the medical care provided in NICUs (Maria *et al.,*1997). Almost all babies who are born preterm require extra medical and nursing care as newborns. In addition, those who are less than 37 weeks usually spend time in a neonatal intensive care unit for a few days or several weeks. During the course of his/her lifetime, it is estimated that each preterm low birth weight baby will use about $676,800 (1995 Canadian dollars) in health care. With the existing number of preterm low birth weight babies, the total lifetime health care costs are likely to exceed $8 billion dollars (Moutquin *et al.,* 1998).

Promising strategies for reducing morbidity and mortality associated with preterm birth involves promoting early detection and appropriate response to preterm. Prompt recognition of the signs and symptoms of preterm is essential if treatment with corticosteroids is to begin early enough to have an optimum effect. Antenatal treatment of the mother with one full-course of corticosteroids (two doses, 24 hours apart) is known to make a difference in neonatal morbidity and mortality for infants of 24-34 weeks gestation (National Institutes of Health, 1994).

2.2. Burden of premature infants

Preterm birth is a significant global burden with 15.1 million babies born before 37 weeks of pregnancy every year across the world, which represents one in ten babies. Of these, 790,400 are born extremely preterm, that is before 28 weeks of pregnancy are completed and 1.1 million babies die from preterm birth complications every year. Babies born too soon are between 6 and 26 times more likely to die during the first four weeks of their lives than babies

born at term (MDG, 2012). While Preterm birth complications account for a third (34%) of all the world's 2.9 million newborn deaths worldwide, this makes prematurity the leading direct cause of newborn mortality. Approximately one out of eight babies is born prematurely with 1305 premature babies are born every day in the United State .With advances in Obstetric practice which have led to increased medical surveillance to identify and prevent progression of maternal and fetal complications, the incidence of preterm deliveries have also increased (Engle *et al*., 2008). In 2005, the World Health Organization (WHO) estimated 9.6% of all births worldwide to be preterm {Beck *et al*., 2010).

One of the main factors contributing to the rising of premature birth is due to a large number of multiple births in recent years and birth weight. For birth weight, we shall categorized babies with Extremely Low Birth Weight (ELBW), Very Low Birth Weight (VLBW), Low Birth Weight (LBW) and Normal Birth Weight (NBW). ELBW babies are babies weighing less than and equal to 1000 grams while VLBW babies weighing greater than 1000 grams but less than and equal to 1500 grams. These two categories of preemies are also considered as High infants Risk (HlR) and therefore the survival of this group shall be observed predominantly. Next Birth weight group is LBW babies who are weight greater than 1500 grams but less than and equal to 2500 grams. Normal Birth Weight group consist of babies who weights are greater than 2500 grams. Study in University of California (2004) has shown the rate of Very Low Birth Weights (VLBW) increase mainly due to the increase in prematurely-born multiple gestations. Preterm birth is a major public health burden whose prevalence continues to rise. The rate of preterm birth in the U.S. is 12.7% (Hamilton *et al*.,2005),it is the single most common causes of perinatal mortality in Europe and North America.

2.2.1. Literature on the risk factor of premature infants

In Scotland, a retrospective cohort study showed that a short inter pregnancy interval (< 6 months) was an independent risk factor for extremely preterm birth (6.1%), moderately preterm birth (3.9%) and neonatal death (13.8%) (Gordon *et al*.,2003). Prematurity remains the most significant cause of neonatal morbidity and mortality. A comparative cross sectional study conducted in Qom Hospital showed that frequency of preterm delivery among live

births was 5.6%. Increasing maternal parity, short inter pregnancy interval, low socioeconomic state, emotional stress, lack of regular antenatal care, ante partum hemorrhage, had significant relationship with preterm labor (Gholamreza *et al*.,2011). Early neonatal death, which attributes to most perinatal deaths, is caused by preterm birth and low birth-weight (Behrmanet, 2004).

According to retrospective study of Azizah (2009), on survival analysis of premature babies in university Malaya medical center, showed on his study to determine the survival to discharge of preterm infants who were admitted to NICU in University Malaya Medical Center (UMMC). Survival analysis was done using a non-parametric approach called Kaplan-Meier to estimate survivor function. The event of interest is death of preterm infants during their six months stay in NICU and those who survive to discharge or lost to follow up are considered as right-censored observations. There are six potential influential factors were taken into account which are gender, gestational age, birth weights, ethnic groups, multiple births and mode of delivery. By using log-rank test the survival function of preterm infants in two or more groups can be compared. The hazard function model was developed by using Cox's Regression model which consists of significant factors and the baseline hazard element. Several methods have been used to test the adequacy of fitted model which ensure well functioning and the PH assumption is not violated.

Based on the analysis on causes of neonatal mortality conducted by Lawn *et al*., (2005), preterm birth (27%), sepsis/pneumonia (26%), asphyxia (23%) and tetanus (7%) were the main causes of neonatal mortality at the global level in 2000. Preterm neonates are at high risk of death due to hypothermia, infection such as sepsis, pneumonia, tetanus and diarrhea. The lack of simple care and access to health care including skilled birth attendants intensifies the situation which is described as an underlying factor of neonatal death (Lawn *et al*., 2005). The proportion of child deaths that occurs in the neonatal period is increasing and The Millennium Development goal for child survival cannot be met without substantial reduction in neonatal mortality. Of the estimated 130 million infants born each year worldwide, 4 million die in the first 28 days of life. Almost two-thirds of infant deaths occur in the first month of life, of these, more than two-thirds die in their first week and among those, two thirds die in their first 24 hours after birth (Lawn *et al.,* 2001).

Jehan *et al.* (2009) have documented direct and indirect determinants of neonatal mortality in a recent population based study from Pakistan, which has the third highest NMR in the world. However, Liu *et al*.(2012), conducted a systematic analysis of causes of child mortality at the global, regional and national levels for 2000-2010. Based on their analysis, the major causes of neonatal deaths were preterm and intrapartum related complications and sepsis/meningitis/ tetanus at global level (preterm birth complications 35%, intrapartum related complications 22%, sepsis/meningitis/tetanus 15%, congenital abnormalities 10%, pneumonia 10% and diarrhea 3%). NMR caused by preterm birth and intrapartum complications varied between 2000 and 2010. Also the study conducted in Fawzy Moaz Hospital in Egypt, were reported that 48% death of preterm infants admitted to NICU (Fakher *et al*., 2005). Luiz Fernando *et al.* (2010), showed that, during the study period 495 newborns, with 129 deaths (26.1%) and the variables of corticosteroid use (HR=1.64, 95% CI 1.02-2.70), malformation (HR=1.93, 95% CI 1.05-2.88), very low birth weight (HR= 4.28, 95% CI 2.79-6.57) , phototherapy (HR= 0.34; 95% CI 0.22-0.53) and intubation (HR=2.28, 95% CI 1 .41-3.70) were significantly associated to preterm infants survival rate.

According to Srinivas *et al*.(2015) study done in Australia when 4454 infants included, hospital survival rates based on gestational age alone were 27%, 59%, 76%, 85%, 91% and over 95% at 23, 24, 25, 26, 27 and 28–31 weeks, respectively. Survival rates for each week up to 29 weeks gestation differed by at least 5% when perinatal risk factors including birth weight percentile, exposure to antenatal steroids, birth outside a tertiary hospital and gender were included in the survival estimation. The study done on the Morbidity Pattern of Sick Hospitalized Preterm Infants in Karachi, Pakistan who reported that jaundice and sepsis as the commonest morbidities in their preterm patients. (Khan *et al.,2012). On other hand study carried out* by (Onwuanaku *et al*., 2011), in Jos University Teaching Hospital Nigeria, however reported sepsis as the commonest morbidity, followed by jaundice.

Infant and under-five mortality rates remain very high in Ethiopia. One in every 13 babies born in Ethiopia did not survive to celebrate its first birth day and one in every eight children died before its fifth birth day. Significant number of mothers gave preterm and still birth deliveries in this study. More than third of the mothers had more than five births and more than one in four mothers had pregnancy interval less than 24 months. One in three

pregnancies was unplanned. Considerable mothers had no ANC follow up and one third of the mothers were not vaccinated for Tetanus at all and also significant numbers of mothers were pre eclampic/eclampic. Various factors have been identified as being associated with preterm delivery; many of them are avoidable. A strategy must now be implemented to alter this outcome and combat what is one more silent epidemic (Belyanew *et al*.,2015).

According to the global estimates by (Black *et al*., 2010), the most important causes of neonatal mortality were complications of preterm birth (12%), birth asphyxia (9%), tetanus (7%), sepsis (6%) and pneumonia (4%). The causes of neonatal death may also vary with in the same country depending upon the socioeconomic status of the regional population and access to health care services. Special medical care and treatment are usually provided in order to support the premature babies live. Preemies are normally transferred to NICU right after birth and be placed in the incubator. The NICU is designed to provide an atmosphere that can reduce stress to the baby and meets basic needs of warmth, nutrition and protection to assure proper growth and development of babies.

In Ethiopia, retrospective cohort study of preterm infants admitted from July 1, 2011 to June 30, 2012 G.C 3,277 neonates were admitted to Tikur Anbessa hospital. Out of this, 855 newborns were found to be preterm babies. This study found that 30.9% of all preterm births admitted to TAH in the specified period were died and around 88.3% of the preterm infants admitted during the study period were hypothermic at admission, 95.9% of the preterm infants who died were hypothermic. Birth weight, gestational age, gender, Hypothermia, type of gestation and place of delivery are major factors associated with survival of preterm infants in this study (Merertu *et al*.,2011).

2.3. Survival models

Survival models: Survival analysis is a collection of statistical method to analyze time-to-event data where the outcome variable of interest is the "time to the occurrence of an event" This variable is also often called "survival time" The "survival time" refers to a number of years, months, weeks or days from the beginning of the patient observance till the occurrence of an observed event in this particularly study case time to-death (Hosmer *et al.*,1999). Hence,

survival analysis is also referred to as "time-to-event analysis", which is applied in a number of applied fields, such as medicine, public health, social science, and engineering.

One of the oldest and most straightforward non-parametric methods for analyzing survival data is to compute the life table, which was proposed by Berkson and Gage (1950) for studying cancer survival. Kaplan and Meier (1958) obtained one important development in nonparametric methods. The Cox PH model is more popular than parametric methods to analyze time-to-event data because no assumption is needed about the shape of the underlying hazard of the event over time. Examples of hazard distributions include exponential, Weibull, and log-logistic. It is the most popular and commonly used model by researchers in medical sciences mainly because of its simplicity, and not being based on any assumptions about the survival distribution Therneau, 2000. However, Cox PH model has the restriction that proportional hazards assumption holds with time-fixed covariates; and it may not be appropriate in many situations and other modifications such as stratified Cox model or Cox model with time-dependent variables are required (Collett *et al.*, 2003).

Cox (1972) introduced a semi parametric survival model. This model is based on the assumption that the survival times of distinct individuals are independent of each other. Proportional hazard modeling is the most frequently use type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (Stevens and Hollis, 1989) and employee turnover (Morita *et al.,* 1989), Under certain circumstances parametric models may offer advantages over Cox's model. The results of data analysis using parametric models are similar to the Cox regression. Although the hazard ratio in Cox and parametric models are approximately similar but the Weibull and Exponential models are the most favorable for survival analysis of the data (Dehkordi *et al.,* 2008). The parametric models such as Weibull, Exponential, and Log-logistic are common and provide the interpretation based on a specific distribution for duration times without need to proportional hazard assumptions (Pourhoseingholi *et al.,* 2010). Therefore, data on premature infants include censored data are not compatible with standard statistical models. This study uses the PH model and parametric regression models to identify risk factors that are assumed to have influence on the survival time of premature infants.

3. DATA AND METHODOLOGY

3.1. Source of data

The premature infants' data for this study were obtained from neonatal intensive Care Unit in Jimma University Specialized Hospital neonatology clinic, South West of Ethiopia. The data in this thesis were extracted from the neonatal chart which contains epidemiological, laboratory and clinical information of all preterm infants including time-to- death history. All medical records of preterm infants those who were admitted to Neonatal Intensive Care Unit from January, 2013 to December, 2015 was collected. The study was a retrospective study (i.e. all the events-exposure had already occurred in the past), which reviews the preterm cards and preterm's information sheet. The total numbers of preterm infants admitted to NICU during this study were 552. Among the total of 552 of preterm registered in the given year, only 490 premature infants whose card had full information, satisfy inclusion criteria and hence are included in this study. And the data were analyzed using the STATA (version 12.0) soft wares.

Inclusion and Exclusion criteria. Preterm infants admitted to NICU with a gestational age of 26 weeks or greater and less than 37 completed weeks were included. But Preterm infants with gross congenital malformation are excluded because they have different mortality and morbidity risk. Also, Term infants who born with in gestational age of greater than 37 weeks are excluded.

3.2. Study Population and period

The study population includes all premature infants. All preterm infants those who were admitted to Neonatal Intensive Care Unit (NICU) at Jimma University Specialized Hospital from January, 2013 to December, 2015 G.C were eligible to be included in the study

3.3. Variables of the study

The response (dependent) variable is continuous and describes the length of hospital stay time in days. The explanatory (independent) variables of interest in this analysis include epidemiologic, health conditions and laboratory factors.

3.3.1. The response variable

The response variable for the i^{th} individual is represented by Y_i and it measures duration to event and it is defined by status variable (event or censoring variable). Survival time measures the follow-up of time from a defined starting point to the occurrence of a given event. This observation time has two components, the beginning point of the study time and the observation of time to the end. In survival analysis, the outcome of interest (death in this study) is the duration of time until death occurs measured in days.

3.3.2. Explanatory variables

The predictor variables in survival data analysis are called covariates and are either categorical or continuous. Several factors were considered in this study to investigate the determinant factors for the time-to-death in premature infants.

Table 3.1 The Operational definition and categorization of the factors variables

No	Variables/factors	Definition and categorization
1	Neonatal sex admission	Neonate sex at admission (0=female, 1=male)
2	Antenatal Care Visit	0=no,1=yes
3	gestational age at birth	Measured in weeks (0=(26-28],1=(28-30],2=(30-32], 3=(32-34],4=(34-37))
4	multiple pregnancy	0=no ,1=yes
5	weight of infants	Weight at admission in gram (0=<=1600,1=(1600-2500)
6	mode of delivery	0=cesearem section,1=Spontaneous vertex delivery
7	Sepsis	0=no , 1= yes
8	Jaundice	0=no,1=yes
9	Temperature at admission	Continuous measured in degree centigrade
10	Hyaline membrane disease	0=no,1=yes
11	Place of residence	0= rural ,1=urban
12	Hypothermia	0= no, 1=yes
13	Hypoglycemia	0= no, 1= yes
14	Age at	Continuous measured in hours
15	Respiratory distress syndrome	0=no , 1=yes
16	Gestational Age vs weight	0=SGA,1=AGA
17	Prenatal Asphyxia	0= no, 1=yes

3.4. **Method of data analysis**

3.4.1. Survival Data Analysis

Survival analysis is the phrase used to describe the analysis of data in the form of times from a well-defined time origin until the occurrence of some particular event or end-points. In medical research, the time origins were often correspond to the recruitment of an individual into an experimental study, such as a clinical trial to compare two or more treatments. If the end- points is the death of a patient, the resulting data are literally survival times. However, data of similar form can be obtained when the end-points is not fatal, such as the relief of a pain, or the recurrence of symptoms .In this case the observations are often referred to as time to event data. The reasons why survival data are not amendable to the standard statistical procedures used in data analysis are given as follows. The main feature of survival data that renders standard methods inappropriate is that survival times are frequently censored. The survival time of an individual is said to be censored when the end-point of interest has not been observed for that individual. The second reason is that survival data are generally not symmetrical distributed, this implies it will not be reasonable to assume that data of this type have normal distribution. (Collet, 2003).

By time, we mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs; alternatively, time can refer to the age of an individual when an event occurs. By event, we mean death, disease incidence, relapse from remission, recovery (e.g., return to work) or any designated experience of interest that may happen to an individual. The use of survival analysis, as opposed to the use of other statistical method, is most important when some subjects are lost to follow up or when the period of observation is finite certain patients may not experience the event of interest over the study period. In this latter case one cannot have complete information for such individuals. These incomplete observations are referred to as being censored. Most survival analyses consider a key analytical problem of censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly. In reality such event can occur due to one of the following reasons:- a person does not

experience the event before the study ends, a person is lost to follow-up during the study period and a person withdraws from the study for unknown/known reasons.

Right censoring:-Survival time is said to be right censored when it is recorded from its beginning, say t=0 and terminates before the outcome of interest is observed. Since the incomplete nature of the observation occurs in the right tail of the time axis, such observations are said to be right censoring. There are some reasons why right censoring may occur, for example, no event before the study ends, loss to follow-up during study period, or withdrawal from the study because of some reasons. The last reason may be caused by competing risks. This type of censoring is commonly recognized survival analysis and also considered in this study. The other mechanism that can lead to incomplete observation of time is truncation. A truncated observation is one which is incomplete due to a selection process inherent in the study design (collet *et al*.,2003 and Hosmer *et al*.,1999).

There are obviously many potential life models that overcome such incomplete observations. In some situations there may be reasons to select a particular family of models; the model may fit data on hand well, past experience may have shown the model to give a good description of lifetime distribution from similar populations, there may be a knowledge of the underlying aging or failure process that suggests the validity of the model, and so on. In situations in which no family of models is singled out as being particular appropriate, the choice of the model is frequently made on the basis of considerations such as: the convenience of mathematically handling the model, the statistical methods available in connection with the model and the degree of complication of calculations involved in using the model.

Once this method was developed in modeling human life time where the target event is death, it has been serving as a powerful methodology that appropriately uses data from all observations. It does not matter whether the data is uncensored or censored. Data collection can be prospective or retrospective, experimental or observational. Time can be measured continuously or discretely and explanatory variables can be continuous or categorical. Additional points should be mentioned in connection with the choice of the model. Firstly, for any chosen particular model it has to fit the available data upon appropriate tests. Second, one

should be aware of the consequences of departures from the assumed model on inferences made. Several methods have been developed for the analysis of survival data. Some of these are: Descriptive statistics which include life tables, survival distribution and Kaplan Meier survival function. These are used for the estimation of the distribution of survival time from a sample. Nonparametric tests are available for comparing the survival experience between two or more groups. The most common and widely used of these tests are the log-rank test and Generalized Wilcoxon test.

The multivariate method uses Cox proportional hazards model. It is considered as the most interesting survival modeling in the interest of examining the relationship between survival and one or more predictors. In addition the model has the capability of including both time-dependent and time independent variables. Parametric distributions that justify the use of a fully parametric model to better address the goal of the analysis. Some of the most common parametric survival models are; the exponential regression model, the Weibull regression model and the log-logistic regression model were used for this study.

3.4.2. Descriptive Methods for Survival Data

This method is especially important if individuals are homogeneous at least within groups. In such situation it is appropriate to use the Kaplan-Meier survival estimator.

Kaplan-Meier estimate

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. This description includes survival distribution and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from all of the observations available. The Kaplan-Meier (KM) estimator, or product limit estimator, is the estimator used by most software packages because of the unsophisticated step approach. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. The KM estimator consists of the product of a number of conditional probabilities resulting in an estimated survival function in the form of a step function.

Survivor function S(*t*):-The survivor function is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time. Thus, it gives the probability that an individual surviving beyond a specified time.

Hazard function h(t):-The hazard function $h(t)$ gives the instantaneous potential for failing at time t, given that the individual has survived up to time t. In contrast to the survivor function, which focuses on failing, the hazard function focuses on not failing, that is, on the event occurring. Thus, in some sense, the hazard function can be considered as giving the opposite side of the information given by the survivor function (David *at al.*, 2005).

Suppose there are n observations,$t_1, t_2,\ldots,t_n$, with corresponding censoring indicators,$\delta_1, \delta_2,\ldots, \delta_n$. Let the number of distinct event times be r ($r \leq n$), with the ordered event times given by $t_{(1)}<t_{(2)}<,\ldots,<t_{(r)}$ and corresponding number of events $d_{(1)}, d_{(2)}\ldots\ldots d_{(r)}$ And also let $R(t_{(j)})$ denote the risk set at the event time $t_{(j)}$, i.e., the set of subjects that did not yet experience the event and were not yet censored before time $t_{(j)}$ and thus still at risk for the event at that time. Therefore, the Kaplan-Meier estimate of the survival function at time t is given by:

$$\hat{S}(t)=\prod_{j=1}^{k}\left(\frac{R(t(j))-d(j)}{R(t(j))}\right), \text{ for } t_{(j)}<t<t_{(j+1)}, \; k=1,2,...,r \qquad \ldots\ldots\ldots\ldots\ldots\ldots[1]$$

$\hat{S}(t)=1$ for $t < t_{(j)}$ and the assumptions that $\hat{S}(t)=0$ for $t=\infty$.

Where d_j is the number of individuals who experience the event at time t_j (in this study preterm who experience the event), and n_j is the number of individuals or premmies who have not yet experienced the event at that time.

After providing a description of the overall survival experience in the study, we usually turn our attention to a comparison of the survivorship experience in key subjects in the data. The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for these groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due merely to

chance variation. Assessing whether or not there is a real difference between groups can only be done using the log rank test and Wilcoxon test.

3.4.3. Comparison of Survivorship functions

The estimated Kaplan- Meier survival curves shows the pattern of one survivorship function lying above another, this means the group defined by the upper estimated curve lived longer, or had a more favorable survival experience than the group defined by the lower estimated curve. But, whether this observed difference is statistically significant requires a formal statistical test such as Cochran-Mantel-Haenszel Log Rank and generalized Wilcoxon test were used for this thesis The general form of this test statistic is given by:

$$Q = \frac{\sum_{i=0} w_j (d_{1j} - \hat{e}_{1j})}{\sum_{j=0}^{m} w_j{}^2 \hat{v}_{1j}} \quad \text{.....[2]}$$

Where $\hat{e} = \frac{n_{li} d_i}{n_i}, \hat{v} = \frac{n_{1i} n_{0i} (n_i - d_i)}{n_i{}^2 (n_i - 1)}$, n_{0i} is the number at risk at observed survival time $t(i)$ in group 0, n_{1i} is the number at risk at observed survival time $t(i)$ in the group 1, d_{oi} is the number of observed deaths in group 0, d_{1i} is the number of observed deaths in group 1, ni is the total number of individuals or risk before time $t(i)$ and d_i is the total number of deaths at $t(i)$. The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times. Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, *Q* follows a chi-square distribution with one degree of freedom. We can also use the above test to compare *k* groups. In this study we use the log rank test and generalized Wilcoxon test which are special cases of *Q*.

i. The Cochran-Mantel-Haenszel Log Rank test

The log-rank test were compare the observed number of deaths with the expected number of deaths for group *i.* Consider the null hypothesis $S_1(t) = S_2(t)$, i.e. there is no difference d_{1j} between survival curves in two groups. Given r_j and dj, the random variable has the hyper geometric distribution.

Under the null hypothesis, the probability of death at $t_{(j)}$ does not depend on the group, i.e. the probability of death at $t_{(j)}$ is $\frac{dj}{rj}$. Assuming that the contingency tables at different death times are independent, the log rank test is given by. $T=\frac{UL}{VL} \sim \chi_1^2$ (Under the null hypothesis).

These tests may be defined in general as follow

$$Q_{LR} = \frac{\sum_{i=0} w_j (d_{1j} - \hat{e}_{1j})}{\sum_{j=0}^{m} \hat{v}_{1j}}$$,Where $w_{j=}$.....……...[3]

ii. The Generalized Wilcoxon test

The Wilcoxon test uses weights equal to risk size at t(j),$n_j=r_j$ This gives less weight to longest survival times Early failures receive more weight than later failures. The Wilcoxon test places more emphasis on the information at the beginning of the survival curve where the number at risk is large. This type of weighting may be used to assess whether the effect of treatment on survival is strongest in the earlier phases of administration and tends to be less effective over time. Therefore, Wilcoxon statistic is less sensitive than the log-rank statistic to difference of d_{1j} from e_{1j} in the tail of the distribution of survival times.

$$Q_{GWt} = \frac{[\sum_{i=0} n_j (d_{1j} - \hat{e}_{1j})]^2}{\sum_{j=0}^{m} n_i^{\,2} \hat{v}_{1j}}$$,.....……...[4]

3.4.4. Regression Models for survival data

One very popular model in survival data is the Cox proportional hazards model, which is proposed by Cox (1972). The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients β, hazard ratio, and adjusted hazard curves.

3.4.5. The Cox Proportional Hazards Regression Model

The Cox Proportional Hazard (PH) Model is a multiple regression method and is used to evaluate the effect of multiple covariates on the survival. Cox (1972) proposed a semi-parametric model for the hazard function that allows the addition of covariates, while keeping the baseline hazards unspecified and can take only positive values. David Cox's (1972) paper took a different approach to standard parametric survival analysis and extended methods of the nonparametric Kaplan-Meier estimates to regression type arguments for life table analyses. Cox advanced to prediction of survival time in individual subjects by only utilizing variables covering with survival and ignoring the baseline hazard of individuals. Cox did this by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time. Researchers favor Cox's proportional hazards modeling because of the robust semi-parametric method of calculating the probabilities of survival while simultaneously adjusting for other possibly influential variables. Other attractive features of Cox modeling include: the relative risk type measure of association, no parametric assumptions, the use of the partial likelihood function, and the creation of survival function estimates.

Another feature of Cox regression is that, it does not choose the density function of a parametric distribution. This means that Cox's semi-parametric modeling allows for no assumptions to be made about the parametric distribution of the survival times, making the method considerably more robust. Instead, the researcher must only validate the assumption that the hazards are proportional over time. The proportional hazards assumption refers to the fact that the hazard functions are multiplicatively related. That is, their ratio is assumed constant over survival time. In other words, the Cox proportional hazards model assumes that

changes in the hazard of any subject over time may always be proportional to changes in the hazard of any other subject and to changes in the underlying hazard over time.

The Hazard Function

Cox proportional hazard model is usually written in terms of the hazard model formula. This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by $\boldsymbol{X}$ and it is generally given by:

$h(t/x)=h_o(t)\exp(\beta' Xi)$ for i=1,2,….k ....[5]

where h(t) is the baseline hazard function that characterizes how the hazard function changes as a function of survival time, $\boldsymbol{X_i}$ is the vector of values of the explanatory variables for the ith individual at time t and $\boldsymbol{\beta}$ is the vector of unknown regression parameters that are assumed to be the same for all individuals in the study, which measures the influence of the covariate on the survival experience.

The survival time of each member of the sample is assumed to follow its own hazard function. In such a case, the above model can equivalently be written as

$h(t/x)=h_o(t)\exp(\beta_1 x_1+\beta_2 x_2 + \cdots + \beta_k x_k)$ for i=1,2,….k ....[6]

i=1,……n, where n is total number of premature infants that are included in the study, $\boldsymbol{X}i=Xi1+\cdots+Xip$ is a column vector of measured covariates for the i^{th} individual (preterm) which are expected to affect the survival probability.

A smart property of the Cox model is that, even though the baseline hazard part of the model is vague, it is still possible to estimate the β's in the exponential part of the model. So, it can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$\log\left\{\frac{h(t,x,\beta)}{h_0(t)}\right\}=\beta' X \qquad \text{……..…[7]}$$

The cumulative hazard function is given by:

$h(t) = h_o(t)\, exp(\boldsymbol{\beta}'\boldsymbol{X})$ ..……..……..……[8]

The corresponding survival functions are related as follows:

$S(t/x) = [S_o(t)]^{\exp\{\sum_{i=1}^{p}\beta_i X_i\}}$, …….. …...…..[9]

where, $S_o(t)$ is the baseline survival function.

3.4.6. Fitting the Proportional Hazard Model

As with logistic regression, the Maximum Likelihood estimates of the Cox model parameters are derived by maximizing a likelihood function usually denoted as *L*. The likelihood function is a mathematical expression which describes the joint probability of obtaining the data actually observed on the subjects in the study as a function of the unknown parameters (the $\boldsymbol{\beta}$'s) in the model being considered. *L* is sometimes written rotationally as ($\boldsymbol{\beta}$) where $\boldsymbol{\beta}$ denotes the collection of unknown parameters. The formula for the Cox model likelihood function is actually called a "partial" likelihood function rather than a (complete) likelihood function. The term "partial" likelihood is used because the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who are censored. Thus the likelihood for the Cox model does not consider probabilities for all subjects, and so it is called a "partial" likelihood.

In particular, the partial likelihood can be written as the product of several likelihoods, one for each of, say *k* failure times. Thus, at the jth failure time, l_j denotes the likelihood of failing at this time, given survival up to this time. Note that the set of individuals at risk at the j^{th} failure time is called the "risk set," (( )), and this set may change actually get smaller in size as the failure time increases.

$$l(\beta) = \prod_{j=1}^{k} l_j \qquad \text{……...……[10]}$$

Here l_j is the j^{th} failure time given the risk set (( ))

In a very general sense, the partial likelihood is given by the expression

$$l_p(\beta) = \prod_{i=1}^{m} \left[\frac{e^{x_i \beta}}{\sum_{j \in R(t_i)} e^{x_j \beta}} \right]^{c_i} \quad \text{.....[11]}$$

Where the summation in the denominator is over all subjects in the risk set at time ti denoted by $R\ (ti)$, the expression in (2**) assume that there are no tied times, and it is often modified to exclude terms when ci=0, yielding

$$l_p(\beta) = \prod_{i=1}^{m} \left[\frac{e^{x_i \beta}}{\sum_{j \in R(t_i)} e^{x_j \beta}} \right] \quad \text{.....[12]}$$

Where the product is over the m distinct ordered survival time and (i) denoted the value of the covariance for the subject with ordered survival time (i). Once the likelihood function is formed for a given model, the next step for the computer is to maximize this function. This is generally done by maximizing the natural log of *L*, which is computationally easier. The log partial likely function is given by:

$$l_p(\beta) = \sum_{i=1}^{m} \left[\beta' xi - \ln\left(\sum_{jERti} \exp(\beta' xj) \right) \right] \quad \text{.....[13]}$$

The maximization process is carried out by taking partial derivatives of log of *L* with respect to each parameter in the model and then solving a system of equations as shown here. This solution is carried out using iteration**.** That is, the solution is obtained in a stepwise manner, which starts with a guessed value for the solution, and then successively modifies the guessed value until a solution is finally obtained. Thus, it requires special methods for their solution. These methods are iterative (like Newton Raphson) in nature and have been programmed into available statistical packages like SPSS, SAS and STATA. The partial likelihood derived above is valid when there are no ties in the data set. But in most real situations tied survival times are more likely to occur. In addition to the possibility of more than one death at a time, there might also be more than one censored observations at a time of death. To handle this

real world fact, partial likelihood algorithms have been adopted to handle ties. There are three approaches in common to estimate regression parameters when there are ties. The most popular and easy approach is Breslow's approximation.

The Breslow Approximation

This approximation is proposed by Breslow and Peto to modify the partial likelihood and has the form

$$\mathbf{L_B(\beta)} = \prod_{i=1}^{m} \frac{\exp(\beta' s_i)}{\left[\sum_{l \varepsilon R_{ti}} \exp(\beta' x_l)\right]^{d_i}} \qquad \text{.....[14]}$$

Where d_i the number of deaths occurred at time t_i

s_i the sum of covariates over d_i subjects at time t_i

Then, the partial log likelihood function of equation (14) is given as

$$l_B(\beta) = \sum_{i=1}^{m} \left[\beta' s_i - d_i \ln\left(\sum_{l \varepsilon R_{ti}} \exp(\beta' x_l)\right) \right] \qquad \text{.....[15]}$$

Breslow maximum partial likelihood estimator, adjusted for tied observation is obtained, by differentiating equation (14) with respect to the components of β and setting the derivative equal to zero and solving for the unknown parameters.

Assumption of Cox Proportional Hazard Model

Though the Cox model is nonparametric to the extent that no assumptions are made about the form of the baseline hazard, there are still a number of important issues which need to be assessed before the model results safely applied.

Primarily, it is the issue of non-informative censoring. To satisfy this assumption, the design of the underlying study must ensure that the mechanisms giving rise to censoring of

individual subjects are not related to the probability of an event occurring. That is, those still at risk at time t in the study are a random sample of the population which again should be at risk after a time t, for all t this assumption means that the hazard function, h(t), can be estimated in a fair/unbiased/valid way.

The hazard ratio i.e. ($\boldsymbol{\beta' X}$) depends on the covariate $\boldsymbol{X}$ but not on time t. The effects of covariates are additive and linear on the log risk scale on which their values should not vary depending on the study time.

The other key assumption in the Cox model is that of proportional hazards. In a regression typesetting this means that the survival curves for two subjects must have hazard functions that are proportional over time (i.e. constant relative hazard). The relative risk between the hazard rates for two subjects is constant over the time. In other words, there is no time interaction with the covariates. It can express mathematically as follows:

$h(t, X_i, \boldsymbol{\beta}) = h_o(t)\exp(\beta' X i)$ for i=1,2,….k ....[16]

Then the hazard ratio becomes,

$$\widehat{HR} = \frac{ho(t)\exp(\beta' \mathrm{X})}{ho(t)\exp(\beta' x^*)} = \exp\{\sum \beta'(x - x^*)\}$$ constant over time)

This shows that the ratio of the hazard functions for two individuals with different covariate values does not vary with time.

3.4.7. Model Building Strategies

In modeling with many independent variables, one is usually concerned with the goal of selecting those variables that result in the "best" model within the scientific context of the problem. Having a basic plan to follow in selecting the variables for the model and assessing the adequacy of the model both in terms of the individual variables and from the point of view of the overall fit of the model is required for achieving this "best" model. It is also highlighted in (Hosmer and Lemeshow, 1998) that successful modeling of a complex data set is part science, part statistical methods, and part experience and common sense.

In this study, model building starts from single covariate analysis as suggested by Collet (1994), Collet recommended the approach of first doing a single covariate analysis to "screen" out potentially significant variables for consideration in the multi covariate model in order to identify the importance of each predictor. All variables that are significant at 5% level, the modest level of significance from one explanatory single covariate regression model are taken into multiple covariate models. The purely statistical method is to use an automatic process (stepwise" regression), which can be "forward": the variables are added successively (the most significant at each step) until no variable adds significant information. Finally, the importance of each variable included in the multiple covariate models should be verified by different model assessment techniques.

3.4.8. Assessing Model Adequacy

Once a model has been developed through the various steps indicated in the above section, we now would like to know how effective the model is in describing the outcome of the variable. So, we need to assess the goodness of fit of the model (Agresti, 1996). Some of the methods for the assessment of a fitted proportional hazards model can equally used for parametric regression models. There are basically a requirements for model adequacy considered in this study. They are:-

i. Checking for Proportionality Assumption

In order to use the Cox model, we must check the assumption of whether the effects of covariates on hazard ratio remain constant over time. This is a critical assumption of proportional hazards model and must be checked for each covariate. Different studies suggest that several tests and graphical techniques can be used to assess proportionality assumptions in fitting the Cox model. The Grambsch-Therneau test of non-proportionality uses partial residuals for the test of proportional hazards assumption. In order to use this test for the i^{th} covariate Grambsch and Therneau (1994) propose a time-varying coefficient as

$$\beta_i(t) = \beta_i + \gamma_i g_i(t) \quad \text{..} [17]$$

Where (t) is time varying coefficient, β_i is constant, $g_i(t)$ is some specified function of time, usually $g_i(t)=\ln(t)$. The Cox proportional hazard model for time varying coefficient with $g_i(t)=\ln(t)$ becomes

$h(t, X_i, \beta_i(t)) = h_0(t)\exp(\beta_i(t)X)$

Substitute $\beta_i(t) = \beta_i + (t)$ gives

$= h_0(t)\exp(\beta_i + \gamma_i g_i(t)X)$

$= h_0(t)\exp((\beta_i + \gamma_i \ln(t))X)$

$= h_0(t)\exp(\beta_i X + \gamma_i ln(t)X)$ ……..[18]

This looks like the proportional hazards model where the interaction term, $Xln(t)$ is included in the model in addition to the main effect X_i. To test the significance of the interaction term $X_i ln(t)$, that is, $Ho: \gamma = 0$ against $H1: \gamma \neq 0$ we can use likelihood based tests like Wald test. If $\gamma = 0$ is not rejected, $\beta_i's$ are not time varying coefficients and hence the proportional hazards assumption is satisfied. If $\gamma = 0$ is rejected then the proportional hazards assumption is not satisfied and we have to look for another model.

The Schoenfeld residuals graphical technique can be used to assess Cox model assumptions. The technique is based on individual contributions to the log-partial likelihood and measures the difference between the covariate for the i^{th} individual and a weighted average of the covariate over the risk set at the time the i^{th} individual event (Schoenfeld, 1982). For greater diagnostic power the scaled Schoenfeld residual s are considered, the scaling can be done on the variance of the i^{th} subject Schoenfeld residuals.

To check the proportionality to check the proportionality assumption for each covariate, we plot the scaled Schoenfeld residuals on the Y-axis against log of survival time on the X-axis. If the proportional hazards assumption is satisfied, the distribution of residuals over time is random, that is, does not show a particular trend, and the smoothed plot called Locally Weighted polynomial regression (Lowess) line summarizing the residuals should be a straight

line and close to the horizontal reference line. Otherwise, a plot of scaled Schoenfeld residuals for a given covariate may reveal a violation of the proportional hazards assumption.

Goodness-of-Fit test

One method of checking goodness of fit of the model is to use R^2. In proportional hazards regression model as in all regression analyses there is no single, simple method of calculating and interpreting R^2, because in this model, R^2 depends on the proportion of the censored observations in the data. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to high percent of censored data (Hosmer and Lemeshow, 1998). Cox and Snell (1989) proposed model assessment using R^2 similar to the one used in linear regression which is given by:

$$R^2=1-\exp[\frac{2}{n}(\mathbf{LLo}-\mathbf{LL_{\beta}})] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots[19]$$

Where **LLo** is the log likelihood for zero models or without covariates, $\mathbf{LL}_{\beta}$ is the log likelihood including covariates, n is the number of subjects included in the study. To check the measure of goodness of fit for the final model in addition to R^2 we use tests like: the partial likelihood ratio, Wald and Score tests.

The Partial Likelihood Ratio (LR) test

To use this we need to fit both the unrestricted and the restricted models. We shall obtain the value of the log-partial likelihood function $LL_p(\beta)$ in the unrestricted model and **LLp** $\beta=0$ when the model imposes the restrictions under H_0. The test statistic for $H0$ is based on the difference of the log likelihood values. Under H_0, the statistic is asymptotically distributed as x^2 with P degrees of freedom.

$$\boldsymbol{QLR=2[LL((\beta)-LL(\beta=0)] \sim x^2_{(p)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots[20]$$

The Wald and Score test

For testing the hypothesis that the model fits the data, other two common approaches are the Wald (QW) and Score tests (QS). Under H_0, the statistic is asymptotically distributed as x^2 with P degrees of freedom. If chi-square is significant, the variable is considered to be a significant predictor in the equation. The test statistics are:

$$Qw = \beta' I^{-1}{}_{pxp}(\beta)\beta \sim \chi^2{}_P$$
$$Qs = U'_{Ho}\, I^{-1}{}_{pxp}(\beta = 0) U_{Ho} \sim \chi^2{}_P \qquad \text{….. ……[21]}$$

Where $I^{-1}{}_{pxp}(\beta)$ and $I^{-1}{}_{pxp}(\beta = 0)$, indicate the matrix of dimension $p \times p$, extracted from the inverse of the Observed information matrix evaluated at β and $\beta = 0$ respectively and U_{H0} is the score function under $H0$. Both Qw *and* Qs have approximately x^2 distribution with P degrees of freedom.

3.4.9. Checking for Influential and Linearity of Covariates

Furthermore, a thorough evaluation of regression diagnostic statistic to identify, if any, subjects: either have unusual configuration of covariates, exert an undue influence on the estimate of the parameters and have an undue influence on the fit of the model.

Statistics similar to those used in linear and logistic regression are available to perform these tasks with a fitted proportional hazards model. There are some differences in the types of statistics used in linear and logistic regression and proportional hazards regression, but the essential ideas are the same in all the three settings Hosmer and Lemeshow (1998).

Leverage is a diagnostic statistic that measures how "unusual" the values of the covariates are for an individual. In linear and logistic regression leverage is the distance of the value of the covariates for a subject to the overall mean of the covariates. Leverage is not easily defined nor does it have the same nice properties in proportional hazards regression. This is due to the fact that subjects may appear in multiple risk sets and thus may be present in multiple terms in the partial likelihood.

The score process residual for the i[th] subject on the k[th] covariate may be expressed as

$$L_{ik} = \sum_{i=1}^{n} (X_{ik} - \bar{X}_{wjk}) dM_i(t_i) \quad \text{.....[22]}$$

It is a weighted average of the distance of the value X^{ik} to the risk set means X . Where the weights are the change in martingale residual (( )) defined as.

$$M(t_j) = dN_i(t_i) - Y_i(t_j)\, exp(\beta' x_i) h_o(t_j) \quad \text{..... [23]}$$

Where (t_j) is the change in the count function for the i^{th} subject at time t_j , always equal to zero for censored subjects and one for uncensored subjects, at actual observed survival time. The function is called the risk process and defined as zero $t_i \leq t_j$ if and one if , $t_i >= t_j$, $h_o(t_j)$ have **the value of** $\frac{\delta_i}{\sum_{j \in R(t)} \exp(x_j'\beta)}$ **evaluated at t_j.** The net effect is that, for continuous covariates, the score residuals have the linear regression leverage property that the further the value is from the mean the larger the score residual is, but "large" may be either positive or negative. Thus, the score residuals are sometimes referred to as the leverage or partial leverage residuals. We plot score residuals against each continuous covariates to observe if there is individuals far away from the mean. Finally, nonlinearity, that is, an incorrectly specified functional form in the parametric part of the model, is a potential problem in Cox regression as it is in linear and generalized linear models (Fox, 2003). In order to assess the linearity assumption on the part of the covariates, we use plot of martingale residuals.

Martingale Residuals

As far as one event models are concerned, martingale residual for the i[th] subject at the moment t is defined as follows $M_i(t) = \delta_i(t) - H(t, Z_i)$ and is interpreted as a difference between (observed) and expected (resulting from the model) number of event occurrence till the moment t. It is calculated for the given subject, at the given time point t. With, δ_i (t) being the dummy variable to indicate that if $\delta_i = 1$ for uncensored observation and $\delta_i = 0$ for censored observation. Usually martingale residuals are subject - specific and are calculated at the end of the study. As residuals of this type do not have symmetric distribution, they can be

transformed into deviance residuals that are supposed to have a symmetric distribution with the mean equal to zero, assuming proper specification of the model. Martingale residuals are useful while examining assumption of linear effect of covariates on logarithm of hazard.

3.5. Parametric Regression Modeling

In previous topics it was focused entirely on the use of semi-parametric model of proportional hazards Cox regression model, in the analysis and prediction of the survival time of infants with prematurity. The basis of this method was to avoid having to specify the hazard function completely. However, there may be settings in which the distribution of the survival time is in specific parametric distribution that justifies the use of a fully parametric model to better address the goal of the analysis. A parametric survival model assumes that the survival time follows a known distribution. Many models using different distributions have been developed. Some of most common survival models are: exponential, weibull and log logistic distribution in this study were used.

I. The Exponential Regression Model

The exponential distribution, with only one unknown parameter and it is the simplest of all life distribution models. In the exponential model, the conditional probability is constant over time. In other words, the main feature of exponential distribution is that the instantaneous hazard does not vary over time. Modeling the dependency of the hazard rate on covariates entails constructing a model that ensures a non-negative hazard rate (or non-negative expected duration time it is constant over time.

For the time data and skewed to the right, with distribution of the time is exponential, the time of survival for a single covariate x, which is called, accelerated failure time, expressed as:

$T=\exp(\beta_0+\beta_1 x+\varepsilon)$

This model can be linearized by taking the natural log of each side of the equation above as:

$lnT=\beta_0+\beta_1 x+\varepsilon*$ where, $\varepsilon*$ is the error component

The exponential model ($t\sim(\alpha)$) is the simplest parametric model and assumes a constant risk or hazard over time, which reflects the property of the distribution appropriately called „lack of memory". The survivorship function may be obtained by expressing in terms of time as:

$$S(t,x,\beta)=\exp(\frac{-t}{e^{\beta_0+\beta_1 x}}).....\quad\ldots\ldots\ldots[24]$$

And the hazard function of the exponential regression model is:

$$h(t,x,\beta)=e^{-\beta_0+\beta_1 x}.....\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots[25]$$

The exponential regression model for the k covariates and i^{th} individual premature infants' is expressed as:

$$h(t,x_i,\beta)=h_0(t)exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik}) \quad . \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots[26]$$

For the exponential regression survival models the hazard ratio for the dichotomous covariate is $(x=1,=0)=e^{-\beta_1}$.

II. The Weibull regression model

Weibull distribution (introduced by Waloddi Weibull in 1939) is one of the parametric distributions which are used for the analysis of life time data and mostly used in literature for modeling life time data. The Weibull distribution is more general and flexible than the exponential distribution and allows for hazard rates that are non-constant but monotonic. It is a two-parameter model ($\lambda\ and\ \rho$), where λ is the scale parameter and ρ is the shape parameter because it determines whether the hazard is increasing, decreasing, or constant over time i.e., the hazard rate increases when, $\rho > 1$ and decreases when $\rho < 1$ as time goes on. When $\rho = 1$, the hazard rate remains constant, which is the special case of exponential.

The pdf for Weibull distribution is given by;

$$f(t) = \lambda\rho t^{\rho-1}\exp(-\lambda t^{\rho}),\ where;\ \lambda,\rho > 0 \quad \text{……..……[27]}$$

And the corresponding survival function and hazard for Weibull distribution are given as;

$$S(t) = \exp(-\lambda t^{\rho}) \quad \text{… ……..……[28]}$$

$$h(t) = \lambda\rho t^{\rho-1} \quad \text{……[29]}$$

Under the Weibull PH model, the hazard function of a particular individual with covariates (x_1; x_2, . . ., x_p) is given by

$$h(t|x)=\lambda\gamma t^{\gamma-1}\exp[\beta_1x_1+ \beta_2x_2+\ldots\ldots \beta_jx_j+\ldots..+\beta_px_p]= \lambda\gamma t^{\gamma-1}\exp(\beta`x) \quad \text{……..……...[30]}$$

We can see that the survival time of this preterm has the Weibull distribution with scale parameter λexp(β`x) and shape parameter γ:- Therefore the Weibull family with fixed γ possesses PH property. This shows that the effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains constant.

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

III. Log-logistic régression model

The log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions. According to the study of (Gupta *et al*, 1999), the log-

logistic distribution is proved to be suitable in analyzing survival data conducted by Cox (1972), Cox and Oakes (1984), Bennet (1983) and O'Quigley and Stare (1982). Gupta, etal.(1999) used log-logistic distribution in survival analysis on lung cancer data in their studies.

The log-logistic distribution is very similar in shape to the log-normal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameter λ is the scale parameter and ρ is the shape parameter. Its probability density function is given by;

$$\mathbf{f(t)} = \frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^{\rho})^2} \quad \text{…… ……..….....[31]}$$

The corresponding survival and hazard functions are given by;

$$S(t) = \frac{1}{1+\lambda t^{\rho}} \quad \text{…… ……..…….[32]}$$

$$h(t) = \frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}, \quad \text{… ……..…….[33]}$$

Where; $\lambda \in R, \rho > 0$

For a Single covariate log-logistic accelerated failure time may be expressed as:

$$\ln T = \beta_0 + \beta_1 X + \sigma\varepsilon \quad \text{……..……[34]}$$

The survivorship function for the model form [33] is

$$S(t, x, \beta, \sigma) = [1 + \exp(z)]^{-1} \text{.....[35]}$$

Where z is the standardized log-time outcome variable, that is : $Z = \frac{(y - \beta_0 - \beta_1 X}{\sigma}$ and y=ln(t).

The odds of a survival time of at least t are, $OR = \frac{S(t, x, \beta, \sigma)}{1 - S(t, x, \beta, \sigma)} = \exp(-Z)$, assumes that the covariate is dichotomous and coded 0 or 1. The odds- ratio at time t from the ratio the odds of a survival time evaluated at x= 0 and x= 1 is:

$$OR\ (x=1,x=0) = \frac{\exp(\frac{-(y-\beta_0-\beta_1 x1)}{\sigma})}{\exp(\frac{-(y-\beta_0-\beta_1 x0)}{\sigma})} = \exp\left(\frac{\beta}{\sigma}\right) \quad \text{.....[36]}$$

This is independent of time.

Then the hazard rate iss given as follows:

$$h(t, x,\beta) = \frac{\lambda p t^{p-1}\exp(X\beta)}{1+\lambda p t^{p}\exp(X\beta)} \quad \text{.....[37]}$$

Parameterization

When we say proportional hazards (PH) it means that the hazard function of a group is proportional to the hazard function of the other group, i.e., the hazard ratio is constant over time (Klein, 1992). The hazard ratio is hence given by;

HR = exp(β`$X_{ij}$), where β` = ($\beta_1, \beta_2, \ldots.\beta_p$) is a vector of regression coefficients and X_{ij} is the vector of covariates for subject j in cluster i. On the other hand, the accelerated failure-time (AFT) model describes stretching out or contraction of survival time as a function of predictor variables. The acceleration factor which is usually denoted by ϕ is given by exp(α'X_{ij}) where α'= (α_1, α_2... α_p) is a vector of regression coefficients in case of AFT model. For the exponential, Weibull and log logistic survival model, the relationship between α and β is given by (Hougaard, 2000)

a) For exponential **β_j=- α_j,** the exponential PH and AFT are in fact the same model, except that the parameterization is different, and hence HR=**exp(- α_j)** is the hazard ratio of the j^{th} group with the reference groups.

b) For Weibull, **β_j=-$\alpha_j\rho$**, where ρ is the shape parameter and hence, **HR=exp (-$\alpha_j\rho$)** is the hazard ratio of the j^{th} group with the reference groups.

c) For log-logistic $\boldsymbol{\beta_j=-\alpha_j\rho}$, where ρ is the shape parameter and **OR = exp(-$\alpha_j\rho$)** indicates the failure odds ratio of the j^{th} group with the reference groups. The log-logistic model is a proportional odds (PO) model, i.e. it has constant OR for two groups.

3.5.1. Model Development

The methods of selecting a subset of covariates in a PHs regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections. Survival analysis using Cox regression method begins with a thorough univariate analysis of the association between survival time and all important covariates (Hosmer and Lemeshow,1999).

Recommendable procedure in selecting variables in the study

According to Hosmer and Lemeshow (1998), it is recommended to follow the steps given below.

i. Include all variables that are significant in the univariate analysis at relaxed level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.
ii. The variables that appear to be important from step one are then fitted together in a model. In the presence of certain variables others may cease to be important. As a result, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.
iii. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method. This process may result in terms in the model determined at step 2 ceasing to be significant.

3.5.2. Model Selection

To select the model that can predict the survival time of diabetic patients, we have two methods. The first is graphical approach. For this method the Cox-Snell residual plot is the common one. It is used to determine how well a specific distribution fits to the observed data. This plot will be approximately linear if the specified theoretical distribution is the correct model. Easy fit displays the reference diagonal line along which the graph points should fall along with the goodness of fit tests; the distribution plots can be helpful to determine the best fitting model. The fundamental difference of this approach is that it is quite subjective to come on conclusion while the goodness of fit tests are "exact" in the sense that the results do not depend on the researcher (provided that the tests are performed correctly), using plot is a more empirical way to use in model selection.

Akaike's Information Criterion (AIC)

To select the model that can predict the survival time to death in premature infants, we will use Akaike information criterion (AIC). Akaike's proposed an informative criterion (AIC) statistic to compare different non nested models. For survival model the value of AIC is computed as:

AIC = −2LogL+2(k+c+1), **……[38]**

Where k is the number of covariates and c the number of model specific distributional parameters. This research will use the AIC to compare various candidates of non- nested parametric models. The preferred model will be the one with the minimum value of the AIC. (Akaike, 1974).

Likelihood Ratio test (LRT)

The likelihood ratio test (LRT) statistic is an adequate test as the new model is nested in the previous model. Suppose there are (p+q) explanatory variables measured: $x_1, x_2, ..., x_p$, $x_{p+1}, ..., x_{p+q}$. and proportional hazards are assumed. Consider the following models.

Model 1: contains only the first p-covariates $\frac{h_i(t,x)}{h_o(t)} = \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)$.

Model 2: contains all (p+q) covariates $\frac{h_i(t,x)}{h_o(t)} = \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_{p+q} x_{p+q})$

These are nested models. For such nested models, we can construct a likelihood ratio test of

$H_o = \beta_{p+1} = \beta_{p+2} = ...\beta_{p+q} = 0$ as $\chi_{LR}^2 = -2[\log(\hat{L}(1) - \log(\hat{L}(2))]$ under H_o. This test statistic is approximately distributed as χ^2 with q degree of freedom.

Checking the Adequacy of Parametric Model

The graphical methods can be used to check if a parametric distribution fits the observed data. The appropriateness of model with the exponential baseline can graphically be evaluated by plotting $-\log(\hat{S}(t))$ versus t where $\hat{S}(t)$ is Kaplan-Meier survival estimate. This plot should be linear (Klein, 1992). Because for exponential distribution, S(t) = exp (- λt), and hence, -log(S (t)) = λt is linear with time.

Model with the Weibull baseline has a property that the log (-log(S(t)) is linear with the log of time, where $S(t) = \exp(-\lambda t^{\rho})$. Hence, $\log(-\log(S(t))) = \log(\lambda) + \rho\log(t)$. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting $\log(-\log(\hat{S}(t)))$ versus log(t) where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Kleinbaum D, Klein M., 2005).The log-failure odd versus log time of the log-logistic model is linear. Where **t**he failure odds of log-logistic survival model can be computed as:

$$\frac{1-s(t)}{s(t)} = \frac{\frac{\lambda t^{\rho}}{1+\lambda t^{\rho}}}{\frac{1}{1+\lambda t^{\rho}}} = \lambda t^{\rho}. \qquad \text{.....[39]}$$

Therefore, the log-failure odds can be written as:

$$\text{Log}\left(\frac{1-S(t)}{S(t)}\right) = \log(\lambda t^{\rho}) = \log(\lambda) + \rho\log(t) \qquad \text{.....[40]}$$

Therefore the appropriateness of model with the log logistic baseline can graphically be evaluated by plotting $\log(\frac{\hat{S}(t)}{1-\hat{S}(t)})$ versus log time where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Hosmer DJ, Lemeshow S, 1999).

3.5.3. Model Assessment

Cox-Snell residuals

The Cox-Snell residual is given by Cox and Snell (Klein J, Moeschberger M.,1997). The Cox-Snell residual for the i^{th} individual with observed survival time t_i is defined as;

$$r_{ci}= \hat{H}_o(t_i)\exp(\sum_{i=1}^{p}(Z_{ik}\hat{b}_k) = H_i(t_i) = -\ln(\hat{s}(t_i) \quad \text{.....[41]}$$

where $\hat{H}_O(t_i)$ is an estimate of the Breslow baseline cumulative hazard function at time ti; which is given by

$$\hat{H}_o(t_i) = \sum_{t_i \le l}^{p} \frac{d_i}{\sum_{j \le R(t_i)} \exp(\sum_{i1}^{p} Z_{jk}\hat{\beta}_k)} \quad \text{.....[42]}$$

Let Z = H(T) be the transformation of T based on the cumulative hazard function. Then the survival function for Z is:

$S_z(z)=P(Z>z)=P(H(t)>z)$ .....[43]

$P(T>H_T^{-1}(z))=S_T(H_T^{-1}(z))$ .....[44]

$\exp(-H_T(H_T^{-1}(z)))=\exp(-z)$ . .....[45]

this was derived by Kalbfleisch and Prentice (1973). This residual is motivated by the following result:

Let T have continuous survival distribution S(t) with the cumulative hazard H(t) = -log(S(t)). Thus, $S_T(t) = \exp(-H(t))$.

Thus, regardless of the distribution of T, the new variable Z= H(T) has an exponential distribution with unit mean. If the model was well fitted, the value $\hat{s}_i(t_i)$ would have similar properties to those of $S_i(t_i)$: So $r_{ci} = -\log S_i(t_i)$ will have a unit exponential distribution with $f_R(r) = \exp(-r)$. Let $S_R(r)$ denotes the survival function of Cox-Snell residual r_{ci}. Then

$$S_R(r) = \int_r^\infty f_R(z)dz = \int_r^\infty \exp(-z)dz = \exp(-r) \text{ and HR } (r) = -\log S_R(r) = -\log(\exp(-r)) = r \qquad [46]$$

Therefore, we will use a plot of $H(r_{ci})$ versus r_{ci} to check the fit of the model. Thus, If the final proportional hazards model is correct and the estimated regression coefficients are close to the true values, the Cox-Snell residuals r_{ci} can be regarded as a sample from a unit exponential distribution, and therefore, the plot of $H(r_{ci})$ against r_{ci} should be a 45°-line through the origin. But the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

4. RESULT AND DISCUSSION

4.1. Preterm Base line Characteristics

The study was based on 490 premature infants from a total of 552 who were admitted to neonatal intensive care unit in Jimma University Specialized Hospital from 1st January, 2013 to 31st December, 2015 whose medical cards had full information for this study. Of 490 premature infants 319(65.1%) were discharged at the end of the follow up and 171(34.9%) premature infants were died.

Plots of the Kaplan Meier curves to the survival experience and hazard rate of time to death of premature infants are shown in figure 4.1 below. The survival plot decreases at increasing length of hospital stay and the estimate of overall Kaplan-Meier survivor function showed that most of the deaths occurred in the shorter length of hospital stay and it declined in the later days of follow up. This indicates the neonatal infant have long survival time as the day of hospital stay increases during the follow up in the neonatal intensive care unit.

The hazard rate increases as the length of hospital stay increases or survival probability decreases. This implies that most of the premature infants were died in shortest stay of hospital after admitted to neonatal intensive care unit.

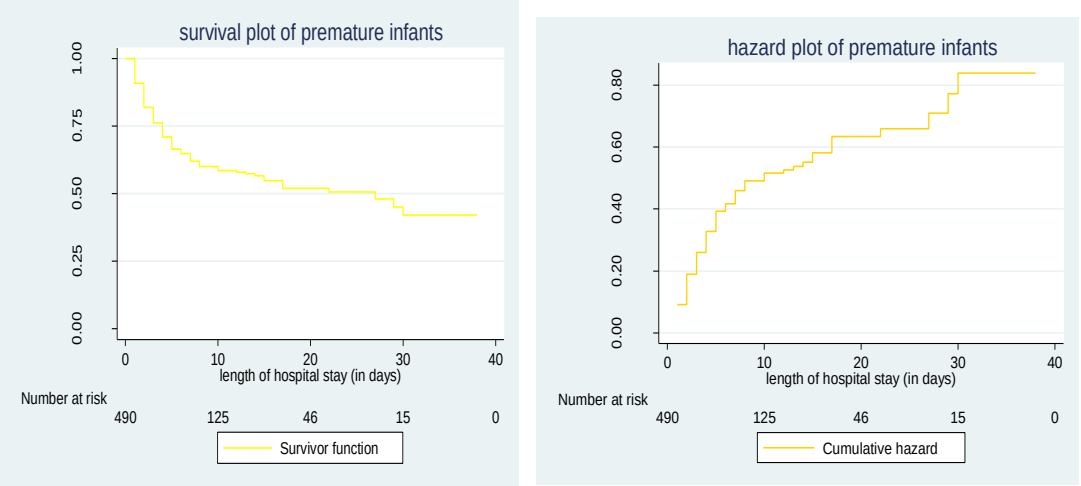


Figure 4.1. The Kaplan-Meier plots of Survival functions and hazard rate of premature infants.

The descriptive statistics was presented in the Table 4.1 below. The medical cards of 490 preterm were reviewed, out of which 45.3% were female having death proportion of 15.5% seems, lower than males 19.4% respectively. There are 75.9% premature infants from mother of rural residences, the death proportion of infants from mother of rural residents were 27.1% which is higher than mother of urban residents (7.8%). Regarding to mode of delivery 33.3% were delivered by cesearem section, of which 11.4% infants died who delivered through cesearem section. Among the premature infants considered during study period 94.1% had appropriate gestational age, of which their death proportion is 33.3% higher than premature infant having small gestational age. The other reason for premature infant admission were having prenatal Asphyxia (8.6%), hyaline membrane disease (38.2%), hypothermia (47.3%), hypoglycemia (32.9%), jaundice (24.1%), sepsis (22.4%),multiple pregnancy of mother (21.6%), respiratory distress syndrome(46.1%) and Antenatal care visit of mother (70.6%).

From all premature infants included in this study, the highest proportion 64.1% of the preterm had birth weight in the interval of 1600-2500 in grams, of these 18.8% of premature infants died during follow up period. The gestational age which measured in weak for premature infant admitted were 4.7%, 12.65%, 18.2% , 23.67 % and 40.8% for weak category of (26-28],(28-30], (30-32], (32-34] and (34-37) respectively.

The preterm were followed up for a median of 27 (95% CI: [15,.]) days with standard error of 4.532. The minimum follow up time was 1 day and the maximum was 38 days .The overall mean estimated survival time of preterm under the study was 21.23 (95% CI: 19.22-23.235) days with standard error of 1.024. Lastly, the minimum age of premature infant admitted to neonatal intensive care unit was 0.13 hours old and maximum of 144 hours with mean and standard error 64.05 hours and 7.622 respectively. The median age of premature infants was 48 hours with standard error 11.061 at the 95%CI (48, 96) hours. The body temperature of premature infants during admission to intensive care unit was minimum 32 and maximum 39 degree centigrade with mean and standard error of 36.908 degree centigrade and 0.112 respectively for admitted premature infants. The median temperature was about 37.1.

Table 4.1.Distributions of death status of premature infants who were admitted to neonatal intensive care unit at JUSH, during 2013-2015.

		Status		
Covariates	Category	Death (%)	Censored (%)	Total (%)
Sex neonate	Female	76(15.5)	146(29.8)	222(45.3)
	Male	95(19.4)	173(35.3)	268(54.7)
Mother residence	Rural	133(27.1)	239(48.8)	372(75.9)
	Urban	38(7.8)	80(16.3)	118(24.1)
Mode of delivery	Cs	56(11.4)	107(21.8)	163(33.3)
	SVD	115(23.5)	212(43.3)	327(66.7)
GAvsweight	SGA	8(1.6)	21(4.3)	29(5.9)
	AGA	163(33.3)	298(6.08)	461(94.1)
PNA	No	144(29.4)	304(62.0)	448(91.4)
	Yes	27(5.5)	15(3.1)	42(8.6)
HMD	No	64(13.1)	239(48.8))	303(61.8)
	Yes	107(21.8)	80(16.3)	187(38.2)
Hypothermia	No	95(19.4)	163(33.3)	258(52.65)
	Yes	76(15.5)	156(31.8)	232(47.35)
Hypoglycemia	No	118(24.1)	211(43.1)	329(67.1)
	Yes	53(10.8)	108(22)	161(32.9)
Jaundice	No	81(16.5)	291(59.4)	372(75.9)
	Yes	90(18.4)	28(5.7)	118(24.1)
Sepsis	No	119(24.3)	261(53.3)	380(77.6)
	Yes	52(10.6)	58(11.8)	110(22.4)
Multiple pregnancy	No	130(26.5)	254(51.8)	384(78.4)
	Yes	41(8.4)	65(13.3)	106(21.6)
Gestational age	(26-28]	16(3.3)	7(1.4)	23(4.7)
	(28-30]	40(8.2)	22(4.5)	62(12.7)
	(30-32]	33(6.7)	56(11.4)	89(18.2)
	(32-34]	32(6.5)	84(17.1)	116(23.7)
	(34-37)	50(10.2)	150(30.6)	200(40.8)
weight at birth	<=1600	92(18.8)	84(17.1)	176(36)
	(1600-2500)	79(16.1)	235(48)	314(64)
Antenatal care visit	No	57(11.6)	87(17.8)	144(29.4)
	Yes	114(23.3)	232(47.3)	346(70.6)
RD S	No	46(9.4)	218(44.5)	264(54)
	Yes	125(25.2)	101(20.6)	226(46)

	Minimum	Maximum	Median	SE	95%CI	Mean	SE	95% CI
Age at admission	0.13	144	48	11.061	[48,96]	64.05	7.62	[49.11,79]
Temperature	32	39	37.1	0.1682	[36.9,38)	36.91	0.11	[36.69,37.13]
Length of stay hospital	1	38	27	4.5324	[15,….]	21.23	0.024	[19.22,23.24]

*PNA=Prenatal Asphyxia, HMD=Hyaline membrane disease, RDS=Respiratory distress syndrome

4.2. Comparison of survival experiences of premature infants using demographic, health and risk behavior variables.

The estimated mean survival time (95% confidence interval), log rank and Breslow (generalized Wilcoxon) for time-to-death with different covariates characteristics are summarized in Table 4.2 below. The mean survival time of time- to-death for female preterm infant was 21.374 [95% CI: 18.45, 24.29] day higher than the male which is 20.44 [95% CI: 17.95, 22.92] day. The infant whose mother came from rural had mean survival time of 20.484[95% CI: 17.98, 22.99] day less than urban mother which is 23.403 [19.75, 27.05] day. The mean survival time of premature infants who born by cesearem section delivery was less than infants who born through spontaneous vertex delivery which is 21.448 [95% CI: 18.03, 24.87] and 19.842 [95% CI: 17.66, 22.23] days respectively. Infants with gestational age vs weight of AGA had mean survival time of 19.166 day [95% CI: 14.09, 24.24]. Premature Infants who had Prenatal asphyxia, Hyaline membrane disease, Hypothermia, hypoglycemia, sepsis, jaundice, and Respiratory distress syndrome had mean survival time of 10.054, 12.319, 22.192, 19.872, 16.194, 6.715 and 11.164 days which is lower than the premature infants who hadn't prenatal asphyxia, hyaline membrane disease, Hypothermia, hypoglycemia, sepsis, jaundice, and Respiratory distress syndrome respectively.

The highest mean survival time of hospital stay was 23.145 [95% CI: 18.98, 27.31] days for infants in the gestational age (32, 34] weeks and those born in the first two category have the smallest mean survival length of hospital stay for premature infants admitted to neonatal intensive care unit. The infants whose mother had no antenatal care visit have mean survival length of hospital stay 15.653 [12.56, 18.74] days lower than infants whose mother had antenatal care visit. The mean survival time and the corresponding 95% confidence interval for the rest categorical variables are listed in table 4.2 below.

From log-rank test indicated in table 4.2 showed that, there was no significant difference in survival experience between the categories of neonatal sex, mother residence, GA vs weight, infants with hypothermia, hypoglycemia , mother of multiple pregnancy and mode of delivery at 5% level of significance and we have no enough evidence to say that the premature infants admitted to neonatal intensive care unit survival curves are different or the Kaplan Meier

curves are statistically equivalent with respect to categories of these covariates. But there is a significant difference of survival experience among groups of prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, Antenatal care visit, gestational age, respiratory distress syndrome and weight of infant.

Similarly the results of Breslow test also show that there were significant differences among premature infants survival experience of groups prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, gestational age, respiratory distress syndrome and weight of infant except for Antenatal care visit which means there is no differences in survival experience of premature infants whose mother had Antenatal care visit and had no Antenatal care visit have at the earlier phases where the number at risk is large. The Kaplan meier plot below in Figure 4.2, shows that the survival experience between premature infants whose mother had Antenatal care visit and those whose mother hadn't groups have no differences in survival experience at the starting time of follow up. Also, from figure in Annex-II, figure 4.3 ,we have seen that the Kaplan Meier curve for infants those who had prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, Respiratory Distress syndrome, weight less than or equal to 1600 gram and gestational age of (26-28], (28-30] weeks were consistently lower than the Kaplan Meier curve for infants those who had no prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, Respiratory Distress syndrome, weight greater than 1600 and less than 2500 grams and gestational age of (32-34], (34-37) weeks respectively. Therefore log-rank, Breslow test and Kaplan Meier curves suggest that there is significant difference of survival experience among groups prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, Antenatal care visit, gestational age, respiratory distress syndrome and weight of infant .

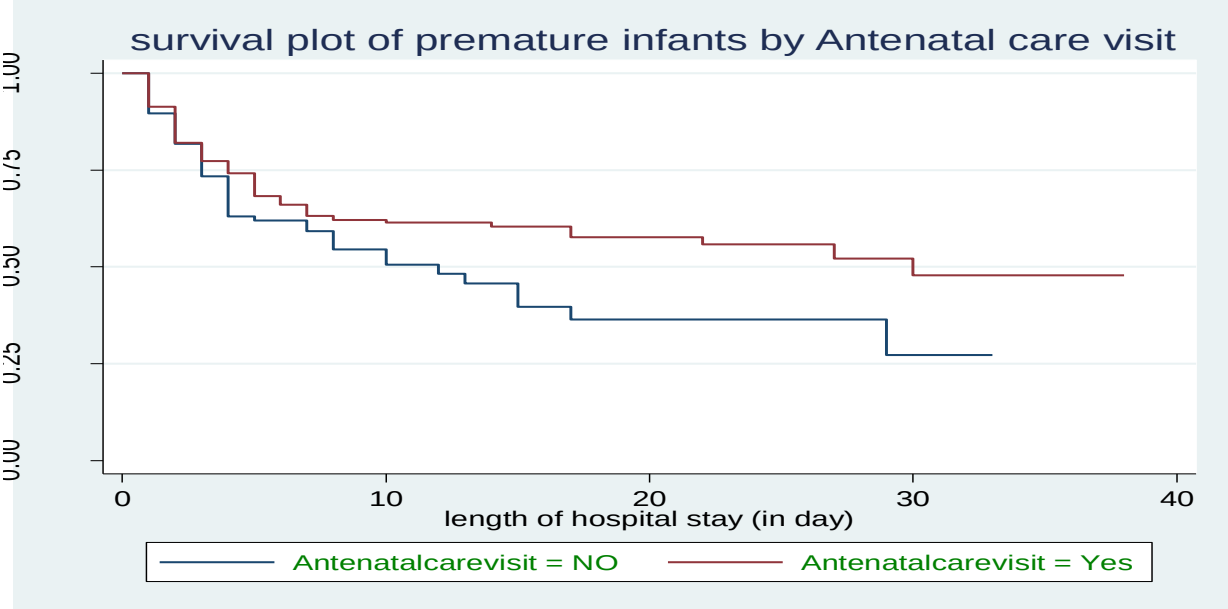


Figure 4.2 kaplan meier plots of survival function for Antenatal care visit of preterms'mother

Table 4. 2 Comparisons of survival experience of premature infants using demographic, health and risk behavior variables.

Test for equality of survivor functions over group

	Mean survival time (in days)		Log-rank (mantel cox)			Breslow (generalized Wilcoxon)		
Variable	Mean	95% CI	Chi-square	Df	Pr>chi-square	Chi-square	Df	Pr>chi-square
Sex of neonate								
Female	21.374	[18.45,24.29]						
Male	20.44	[17.95,22.92]	0.0631	1	0.802	0.16	1	0.6872
Mother Residence								
Rural	0.484	[17.98,22.99]						
Urban	3.403	[19.75,27.05]	0.90	1	0.343	0.08	1	0.7733
Mode of delivery								
CS	21.447	[18.03,24.87]						
SVD	9.842	[17.66,22.23]	0.0000	1	0.862	0.06	1	0.8059
Hypothermia								
No	20.762	[18.12,23.40]						
Yes	22.192	[19.19,25.19]	0.113	1	0.737	1.00	1	0.3169
Hypoglycemia								
No	21.719	[19.36,24.08]						
Yes	9.872	[16.00,23.75]	1.3	1	0.258	1.05	1	0.3057

Multiple pregnany								
No	21.173	[18.82,23.52]						
Yes	0.800	[17.30,24.29]	0.0000	1	0.983	0.01	1	0.9252
GA vs weight								
SGA	19.166	[14.09,24.24]						
AGA	20.962	[18.89,23.03]	0.5	1	0.484	0.10	1	0.7534
Perinatal asphyxia								
No	2.541	[20.46,24.62]						
Yes	10.054	[6.53,13.58]	16.13	1	0.0001	11.37	1	0.0007
Hyaline membrane disease								
No	27.443	[25.06,29.82]						
Yes	12.319	[10.08,14.55]	53.56	1	0.0000	37.38	1	0.0000
Jaundice								
No	27.335	[25.23,29.44]						
Yes	6.715	[6.74,10.69]	96.6	1	0.0000	65.18	1	0.0000
Sepsis								
No	22.162	[19.79,24.53]						
Yes	16.194	[13.49,20.34]	6.35	1	0.018	7.63	1	0.0057
Gestational age								
(26-28]	11.194	[4.36,18.03]						
(28-30]	10.331	[7.25,13.42]						
(30-32]	21.724	[17.97,25.48]	45.67	4	0.0000	51.06	4	0.0000
(32-34]	23.145	[18.98,27.31]						
(34-37)	20.510	[17.58,23.45]						
Weight of infants								
<=1600	17.208	[14.49,19.93]						
(1600-2500)	23.060	[20.14,25.98]	17.7	1	0.0000	15.63	1	0.0001
Antenatal care visit								
No	15.653	[12.56,18.74]						
Yes	22.797	[20.48,25.11]	4.2	1	0.043	1.68	1	0.1945
Respiratory distress Syndrome								
No	9.779	[27.54,32.02]						
Yes	11.164	[9.20,13.12]	84.2	1	0.0000	60.65	1	0.0000

*REF=References

*Source Jimma university specialized hospital NICU from Jan 2013 to Dec 2015 , Jimma, Ethiopia.

4.3. Cox Proportional Hazards Model

4.3.1. Univariate Analysis

The Univariate analysis indicates that not all of the explanatory variables are statistically important to be included in the multiple covariate analysis. As the univariate Cox regression analysis results in Table 4.3 found at ANNEX-I showed that the covariates prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice, Antenatal care visit, respiratory distress syndrome, temperature ,Age, weight of infant and gestational age were significant. Thus, these covariates together with the variables found significant in the Kaplan-Meier survival analysis are candidate covariates for the multiple covariate Cox regression model. In this study, a model that contains all variables that are significant in the Univariate analysis in relation to time to death of premature infants at the 5 percent level of significance is used in multiple covariate and fit the full multiple covariate Cox PH model including all the significant risk factors.

4.3.2. Multiple Covariate Analysis of Proportional Hazard Models

The aim of model development is to obtain a model that satisfactorily describes the data at hand. For the same purpose, the first step is to select covariates which are important in the study at some relaxed level of significance. In this study, a model that contains all variables that are significant in the Univariate analysis at the 5% percent level of significance is used for multiple covariates. So that these significant variables will be included in the multi-variable model and the remaining covariates not included. Then the full multi-variable Cox proportional hazard model is fitted including all the potential covariates which are significant at 5% level, at the Univariate levels with stepwise selection (conditional LR) method.

Results presented in Table 4.4 indicate the parameter estimates of coefficients $\boldsymbol{\beta}_i$ for the covariates in the final model along with the associated standard error, Wald-statistic, significance level, hazard ratio and 95% confidence interval for the hazard ratio. In order to decide whether or not a variable is significant, the p-value associated with each parameter has been estimated and variables that have p-value less than 0.05 are considered as important variables and hence, interpretable. As can be seen from Table 4.3 in Annex-I, the covariates Prenatal asphyxia, Hyaline membrane disease, Sepsis, Jaundice, gestational age, weight of

infants, Age at admission, Respiratory distress syndrome, Temperature and Antenatal care visit were passed the first filtration of variables for multiple covariate analysis. Survival of premature infants was significantly related with having sepsis, jaundice, prenatal asphyxia, hyaline membrane disease, respiratory distress syndrome, gestational Age at birth and temperature during admission. The values of the Wald-statistic for individual β coefficients support that the estimated values β_i's are significantly different from zero at 5% level of significance for all the covariates in table 4.4. The remaining variables which were used in the single covariate analysis such as antenatal care visit of mother, weight of infants and Age at admission found to be non-significant. This implies that the covariates antenatal care visit of mother, weight of infants and Age at admission were no jointly effect on the time to death of premature infants admitted to NICU rather individual effect on the survival of preterm. The formal tests are applied to the model adequacy and the results are displayed in the section 4.3.4.

Table 4. 4. The parameter estimates, standard errors and the hazard ratios of the multiple analysis of Cox proportional hazards model for selected covariates.

Variables	$\boldsymbol{\beta}$	**SE**	**Wald**	**Df**	**Sign.**	**HR**	**95% CI. for HR**
Sepsis Ref(no)							
yes	0.57454	0.19024	9.12	1	0.003	1.7763	(1.2234, 2.5790)
Jaundice Ref(no)							
Yes	0.8734	0.1728	25.556	1	0.000	2.3950	(1.7071, 3.3603)
Gestational Age Ref(26-28]			10.136	4			
(28-30]	-0.5126	0.3147	3.384	1	0.103	0.5989	(0.3232, 1.1097)
(30-32]	-0.6565	0.3207	0.113	1	0.041	0.5186	(0.2766 , 0.9725)
(32-34]	-0.9711	0.3151	0.074	1	0.002	0.3787	(0.2042, 0.7021)
(34-37)	-0.5927	0.3222	2.409	1	0.066	0.5528	(0.2940 , 1.0395)
Perinatal asphyxia Ref(no)							
Yes	0.4946	0.2277	4.717	1	0.03	1.6399	(1.0495,2.5625)
Hyaline membrane D Ref(no).							
Yes	0.6298	0.1903	10.947	1	0.001	1.8772	(1.2927, 2.7261)
Temperature	-0.1563	0.0635	6.058	1	0.014	0.8553	(0.7552,0.9687)
Respiratory distress S .Ref(no)							
Yes	0.7992	0.1953	16.75	1	0.000	2.2237	(1.5166,3.2605)

*Ref=References, SE=Standard error, HR=Hazard Ratio, CI=confidence interval, D=disease , $\boldsymbol{\beta}$=coefficient

4.3.3. **Checking For the Linearity of Covariates in the Model**

Our next step is to examine the scale of continuous covariates in the preliminary main effects model. In order to examine the scale of continuous covariates, many of techniques are available, all of which is designed to determine whether the data support the hypothesis that the effect of the covariate is linear in the log hazard. Among the techniques, the martingale residuals plotted against covariates to detect for the correctness of the functional form. And, if the linearity assumption is failed, then we need to look transformations that the covariate is linearized in the log hazard. As a result in this study graphical technique of the plots of the martingale residuals are used to assess the linearity of relation to continuous covariate in which the correct functional form is understood. Here, the plot of the continuous variables in the model, Temperature is shown in the Figure 4.4 below. From the plot of martingale residuals versus covariate temperature, the plots do not show systematic patterns or trend, the resulting smoothed plots were approximately horizontal straight lines. Therefore the plots of martingale residual confirm that temperature of premature infants have an approximate linear relationship with the survival time.

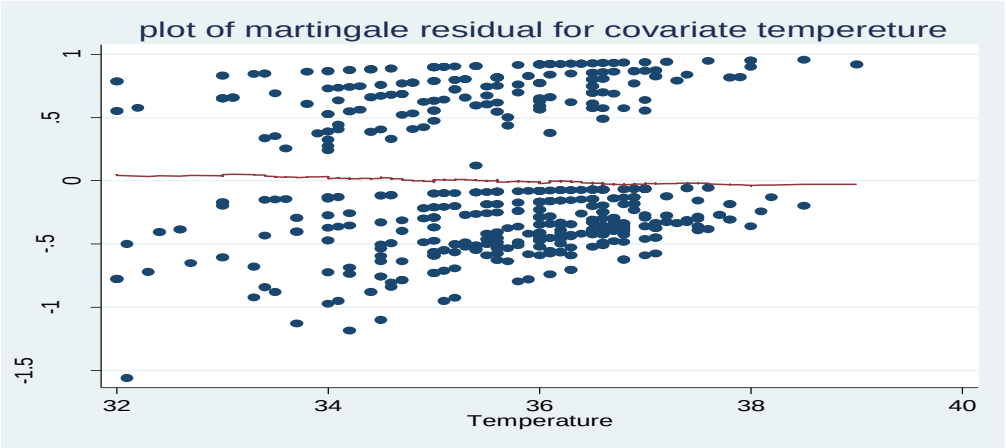


Figure 4. 4. Plots of Martingale residuals for the continuous covariates Temperature

4.3.4. **Assessment of Model Adequacy**

The formal test applied to the model presented in Table 4..5 at ANNEX-I, shows the time-dependent covariates (interaction of covariates with logarithm of time) were not significant for Sepsis, Jaundice, gestational Age, perinatal asphyxia, hyaline membrane disease, temperature and respiratory distress syndrome which justifies the proportional hazard assumption holds at 5% level of significance.

The scatter plots of Scaled Schoenfeld residuals in ANNEX-II Figure 4.5 indicates the scatter of residuals of the covariates distributes in nonsystematic way about the reference line (without definite increment or decrement) and the Loess curve connecting the values of the smoothed residuals is approximately horizontal. This shows that there is no evidence of a departure from the proportional hazards assumption for the covariates that are included in the model. In addition to Schoenfeld residuals plot, the log (-log (survival)) plot versus survival time to check the proportional hazard assumption for all the categorical variables included in the model in ANNEX-II figure 4.6 (a-e), showed that the graphs for each of the categorical variable display lines that appeared to be parallel implying that the proportional-hazards assumption among categorical variable such as prenatal Asphyxia, hyaline membrane disease, sepsis, jaundice and respiratory distress syndrome has not been violated.

Test of proportional-hazards assumption by formal statistical test.

The p-value for testing whether the correlation is zero is the p-value for the statistical test that PH assumption is violated. The null hypothesis is that proportional hazard assumption is not violated. The Table 4.6, shown below indicates the correlation between those significant predictors and rank time. The association between rank time and the covariates were not statistically significant, all the seven covariates (Prenatal asphyxia ,hyaline membrane disease, Sepsis, Jaundice, gestational age, respiratory distress disease ,temperature and gestational Age) of premature infants have p-value greater than 5% indicating that all the covariates satisfy the proportionality assumption at 0.05 level of significance.

Table 4. 6 Test of proportional hazards assumptions

Covariates	Rho	Chi2	Df	Prob>chi2
PNA REF(NO)				
Yes	0.08025	1.18	1	0.2764
HMD REF(NO)				
Yes	0.04938	0.43	1	0.5137
Sepsis REF(NO)				
Yes	-0.05289	0.52	1	0.4728
Jaundice REF(NO)				
Yes	0.15441	4.90	1	0.269
Gestational age REF((26-28])				
(28-30]	0.08628	1.36	1	0.2439
(30-32]	0.02952	0.15	1	0.6986
(32-34]	0.11396	2.18	1	0.1394
(34-37)	0.13836	3.29	1	0.0699
Temperature	0.06424	0.81	1	0.3674
RDS REF(NO)				
Yes	0.0806	1.34	1	0.2469
Global test		20.15	12	0.0643

* REF=References , Rho=correlation

Identification of influential and poorly fit subjects

The residuals plots can be used to check the linearity assumption and to check the influential and outlier observations. As can be observed in Figure 4.7 below some infants have a large spike and these infants are suspected to have undue influence on the parameter estimates. To check their influence the suspected subjects were removed one at a time and model is refitted. There were no large change in the model estimates and hence these infants are not as such influential outliers and then retained in the model.

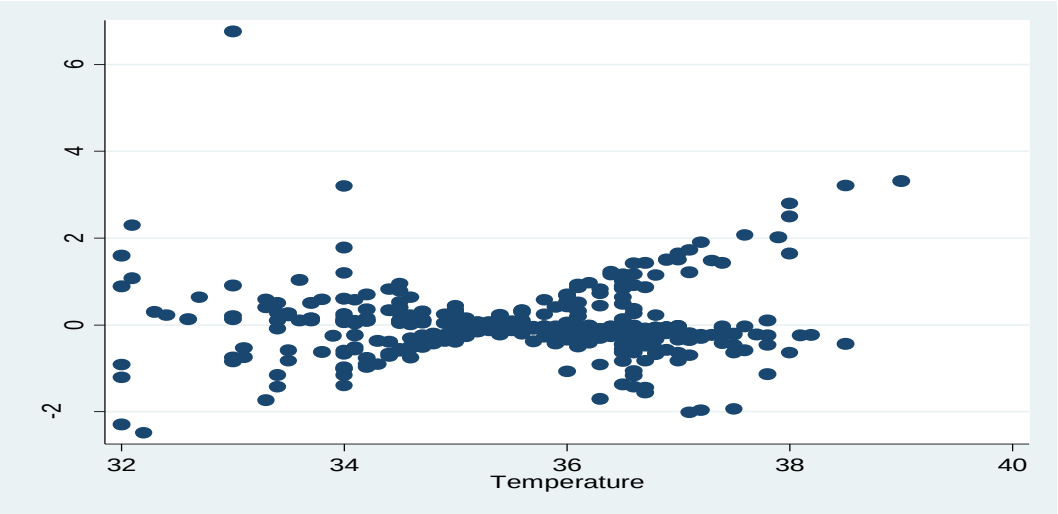


Figure 4 7 Plots of score residuals for temperature of infants at admission to detect the existence of influential observation in Cox proportional hazards model.

4.3.5. **Checking for Overall goodness of fit**

In the final model with default method the initial log likelihood function was -2log likelihood =1936.112 and after the covariates are incorporated the log likelihood function becomes -2loglikelihood =1780.659. From the overall tests of model coefficients in table 4.7 the overall(score) provides score test for simultaneously assessing the effects of the parameters in the model. It was found that the seven covariates contribute significantly in explaining the variability in the survival of premature infants admitted to NICU (p-value=0.000). The likelihood of the data and the null model compared via chi-square statistics. The LL_o of the base line or null model =-0.5*1936.112 has significant improvement to the final model with LL_β of the final model=-0.5*(1780.659).The R^2 is calculated as:-

$R^2 = 1 - \left\{ Exp\left[\frac{2}{n}\left(LL_0 - LL_\beta \right) \right] \right\}$ where LL_0-log likelihood without covariate and LL_β –log likelihood for the respective covariates.

$$R^2 = 1 - \left\{ Exp\left[\frac{(1780.659 - 1936.112)}{490} \right] \right\} = \mathbf{0.272.}$$

A perfectly, adequate model has low R^2 due to high percent of censored data (cox, 1972). Thus, the model fitted in this study the value R^2 statistic is 0.272 implying a good fit of the model. In addition to R^2, the results of likelihood ratio test (chi-square=155.45, p<0.00001) , score test (chi-square=176.36, p<0.00001) and plot of cox-Snell residuals for assessing the fit of cox model suggests that model is in good fit, i.e significant at 5%level of significance and close to straight line respectively. Thus, all in all we can say that our model fits the data very well.

Table 4. 7 Overall tests of goodness of fit.

-2loglikelihood	Overall (score)			Likelihood ratio test		
(model)	Chi-square	Df	p-value	Chi-square	Df	p-value
1780.659	176.36	10	0.0000	155.45	10	0.000

-2loglikelihood of null model=1936.112

The plot of the Nelson-Aalen estimate of the cumulative hazard function of the cox-Snell residual against the cox-Snell residuals is presented in figure 4.8 below. It can be seen that the plot of estimated cumulative hazard rates versus cox Snell is fairly close to the 45^0 straight line through the origin. Thus, the plot is evidence that the model fitted to the data is satisfactory.

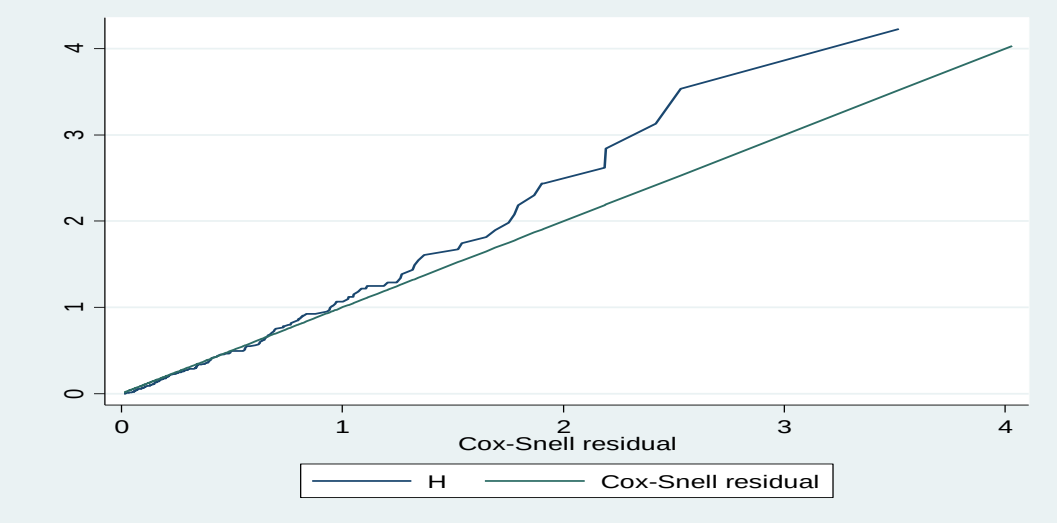


Figure 4 .8 Cox-Snell Residuals for assessing the fit of a cox model

4.3.6. **Interpretation and Presentation of the Final Cox Proportional Hazard Model**

The model that fit to the premature infants' data in table 4.4 has one continuous linear covariate (temperature) and six categorical covariates (Sepsis, jaundice, gestational age, prenatal asphyxia, hyaline membrane disease and respiratory distress disease) of premature infants. The model adequacies are presented in section 4.3.4. It suggested that the model is in good fit. Thus, the cox regression coefficients and hazard ratio in the final model are interpreted as follows.

Hazard ratio having 95% CI for premature infants admitted to NICU who had Sepsis, Jaundice, Prenatal asphyxia, Hyaline membrane disease and Respiratory compared to those infants who hadn't Sepsis, Jaundice, prenatal asphxia ,Hyaline membrane disease and Respiratory distress syndrome were 1.7763 (1.2234,2.5790), 2.3950(1.7071,3.3603), 1.6399(1.0495,2.5625), 1.8772(1.2927, 2.7261) and 2.2237 (1.5166,3.2605), respectively. That is, the risk of death for premature infants those who had Jaundice and Respiratory distress syndrome were 2.3950 and 2.2237 times higher than those infants without jaundice and Respiratory distress syndrome respectively. Premature infants who had Sepsis, Prenatal asphyxia and Hyaline membrane disease were 77.63%, 63.99% and 87.72% more likely to die than those infants without Sepsis, Prenatal asphyxia and Hyaline membrane disease respectively.

The hazard ratio (95% CI) for premature infants admitted to NICU those who were born in between gestational age (28-30], (30-32], (32-34] and (34-37) weeks compared to those who were born at gestational age (26-28] were 0.598 (0.3232, 1.1097), 0.5186 (0.2766 ,0.9725), 0.3787(0.2042,0.7021) and 0.5528(0.2940 , 1.0395), respectively. That is, Premature infants who were born in between gestational age (28-30], (30-32], (32-34] and (34-37) weeks were 40.2%, 48.14% ,62.13% and 44.72% less likely to die than to those infants who were born in the interval (26-28] weeks of gestational age respectively.

Moreover, by letting other covariates constant, the hazard ratio (95%CI) of temperature at admission for premature infants were 0.8553(0.7552, 0.9687) at NICU. This means, the hazard ratio for a one unit increase in temperature is around 85.53%, so that increasing infants

temperature ,such that infants temperature goes up by one leads to a reduction in risk of death of 14.47%% among premature infants admitted to NICU survivors.

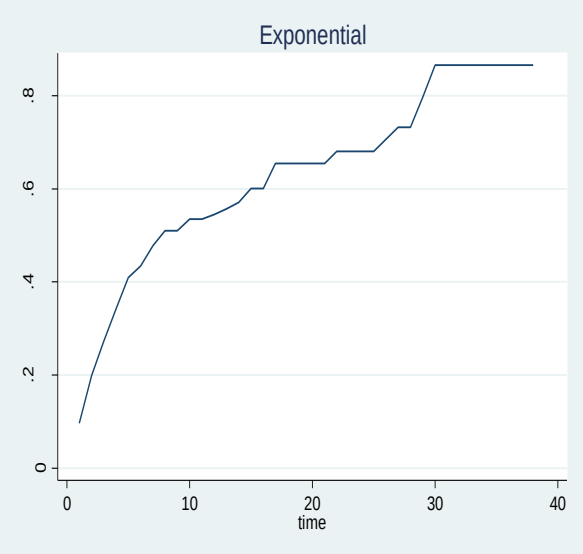
4.5. Parametric Regression Modeling For Time to death of Premature infants

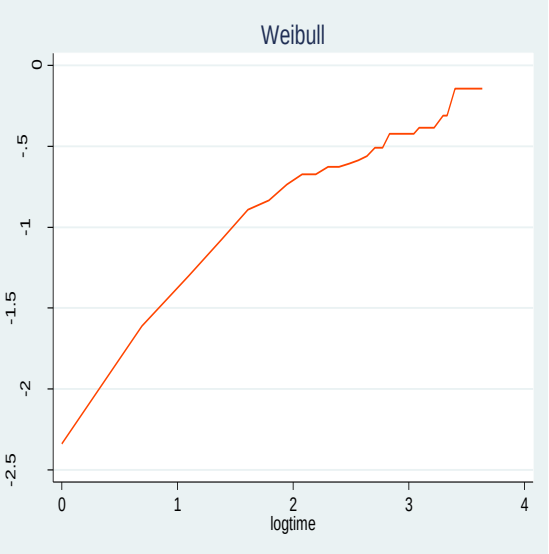
4.5.1. Model selection for survival time of premature infants

For the data on premature infants the parametric regression models were fitted. The graphical assessment of the parametric baseline distribution assumption of Figure 4.9 found below, indicates that the log logistic regression model of log failure odds with the logarithm of time relatively linear than other graphs. But graphical methods may not assure the result. The common applicable criterion to select the model is the Akaikie information criterion (AIC) statistic proposed by Akaikie (1974). From Table 4.8, the log logistic regression model had the least AIC and BIC value which shows that the log-logistic baseline distribution fit the data of premature infants in neonatal intensive care unit.

Table 4. 8 The AIC and BIC value for different parametric regression model

Model type	Exponential	Weibull	Log-logistic
Log-likelihood	-413.043	-412.295	-408.861
AIC Value	848.087	848.589	841.722
BIC Value	894.225	898.922	892.055





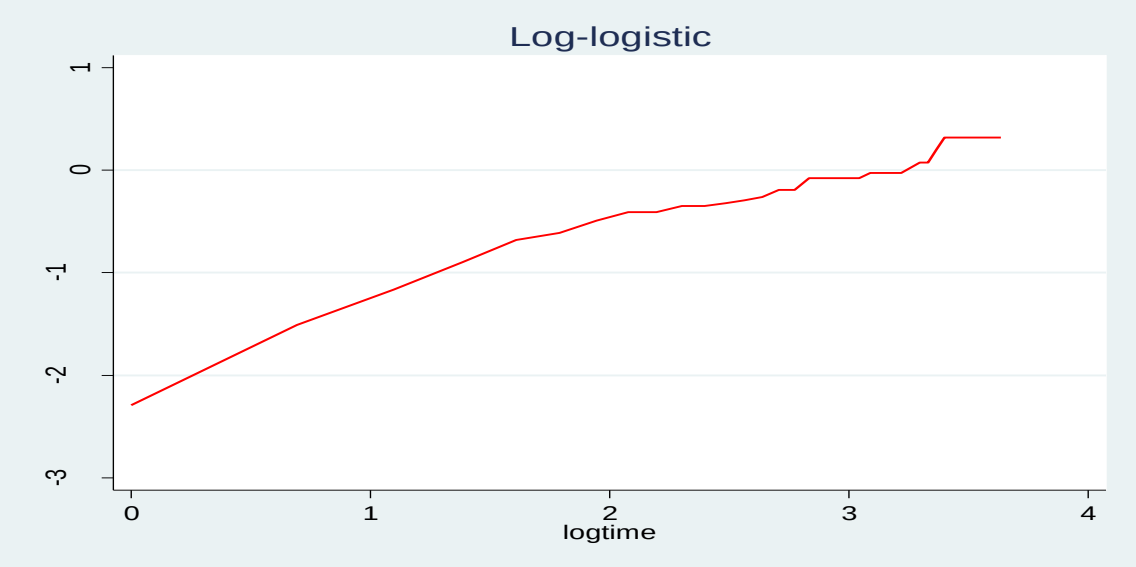


Figure 4. 9 Graphical evaluations of the exponential, Weibull and log-logistic assumptions

From the ANOVA of Table 4.9 below, the risk factors for the survival of infants with prematurity are Sepsis, Jaundice, Gestational age, prenatal asphyxia, hyaline membrane disease, Temperature and respiratory distress syndrome using a level of significance 5%.

Table 4. 9 ANOVA table for covariates in the log logistic regression model

Covariate	**Df**	**Deviance**	**Res.Df**	**-2*LL**	**P(>chi2)**
Null	NA	NA	488	1336.682	NA
Sepsis	1	6.7585	487	1329.923	0.0093
Jaundice	1	73.789	486	1256.134	0.0000
Gestational age	4	20.808	482	1235.326	0.0003
PNA	1	4.8663	481	1230.460	0.027
Temperature	1	9.011	480	1221.449	0.0026
HMD	1	27.7012	479	1193.747	0.000
RDS	1	23.6381	478	1170.109	0.0000

NA=not applicable, Df=degrees of freedom, LL=Log Likelihood

4.5.2. Multiple covariate Analysis of log logistic Regression Model

In order to decide whether or not a variable is significant, the p-value associated with each parameter has been estimated and variables that have p-value less than or equal to 0.05 are considered as important variables and hence, interpreted. The relationship between covariates and survival probability of premature infants modeled by log logistic regression model are presented in Table 4.10 below. As can be seen from this Table, survival of the infants was significantly related with having sepsis, Jaundice, prenatal Asphyxia, hyaline membrane disease, respiratory distress syndrome, with gestational age at birth and temperature at admission. The Wald test for the parameter estimates indicates that the coefficient of the parameters in each covariate is significantly different from zero at 0.05 levels of significance. The formal tests are applied to the model adequacy and the results are displayed in section 4.5.3.

Table 4. 10 Parameter estimates, standard errors and the hazard ratios in the final log logistic regression model

Covariates	**β**	**SE**	**Wald**	**OR**	**Sign.**	**95% CI for OR.**
PNA yes	0.9077	0.3538	6.554	2.479	0.01	(1.239,4.9585).
HMD yes	0.9694	0.2557	14.364	2.636	0.0001	(1.5971,4.3517)
Jaundice yes	1.007	0.2376	17.893	2.737	0.000	(1.7182,4.361)
Sepsis yes	0.7286	0.2614	7.784	2.072	0.005	(1.2415,3.4587)
Gestational age (28-30]	-0.6724	0.4586	2.161	0.510	0.143	(0.2078,1.2543)
Gestational age (30-32]	-1.09	0.4559	5.712	0.336	0.017	(0.1376,0.8217)
Gestational age (32-34]	-1.420	0.4541	9.797	0.241	0.002	(0.0993,0.5886)
Gestational age (34-37)	-0.8323	0.4519	3.386	0.435	0.065	(0.1794,1.0549)
Temperature	-0.2092	0.0881	5.627	0.811	0.018	(0.6826,0.9642)
RDS yes	1.1901	0.2451	23.620	3.287	0.000	(2.033,5.3148)
Log(scale)	0.2772	0.0780	12.674	1.3194	0.000	(1.1324.,1.5374)
Scale	0.801	0.0779				(2.012,2.4571)

Using the regression model of equation (37) and with the parameters found, the survival time of premature infants admitted to NICU have log logistic distribution, The log-logistic

distribution is very similar in shape to the log-normal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameter λ is the scale parameter and ρ is the shape parameter, which can be expressed as time ~Log-logistic (λ, p). The default output for parametric provide maximum likelihood estimates of intercept v, and scale parameter δ, associated with the logistic distribution. The parameters of the underlying log logistic distribution are the functions of these extreme value parameters, λ=exp(-μ/p)=0.952245 and p=1/α =1.248439 where μ=exp(v)=0.06081. Then time ~log-logistic (0.952245, 1.248439) have hazard rate of $h_o(t) = \frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}} = \frac{1.18882t^{0.248439}}{1+0.952245t^{1.248439}}$ .The log-logistic regression model that predicts the survival of premature infants admitted to neonatal intensive care unit with identical data settings were:

$$h(t, x,\beta) = \frac{1.18882t^{0.248}\exp(X\beta)}{1+0.952245t^{1.248\exp(X\beta)'}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.[47]$$

In parametric settings, except for exponential regression models the baseline function is not proportional for all subjects as a case of cox regression model. For the log logistic regression model the base line hazard will vary with $h_o(t) = \frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$. Therefore the base line hazard function of premature infants admitted to NICU at Jimma university specialized hospital was with formula of (33) in every increase in time measured in days:

$$h_0(t) = \frac{1.18882t^{0.248}}{1+0.952245t^{1.248}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..[48]$$

4.5.3.Assessment of Adequacy of the log-logistic Regression Model

I. Graphically assessment of log-logistic parametric regression model Assumptions

The log-logistic assumption can be graphically evaluated by plotting ln(1-S(t))/(S(t)) against log of time where S(t) are the Kaplan–Meier survival estimates. From the figure 4.10 below, the survival time of premature infants admitted to NICU follows a log logistic distribution since the resulting plots approximately a straight line with positive slope. Hence, the log logistic assumption was not violated. Also, from the figure 4.11 found ANNEX-II showed the graphical assessment of proportional odds that indicates parallel curves support the

proportional odds (PO) assumption, and. If the log-logistic and PO assumptions hold, then the AFT assumption also holds.

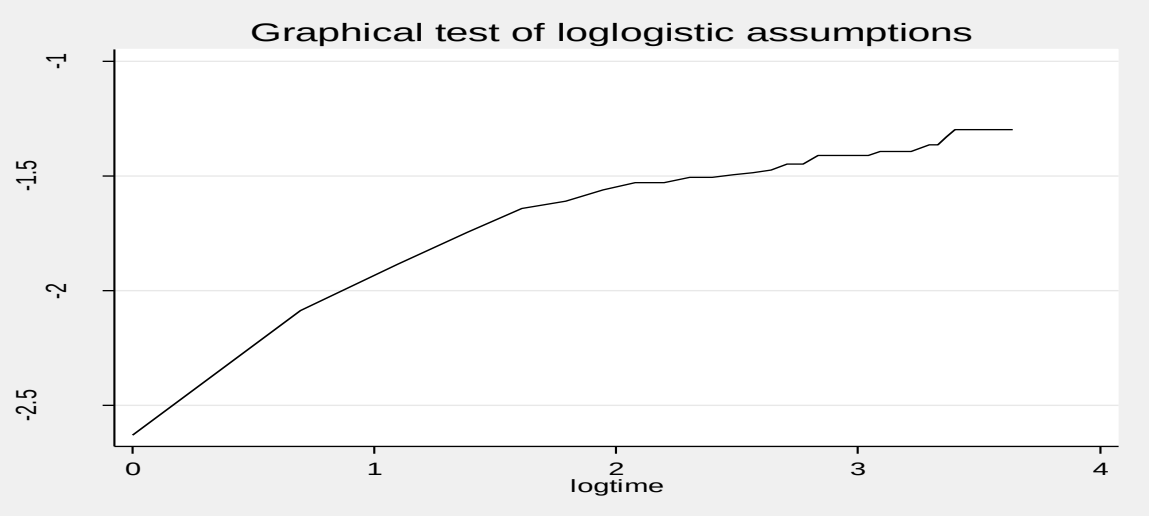


Figure 4 10 Graphical evaluations of log-logistic assumptions

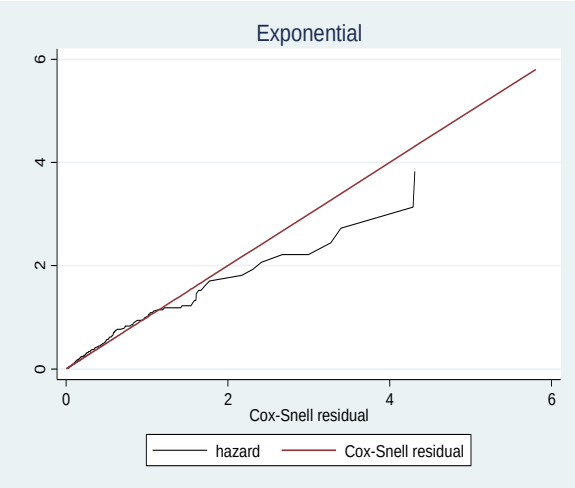
From the likelihood ratio test Table 4.11 below, it can be seen that the model is significant and in using the log likelihood values of the null model and the full model it can be seen that the model has a significant improvement after the covariates are added in the model.

Table 4. 11 The likelihood ratio and significance of the log logistic regression model

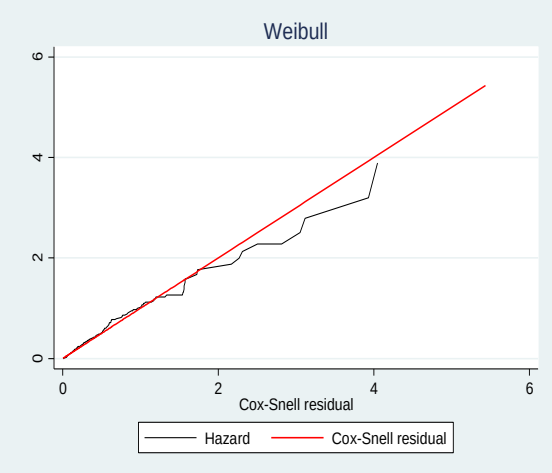
Log-lik(intercept only)	Loglik (model)	Chi-square	Df	Sign.	Scale	Intercept
-668.3	-585.1	166.57	10	0.0000	0.801	-2.800

II. The Cox Snell Residual Plots

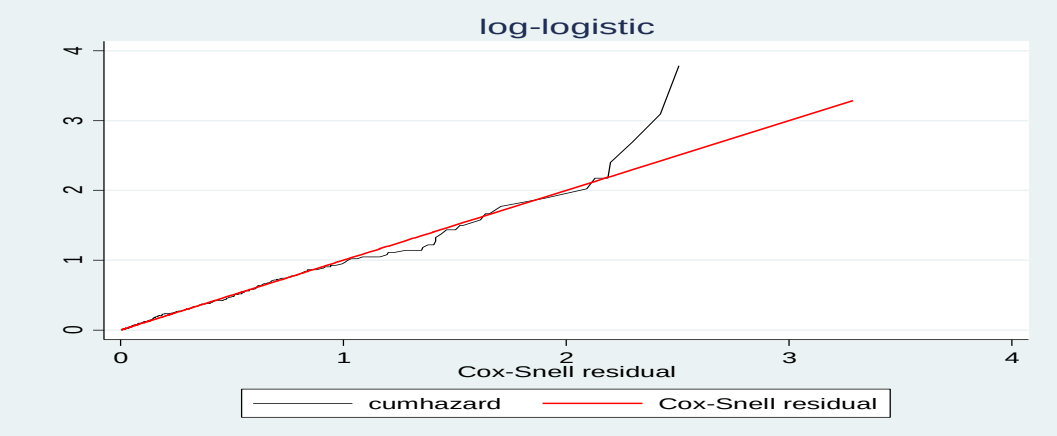
The Cox-Snell residuals using the exponential, weibull and log-logistic models to our data via maximum likelihood estimation found in Figure 4.12 below. The plot shows that the line related to the Cox-Snell residuals of the log-logistic models were nearest to the line through the origin, again indicating that this model describes the premature infants' dataset well. This result support the result obtained from the log failure odd plot in figure 4.10.



a) The plot of Cox–Snell residuals for exponential survival model



b) The plot of Cox–Snell residuals for Weibull survival model



c) The plot of Cox–Snell residuals for Log-logistic survival model

Figure 4 12 Plots of parametric survival models to examine models that fit the data better.

4.5.4. Interpretation of the log logistic Regression Model

Results presented in Table 4.10 before , indicates the parameter estimates of coefficients β_i for the covariates in the final log-logistic regression model along with the associated standard error, significance level, odds ratio and 95% confidence interval for the odds ratio. Survival time of premature infants were significantly related with having sepsis, Jaundice, prenatal Asphyxia, hyaline membrane disease, respiratory distress syndrome, with gestational age at birth and temperature at admission as shown on the same table. The Wald statistics for the parameter estimates indicate that at least one of the parameters in each covariate level is

significantly different from zero at 0.05 levels of significance. The estimated value of p=1/0.801=1.248, greater than 1, indicates a non-monotonic hazard function, namely among the premature infants death, the hazard rate tends to increase initially and then decrease as time progresses. The odds ratio (OR) of each covariate is the multiplicative effect on the odds of survival among the premature infants admitted to NICU.The relative odds (95% CI) of premature infants who had prenatal asphyxia (PNA), hyaline membrane disease (HMD), Jaundice, Sepsis and Respiratory distress syndrome as compared to premature infants who had no Prenatal asphyxia, Hyaline membrane disease, Jaundice,Sepsis and Respiratory distress syndrome were 2.479 (1.239,4.9585),2.636 (1.5971,4.3517), 2.737 (1.7172,4.361), 2.072(1.2415,3.4587) and 3.287(2.033,5.3148) respectively. That is, premature infants who had prenatal asphyxia, Hyaline Membrane distress, Jaundice, Sepsis and Respiratory distress syndrome have 2.479,2.636, 2.737, 2.072 and 3.287 times higher odds of dying than premature infants who had no prenatal asphyxia, Hyaline membrane disease, Jaundice, Sepsis and Respiratory distress syndrome respectively.

The odds ratio (95% CI) premature infants who born at gestational age group of (30-32]and (32-34] weeks as compared to premature infants born at gestational age of (26-28] weeks were 0.336(0.1376,0.8217) and 0.241 (0.0993,0.5886) respectively. This indicates the relative odds of survival for premature infants born at gestational age (30-32] and (32-34] weeks have 66.4% and 75.9% times lesser odds of dying than premature infants born in 26-28 weeks.

A one degree increase in temperature would lower the individual odds of survival by 18.9%, other covariates being constant. From the log logistic regression model, temperature of the premature infants admitted to NICU decreases the Odds ratio of the premature infants by 81.11%(OR=0.811,95%CI=0.6826,0.9642), that is, for every one degree centigrade increment in the temperature of preterm, the survival odds reduced by 18.9% controlling the effects of all other covariates in the model.

4.6. DISCUSSION ON THE RESULTS

The main aim of the study was modeling the determinants of time-to-death of premature infants admitted to neonatal intensive care unit in 2013-2015 using cox proportional hazards and parametric with suitable three baselines parametric distributions. Also, the study tries to estimate and compare the survival time of premature infants in Jimma University Specialized Hospital and to determine the important predictive factors on premature infants using the 2013-2015 hospital data. From the estimates in Univariate analysis, we found that the survival of an infants is significantly related with prenatal Asphyxia, Sepsis, hyaline membrane disease, jaundice, Gestational age, weight at admission, antenatal care visit of mother, age at admission, temperature and respiratory distress syndrome. Then this variable is the candidate for the multiple covariate analysis that jointly serve as predictive factors on the survival of premature infants.

The Cox's proportional hazard model fitted using complete case analysis found seven variables that jointly serve as predictive factors on the survival of premature infants From both semi-parametric (cox PH) and parametric models i.e., the three baseline parametric distribution (exponential, Weibull and log-logistic) were employed to examine the factors that determine survival of premature infants. Factors that are concerned for our study were prenatal Asphyxia, Sepsis, hyaline membrane disease, jaundice, Gestational age, temperature and respiratory distress syndrome. The multiple covariate analysis given in table 4.5 revealed that all of these factors were significantly related to death of premature infants. But weight at admission, antenatal care visit of mother and age at admission were non-significant. This means the weight at admission, antenatal care visit of mother and age at admission have no jointly predictive effect on the survival time of premature infants admitted to NICU. The comparison of base line distributions of the models was done using the AIC criteria, where a model with minimum AIC is accepted to be the best (Akakie, 1974). Accordingly, model with log-logistic baseline distribution which had AIC value of 841.722 was the most appropriate model over exponential and weibull to describe the premature infants' data set and graphical evidence (figure 4.12).

The overall mean and median survival time of premature infants admitted to neonatal intensive care unit were 21.227 and 27 days respectively, where 490 infants admitted to NICU of 171(34.9%) deaths in Jimma University Specialized Hospital and that confirm in comparison to the study done by Fakher *et al.,* (2005), 48% death of preterm infants admitted to NICU in Fawzy Moaz Hospital, Egypt. Luiz Fernando *et al*., (2010), reported from 495 newborns, with 129 deaths (26.1%) for infants admitted to neonatal intensive care unit of Taubate University Hospital, Brazil. Of the 138 preterm, 76 (55.1%) were discharged, 47 (34.1%) died, while 15 (10.9%) were discharged against medical advice (DAMA). Twenty six of the 47 patients who died were males while 21 females with a male to female ratio of 1.3:1 by Kunle *et al.,* (2014), southern Nigeria. In our country Ethiopia, it is a serious problem as compared in developed countries. Merertu *et al.,* (2013), 30.9% deaths of all preterm births admitted to Tikur Anbessa Hospital, Addis Ababa and Habtamu *et al.,(*2013) indicates that the mortality rate was 15.9% in Jimma University specialized Hospital in the period of Jan 2012 to Dec 2012.

The first factor that affects survival time of premature infant admitted to neonatal care unit is prenatal asphyxia of neonate. As it was indicated both in cox proportional hazard models and log-logistic regression models the hazard rate of premature infants admitted to NICU who had prenatal Asphyxia is about HR=1.6399 and OR=2.479 times higher than premature infants who had no prenatal Asphyxia respectively. This result is in accordance with the studies by lawn *et al*. (2005). Other studies like studies of Black *et al.* (2010).

Early onset of disease (Sepsis) of premature infants is a prognostic factor that significantly predicts survival of time to death of premature infants. The multiple covariate analysis of cox proportional hazard showed that, hazard rate of having Sepsis is 77.63% higher than premature infants who had no Sepsis and log logistic parametric model also indicates that the proportional odds of having sepsis is much higher for infants those had Sepsis.(OR=2.072). The result is comparable with earlier study (Liu *et al*., 2012; Black *et al*., 2010) at global level. Habtamu *et al*.(2013).

Our findings showed that Gestational age levels are significant effects on time to death of premature infants admitted to neonatal intensive care unit in JUSH. The results of log rank

and generalized Wilcoxon test showed that there is a survival experience differences among the category of gestational age for premature infants. The hazard rate from cox PH and proportional odds from log-logistic suggested that the lower gestational age have shortened survival time than others. This result consistent with Merertu *et al*., (2011), Azizah et al (2009) in university Malaya medical center, found that gestational age is one of the potential factors for survival of time to death of premature infants admitted to NICU. The study carried out by Srinivas *et al*.,(2015) reported that hospital survival rates based on gestational age alone were 27%, 59%, 76%, 85%, 91% and over 95% at 23, 24, 25, 26, 27 and 28–31 weeks, respectively in Sydney, Australia and Behnaz *et al.* (2015) neonatal death rate was overall 27.4% which was significantly higher in gestational age of less than 28 weeks compared with other gestational age subgroups in Fatemieh Hospital, Hamadan, Iran.

The results of this study suggested that hyaline membrane disease was significant predictive factor for time-to-death of premature infants admitted to neonatal intensive care unit. The mean survival time of premature infants who had hyaline membrane disease were 21.319 [10.08,14.55] days. Infants who had hyaline membrane disease have hazard ratio 1.8772(1.2927, 2.7261), that is the risk of death for premature infants who had hyaline membrane disease was 87.72% higher than infants who hadn't hyaline membrane disease admitted to NICU. The log-logistic parametric regression model also suggested that the proportional odds of survival were 2.636 times higher odds of dying than premature infants who had no hyaline membrane disease (or 37.94% less likely of risk of dying for premature infants who had no hyaline membrane disease) admitted to NICU. This result supported by study carried out in ST. George , Germany on prevention of hyaline membrane disease in preterm infant conclude that Hyaline membrane disease is the leading cause of death for infants.(Anna *et al.*,2016.).

From this study, having Jaundice disease is an important predictor of time to death of premature infants admitted to NICU. The log rank and generalized Wilcoxon showed that there is a significant difference between premature infants of having jaundice disease and those who hadn't the disease. Both cox proportional hazard and log logistic parametric models suggest that the jaundice disease have significantly associated with premature infants' death. Premature infants who had jaundice have hazard rate of 2.395 (CI=1.7071, 3.3603),

higher than Premature infants who hadn't jaundice disease. The estimated mean survival time preterm were 6.715 (CI=6.74, 10.67) days which is less than the premature infants who had no jaundice disease. This is similar to reports by Khan *et al*.,(2012) in Karachi, Pakistan who reported jaundice and sepsis as the commonest morbidities in their preterm patients. However, Onwuanaku *et al (2013)* in Jos University Teaching Hospital Nigeria, reported sepsis as the commonest morbidity, followed by jaundice. They recommended that is an urgent need for the prevention and adequate management of jaundice in this vulnerable group.

In our study in both models, cox proportional and log-logistic regression models, the initial body temperature infants was significantly associated with reducing the survival probability of premature infants. The estimated median survival time of premature infants' temperature was 37.1 with maximum of 39 degree centigrade for premature infants' admitted to neonatal intensive care unit. Cox proportional hazard and log-logistic parametric regression model were suggested the hazard ratio and odds ratio of 0.8553 and 0.811 respectively. This indicates that one unit increase in initial temperature of premature infants reduces the risk of death by 14.47% and 18.9% respectively. This result confirmed with study carried out by Merertu *et al*.,2013 reported that Temperature at admission has strong association with survival in a multivariate model adjusted for gestational age and birth weight of neonatal and With the neonates whose temperature at admission was less than or equal to 33°c have a 5.43 times higher risk of death compared to those with T° at admission was between 36.50 - 37.5°c.

Respiratory distress syndrome is one of the potential risk factors that affect the survival time of premature infants admitted to NICU The result of mantel-heanszal log-rank and generalized Wilcoxon test showed that there is a significance survival probability differences with premature infants' who had respiratory distress syndrome disease and hadn't the disease(p-value=0.000). Results of cox proportional hazard and log-logistic parametric regression model showed that the respiratory distress syndrome disease have significantly associated with survival time of premature infants admitted to NICU. This study confirmed with the study carried out by Behzan, (2015); concluded that the death rate of respiratory distress syndrome (73.8%) is higher in premature infants who had the disease, which predict occurrences of death in premature infants.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The objective of the study was to identify significant risk factors that affect survival of premature infants who admitted to neonatal intensive care unit at Jimma University Specialized Hospital. For determining the risk factors for the survival of premature infants and modeling the survival time, a total of 490 preterm were included in the study out of which 45.3% were females and 54.7% were males. Among those preterm 34.9% were died and the rest were censored. The estimated median survival time of hospital stay for premature infants was 27 days. Both Cox proportional hazard and parametric model with baseline log-logistic distribution analysis showed that the major factors that affect the survival to time of preterm are prenatal Asphyxia, Sepsis, hyaline membrane disease, jaundice, Gestational age, temperature and respiratory distress syndrome for time to death of premature infants. Preterm infants having prenatal asphyxia, Sepsis, Jaundice, Hyaline membrane and respiratory distress syndrome have higher death rate. Similarly, preterm with poor health indicators like lower gestational age (26-28) weeks and initial temperature, were less likely to survive.

To predict and model the survival time of premature infants, various baseline parametric regression models were applied. Based on the log-likelihood, AIC ,BIC and R^2 values the parametric model with log-logistic baseline distribution is better fits to predict the survival time of the premature infants' for the data of preterm at Jimma University Specialized Hospital than the other parametric models.

5.2 Recommendations

Based on the result of the study different factors are identified for the death of premature infants, The following recommendations are made for health policy makers, clinicians and the public at large:

i. According to the results of this study the main predictive factors for the survival time of premature infants are more of clinical variables. So, health workers should be cautious when mother's born preterm which has perinatal asphyxia, Sepsis, Jaundice, hyaline membrane disease and respiratory distress syndrome.
ii. The medical managements better to arrange a program for continuous training to the medical staff for better assessment, diagnosis and management of premature cases.
iii. There medical managements should facilitate more research to find out more precise diagnosis of causes of premature infants and maternal adverse outcome with better computerized recording system.
iv. The log-logistic regression model provides better predictions to the survival probability of premature infants'. So, future researchers could make use of this model.
v. Further studies should be conducted in each Hospital of Ethiopia and identify other factors that are not identified in this study.

Limitation of the study

This study had some limitations: the first is that the study used data from single hospital. Thus, the findings of this study should be interpreted very carefully when they are inferred to the national level. The second limitation is lack of published literature on our country related to the survival time of premature infants' the references are more of other countries outcome. Finally as different literature pointed out, there are different factors that are assumed to have impacts on the survival of premature infants related to mother of preterm such as parity, gravidity, age at marriage , educational level of mother and HIV status. However, data on these variables could not be available in neonatology clinic since the neonatology clinic is separated from maternity ward, so these variables were not integrated in this study.

REFERENCES

Aalen, O. (1994). Effects of Frailty in Survival Analysis. *Statistical Methods in Medical Research. Biometrics 3:* 227-243.

Aalen, O.,2008. Survival and event history analysis:. *a process point of view. Springer* Science *& Business Media*.

Abdulkarimova, U. (2013). Frailty Models For Modelling Heterogeneity. *Digital commons at Mcmaster Open Access Dissertations and Theses 1:*1-75.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on* Automatic Control *19* , 716–723.

Anna Mihaylova, Ekaterina Uchicova, Nikoleta Parahuleva, Mariana Parahuleva (2016) prevention of hyaline membrane disease (HMD) in preterm infants, St. George "University Hospital Plovdiv, Germany.

Beck, S., Wojdyla, D., Say, L., betran, A.P., Merialdi, M., Requejo, J.H., *et al.* (2010) The Worldwide Incidence ofPreterm Birth: A Systematic Review of Maternal Mortality and Morbidity. *Bulletin of the World Health Organization*, **88**, 31-38.

Behnaz B.; Farzaneh E.and Maryam Sh., (2015),Neonatal Mortality and its Main Determinants in Premature Infants Hospitalized in NICU in Fatemieh Hospital, Hamadan, Iran.

Behrmanet. Al., N. (2004). *text book of pediatrics, 17th edition.* Saunders.

BelaYnew W., Teumay Abraha , Getachew G.and Mohamed Kedir (2015). *Effects of* inter pregnancy interval on preterm birth and associated factors among postpartum mothers who gave birth at felegehiowt referral hospital.

Black RE, Cousens S, Johnson HL, *et al*(2010): Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375: 1969-1987.

Breslow, N. E. (1974). *Covariance analysis of censored survival data. Biometrika 30,*89-100.

Clayton, DG(1978), Klein *et al*.: Nielsen *et al*(1992)., Hastie and Tibshirani(1993). (u.d.). A Model For Association In Bivariate Life Tables And Its Application. *In Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika, 65: 141-151.*

Collett, D. (2003). *Modelling survival data in medical research (Second edition). Chapman and Hall/CRC,*. London.

Cox, D. (1972). Regression models and life-table. *Journal of Royal Statistical Society B. 34:* 187-220.

Cox, D. R & Oakes, D. (1984.). *Analysis of Survival Data.*: Chapman and Hall. London.

Dehkordi, B.M., Safaee, M. A., Pourhoseingholi, R., Fatemi, Z., Tabeie and Zali, M. R. (2008) Statistical Comparisons of Survival Models for Analysis of Cancer Data. *Asian Pacifc Journal of Cancer Prev.,* Vol 9: P 417-420.

EDHS. (2011). *Central Statistical Agency, (2011). Ethiopian Demographic and Health Survey, Addis Ababa, Ethiopia,*.

Efron, B. (1977). *The efficiency of Cox's likelihood function for censored data. 72:* 557-565.

Engle, W. A., & Kominiarek, M. A. (2008). Late preterm infants, early term infants, and timing of elective deliveries. *Clinics in perinatology*, *35*(2), 325-341.

Fakher M, Shaaban W, Monein AA, Hassan Z, Fikry MM (2005) Statistical Study Of Preterm Infants Admitted To NICU In Fawzy Moaz 19: 155-158.

Fan, J. and Li, R. 2001. Variable selection via non concave penalized likelihood and it oracle properties, Journal of American Statistical Association. 96, 1348-1360.

FMOH. (2011). Federal Ministry of Health (2011) Saving Newborn Lives in Nigeria. *Newborn Health I the Context of the Integrated Maternal, Newborn and Child Health Strategy. Revised 2nd Edition.*

Gehan, E. A. (1965). *A generalized Wilcoxon test for comparing arbitrarily singly censored samples. Biometrika 52,*, 203.223.

Gholamreza J, Zohreh K, Parvaneh Sadeghi M. Contribution of Maternal Demographic and Medical Factors tothe Risk of Prematurity in Qom Hospitals. Macedonian Journal of Medical Sciences, 2011; 4(1): 60-63.

Goldenberg, R.L., Culhane, J.F., Iams, J.D. and Romero, R. (2008) Epidemiology and Causes of Preterm Birth. *Lancet*, **371**, 75-84.

Gordon C S Smith , Jill P Pell and Richard Dobbie (2003), Inter pregnancy interval and risk of preterm birth and neonatal death: retrospective cohort study, BMJ; 327: 313.

Gupta, R., and Kundu, D.(1999). Generalized exponential and log-logistic distributions.

Habtamu Sime , Netsanet Workneh and Eshetu Girma (2013) Morbidity and Mortality of

Neonates Admitted in Jimma University Specialized Hospital Paediatrics Neonatal Ward: A One Year Retrospective Analysis

Hamilton BE, Martin JA, Ventura SJ(2006) . Births: preliminary data for 2005. Natl Vital Stat Rep; 55: 1–18.

Hosmer DJ, Lemeshow S. (1999). *Applied Survival Analysis:* Regression Modeling of Time to Event Data. (2nd Edition), John Wiley & Sons, Inc,Canada.

Hosmer, D.W. and Lemeshow S. *et al*. (1999). *Applied Survival Analysis. John Wiley and Sons, Inc.*,New York.

Hougaard, P. (2000). *Analysis Of Multivariate Survival Data. Springer,1, 255-273.* Newyork.

Howson CP, Kinney MV, Mcdougall L (2013) Born Too Soon: Preterm Birth Matters. Reproductive Health Journal 10: 1-9.

Jehan *brown, Joy* (2009). Neonatal mortality, risk factors and causes: . *a prospective population based cohort study in urban Pakistan. Bull WHO, Vol. 87,*, pp 130-138.

Jiezhi Qi, 2009 comparison of proportional hazard and Accelerated Failure time ,Journal of the Royal Statistical society series , volume 32, pp 165-171.

Kalbfleisch JD, Prentice RL. (1980.). *The statistical analysis of failure time data.* . New York: Wiley,

Kalbfleisch, J. D., and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60. 267.278.

Kaplan, E., and Meier, P. (1958). Nonparametric estimation from incomplete observations. . *Journal of American Statistical Association 53* , 457.481 .

Khan, M.R., Maheshwari, P.K., Shamim, H., Ahmed, S. and Ali, S.R. (2012) Morbidity Pattern of Sick Hospitalized Preterm Infants in Karachi, Pakistan. *Journal of Pakistan Medical Association*, **62**, 386-388.

Klein J, Moeschberger M. (1997.). *Survival analysis:*Techniques for censored and truncated data;. Springer-Verlag.: New York.

Klein, J. (1992). *Semiparametric estimation of random effects using the Cox model based on the EM algorithm. Biometrics 48, 795-806.*

Kleinbaum D, Klein M. (2005). *Survival analysis:* A self- learning text, Springer-Verlag.: New York.

Lancaster T. (1979). Econometric Methods For The Duration Of Unemployment.

Econometrica 47:. 439-454.

Lawn J. E. ,Cousens S. ,Zupan J.(2005), *4 million neonatal deaths: when? Where? Why? The Lancet* 365: 891-900

Lawn J. E., Wilczynska-Ketende K. and Cousens S. N. (2006), Estimating the causes of 4 million neonatal deaths in the year 2000, Int. J. Epidemiol; 5 (3): 706-718.

Lawn JE, Gravett MG, Nunes T, Rubens E, Stanton C, *et al*. (2010) Global Report on Preterm Birth and Stillbirth (1 of 7): Definitions, Description of the Burden and Opportunities to Improve Data. Bmc Pregnancy and Childhood 10: 1-22.

Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., *et al.* (2012). Global, regional, and national causes of child mortality in 2000-2010: an updated systematic analysis. *The Lancet*, in press.

Luiz Fernando Paula Risso and C. Nascimento2 (2010) , Risk factors for neonatal death in neonatal intensive care unit according to survival analysis 2005-2007, Taubate University Hospital, Brazil.

Maria Teresa Zullini, Maurizio Bonati and Elena Sanvito (1992) on behalf of the Pakistan Collaborative Group on Neonatal Care, Survival at nine neonatal intensive care units in São Paulo, Brazil.

MDG. (2012). Report , AUC, AfDB, UNECA and UNDP, Assessing Progressin Africa toward the Millennium Development Goals.

Merertu.T. Bogale .W, Yonas. R and Amha.M (2011). Survival Of Preterm Infants Admitted To Tikur Anbessa Hospital Nicu, Addis Ababa.

Mohammad, T., Khattak, A.A., Shafiq-Ur-Rehman. (2009) Mortality and Morbidity Pattern in Small-For-Gestational Age and Appropriate-For-Gestational Age Very Preterm Babies: A Hospital Based Study. *Journal of Ayub Medical College Abbottabad*, **21**, 16-21.

Morita, J.G.,T.H. & Mowday, R.T. 1989. Introducing survival analysis to organizational researchers: A selected application to turnover research, journal of Applied Psychology, 74:280-292

Moutquin JM, Milot-Roy V, Irion O (1996). Preterm birth prevention: Effectiveness of current strategies. Journal SOGC;18:6:571-585.

National Institutes of Health (1994) Effect of corticosteroids for fetal maturation on perinatal outcomes. NIH Consensus Statement;12:2:1-23.

Offiah I, O'Donoghue K, Kenny L (2012) Clinical Risk Factors for Preterm Birth 3: 74-91.

Onwuanaku, C.A., Okolo, S.N., Ige, K.O., Okpe, S.E. and Toma, B.O. (2011) The Effects of Birth Weight and Gender on Neonatal Mortality in North Central Nigeria. *BMC Research Notes*, **4**, 562.

Pennell, C.E., Jacobsson, B., Williams, S.M., Buus, R.M., Muglia, L.J. and Dolan, SM. (2007) Genetic Epidemiologic Studies of Preterm Birth: Guidelines for Research. *American Journal of Obstetrics & Gynecology*, **196**, 107-118.

Pourhoseingholi, M.A., Vahedi, M., Dehkordi, M.B., Safaee, A. and Ashtari, S. (2010). Alternative for the Cox Regression Model: Using Parametric Models to Analyze the Survival of Cancer Patients. *Iran J. Cancer Prev.*, Vol 1: P 1-9.

Rehana T.(2006). Preterm Delivery. *A Major Predictor of Perinatal Morbidity and Mortality. Jpm 20,* : 279-283.

Risso SDP, Nascimento LFC (2010) Risk Factors for Neonatal Death in Neonatal Intensive Care Unit According to Survival Analysis. Universidade Detaubate Unitau Taubaté (Sp) Brazil 22: 9-26.

Santa EJ, Cox T. (2006). *Preterm Birth: Causes, Consequences, and Prevention.*

Schoenfeld, D. (1982). *Partial residuals for the proportional hazards regression model.*

Srinivas Bolisetty, Nele Legge, Barbara Bajuk3 and Kei Lui1, (2015), Preterm infant outcomes in New South Wales and the Australian Capital Territory. Sydney, New South Wales, Australia.

Stevens, V. J., Hollis, J. F. 1989. Preventing smoking relapse using an individually tailored skills training technique. Journal of Consulting and Clinical Psychology; 57:420–24.

TherneauT. and P.M. Grambsch. (2000). *Modelling Survival Data:* Extending the Cox Model. Springer-Verlag, New York.

Tigist B ,Abdela A. and Gebreslasie KZ (2013).Preterm Birth and Associated Factors among Mothers Who gave Birth in Debremarkos Town Health Institutions, Ethiopian

UNICEF. (2012). *Maternal, Newborn and Child Survival, Country Profile, Ethiopia. .* Statistics and Monitoring Section /Policy and Practice.

World Health organization (WHO), 2012, Born Too Soon: In The Global Action Report On Preterm Birth.: Howson, Kinney M. & Lawn E. (Eds.). Geneva.

ANNEXIES

ANNEX-I

Table 4.3. Univariate analyses of cox regression for preterm survival time on important socio-demographic, health and clinical characteristics of premature infants in JUSH, 2013-2015.

Variable /levels	Coef	Wald test	SE (coef)	Z	95% CI	p-value
PNA REF(no)						
Yes	0.8281	15.53	0.2101	3.941	(1.516 3.455)	0.0001*
HMD REF(no)						
Yes	1.1053	48.54	0.1586	6.967	(2.213 4.121)	0.0000 *
Jaundice REF (no)	1.4092					
Yes		84.02	0.1537	9.166	(3.028 5.532)	0.0000*
Sepsis REF (No)						
Yes	0.4207	6.39	0.1664	2.529	(1.099 2.11)	0.0114 *
Gestational Age						
REF (26-28]						
GAge[28-30)	-0.3778		0.2962	-1.275	(0.3836,1.2248)	0.2022
GAge[30-32)	1.1442	43.38	0.3052	-3.749	(0.1751,0.5793)	<0.0002 *
GAge[32-34)	-1.3986		0.3067	-4.560	(0.1354 ,0.451)	<0.0000 *
Gage(34-37)	-1.3594		0.2881	-4.719	(0.146 ,0.4517)	<0.0000 *
Temperature	-0.25942	19.94	0.05809	-4.466	(0.6885,0.865)	0.0000*
Weight						
REF (<=1600)	0.6408					
(1600-2500)		17.17	0.1547	4.143	(0.3891 ,0.7134)	0.0000*
Antenatal care visit						
REF (no)				-		
Yes	0.3263	4.03	0.1625	2.009	(0.5248 ,0.992)	0.0446 *
Respiratory distress S.						
REF (no)						
Yes	1.4665	71.13	0.1739	8.434	(3.082 , 6.094)	<0.0000**
Age at admission	0.0095	4.66	0.0044	2.158	(1.001,1.018)	0.31

*REF=References, PNA=prenatal asphyxia, HMD=hyaline membrane disease, ANC=Antenatal care visit, RDS=Respiratory distress syndrome.

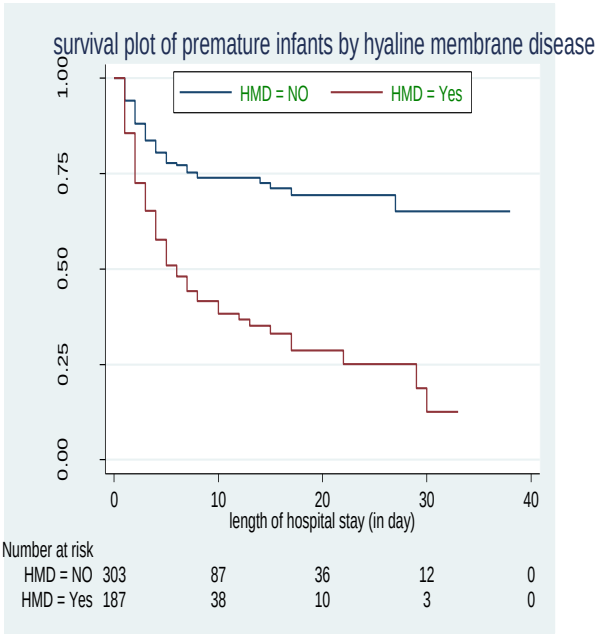
Table 4. 5 Statistical test for proportional hazards assumption of the covariates interaction with logarithms of time.(at jimma university specialized hospital, 2013-2015)

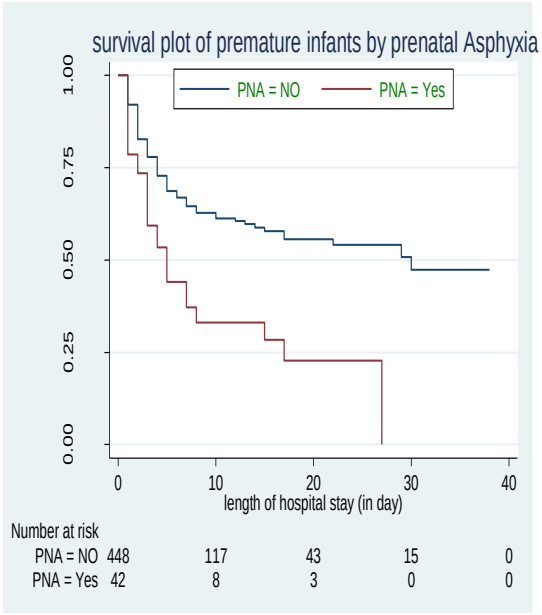
Variables	β	**SE**	**HR**	**Wald**	**Sign.**	**95% CI for HR**
Sepsis yes	0.428	0.314	1.53	1.8496	0.173	[0.8294,2.839]
Jaundice yes	0.452	0.274	1.57	2.7225	0.099	[0.9185,2.6887]
Gestational Age(26-28]						
Gestational Age (28-30]	-0.282	0.470	0.754	0.0036	0.548	[0.3002,1.895]
Gestational Age (30-32]	-0.119	0.468	0.888	0.0625	0.799	[0.3548,2.222]
Gestational Age (32-34]	-0.662	0.496	0.516	1.7689	0.182	[0.1951,1.3637]
Gestational Age (34-37)	-0.650	0.443	0.522	2.1609	0.143	[0.2191,1.244]
Perinatal asphyxia yes	0.639	0.371	1.89	2.9584	0.085	[0.9156,3.9203]
Hyaline membrane disease yes	0.719	0.306	2.05	5.5225	0.019	[1.127,3.7347]
Temperature	0.0435	0.0994	1.04	0.1936	0.662	0.8596,1.269]
Respiratory distress syndrome yes	0.0829	0.322	1.09	0.0676	0.797	0.5781,2.0417]
Ln.time*Sepsis yes	0.00504	0.274	1.01	0.0004	0.985	0.5870,1.721]
Ln.time*Jaundice yes	0.302	0.218	1.35	1.9321	0.165	0.8831,2.073]
Ln.time*Gestational Age(26-28]						
Ln.time*Gestational Age (28-30]	0.198	0.474	1.22	0.1764	0.677	[0.4810,3.0864]
Ln.time*Gestational Age (30-32]	-0.0198	0.468	0.98	0.0016	0.966	[0.3918,2.4530]
Ln.time*Gestational Age (32-34]	0.1690	0.473	1.18	0.1296	0.720	[0.4685,2.9953]
Ln. time* Gestational Age (34-37)	0.273	0.459	1.31	0.3481	0.553	[0.5340,3.230]
Ln.time*Perinatal asphyxia yes	-0.149	0.283	0.862	0.2809	0.599	0.4946,1.500]
Ln.time*hyaline membrane D yes	-0.0997	0.240	0.905	0.1764	0.678	[0.5656,1.4482]
Ln.time*Temperature	-0.0818	0.0861	0.921	0.9025	0.342	[0.7783,1.0908]
Ln.time*Respiratory distress S .yes	0.297	0.251	1.35	1.3924	0.236	[0.8231,2.2026]

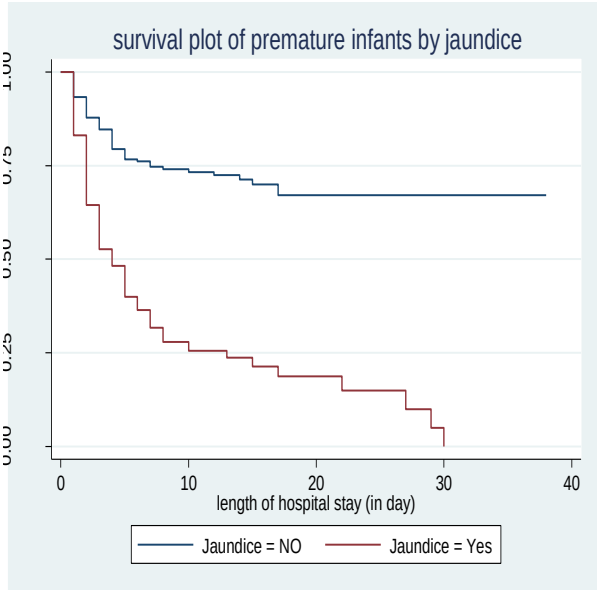
**For the above table , each of the first level of variables is as references corresponding to each variable .

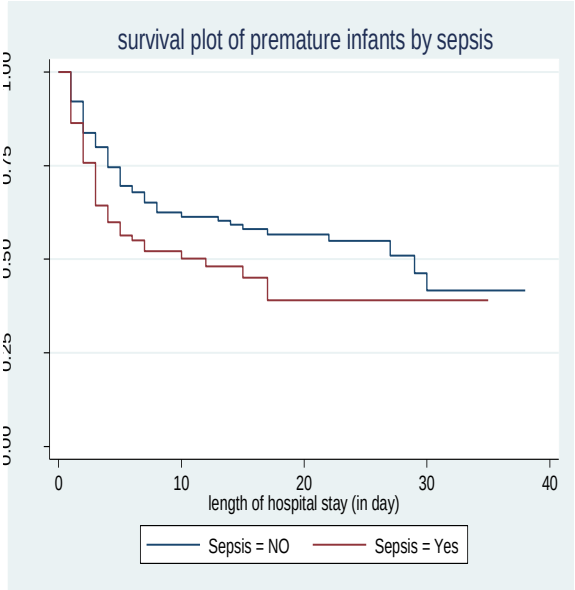
ANNEX -II

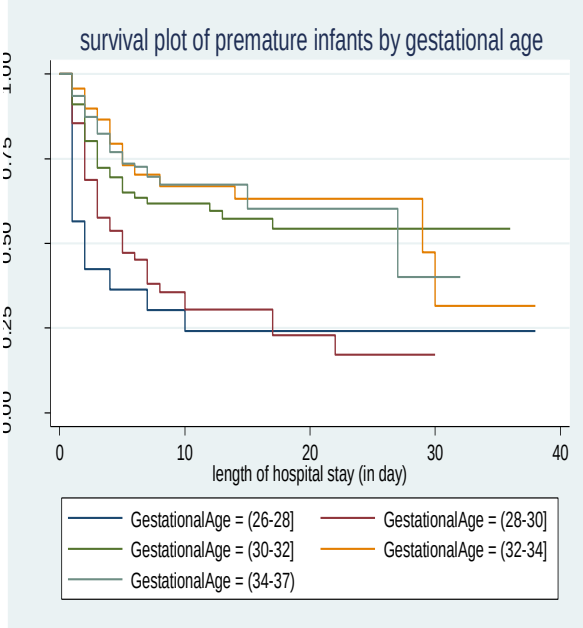
Figure 4.3 Kaplan Meier for Comparison of survival experience on premature infants using demographic, health and risk behavior variables.

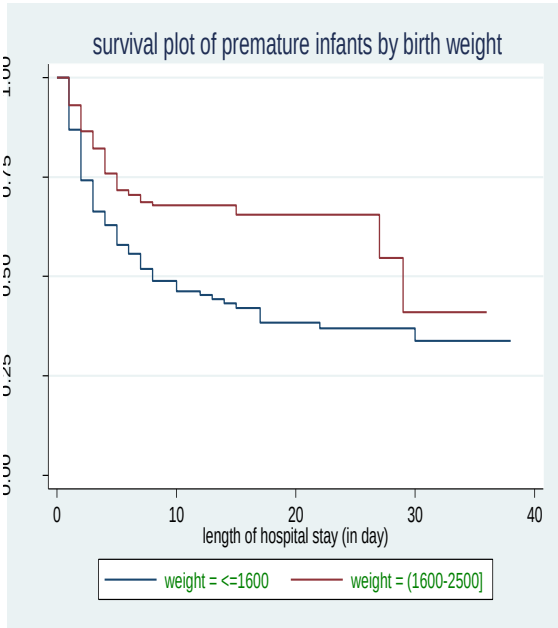









survival plot of premature infants by gestational age
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
length of hospital stay (in day)
GestationalAge = (26-28]
GestationalAge = (28-30]
GestationalAge = (30-32]
GestationalAge = (32-34]
GestationalAge = (34-37)


survival plot of premature infants by birth weight
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
length of hospital stay (in day)
weight = <=1600
weight = (1600-2500]

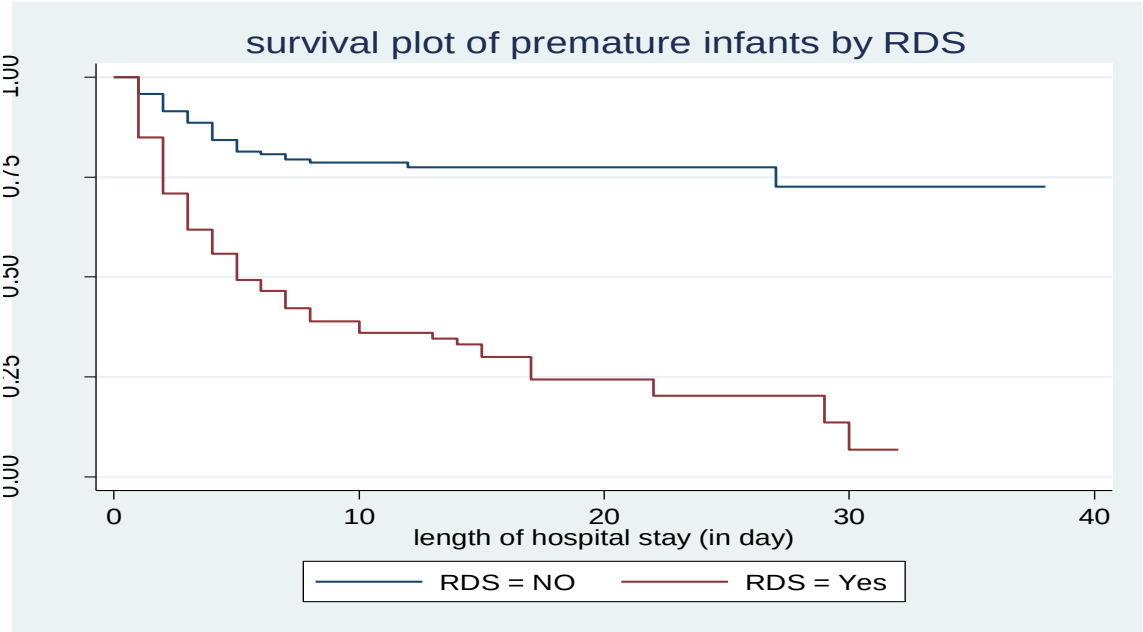
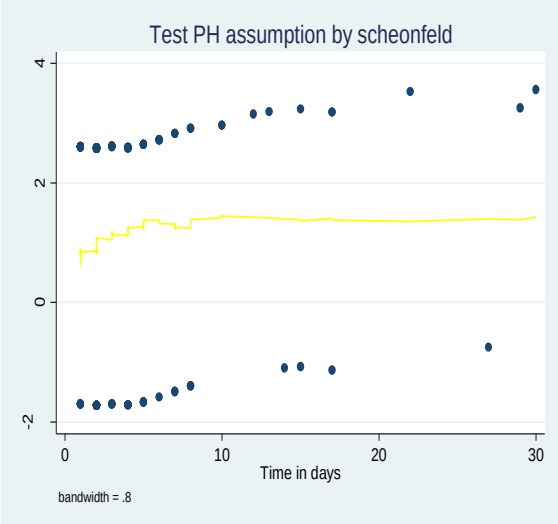
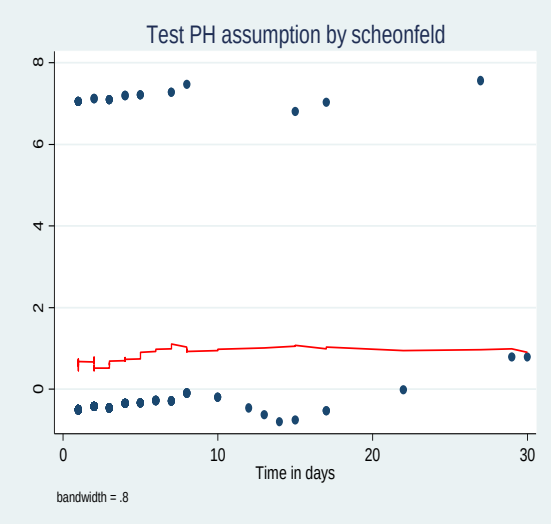

survival plot of premature infants by RDS
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
length of hospital stay (in day)
RDS = NO
RDS = Yes

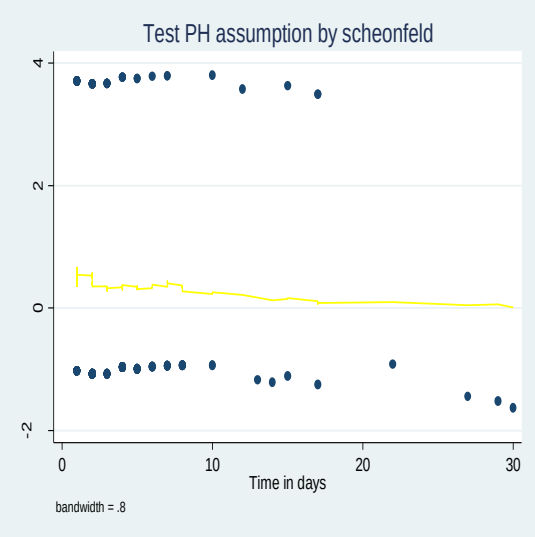
Figure 4 .5 Plots of Scaled Schoenfeld Residuals against length of hospital stay for Each Covariate in Cox Proportional Hazards Model Fit of premature infants.



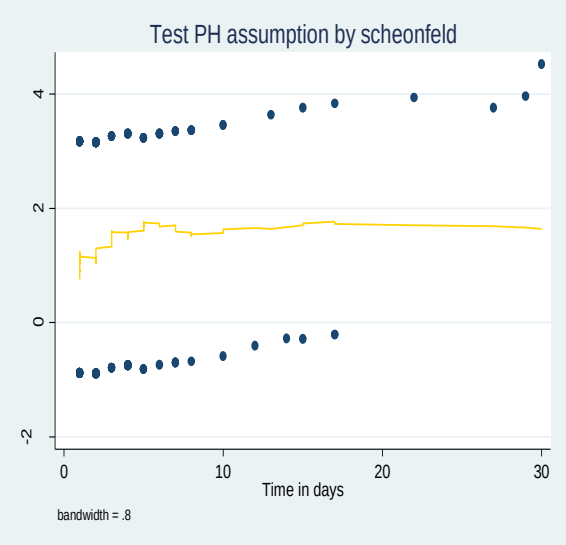
a) The plot of Scaled Schoenfeld residual for hyaline membrane disease



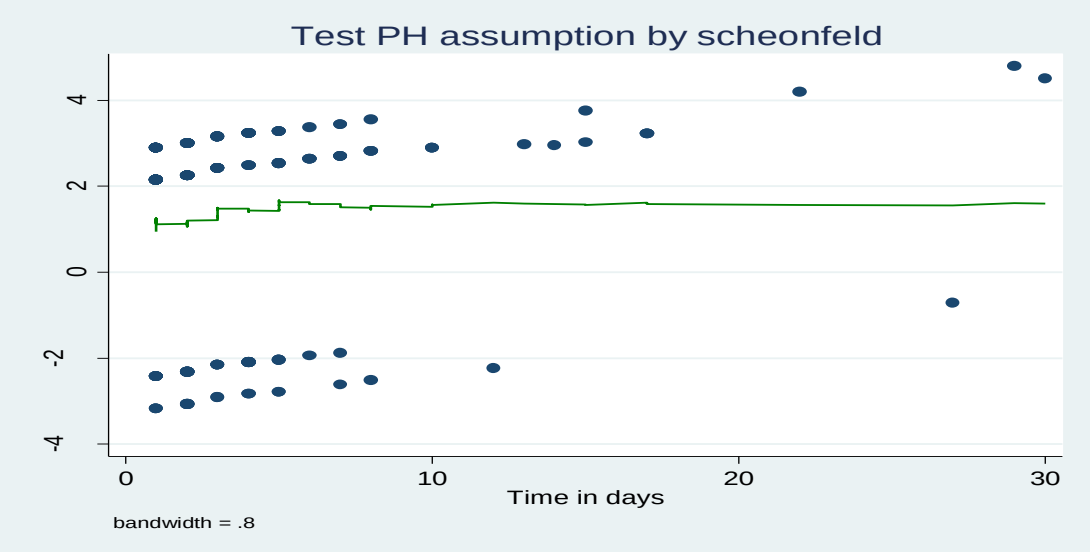
b)The plot of Scaled Schoenfeld residual for Prenatal Asphyxia



c) The plot of Scaled Schoenfeld residual for sepsis

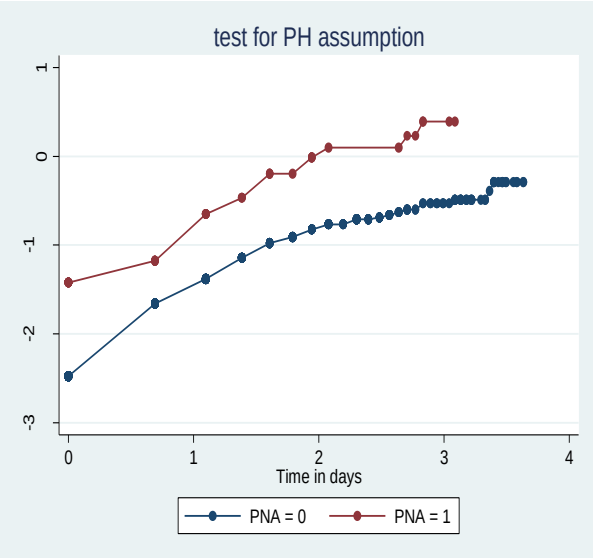


d) The plot of Scaled Schoenfeld residual For jaundice

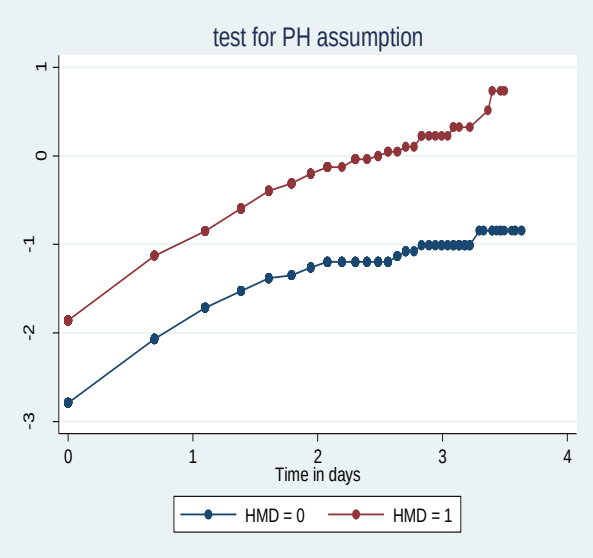


d) The plot of Scaled Schoenfeld residual for Respiratory distress syndrome

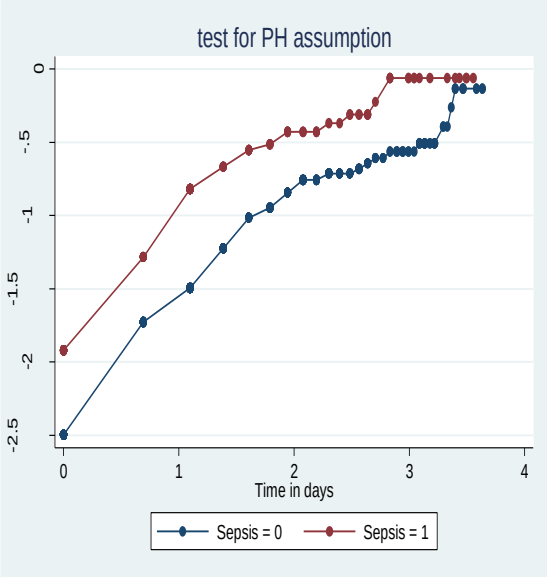
Figure 4 .6 Graphical assessment of proportional assumption checking by log(-log(survival probability)) versus time of hospital stay in days.

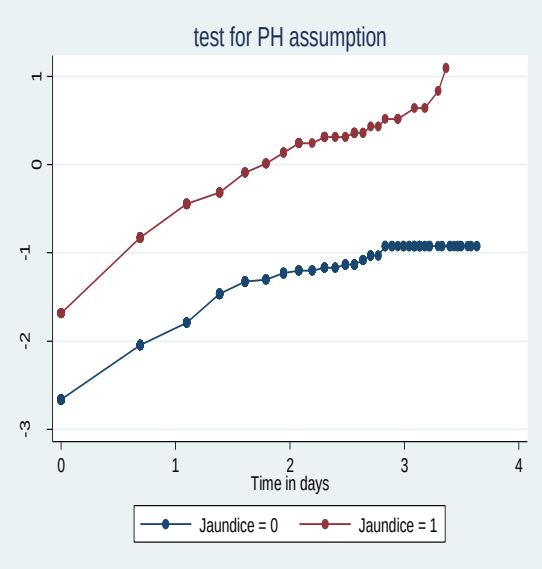


a) plot of log[-log(survival probability)] versus covariate prenatal Asphyxia



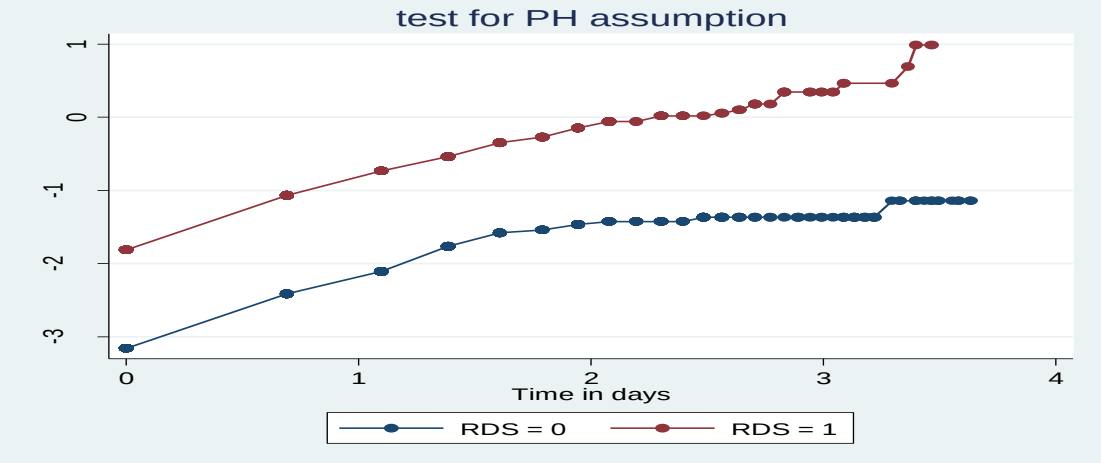
b) Plot of log[-log(survival probability)] versus covariate Hyaline membrane disease





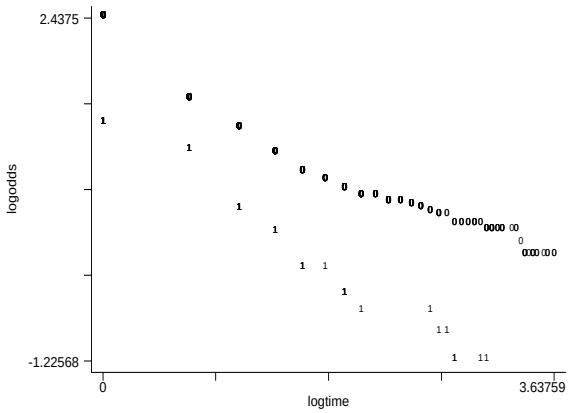
c) Plot of log[-log(survival probability)] versus Covariate jaundice disease

d) Plot of log[-log(survival probability)] versus covariate Sepsis

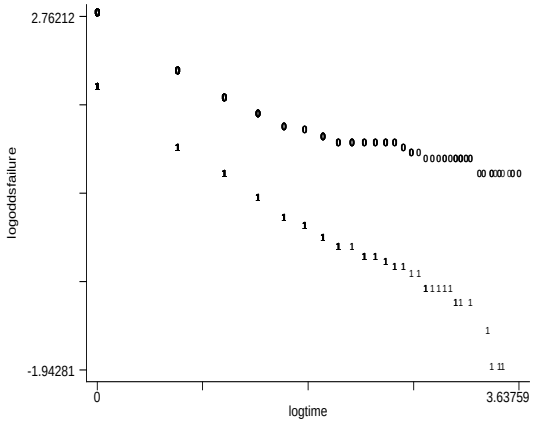


e) Plot of log[-log(survival probability)] versus covariate respiratory distress syndrome

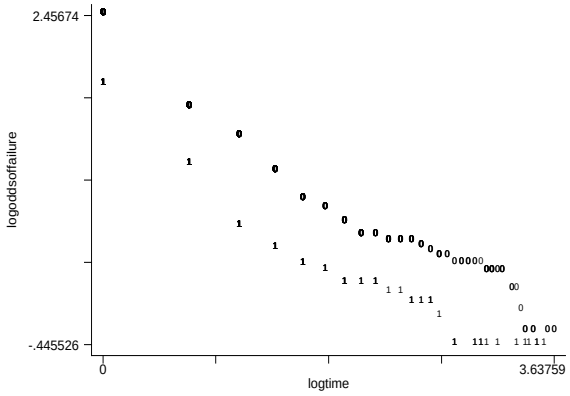
Figure 4 .11 Graphical assessments of log-logistic proportional odds assumptions using log odds of failure with logarithms of time.



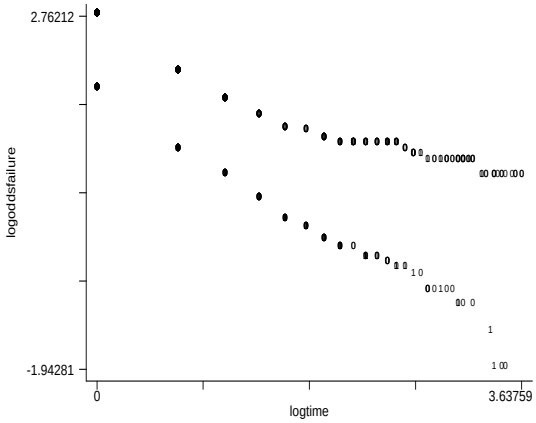
a) Graph of log odds of prenatal asphyxia versus log of time



b) Plots of log odds of hyaline membrane disease versus log of time.



C) Plots of log odds of Sepsis versus Log of time



d) Plots of log odds of Jaundice versus log of time