BAYESIAN SURVIVAL ANALYSIS OF UNDER-FIVE PNEUMONIA PATIENTS IN TERCHA GENERAL HOSPITAL, DAWRO ZONE, SNNPR, SOUTH WEST ETHIOPIA

MSc THESIS

By:

LEMA ABATE

A THESIS SUBMITTED TO

SCHOOL OF GRADUATE STUDIES IN JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

AS A PARTIAL FULFILLMENT FOR THE REQUIREMENT OF THE DEGREE OF

MASTER OF SCIENCE IN BIOSTATISTICS

NOVEMBER, 2018

JIMMA ETHIOPIA

BAYESIAN SURVIVAL ANALYSIS OF UNDER-FIVE PNEUMONIA PATIENTS IN TERCHA GENERAL HOSPITAL, DAWRO ZONE, SNNPR, SOUTH WEST ETHIOPIA

MSc THESIS

By:

LEMA ABATE

ADIVISOR: PROF. KADI ADIVEPPA

CO-ADIVISOR: MEGERSA TADESSE (MSc.)

NOVEMBER, 2018

JIMMA ETHIOPIA

STATEMENT OF AUTHOR

I hereby declare that the thesis is my original work and all sources of materials used for the thesis have been duly acknowledged with proper citation. This thesis has not been submitted to any other institution for award of any academic degree, diploma or certificate. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics.

The thesis can be deposited in the university library to be made available to borrowers for reference. Brief quotations from this thesis are allowed without requiring special permission if 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 her or his judgment the proposed use of the material is for a scholarly interest. In all other instances, however permission must be obtained from the author.

Name: Lema Abate

Date: _____

Signature: _____

Jimma, Ethiopia

APPROVAL SHEET

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis advisors, we here by certify that we have read the thesis prepared by **Lema Abate** under our guidance, which is entitled **"Bayesian Survival Analysis Of Under-Five Pneumonia Patients in Tercha General Hospital, Dawro Zone, SNNPR, South West Ethiopia",** 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 ready for submission to the university library.

Prof. Kadi Adiveppa _____ _____

Advisor Signature Date

Megersa Tadesse (M.Sc.) _____ _____

Co-advisor Signature Date

As the members of the board of examiners of M.Sc. 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 would like to thank the almighty God for the gift of health and strength throughout the process of carrying out this thesis and for seeing me through all. I wish to acknowledge the support of all who contributed in various ways to make this study success.

I want to express a sincere acknowledgement to my advisor Prof. Kadi Adeveppa for his advice and guidance that contributed to the successful realization of the thesis.

My deep gratitude goes to my co-advisor, Mr. Megersa Tadesse for his valuable suggestions and comments to the successful realization of this thesis. It is a great pleasure working with him. I hope that his cooperation will keep on going in the future.

I would like to acknowledge the Dawro zone department of Health staffs at the Tercha General Hospital to undertake this study with their cooperation and permission in using the data with special thanks for Mr. Abashaw Ayele, Dr. Yonatan Fikadu and Dr. Tamirat Tekle for their willingness to help me.

I would like to acknowledge Jimma University and all my instructors and staffs of Statistics Department at Jimma University for their cooperation from starting title selection up to the completion of this thesis. And also I would like to be grateful for my sponsor Mizan-tepi University for the contribution in my study. Finally, I am also very grateful to all my lovely parents and friends special to Sintayehu Niguse whose presence and support have always been important to me.

DEDICATION

This thesis is dedicated to my family especially to my father Abate Adulo, my mother Gebegnesh Getisa, my brother Adamu Abate and Tarekegn Abate.

ABSTRACT

Background: *Pneumonia is the number one largest infectious cause of death in children worldwide. More than 150 million cases of pneumonia occur in each year and it kills about 2,500 children every day. It is most prevalent in South Asia and sub-Saharan Africa. In Ethiopia, pneumonia is a leading single disease killing under-five children. Pneumonia is ranked as first cause of morbidity and mortality of children in Dawro zone and its prevalence is 27.81%. The aim of this study was to investigate the survival rate of under-five pneumonia patients in Tercha General Hospital using Bayesian and classical survival analysis.*

Methodology: *Retrospective study was conducted in Tercha General Hospital from September 2016 up to August 2017. Children whose age greater than 29 days and less than five year were included in the study and Patients with insufficient information were excluded from the study. Bayesian Survival analysis is a statistical method for data analysis of time to event data by introducing external information in terms of the prior distribution. The Semi-parametric, classical parametric models and Bayesian parametric models are used for the analysis.*

Result: *The Weibull Accelerated failure time model is good model compared to lognormal and Log-logistic models in both Classical and Bayesian approach based on AIC and DIC evidence respectively. The results implied that patients whose residence were urban and patient nurse ratio (PNR) were prolong timing death of under-five pneumonia patients, while season of diagnosis were Spring and summer, patients with comorbidity and patients with severe acute malnutrition (SAM) were statistically significantly shorten timing of death of under-five pneumonia in Tercha General Hospital in both Classical and Bayesian approach analysis.*

Conclusion: *Finally, the results from both classical and Bayesian approach analysis showed that sex, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patients refer status and patient nurse ratio (PNR) were found to be significant predictors for survival time of patients in Tercha General Hospital. The researchers who are interested to investigate on similar area recommended applying Bayesian analysis by introducing frailty modelling.*

Key words: Pneumonia, under-five, Survival analysis, Bayesian, Mont Carlo simulation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONOMY

ADIS	Acquired Immune Deficiency Syndrome
AFT	Accelerating Failure Time
AIC	Akaike's Information Criterion
ARI	Acute Respiratory Infection
CAP	Community acquired pneumonia
CSA	Central Statistical Agency
CHERG	Child Health Epidemiology Reference Group
DIC	Deviance Information Criteria
DRC	Democratic Republic of Congo
DZAR	Dawro Zone Annual Report
EDHS	Ethiopia Demographic Health Survey
GAPP	Global Action Plan for Prevention & Control of Pneumonia
HIV	Human immune Virus
JUSH	Jimma University Specialized Hospital
LOS	Length of hospital stay
MCMC	Markov Chain Monet Carlo
NPR	Nurse to Patient Ratio
PCVS	Pneumococcal Conjugate Vaccines
SNNPR	Southern Nations Nationalities and People's Region
TGH	Tercha General Hospital
UNDP	United Nations Development Program
UNICEF	United Nations Children's Fund
WHO	World health organization

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Pneumonia is an acute illness in which the alveolar air spaces of the lung become inflamed and filled with fluid and white blood cells, giving rise to the appearance of consolidation on the chest radiograph. It is the single largest infectious cause of death in children worldwide and accounts for 16% of all deaths of children under five years old and also it affects children and families everywhere, but is most prevalent in South Asia and sub-Saharan Africa (WHO, 2016). It can be caused by bacterial, viral, or parasitic infection as well as by noninfectious agents and Most severe cases of pneumonia are caused by bacteria, of which the most important are Streptococcus pneumoniae (pneumococcus) and Haemophilus influenza (Williams, 2002).

The mortality rates of children under the age of five years in most developing countries ranges from 60 to 100 per 1000 live births, one fifth of these deaths are due to pneumonia (WHO, 2016). The incidence of pneumonia in children under the age of five years is 0.29 episodes per child year, which equates 151.8 million cases annually in developing countries, a further 4 million cases occur in developed countries. Fifteen countries contribute 74% of the world's annual pneumonia cases (Rudan I, 2008). And also According to estimates from the World Health Organization Pneumonia kills about 2,500 children every day and more than 150 million cases of pneumonia occur in children under-five in each year, of which 20 million cases require hospitalization (Leung, 2016).

Morbidity and mortality from pneumonia is greater in low and middle income countries (LMIC). An estimate from the Child Health Epidemiology Reference Group (CHERG) puts the total number of pneumonia deaths worldwide in children under-five at 935,000 (Liu, 2016). Sub-Saharan Africa takes the lead in having half of its under-five deaths resulting from pneumonia compared to other regions. Also, regional disparities exist in the percentage of under-five deaths resulting from pneumonia with 5% of deaths occurring in developed regions and 17% of deaths in Sub-Saharan Africa (Liu, 2016). The child mortality due to pneumonia in least developing, developing and industrialized countries is 545,000(39%), 1,390,000 (>99%) and 2,000 (<1%),

respectively and its prevalence in sub-Saharan Africa countries alone is 46 % (CSA, 2012). The African Region has, in general, the highest burden of global child mortality. It has about 45% of global under-5 deaths and 50% of worldwide deaths from pneumonia in this age group (World health statistics, 2007). By contrast, less than 2% of these deaths take place in the European Region and less than 3% in the Region of the Americas. More than 90% of all deaths due to pneumonia in children aged less than 5 years take place in 40 countries. Even more striking is the fact that, according to the official estimates from WHO for the year 2004, two-thirds of all these deaths are concentrated in just 10 countries: India (408, 000 deaths), Nigeria (204, 000), the Democratic Republic of the Congo (126, 000), Ethiopia (112, 000), Pakistan (91, 000), Afghanistan (87, 000), China (74, 000), Bangladesh (50, 000), Angola (47, 000) and Niger (46, 000) (World health statistics, 2007).

According to 2012 central statistical agency report there is high burden of pneumonia in Ethiopia that is 88 in 1,000 children under age 5 die before their fifth birthday (CSA, 2012). Acute respiratory infection (ARI), and particularly pneumonia, accounts for 18% of death in Ethiopia; improving early care is a key strategy for early diagnosis and treatment (UNICEF, 2014). Integrated management of common childhood illness and community case management are among the program initiatives scaled up nationally to address ARI (Miller, 2014).

The study conducted at Hawassa city reported that the community acquired pneumonia result in the death of 16.34% patients and identified the factors that causes the under-five aged children mortalities using multilevel logistic regression model (Tariku T., 2017). Dawro Zone Health Office was ranked top 10 series burden disease in under five aged children according to its severity; from these ten top diseases pneumonia is ranked as number one cause of morbidity and mortality of children in Dawro zone its prevalence is about 27.81% (Dawro zone annual report, 2017).

Few studies in this area was tried to use statistical models like Binary logistic regression model and multilevel logistic regression models to identify the determining factors of pneumonia and several studies were about the prevalence of the pneumonia. But those studies were not focused on the survival time of patients hospitalized with pneumonia and also the hospital level variables were not included. Length of hospital stay (LOS) is one of hospital level variable which is

measured from patient's admission in hospital up to discharge of patients from hospital. Some of literatures used LOS as independent variables and reported that it is not significantly associated with discharge status of CAP under-five patients (Tariku T., 2017). Therefore the researchers were used LOS (survival time of patients) as response variable to identify determinants of survival time of under-five pneumonia patients, so researchers were chosen survival analysis as the good models to investigate this types of data in both classical and Bayesian approaches.
Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event and there are many standard parametric models such as Weibull, Lognormal and log-logistic (Klein, J. and Moeschberger, M. , 2003). The semiparametric and parametric survival models are used to fit the survival time of pneumonia patients. Lately Bayesian methods have been used in many research studies, especially in the field of medicine, as an alternative to classical or frequentist statistical methods. One of the reasons is that classical methods base their maximum likelihood estimations on asymptotic considerations that are usually only valid for a considerable data size (M.L. Calle et al., 2006).

In practice the methods are applied to relatively small data sets where the validity of the asymptotic assumptions is doubtful (M.L. Calle et al., 2006). In the Bayesian approach no assumption is made as to the shape of the percentile distribution, rather the data themselves specify the distribution. Another advantage of the Bayesian approach is the possibility of improving the precision of the results by introducing external information in terms of the priori distribution (M.L. Calle et al., 2006). In this study classical and Bayesian survival Analysis are used to identify important risk factors of under-five pneumonia patients. The response variable is from the admission time until an event (death) due to pneumonia. Kaplan Meier survival curves and log rank test are used to compare the survival experience of different category of patients and Cox PH model and Accelerating failure time (AFT) model were used to identify predictors of mortality of under-five pneumonia patients. There has been a limited literature on the use of classical survival analysis and Bayesian survival analysis of under-five pneumonia disease. Therefore the main aim of this thesis was to investigate survival rate of under-five pneumonia patients in Tercha General Hospital Using classical and Bayesian survival analysis.

1.2 Statement of the problem

Pneumonia is the major killer of children under the age of five years than any other diseases known to affect children, more than the death shares of Acquired Immune Deficiency Syndrome (AIDS), Malaria and Measles combined (WHO, 2016). More than 50% of all new pneumonia cases of the under- five childhood are concentrated in the poorest world's regions, Sub-Saharan Africa and South Asia. In terms of mortality, about 90% of all under five Pneumonia deaths burden is reported to occur in these two regions (WHO, 2012).

In 2016 United Nations Children's Fund (UNICEF) reported that there is 1 out of 6 childhood deaths were due to pneumonia globally in 2015 (UNICEF, 2016). In Sub-Saharan Africa, the proportion of deaths due to pneumonia in children younger than five year is 17-26 percent (Black E, et al, 2003). Nearly 50% of pneumonia deaths take place in only six densely populated and poorest countries: India, Nigeria, Democratic republic of Congo, Pakistan, Angola and Ethiopia (UNICEF, 2014). In Ethiopia pneumonia is a leading single disease killing under-five aged children and it was estimated that 3,370,000 children encounter pneumonia annually which contributes to 20% of all causes of death killing over 40,000 under-five children every year (Fischer W., 2013). According to pneumonia and diarrhea progress report of 2015, Ethiopia is among 15 top under five pneumonia high burden countries. The study conducted by Gilgel Gibe Field Research Center reported that Neonatal and infant mortality rates were respectively 38 and 76.4 per 1000 live births. The two most common causes of death during neonatal period were prematurity (26.4%) and pneumonia (22.6%). Whereas the top causes of death in post-neonatal period were pneumonia (42%), malaria (37%) and acute diarrheal diseases (30%) (Amare D. et al, 2007).

Several studies have been conducted to identify important risk factors of under-five mortality due to Pneumonia. Many scholars used logistic regression and Multi-level logistic regression models to identify the risk factors of pneumonia (Tariku T., 2017) and (Crighton E., et al , 2007). These statistical methodologies are not capable to consider the survival rate of the patients in the hospital and also Logistic regression does not account the censoring observations, that is, it does not hold for time-to-event data. It is necessary to use another model that explores the important risk factors of under-five child survival time due to pneumonia. Therefore; Survival analysis is

introduced in order to investigate the survival time of patients and to consider the censoring observations in the study. According to (Gelfand, A. E., & Mallick, B. K., 2005) study Bayesian approach is the best method to obtain the appropriate estimates of the model. The advantage of the Bayesian approach is the possibility of improving the precision of the results by introducing external information in terms of the priori distribution (M.L. Calle et al., 2006). The study conducted at Beirut Lebanon was reported that the Bayesian approach may have advantages over the frequentist one, particularly in case of a low power of the frequentist analysis (Pascale S., 2014)

According to several literature of pneumonia; pneumonia burden is high in rural part of the country; However, Dawro zone is one of the rural Zone in the SNNPR and According to 2017 annual report of Dawro Zone Health Office, pneumonia is ranked as first cause of morbidity and mortality of children in Dawro zone and its prevalence is 27.81% (Dawro Zone Annual report, 2017). This study is, therefore, intended to investigate the survival rate of under-five aged children hospitalized due to pneumonia observed at Tercha General hospital and to identify risk factors associated with under-five children survival time due to pneumonia using both classical survival models and Bayesian survival models.

Thus, in this study the risk factors for death of under-five aged pneumonia patients are going to be realized. In line with the above reality, these study attempt to come up with possible solutions and recommendations after having clear understanding upon the situation by giving due emphasis to answer the following questions:

1. What are highly determining factors for the survival time of under-five pneumonia patients in Tercha General Hospital?
2. Which survival model is the best to fit survival rate of under-five pneumonia patients in Tercha General Hospital?
3. Which Approach is the best from classical and Bayesian Approaches of survival analysis using hospital based pneumonia data in this study?

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study is to investigate the survival rate of under-five pneumonia patients in Tercha General Hospital using classical and Bayesian survival analysis.

1.3.2. Specific Objective

1. To identify risk factors associated with mortality of under-five aged children due to pneumonia in Tercha General Hospital.
2. To identify the best survival model that fit the survival rate of under-five pneumonia patients in Tercha General Hospital based on the risk factors.
3. To compare classical survival model and Bayesian Survival model using hospital based pneumonia data.

1.4. Significance of study

The result of this study provides information on risk factors of under-five mortality due to pneumonia in the hospital. It helps to reduce the death of under-five aged children by giving awareness for the society on the factors that increase the probability of under-five aged children death due to pneumonia. It serves as stepping-stone for those who are interested to undertake in depth research on issues related to the death of under-five due to pneumonia. Generally:

- ❖ It could provide information to government and other concerned bodies to make enabling environment for the intervention to reduce under-five mortality due to pneumonia.
- ❖ It provides as an input for researcher for further study, analysis and developing appropriate intervention methods for the prevention of morbidity and mortality due to pneumonia.
- ❖ It provides best survival model to fit biological as well as socio- demographic factors for health staffs as well as related researchers.

1.5. Scope of the Study

The study would have been covered under-five pneumonia patients registered in Tercha General Hospital from September, 2016 to August, 2017. The study is used to identify the risk factors and compare different survival models in both classical and Bayesian approaches in Tercha General Hospital under-five pneumonia patients' data set.

CHAPTER TWO

LITERATURE REVIEW

2.1. Overview on Under-Five Pneumonia Case Mortality

According to World Bank report the risk of pneumonia in children in developing countries is 3 to 6 times higher than other children. Not only outbreak of pneumonia, but also the mortality rate of this disease is higher in developing countries (Jamison DT, 2006). Various surveys have shown the nature and importance of pneumonia, many predisposing factors of pneumonia, arising from incorrect caring of infants in family and inadequate knowledge and awareness of mothers about proper infant care practices to this disease that exacerbating the problem in Iran (Ramazani M., 2006). Factors such as low birth weight and its impact and relationship with infection of the lower respiratory tract, the impact of malnutrition on children's impaired immune responses in developing countries and the prevalence of childhood infectious diseases such as diarrhea and pneumonia, lack of breast feeding and its impact on the reduction of passive safety defects in children, micronutrient deficiencies such as vitamin D and vitamin A and its effect on the immune response of children in this countries (Monir R., et al , 2015).

Approximately 20% of the 9 million estimated deaths in children aged less than five years in 2007 were ascribed to pneumonia; again, about 19% of all deaths in children aged less than five years in 2008 were attributable to pneumonia and this figure has reportedly increased to 21% in the 2012 WHO world health statistics report (WHO, 2012). In the 2014 estimates of pneumonia mortality by the UNICEF indicated that the disease was responsible for 15% of under five deaths in 2013 and out of 64.0% of all infectious causes of under- five mortality in 2010, pneumonia still takes the big share of 18.3% worldwide (UNICEF, 2014). The contribution of pneumonia to the deaths of older children was estimated to reach 14.1% with approximately four percent of childhood-pneumonia related death occurred in the first 28 days of life globally (Li Liu, 2012). In 2011, about 1.3 million children aged less than five years died of pneumonia globally. The same report showed that the case fatality ratio of pneumonia reached up to 8.9% worldwide (Christa L, 2013). According to 2012 lancet report, the global estimate of childhood pneumonia deaths was 18%, which can be translated to approximately 1.4 million childhood deaths, roughly

a 100,000 deaths rise from the previous report of 2011 (Li Liu, 2012). The study conducted at Hawassa city reported that the pandemic strain of community acquired pneumonia result in the death of 16.34% patients and had identified the factors that causes the under-five aged children mortalities using multilevel logistic regression model (Tariku T., 2017). Studies conducted in Bushulo Major Health Center, Hawassa, using total sample of 431 patients of pneumonia were considered and Out of which 18.79% death cases occurred and 81.21% were discharged (Zinabu T. et al., 2014).

2.2. Socio demographic characteristics

Socioeconomic factors contribute to high childhood pneumonia rates. In particular, poverty and other factors that inhibit access of care including migrant status, residence in rural areas, and low parental education levels have been shown to be associated with increased incidence of childhood pneumonia in China (Feng X., 2012). Children admitted with severe pneumonia were younger, more likely to be male and admitted during the rainy season than admitted children without severe pneumonia (Betuel S. et al , 2009). Several studies were reported that age of the children and the sex were the risk factors for the mortality of under-five children due to pneumonia. In a birth cohort study in Cape Town, South Africa, it was shown that the majority of the pneumonia burden among children is within the first 2 years of life (Campbell H and Nair H, 2015) and the results of this study indicated that severe pneumonia accounts for the most pneumonia deaths in the first 6 months of life. This relationship of increased pneumonia cases in younger ages has also been demonstrated in other studies (Monto A., 2004). And also other study evaluated that the effect of gender and revealed a male compared to females by gender (Campbell H and Nair H, 2015). A longitudinal cohort study in Pakistan noted similar increased incidences of childhood pneumonia in younger children and males (Khan A., 2009). Study conducted at Wondo Genet district, Sidama zone using multivariable logistics regression reported that children at age range 2-12 months were 4 times more likely to develop pneumonia as compared to older age groups (Teshome A., 2017). Both the incidence of and mortality from pneumonia widely vary across the age of the child where children younger than 2 years of age disproportionately bear about 81% of the overall under- five pneumonia morbidity burden (Christa L, 2013). Where children at age rang 2-11months were 85% higher chance to have pneumonia as compared to older age, 2013 lancet report, which revealed higher occurrence of

pneumonia in children younger than 2 years of age. According to case control study in Pakistan, younger children were found to be at increased risk of pneumonia compared to older children under the age of five years (Fatmi K. and Franklin W. , 2002).There is also studies conducted in Pakistan on the difference in incidence of pneumonia between boys and girls, with the higher episodes of pneumonia occurred among boys (Christa L, 2013). And also Study conducted at JUSH reported that among the children males accounted for 54.2% of the children and male to female ratio are 1.18:1 and children suffering from severe pneumonia in rural area accounted for 79.4% compared to children in urban area (20.6%) (Firaol B., 2017).

2.3. Co-morbidity

Co-morbidity has been found to elevate the risk of pneumonia. Diarrheal diseases are one of the determinants of under-five pneumonia as established by child health epidemiology reference group (CHERG), an academic review group started on by WHO (Fischer W., 2013). Diarrhea caused acute respiratory tract infection including pneumonia in a cohort study among children in Pakistan (Soofi S., 2012). Measles is an established risk factor for pneumonia. Pneumonia mortality caused by measles reached as high as 86%. Measles actually accelerates the fatality rate of pneumonia through immune suppression (Duke T, 2002). Case control study in Pakistan reported that children who had history of measles were susceptible to the development of pneumonia compared to those children who reported no history of measles (Fatmi K. and Franklin W. , 2002). The Child Health Epidemiology Reference Group (CHERG) revealed that other co morbid diseases such as HIV/AIDS, Malaria and Malnutrition were identified to be associated with increased occurrence of pneumonia (Fischer W., 2013).Children who have a concomitant chronic illness may have their immunity lowered making them more susceptible to severe disease. (Suwanjutha S., 2005) Found that children with an underlying heart condition were four times more likely to have severe pneumonia. According to Lancet Infect Dis. 2008 report in cases where the comorbidity happens to be human immunodeficiency virus, studies have shown that children who are human immunodeficiency virus infected are 40 times more likely to get pneumonia than their HIV free counterparts. (Rudan I, 2008) in their meta-analysis lists the presence of concomitant diseases as one of the likely risk factors for pneumonia; most evidence consistently point to the role, but there are some opposing findings. In 2004, World Health Organization recommended the treatment of non-severe pneumonia with oral antibiotics by trained community health workers at the community level.

2.4. Seasonal effect on pneumonia

The study conducted at Hawassa City on under-five aged children mortality found that patients admitted in summer and spring season had high risk of dying from CAP as compared with other seasons (Tariku T., 2017). According to study conducted in ten district Hospitals in Malawi classified the season in to Quarters as July-September, October-Dec, Jan-March and April-June Generally the pattern of pneumonia cases does not vary between the seasons in Malawi except in January through March and slightly peaks up again in the cool/dry season June and July. January through March coincides with the rainy season where there is a peak for both malaria and malnutrition (Ellubey R., 2004). Altitude, annual rainfall, number and nature of the seasons and average monthly temperatures are the factors listed by CHERG as factors of under-five pneumonia (Fischer W., 2013). And also other study conducted at southern Israel Hospital reported that the prevalence of the CAP and nosocomial pneumonia were higher in the spring and summer season compared to that of winter and autumn (Lieberman D and Porath A , 2005). According to study conducted in the Himalayas showed that high altitude is significantly associated with increased pneumonia cases. High altitude is likely to contribute to pneumonia due to lung physiologic compensatory mechanisms such as increased ventilation, increased cardiac output, and a shift in the oxygen-hemoglobin affinity curve. These compensations are delayed in infants who take 3-4 years to adapt fully (Khan A., 2009). Seasonality is another possible risk factor identified by (Rudan I, 2008), likely related to seasonal viruses including RSV and influenza.

2.5. Malnutrition

Protein-energy malnutrition results from inadequate intake, poor utilization of calories or protein in the diet, or from childhood infectious diseases, such as diarrhea and pneumonia (Black E, et al, 2003) and (Brown R., 2004).In epidemiological studies, malnutrition is usually assessed using anthropometrical measurements. A number of studies have examined the relationship between malnutrition, particularly low weight-for-age and incidence of pneumonia (Zafar F., 2002).

Malnutrition Refers to conditions that result from inadequate intake or consumption of energy or protein in the diet and usually is associated with a deficiency of certain vitamins and minerals. In

developing countries, underweight (weight below the proportional weight for age) is as a valid predictor of child malnutrition, so that low birth weight children are the children who have inadequate dietary intake or recurrent infections diagnosed (Brown R., 2004). It is estimated that about 36 percent of children less than 5 year in developing countries, have lower weight than 2 standard deviations in comparison with reference standards. Children with malnutrition have deficient immune responses; consequently these childhood infections are more severe in these children Studies shows children who their weight is less than 70% appropriate weight for their age, compared to other children, increased an 8 times risk of mortality from pneumonia for them (Lehmann D, 2005).

2.6. Treatment of pneumonia

Study conducted at Mozambican reported that children with severe pneumonia or suspected bacteremia/sepsis, empirical antimicrobial therapy with parenteral chloramphenicol or a combination of penicillin plus gentamicin was given. Infants less than 2 months of age and severely malnourished children were treated with ampicillin and gentamicin. Antibiotic therapy was re-assessed based on results of blood cultures. Ceftriaxone was used in cases of multi-resistant bacteria (defined as resistance to two or more antibiotic classes) (Betuel S. et al , 2009). Study conducted at Wondo Genet district, Sidama zone using multivariable logistics regression reported that treatment types taken by pneumonia patients at hospital levels has not significantly associated with mortality under-five children (Teshome A., 2017). There are multiple antibiotics indicated and effective in the treatment of pneumonia. Administration of the most appropriate antibiotic as a first-line medicine may improve the outcome of pneumonia. In order to effectively treat the disease while minimizing antimicrobial resistance and virulence, it is important to know which antibiotics work best for children depending on the severity of the illness (UNICEF, 2014). According to Recommendations for management of common childhood conditions: Newborn conditions, dysentery, pneumonia, oxygen use and delivery, common causes of fever, sever acute malnutrition and supportive care there are four types of antibiotics suggested for treatment of pneumonia are ceftriaxone, ampicillin, cephalosporin, and macrolides (WHO, 2012). Furthermore, there is a need for reformulation of existing, recommended antibiotic treatments for children. The WHO ‘Priority life-saving medicines for women and children 2012’ listed two recommended dosages of gentamicin: 40 mg/ml and 20 mg/ml. The 40mg/ml is an

adult formulation, adaptable to older children but unsuitable for neonates, and the 20 mg/ml formulation is ideal for neonates and children. However, 20 mg/ml of gentamicin is not currently manufactured; as a result, dilutions of the 40 mg/ml formulation will need to be made until that time when the 20 mg/ml formulation is available. Lastly, the worldwide estimate is that 30% of isolates from those with pneumonia are resistant to macrolides, including erythromycin, azithromycin, and clarithromycin. Similarly, 30% of bacterial pneumoniae is now multidrug resistant. The continual rise in antibiotic resistance is a major public health concern that requires keen observation of respiratory illness in children to assess proper treatment options (WHO, 2016).

2.7. Hospital level risk factors of pneumonia

As underlined Rural children experienced a higher rate of pneumonia and a lower rate of care from a trained provider compared to urban children (Miller N., 2014). Nurse-to-patient ratio **(**NPRs) are typically expressed in two ways: the number of nurses working per shift or over a 24 hour period divided by the number of beds occupied by a patient over the same time period; or the number of nursing hours per patient bed days (RCN, 2010). A higher level of nursing staff indicates more nurses (or higher proportion of nurses) for assigned patients. Lower nurse staffing is defined as fewer nurses (or lower proportion) for the number of assigned patients (Penyoyer D., 2010). NPRs are easily and cheaply measured but it is a relatively blunt instrument that can function as one indicator, and can be triangulated with other measurement approaches to establish safe nurse staffing levels. According to the study conducted at Europe, the effect of nurse-to-patient ratios on nurse sensitive patient outcomes in acute specialist units found that a higher level of nurse staffing was associated with a decrease in the risk of in hospital mortality (Andrea D., 2017). A higher level of nurse staffing will lower the risk of in hospital mortality. For every increase of one nurse, patients were 14% less likely to experience in hospital mortality. In addition to nurse patient ratios, it is also important to incorporate skill mix within a critical care unit particularly when planning workforce shifts (Penyoyer D., 2010). Patients will be less likely to experience an adverse event in units with a high nurse to patient ratio. This has important implications for clinical practice and the optimization of patient outcomes. These studies highlight the need for some agreement, at an international level, about the most appropriate way to measure nurse staffing levels (Andrea D., 2017).

According to the study conducted at Hawassa city applying multi-level models, the hospital level variables such as patient to nurse ratio and patient to physician ratio are significantly predicted the discharge status of the CAP patients and the study reported that Patients who admitted during patients to nurse ratio was high had high risk of dying from pneumonia, Since patients in hospital are nurtured by nurses and this has a positive impact on the recovery from their illness. Also in this study an investigators reported that the bed occupancy rate is not significantly predict the discharge status of the patients (Tariku T., 2017).

2.8. Overview on models in the study

Studies conducted at United Arab Emirates on Overview of Frequentist and Bayesian Approach to Survival Analysis was focused on the strength and weakness of both classical and Bayesian approaches using different datasets. From which the data on the survival distribution of patients who have undergone surgery is expressed in terms of parameters, mean survival years and extreme values. The investigators assumed that the mean survival year is 10 years and with range of values of the distribution from 6 to 16 years. The prior distribution is expressed in terms of mean survival years as 8 years and extreme values of the distribution as 3 and 12 years. The posterior distribution will be a synthesis of the prior distribution with the evidence obtained from the data which will be with mean 9 years and extreme values 6 and 12 years discarding the minimum value of prior distribution as it was not supported by the data and discarding the maximum value from the data as it was not supported by the prior. The mean survival rate (9 years) is obtained as the average of the two mean values, prior (8 years) and the data (10 years) (Cluj-Napoca and Romania., 2016). The study conducted at Beirut Lebanon were showed that the Bayesian approach may have advantages over the frequentist one, particularly in case of a low power of the frequentist analysis; the use of informative priors might be particularly useful in narrowing credible interval and precise the choice between the null and alternative hypothesis (Pascale S., 2014). Whenever the frequentist results were clear cut (due to a large sample size or a strong association), performing the MCMC method helped to increase the accuracy of the results by narrowing the credible interval, but did not change the direction of hypothesis acceptance; the Gibbs sampling might give closer results to the truth and using informative priors might further help to improve credible intervals (Hakim E., 2009).

Bayesian survival analysis is more advantageous than classical survival Analysis, in terms of flexibility of model building for complex data. Bayesian survival analysis that used informative and proper prior information was more advantageous than classical survival Analysis. In every condition, informative and proper prior information should be used for analyzing data with Bayesian survival analysis. Bayesian survival analysis showed better performance than classical survival Analysis (Ibrahim J. et al, 2001)and (Wong W. et al , 2005).

Apart from the simplest models, inference within classical statistics is based on large sample approximations while Bayesian methods are exact in the sense that assuming the model assumptions are valid, the posterior distribution do give the right answer. The need for numerical approximations violates this exactness, but such errors are usually of a smaller scale than variability due to data. On the other hand, large sample approximations are usually quite robust to model assumptions. This is also true for Bayesian settings, but with small samples the results will typically rely heavily on the model assumptions. Maximum likelihood methods involve optimization, Bayesian approaches involve integration. Advances in statistical computing, and in particular Monte Carlo methods, have for many problems made the computational challenges easier to handle within the Bayesian framework than in classical settings. Pragmatic considerations have in such cases largely given preferences to the use of Bayesian methods (Geir Storvik, 2014).There is no established method for determining an appropriate number of iterations and burn-in size. Rather, the researcher use a trial-and-error process in which the ultimate goal is to obtain stable parameter estimates that minimize simulation error. As with the computational intensity this steps require more time on the part of the researcher. However, MCMC estimation is indispensable as a tool for handling intractable epidemiological inquiries (Ghassan H., 2013).

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Study Area

The study was taken place in Dawro Zone Tercha General Hospital around 144 km away from Jimma town, 319 km away from the regional city of SNNPR Hawassa and 491 km away from Addis Ababa capital city of Ethiopia. Dawro Zone is one of the 14 Zones in Southern Nations Nationalities and People's Regional state. Astronomical, it is roughly lies between 6059^0-7035^0North Latitude and 3606^0-37035^0East Longitude (Fantahun O. and Abayneh K., 2017). It is bordered with Oromia region in the North West, kambata Tembaro Zone in the North East, Wolaita Zone in the East, Gamo Gofa Zone in the South, and Konta special woreda in the west. According to Dawro zone Annual Statistical Abstract 2016-2017 report the total area of the zone is estimated to be 4436 square kilometer which shares 4.07% of the total area of the region and the population size is 636,218 accounting nearly 3.3% of the total population of the region. The average population density of the zone is 143 people per square kilometer. Based on the geographic administration the zone divided into five woreda (Tocha, Mareka, Loma, Genna Bossa and Issera) and one town administration (Tercha), which sub-divided in to 167 rural and 18 urban kebele. In urban development, there are 16 municipal towns and one town administration. According to 2017 Dawro Zone Health Department annual report Dawro zone has one Zonal General Hospital and Two District Hospitals, 22 health centers and 162 health posts (Dawro zone annual report, 2017).

3.2. Study Design and Study Population

This study is a retrospective study that reviews or visits all under-five aged children cards hospitalized with Pneumonia in Tercha General Hospital. The population of this study is all under-five pneumonia patients registered in Tercha General Hospital from September 2016 up to August 2017 G.C. A total of 1887 under-five pneumonia patients were recorded with full information in given study period of time. The researcher calculated the sample size and selected the representative sample from the population that fulfilled the inclusion criteria in this study.

3.2.1. Inclusion and Exclusion Criteria

Inclusion Criteria

All under-five pneumonia patients registered in Tercha General Hospital with full information in the pediatric registration chart and in the patients' identification card were considered to be eligible for the study.

Exclusion criteria

Patients with insufficient information about one of the vital variables either in the pediatric registration chart or patients' identification cards and patients less than one month were not eligible. Out-patients even if managed as pneumonia and children managed as pneumonia but above 5 years old were excluded.

3.2.2. Data Collection Procedure

The data for this study is secondary data that was recorded on pediatric registration chart and cards via nurses, laboratory technicians, medical doctors and clinicians. The hospital's registry is used to extract data of under-five pneumonia and patients' initial date of admission up to discharge of patients. During the study period, the pediatric registration chart and the patient's identification cards were used to select the variables in the study by trained clinicians. The completed data collection forms are examined for completeness and consistency during data management, storage and analysis. The Data collection process was carried out in time interval of 20/May/2018 – 20/June/2018 G.C.

3.2.3. Sampling Technique and Sample Size Determination

Sampling technique is a method of taking small ratio of observation from a large population with the aim of getting information of those large populations from the sampled observation by using some statistical techniques. In conducting studies the researchers have the stages of deciding the sample size and the decision is important because of taking too large sample implies waste of resources and time while too small sample reduces the usefulness of the results. In this study the researcher used Simple Random sampling technique as an appropriate sampling technique to select a representative sample of the patients by using lottery method. To get optimum sample size, there are several formulas developed for sample size calculation. According to Cochran G., (1977) sample size determination formula the researcher computed the sample size as follows.

$$n_0 = \frac{\frac{z^2 p(1-p)}{d^2}}{1+\frac{1}{N}\left[\frac{z^2 p(1-p)}{d^2}-1\right]} \text{----------(1)}$$

Where, n = the sample size needed, N = the total population size, z is the upper $\alpha/2$ points of standard normal distribution with α=0.05 significance level. Suppose the maximum allowable difference between the maximum likelihood estimate and the unknown population parameter, denoted by d. The specification of d must be small to have a good precision. The estimated proportion of death due to pneumonia disease was p=0.164 According to (Tariku T., 2017) and the researcher have been used a margin of error of 0.04. The sample size would be calculated as follows:

$$n_0 = \frac{(1.96)^2 0.164(0.836)}{0.04^2} = 330$$

If $\frac{n_0}{N} > 5\%$ i.e. $\frac{n_0}{N} = \frac{330}{1887} = 0.175$

But 0.175 is greater than 0.05, therefore the researcher used the next formula to get optimum sample size.

$$n = \frac{n_0}{1+\frac{n_0-1}{N}} \text{----------(2)}$$

$$n = \frac{330}{1+\frac{330-1}{1887}} = 281$$

3.3. Study Variables

3.3.1. Response variable

The response variable is time to death of under-five pneumonia patients in days. The survival time of outcome of interest (death in this study) is the duration of time considered from the day that the children admitted in the Hospital until death occurs. The status variable is coded as 0 for censored and 1 for death.

3.3.2. Independent Variables

The predictor variables (factors) are variables that are assumed to influence the survival time of under-five aged children hospitalized with pneumonia in Tercha General Hospital. These variables are selected based on some previous study conducted at Hawassa city by (Tariku T., 2017) , at Tanzania Lake Zone's public hospitals by (Kristina L., 2017) and at Jimma university specialized hospital by (Firaol B., 2017). The variables that are expected to be the factors/determinants of mortality of under-five pneumonia patients based on different literatures are as follows:

Table 3.1: Description of variables in the study

	Variables	**Coding for Categorical variables**		**Description**
1	Sex	Female= 0	male = 1	Sex of children
2	Age	1-11 = 0	36-47 =3	
		12-23 = 1	48-59 =4	Age of children
		24-35 = 2		
3	Residence	Rural = 0	Urban =1	Residence of children
4	Season	Autumn =0	Winter =1	Season of Diagnosis
		Spring = 2	Summer =3	
5	Co-morbidity	No = 0	Yes =1	Co-morbidity (CAP complicated)
6	SAM	No = 0	Yes =1	Sever Acute Malnutrition
7	Treatment types	Penicillin =0	Ceftriaxone =1	Treatment types taken at time of Diagnosis
		Ampicillin = 2	Combined =3	
8	Patients refer status	No = 0	Yes =1	Patient refer status from other health center

Continuous Variables in the study		
9	Weight	Weight of the children
10	Bed Occupancy Rate (BOR)	The percentage of official beds occupied by hospital Inpatients for a given period of time; $BOR = \frac{(\text{Total length of hospital stay in a given month})*}{30 \times \text{Number of beds in that month}}$
11	Patient: Physician Ratio (PPhR)	Ratio of patient to physician counted in a given month; $PPhR = \frac{\text{Number of patients admitted in a given month}}{\text{Number of physician on service in a given month}}$
12	Patient: Nurse Ratio (PNR)	Ratio of patient to nurse counted in a month $PNR = \frac{\text{Number of patients admitted in a given month}}{\text{Number of nurse on service in a given month}}$

** Length of Hospital stay is the number of calendar days from the days of patient admission to the day of discharged/died*

3.4. Methods of Data Analysis

In the method of data analysis the researcher introduced different survival analysis methods these are nonparametric, semiparametric cox PH model, Classical approach parametric (AFT) and Bayesian approach parametric (AFT) survival models. The software's used in these studies are latest version (R 3.5.0 and Stata13), and WinBUGS for both Classical and Bayesian approach analysis respectively.

3.4.1. Survival Data Analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. By time, mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs.

Survival data are almost always incomplete. The statistical terminology for such data is censoring. Censoring is common in survival analysis and it is considered as an important feature of survival data. The most common form of censoring for incomplete data is right censoring when a subject's follow-up time terminates before the outcome of interest is observed. The second type of censoring is left censoring which is observed when an individual had developed an event of interest prior to the beginning of the study. An observation is categorized into an

interval censored if it is only known that the event of interest occurs within an interval of time without the knowledge of when exactly it occurs (Klein, 2005). The censoring used in this paper is right censoring. Survival time is recorded from the admission of under-five pneumonia patients up to the discharge or death. This type of censoring is commonly recognized survival analysis and also considered in this study. If the event of interest is not occurred before the final day of the study such types of censoring is called right censoring.

3.4.2. Descriptive Methods for survival data

In any applied setting, a statistical analysis should begin with a thoughtful and thorough description of the data. In particular, 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. Routine applications of standard measures of central tendency and variability will not yield estimates of the desired parameters when the data include censored observations. In summarizing survival data, the two common functions are the survivor function and the hazard function (Hosmer and Lemeshow, 1999).

Survival Functions

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 (Klein, 2005). Moreover, the distribution of survival time is characterized by three functions:

i. The survivorship function

ii. The probability density function and

iii. The hazard function.

Let T be a random variable associated with the survival times, t be the specified value of the random variable T and f (t) be the underlying probability density function of the survival time T. The cumulative distribution function F(t) ,which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given by;

$$F(t) = P(T < t) = \int_0^t \mathrm{f(u)du}, \mathrm{t} \geq 0 \text{ ----------} (3)$$

The survival function S(t), is given by:

$$S(t) = P(T \geq t) = 1 - F(t), t \geq 0 \text{ ---------- (4)}$$

From equations (3) and (4) the relationship between F(t) and S(t) can be derived as:

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - S(t)) = \frac{-d}{dt}S(t) \geq 0 \text{ ---------- (5)}$$

Hazard Function

The hazard function h(t) 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 (Klein, 2005).

The hazard function $h(t) \geq 0$ is given as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{\text{an individual fails in the time interval}(t, t + \Delta t) \text{given survives until time } t\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P\{t \leq T \leq +\Delta t / T\}}{\Delta t}$$

By applying the theory of conditional probability and the relationship in equation (3), the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}S(t) \text{ ---------- (6)}$$

The corresponding cumulative hazard function H(t) is defined as:

$$H(t) = \int_0^t h(u)du, = -\ln S(t) \text{ ---------- (7)}$$

Then,

$$S(t) = \exp(-H(t)) \text{ and } f(t) = h(t)S(t) \text{ ---------- (8)}$$

3.5. Non-parametric Survival Methods

Nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution. Survival data are conveniently summarized through estimates of the survival function and hazard function. Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e., if the estimated survival functions for two groups

are approximately parallel. The nonparametric methods used in these studies are the Kaplan-Meier Estimation and Log-rank test method.

3.5.1. The Kaplan-Meier Estimator

The Kaplan-Meier (KM) estimator is the standard non parametric estimator of the survival function, S(t) proposed by Kaplan and Meier (1958) which is not based on the actual observed event and censoring times, but rather on the ordered in which events occur. Kaplan-Meier estimator 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. Let ordered survival times are given by $0 \leq t_1 \leq t_2 \leq t_j \leq \infty$, then:

$$\hat{S}(t) = \begin{cases} 1, \ if \ t < t_i \\ \prod_{t_{i \leq t}} \left(1 - \frac{d_i}{n_i}\right) \end{cases} \text{----------} (9)$$

Where: d_i= number of under-five aged children with pneumonia died at t_i

n_i= number of under-five aged children with pneumonia at risk before t_i

3.5.2. Log-Rank Test

Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests. Among the various non-parametric tests one can find in the statistical literature, the Mantel-Haenzel (1959) test, currently called the "log-rank test" is the one commonly used non-parametric tests for comparison of two or more survival distributions (Mantel, 1959). The log rank test statistic for comparing two groups is given by:

$$Q = \frac{\left[\sum_{j=0}^{r} w_j (d_{1j} - \hat{e}_{1j}\right]^2}{\sum_{j=0}^{r} w_j^2 \hat{v}_1} \text{----------} (10)$$

Where: r is the number of rank-ordered failure times (event times), w_j is the weight for censor adjustment at time t(j), $\hat{e}_{1j} = \frac{n_{1j} d_j}{n_j}$, $\hat{v}_{1j} = \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$, d_{1j} is the observed number of failure (event occur) at time t(j) in group 1, n1j is the number of individuals at risk of event occur in the first group just before time t(j), n2j is the number of individuals at risk in the second group just before time t(j), d_j is the total number of events occurred at t(j), n_j is the total number of

individuals at risk before time t(j). Q follows a chi-square distribution with k-1 degree of freedom.

3.6. Semi-parametric Survival models

Semi parametric survival model asks fewer assumptions than typical parametric methods but more assumptions than those nonparametric methods. In particular, and in contrast with parametric models, it makes no assumptions about the shape of the so-called baseline hazard function. In order to explore the relationship between the survival experience of individual and explanatory variables, an approach based on statistical modeling can be used (Collett D., 2003).

3.6.1. Cox PH Regression Model

The non-parametric method does not control for covariates and it requires categorical predictors. One of very popular model in survival data analysis is the Cox PH model which is introduced by (Cox D., 1972), and is a broadly applicable and the most widely used method of survival analysis. A model based on the exponential distribution may be parameterized as follows:

$$\log h_i(t,x,\beta) = \alpha + \beta_1 X_{i1} + \beta_{21} X_{i2} + ... + \beta_k X_{ik}$$

Equivalently;

$$h_i(t,x,\beta) = \exp(\alpha + \beta_1 X_{i1} + \beta_{21} X_{i2} + ... + \beta_k X_{ik}) = \exp(\alpha)(\beta^t X)$$

In this case the constant α represents the log-baseline hazard since $\log h_i(t) = \alpha$ when all the x's are zero. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time.

$$h_i(t,x,\beta) = h_0(t)\exp(\beta_1 X_{i1} + \beta_{21} X_{i2} + ... + \beta_k X_{ik})$$

$$h_i(t,x,\beta) = h_0(t)\exp(X_i^t \beta) \text{----------} (11)$$

Where, $h_0(t)$is the baseline hazard function; X_i is a vector of covariates and β is a vector of parameters for fixed effects.

The corresponding survival function for Cox-PH model is given by:

$$S(t,X) = [S_0(t)]^{\exp\left\{\sum_{i=1}^{p} \beta^t X_i\right\}} \text{----------} (12)$$

Where: $S_0(t)$ is the baseline survival function.

In this model, no distributional assumption is made for the survival time; the only assumption is that the hazards ratio does not change over time (i.e., proportional hazards) that is why this

model is also known as semi-parametric model. With the Cox proportional hazards model the outcome is described in terms of the hazard ratio.

The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates X and X^* is given by:

$$\hat{HR} = \frac{h_0 \exp(\hat{\beta}^t X)}{h_0 \exp(\hat{\beta}^t X^*)} \quad \text{----------(13)}$$

This hazard ratio is time-independent, which is why this is called the proportional hazards model. The parameter of the Cox proportional hazard model refers to the hazard ratio of one group in comparison to the other groups for categorical covariates and change in hazard ratio with a unit change of the covariate for the continuous variables when other covariates are fixed. The change in hazard ratio for the continuous covariate is given by:

$$\hat{HR} = \frac{h_i(t, X_{k+1})}{h_k(t, X_k)} = \exp(\beta_k) \quad \text{----------(14)}$$

Represent changes in the hazard when there is a unit change in the covariates while other covariate keeps constant.

Partial Likelihood Estimation for Cox PH Model

In fitting the Cox proportional hazards model, the researcher estimates $h_0(t)$ and β. A more popular approach is proposed by (Cox D., 1972) in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters $h_0(t)$ in the Cox PH model. The data in survival analysis based on the sample size n are denoted by the triplet (t_i, δ_i, X_i), i= 1, 2… n where t_i is the time at which the i^{th} individual experience the event (in this study; death), δ_i = 1 if the event has occurred, δ_i = 0 if censored, X_i is the vector of covariate or risk factors for the i^{th} individual. In general partial likelihood function expressed as:

$$L_p(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta^t X_i)}{\sum_{j \varepsilon R_{(t_i)}} \exp(\beta^t X_i)} \right]^{\delta_i} \quad \text{----------(15)}$$

Where: the summation in the denominator is over all subjects in the risk set at time ti denoted by R(ti)over all n subjects that the event occurred.

The corresponding log partial likelihood function is given by:

$$\log(L_p(\beta) = \sum_{i=1}^{m}\left\{\beta^t X_i - \log\left[\sum_{j\varepsilon R_{(ti)}} \beta^t X_i\right]\right\} \text{----------(16)}$$

In general, large sample properties like normality and consistency of maximum likelihood estimators of β based on partial likelihood have been shown to be the same as those of any estimator from complete likelihood (Hosmer and Lemeshow, 1999).

Testing the Assumption of PH Model

The proportional hazards assumptions are vital to use in a fitted proportional hazards model. Variable adds significant information, if the newly added variable is not significant; it can be taken as the proportional hazard assumptions are satisfied. The method of checking the assumption of the Cox proportional hazards model is scatter plots using the Schoenfeld residual (Schoenfeld D., 1982). The residuals constructed for each covariate that are included in the model which are expected to predict the death time of under-five pneumonia patients. Under the proportional hazard assumption for the respective covariate, a scatter plot of Schoenfeld residuals against event times is expected to scatter in a nonsystematic way about the zero line, and the polygon (Lowess curve) connecting the values of the smoothed residuals should have a zero slope and cross the zero line several times (Klein, J. and Moeschberger, M. , 2003). If this plot shows some trend the assumption is violated, where as if the plot demonstrates randomly distributed around the reference line then the assumption is satisfied.

3.7. Classical Parametric survival models

3.7.1. Accelerated Failure Time (AFT) Models

The accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models researchers measure the direct effect of the explanatory variables on the survival time instead of hazard. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time (Kalbfleisch J.and Prentice R., 2002). The AFT model states that the survival function of an individual with covariate X at time t is the same as the survival function of an individual with a baseline survival function at a time $t * \exp(\beta^t \mathbf{X})$, where $\beta^t = (\beta_1, \beta_2, \ldots, \beta_k)$ is a vector of regression coefficients. In other words, AFT model is defined by:

$$S(t/X = S_0\left\{t * \exp(\beta^t X)\right\}, for all X \text{----------(17)}$$

The natural logarithm of the survival time $Y = \log(T)$ is modeled. This is the natural transformation made in linear models to convert positive variables to observations on the entire real line. A linear model is assumed for Y:

$$Y = \log(T) = \mu + \beta_1 X_1 + ... + \beta_p X_p + \sigma W \text{ ----------} \quad (18)$$

Where: $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ are parameters of a px1 vector

$\sigma =$ is scale parameter

W = is the an error term

When S_0 is denote by the survival function with X = 0 then

$$P(t > t / X) = P(Y > \log(t) / X)$$

$$= P[\mu + \sigma W > \log(t) - \beta^t X / X)]$$

$$= P[\exp(\mu + \sigma W > t * \exp(-\beta^t X / X)]$$

$$= S_0[t * \exp(-\beta^t X)]$$

The effect of the covariates on the survival function is that the time scale is changed by a factor $\exp(-\beta^t X)$, and is called as an acceleration factor. The acceleration factor is explained as follows

If $\exp(-\beta^t X) > 1 \rightarrow$ the survival process accelerates.

If $\exp(-\beta^t X) < 1 \rightarrow$ the survival process decelerates.

If X is an indicator variable, then it is equivalent to

$\beta > 1 \rightarrow$ Time shrinks

$\beta < 1 \rightarrow$ Time accelerates

The baseline distributions used in these studies are Weibull, lognormal and log-logistic (Collett D., 2003 and Klein, 2005).

Table 3.2: Parametric distributions for the baseline hazards

AFT models	$f(t)$	$S(t)$	$h(t)$
Weibull	$\lambda\rho t^{\rho-1}\exp(-\lambda t^{\rho})$	$\exp(-\lambda t^{\rho})$	$\lambda\rho t^{\rho-1}$
Lognormal	$\frac{1}{t\sigma\sqrt{2\pi}}\exp(-\frac{[\log(t)-\mu]^2}{2\sigma^2})$	$1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$	$\frac{\Phi\frac{(\log(t))}{\sigma}}{\frac{(\log(t))}{\sigma}\sigma t}$
Log-logistic	$\frac{\rho t^{\rho-1}\lambda^{\rho}}{(1+(\lambda t)^{\rho})^2}$	$\frac{1}{1+(\lambda t)^{\rho}}$	$\frac{\rho t^{\rho-1}\lambda^{\rho}}{1+(\lambda t)^{\rho}}$

Parameter Estimation for classical parametric Survival Models

Parameters in survival regression models can be estimated by maximum likelihood method. Suppose we have a censored sample (Y_1, δ_1) ,…, (Y_n, δ_n)

Where: $Y_i = min(Y_i, C_i)$ and $\delta_i = l(Y_i \leq C_i)$, i=1,2,…, n, with a $T_1, T_2, \ldots, T_n \sim f(t)$ and survival function by S(t), T_i and C_i are independent and let β be the unknown parameter. The likelihood function for right censored data is given by:

$$L(\beta) = \prod_{i=1}^{n} f_i(y_i, \beta)^{\delta_i} * (S(y_i, \beta))^{1-\delta_i} \quad \text{----------(19)}$$

$$= \prod_{i=1}^{n} \left(\frac{f(yi,\beta)}{S(y_i,\beta)}\right)^{\delta_i} S(y_i, \beta)$$

$$= \prod_{i=1}^{n} (h(y_i, \beta))^{\delta_i} S(y_i, \beta)$$

Using logarithm on both sides

$$\log L(\beta) = \sum_{i=1}^{n} \delta_i \log h(y_i, \beta) + \sum_{i=1}^{n} S(y_i, \beta) \quad \text{----------(20)}$$

The maximum likelihood parameters estimates are found by using Newton-Raphson procedure.

3.8. Model Assessment

Every step during model fitting uses the upcoming statistical procedures and later at the end the investigator checks all the assumptions needed for the model. Some of the statistical procedures that are used to assess the final models in these studies are as follows:

3.8.1. Checking the Adequacy of Parametric Baselines

After fitting the model, the model diagnostic checking is used to know how effective the model is in describing the outcome variable. This is referred to as goodness of fit. The graphical methods can be used to check if a parametric distribution fits the observed data. 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. Log-logistic baseline can graphically be evaluated by plotting $\log\left(\frac{S(t)}{1-S(t)}\right)$ versus log(time) where $\hat{S}(t)$ is Kaplan-Meier survival estimate. If the plot is straight

line, log-logistic distribution fitted the given dataset well. If the plot $\Phi^{-1}[1 - S(t)]$ against log (t) is linear, the Log-normal distribution is appropriate for the given data set (Datwyler C. and Stucki T. , 2011).

3.8.2. The Cox- Snell Residuals

The Cox-Snell residuals method is used to any parametric model and the residual plots can be applied to check the goodness of fit of the model. For the parametric regression problem, analogs of the semi-parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein, J. and Moeschberger, M. , 2003). The Cox-Snell residual for the j^{th} individual with observed survival time t_j is given by:

$$r_j = \hat{H}(T_j / X_j) = -\log \hat{S}(T_j / X_j) \quad \text{----------} \quad (21)$$

Where: $\hat{H}$ and $\hat{S}$ are the estimated values of the cumulative hazard and survivor function of the j^{th} subject at time t_j respectively. If the model fits the data, then the $r_j's$ should have a standard (λ=1) exponential distribution, so that a hazard plot of r_j versus the Nelson–Aalen estimator of the cumulative hazard of the $r_j's$ should be a straight line with slope unity and zero intercept. If yes, the fitted model is adequate. In general, Cox-Snell residual that provides a check of the overall fits of the model (Cox and Snell, 1968).

3.9. Bayesian method of survival data analysis

The Bayesian approach analysis considers the parameters of the model as random variables and requires that prior distributions specified for them and data are considered as fixed. The key ingredients to a Bayesian analysis are the likelihood function, which reflects information about the parameters contained in the data, and the prior distribution, which quantifies what, is known about the parameters before observing data. The prior distribution and likelihood can be easily combined to form the posterior distribution, which represents total knowledge about the parameters after the data have been observed (Christensen R., 2011). In the Bayesian inferences the researcher introduces the AFT models used in the classical survival analysis with prior distributions.

3.9.1. Prior Distribution

The prior distribution is a probability distribution that represents the prior information associated with the parameter of interest. It is a key aspect of a Bayesian analysis. There are two types of prior distribution, informative priors and non-informative priors. An informative for θ prior is a prior distribution that is used when information about the parameter of interest is available before the data is collected, and this information is to be included in the analysis. Typically, informative prior distributions are created from historical studies, pure expert knowledge (experience) and a combination of both. A "non-informative" prior distribution is used to express complete ignorance of the value of before the data is collected. In the non-informative sense no value is favored over any other and are also described as diffuse or flat at prior due to this reason and their shape. The most common non-informative prior is the uniform distribution over the range of the sample space for θ (Ibrahim J. et al, 2001). In this study the researcher used normal prior distribution for coefficients with mean zero and variance 1000 and inverse-gamma prior distribution with scale parameter a=0.01 and shape parameter b=0.01 (Ghassan H., 2013). In these models, both of β and σ are unknown, no joint conjugate prior is available. A typical joint prior specification can be expressed as a product of a multivariate normal (for parameter β/σ^2) and an inverse gamma prior (for σ^2), that is

$$\frac{\beta}{\sigma^2} \sim N_p(\mu_0, \sigma^2 V_0); \;\; \sigma^2 \sim IG(a,b)$$

Likelihood Function

A likelihood functions is a function that gives the probability of observing of the sample data given the current parameters. Suppose we observe n independent vectors of (T_i, δ_i, x_i), where T_i is time to the event and δ_i is indicator variable telling us whether T_i is uncensored or censored.

$$\delta_i = \begin{cases} 0 \;\; censoring\, observation \\ 1\, event or\, dead\, (\, failure) \end{cases}$$

The likelihood function of the set of unknown parameters θ in the presence of right censoring is written as

$$L(\theta) = \prod_{i=1}^{n} \left[f(t_i / x_i, \theta)^{I(\delta i=0)} * S(t_i / x_i, \theta)^{I(\delta i=1)} \right] \text{----------(22)}$$

Log-likelihood would be as follows

$$l(\theta) = \log\left\{\prod_{i=1}^{n}\left[f(t_i / x_i, \theta)^{I(\delta i=0)} * S(t_i / x_i, \theta)^{I(\delta i=1)}\right]\right\}$$

$$l(\theta) = \sum_{i=1}^{n}\left[\log f(t_i / x_i, \theta)^{I(\delta i=0)} + \log(S(t_i / x_i, \theta)^{I(\delta i=1)})\right] \text{----------(23)}$$

Where: $f(t_i / x_i)$ and $S(t_i / x_i)$ are the density and survival distributions..

3.9.2. Posterior Distribution

The posterior distribution is obtained by multiplying the prior distribution over all parameters, θ by the full likelihood function $L(\theta / X)$. All Bayesian inferential conclusions are based on the posterior distribution of the model generated. The inference is performed by sampling from posterior distribution until the convergence to the posterior distribution is achieved (Dezfuli H. et al , 2009). The major problem in the Bayesian approach is that in the most cases the full form of the posterior distribution cannot be obtained in closed form, that is, the posterior density may not belong to standard distribution. Such problem cannot be solved easily. In order to solve such problems the researcher used MCMC iteration method. Then the researcher assumed that θ is a random variable and has a prior distribution denoted by π(θ). Inference concerning θ is then based on the posterior distribution, which is obtained by Bayes' theorem. Then posterior distribution of θ is given by:

$$\pi(\theta / X) = \frac{L(X / \theta)\pi(\theta)}{\int L(X / \theta)\pi(\theta)d\theta}$$

$$\pi(\beta_i / X) = \frac{L(X / \beta_i)\pi(\beta_i)}{\int L(X / \beta_i)\pi(\beta_i)d\beta_i} \text{----------(24)}$$

Combining the likelihood function with the prior distribution on (β, σ^2) and the full conditional distributions for unknown parameters, the posterior distribution can be written as:

$$\prod(\frac{\beta}{\sigma^2}, t, x) \propto \prod_{i=1}^{n}\left[f(t_i / x_i, \theta)^{I(\delta i=0)} * S(t_i / x_i, \theta)^{I(\delta i=1)}\right] * \prod(\frac{\beta}{\sigma^2})$$

$$\prod(\frac{\beta}{\sigma^2}, t, x) \propto \prod_{i=1}^{n}\left[f(t_i / x_i, \theta)^{I(\delta i=0)} * S(t_i / x_i, \theta)^{I(\delta i=1)}\right] * \prod(\frac{\beta}{\sigma^2}) * \prod(\sigma^2) \text{----------(25)}$$

The posterior distribution for the model specification above does not have closed form solution for the parameters. For these models, MCMC-Gibbs sampler is implemented using WinBUGS

software. The baseline hazard distributions used in the classical parametric survival analysis such as Weibull, lognormal and log-logistic in table 3.2 above are also used in Bayesian method by introducing prior for each parameters.

3.9.3. MCMC Estimation methods

The Bayesian approach applies probability theory to a model derived from substantive knowledge and theory, deal with realistically complex situations; the approach can also be termed 'full probability modeling'. There has recently been enormous progress in methods for Bayesian computation, generally exploiting modern computer power to carry out iteration known as Markov Chain Monte Carlo (MCMC) methods. The MCMC iteration is used to do the integration numerically rather than analytically by sampling from the posterior distribution of interest even when the form of that posterior has no known algebraic form (Spiegelhalter D., 2004).

3.9.4. Gibbs Sampler

Gibbs sampler is an algorithm that sequentially generates samples from a joint distribution of two or more random variables (Ibrahim J. et al, 2001). The sampler is often used when:
The joint distribution $\pi(\theta/X)$, is not known explicitly. The full conditional distribution of each parameter is not known. Gibbs Sampler Algorithm is written as follows:

1. Choose an arbitrary initial value of $\theta^{(0)} = \left\{\theta_1^{(0)}, \theta_2^{(0)}, \dots \theta_k^{(0)}\right\}$
2. For I = 0,1,2,…,N-1, generate each component of θ as follows:
   a. Draw $\theta_1^{(i+1)}$ from $\pi(\theta_1|\theta_2^i, \theta_3^i, \dots, \theta_k^i, Y, X)$
   b. Draw $\theta_2^{(i+1)}$ from $\pi(\theta_2|\theta_1^{i+1}, \theta_3^i, \dots, \theta_k^i, Y, X)$
   c. Draw $\theta_3^{(i+1)}$ from $\pi(\theta_3|\theta_1^{i+1}, \theta_2^{i+1}, \theta_4^i, \dots, \theta_k^i, Y, X)$
   d. …
   e. Draw $\theta_k^{(i+1)}$ from $\pi(\theta_k|\theta_1^{i+1}, \theta_2^{i+1}, \dots \theta_{k-1}^{i+1}, Y, X)$
3. Repeat step 2 until convergence
4. Return $\theta^{(b+1)} = (\theta_1^{(b+1)}, \theta_2^{(b+1)}, \dots \theta_k^{(b+1)}), \theta_1^{(b+2)}, \theta_2^{(b+2)}, \dots, \theta^{(N)}$

The means of the posterior samples provide point estimates for the model parameters. The 95% credible intervals provide an alternative indication of the covariates' effects along with estimation precision. The MC error is an estimate of the difference between the mean of the sampled values (which researcher use as an estimate of the posterior mean for each parameter) and the true posterior mean or the MC error (SEM) shows how much uncertainty there is about the true posterior mean via the sampled mean. As a rule of thumb, the iteration should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation (Muluneh S. et al , 2011).

3.10. Model Comparison and Selection

Akaikie informative criterion (AIC)

Akaikie (1974) proposed an informative criterion (AIC) statistic to compare classical models with different number of parameters. For each model the value is computed as:

$$AIC = -2\log likelihood + 2(k+c) \quad (26)$$

Where: k is the number of parameters and c the number of model specific distributional parameters. The preferable model is the one with the lowest value of the AIC (Munda M., 2012).

Deviance Information Criteria (DIC)

$$pD = E[-2\log\{f(y/\theta)\}] + 2\log\{f(y/\hat{\theta})\}$$

$$DIC = -2\log\{f(y/\hat{\theta})\} + pD \quad (27)$$

Where: pD is effective number of parameters in the model, $\hat{\theta}$ is maximum likelihood estimate.

DIC is used for Bayesian survival model comparison. The preferable model is the one with the lowest value of the DIC (Spiegelhalter D., 2004).

Bayesian information criteria (BIC)

$$D(\theta) = -2\log(likelihood(\theta/data))$$

$$BIC = E[D(\theta)/data] + p\log(n) \quad (28)$$

Where: $D(\theta)$ is deviance and $E[D(\theta)/data]$ is the posterior mean of the deviance.

The preferable model is the one with the lowest value of the BIC. BIC penalizes models which improve fit at the expense of more parameters, A problem is that the true dimensionality (number of parameters p) of the model is often not known and also that the number of parameters may increase with sample size n (Antonietta M., 2013), Due to this reason the researcher do not use BIC values to compare classical and Bayesian approach analysis; therefore researcher used precision (standard errors) of each of significant parameters to compare both approach analysis. The DIC value is used to compare Bayesian approach models.

3.11. Model Diagnostics

Once a model has been developed, the researcher would like to know how effective the model is in describing the outcome. This is referred to as goodness of fit. The most common ways of checking goodness of fit in Bayesian approach analysis are diagnosis for convergence and mixing. Diagnosis of the convergence is important to answer the questions of how to determine whether the sampler has reached its stationary distribution. To use summary statistics of the estimated posterior distribution for the parameters; the MCMC iteration should converge. To check these investigators have used four convergence checking methods for Bayesian analysis.

Time series or history plot: are commonly used methods to assess convergence (Merke P., 2005). If the plot looks like a horizontal band, with no long upward or downward trends and different independent initial values of the chains should be mix together or overlapped then the researchers have evidence to say that the chains has converged.

Autocorrelation Plot: Another way to assess convergence is to evaluate the autocorrelation between the draws of the Markov chain, which is a measure of dependency among Markov chain samples. The researcher would expect that the k^{th} lag autocorrelation to be smaller as k increases, which mean that our 2^{nd} and 50^{th} draws should be less correlated than our 2^{nd} and 4^{th} draws. If autocorrelation is still relatively high for higher values of k, this indicates a high degree of correlation between our draws and slow mixing.

Gelman-Rubin Statistic: For a given parameter, Gelman-Rubin statistic assesses the variability within parallel chains as compared to variability between parallel chains (Gelman A. and Brooks A., 1998). The model is judged to have converged if the ratio of between to within variability is

close to one. The green line represents the between variability, the blue line represents the within variability, and the red line represents the ratio. Evidence for convergence comes from the red line being close to one on the y-axis and from the blue and green lines being stable (horizontal) across the width of the plot.

Kernel Density plot: The plots for the parameters of predicator variables should be resemble the curves of normal distribution if so the simulated parameter values are converged.

3.12 Ethical Consideration

The Research Ethics Review Board of Jimma University has provided an ethical clearance for the study. The data was collected from Tercha General Hospital, and to do so the department of statistics asked to write an official co-operation letter to the Hospital from where data was obtained. The study conducted without individual informed consent because it relied on retrospective data. In this research, the information obtained from the pediatric registration charts and patients' card kept secured.

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Descriptive Statistics

Descriptive statistics is the beginning of any statistical analysis before proceeding to more complicated models. This study included a total of 281 under-five pneumonia patients fulfilling the inclusion criteria in Tercha General Hospital. Summary results for covariates included in this study are presented in Table 4.1 below.

Table .4.1: Descriptive Summary of Pneumonia Data in TGH (2016-2017)

Status of Patient				
Variable	**Category (codes)**	**Number of Event (%)**	**Number of Censored (%)**	**Total (%)**
Sex	Female(0)	20(42.55%)	106(45.30%)	126(44.84%)
	Male(1)	27(57.45%)	128(54.70%)	155(55.16%)
Age	1-11 (0)	21(44.68%)	119(50.85%)	140(49.82%)
	12-23 (1)	11(23.40%)	51(21.79%)	62(22.06%)
	24-35 (2)	7(14.89%)	25(10.68%)	32(11.39%)
	36-47 (3)	5(10.64%)	24(10.26%)	29(10.32%)
	48-59 (4)	3(6.38%)	15(6.41%)	18(6.41%)
Residence	Rural (0)	34(72.34%)	160(68.38%)	194(69.04%)
	Urban (1)	13(27.66%)	74(31.62%)	87(30.96%)
Season of Diagnosis	Autumn (0)	7(14.89%)	74(31.62%)	81(28.83%)
	Winter (1)	6(12.77%)	27(11.54%)	33(11.74%)
	Spring (2)	16(34.04%)	73(31.20%)	89 (31.67%)
	Summer (3)	18(38.30%)	60(25.64%)	78(27.76%)
Co-morbidity	No (0)	33(70.21%)	137(58.55%)	170(60.50%)
	Yes (1)	14(29.79%)	97 (41.45%)	111(39.50%)
Sever Acute Malnutrition (SAM)	No (0)	30(63.83%)	195(83.33%)	225(80.07%)
	Yes (1)	17(36.17%)	39(16.67%)	56(19.93%)

Treatment types taken by patients	Penicillin (0)	8(17.02%)	45(19.23%)	53(18.86%)
	Ceftriaxone (1)	9(19.15%)	36(15.38%)	45(16.01%)
	Ampicillin (2)	14(29.79%)	78(33.33%)	92(32.74%)
	Combined (3)	16(34.04%)	75(32.05%)	91(32.38%)
Patient refer status	No (0)	40(85.11%)	172(73.50%)	212(75.44%)
	Yes (1)	7(14.89%)	62(26.50%)	69(24.56%)
Continuous variables		Mean	Standard deviation	
	Weight of patient	9.626	3.253	
	BOR	0.513	0.080	
	PPhR	13.900	2.764	
	PNR	3.420	1.130	

From the table 4.1 above the total of 281 patients of pneumonia included in the study, 44.84% of the patients were female and 55.16% male. Among those patients by considering sex, the death proportion for female is 42.55% which is lower than that of male patients which is 57.45%. Considering age groups included in the study total sample of patients 49.82%, 22.06%, 11.39%, 10.32% and 6.41% of patients were from age group 1-11, 12-23, 24-35, 36-47 and 48-59 respectively and the death proportion for the age group were 44.68%, 23.40%, 14.89%, 10.64% and 6.38% respectively. Of the total patients 69.04% were from rural area and 30.96% from the urban. Death proportions of patients with residences were 72.34% and 27.66% respectively. Out of the total patients, 28.83% were in Autumn, 11.74% were in Winter, 31.67% were in Spring and 27.76% patients were in Summer. The death proportions of patients in Autumn, Winter, Spring and Summer patients were 14.89%, 12.77%, 34.04% and 38.30% respectively.

As shown in Table 4.1 above of total patients 60.50% patients were without Co-morbidity and 39.50% were with Co-morbidity. Death proportions among without co-morbidity and with co-morbidity were 70.21% and 29.79% respectively. Similarly in Sever Acute Malnutrition (SAM) case, out of the total patients there were 80.07% patients without Sever Acute Malnutrition and 19.93% were with Sever Acute Malnutrition. Death proportions among without Sever Acute Malnutrition and with Sever Acute Malnutrition were 63.83% and 36.17% respectively. Among under-five aged children included in the study, 18.86% patients took treatment type Penicillin,

16.01% patients took treatment type Ceftriaxone, 32.74% patients took treatment type Ampicillin and 32.38% patients took the Combination of two and above treatments types. The death proportions of patients who took Penicillin, Ceftriaxone, Ampicillin and Combination of two or above were 17.02%, 19.15%, 29.79%, and 34.04% respectively.

Among the total patients included in the study 75.44% patients were not referred from other healthy centers and 24.56% patients were referred from other healthy centers. Death proportion among patients who were not referred from other health center and patients who were referred from other health center were 85.11% and 14.89% respectively. The mean weight of the patients included in the study was 9.626 with standard deviation of 3.253. The mean of Bed occupancy rate (BOR) at the time of study period was 0.513 with the standard deviation of 0.08. The mean of patient to physician ratio (PPhR) at the time of study period was 13.900 with the standard deviation of 2.76. The mean of patient to nurse ratio (PNR) at the time of study period was 3.420 with the standard deviation of 1.130. After the medical cards of pediatric were reviewed among those patients of under-five pneumonia 47(16.73%) died and 234(83.27%) were censored.

4.2 Non-parametric Survival Analysis

Non-parametric methods in survival analysis is very important to visualize the survival time of patients under different groups of covariates; the Kaplan-Meier Estimate and log-rank test are used to compare the survival rates of two or more groups of under-five pneumonia patients. According to Figure 4.1 below, survival probability for patients who were come from urban had better survival time than those who were come from rural. Also the log-rank test in Table 4.2 below demonstrated significant difference between patients who were come from urban and rural (p=0.001) at 5% level of significance.

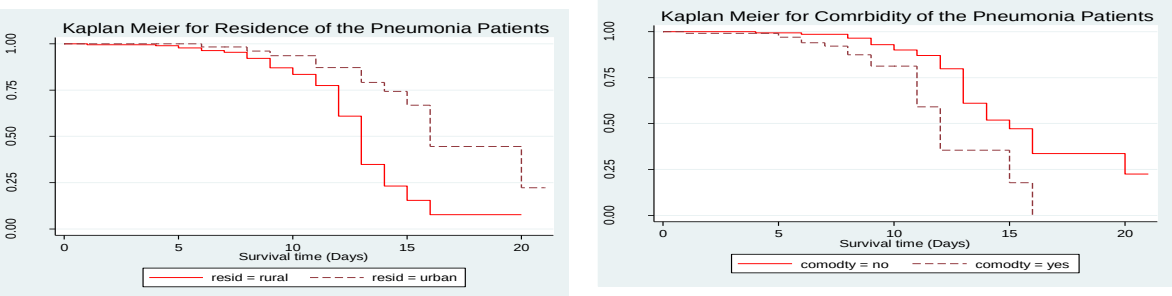


Figure 4.1: K-M plot for Residence and Co-morbidity of pneumonia patients

KM plot for comorbidity in the above figure indicates patients who had not experienced comorbidity had higher survival time than patients who had experienced co-morbidity. The log-rank test also revealed that there is significant difference between two groups (p=0.000) at 5% level of significance.

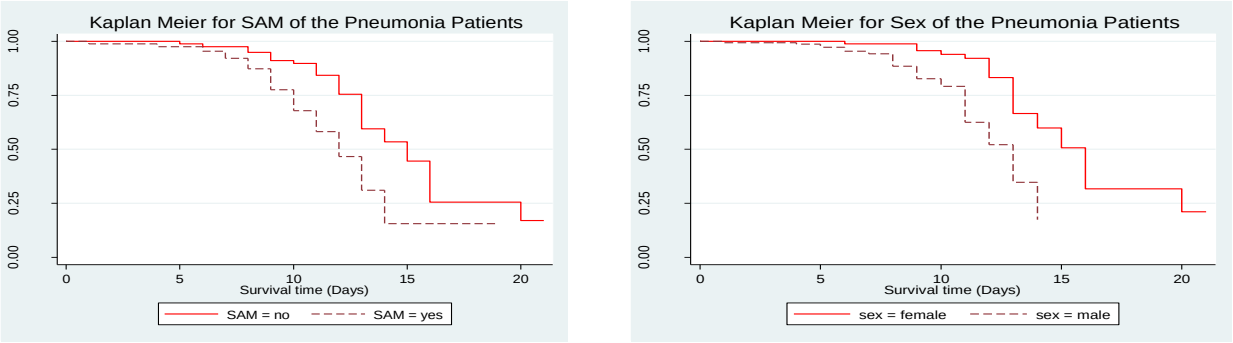


Figure 4.2: K-M plot for SAM and sex of under-five pneumonia patients

The survival time plot by severe acute malnutrition (SAM) is given in figure 4.2 above showed that patients who had not suffered of severe acute malnutrition had better survival time than patients who had suffered of severe acute malnutrition (SAM) and long rank test in the table 4.2 below also revealed that SAM had significant difference among two groups (p=0.000) at 5% level of significance. The K-M survival time plot for sex in figure 4.2 above, showed that female patients had better survival time than male patients and long rank test in the table 4.2 below also revealed that there is significant difference between two groups (p=0.023) at 5% level of significance.

Table 4.2: Log Rank Tests of each Covariate

Variable	Chi-square	Degree freedom (df)	P-value
Sex	5.18	1	0.023
Age	0.996	4	0.71
Residence	10.91	1	0.001
Season of Diagnosis	3.983	3	0.263
Co-morbidity	14.064	1	0.000
SAM	12.784	1	0.000
Treatment types	10.158	3	0.017
Patient refer status	13.361	1	0.000

The survival time plot by patient's age group in appendix A figure 4.10 showed that the risk of death is slightly different for each age group. But the log rank test in table 4.2 above demonstrated that age group had no significant difference (p=0.71) at 5% level of significance. And also the KM plot and the log rank test for patient refer status revealed that there is significant difference between two groups. Similarly the survival time plot by season of diagnosis in appendix A figure 4.11 showed that the risk of death is slightly different for each seasons of diagnosis. But the log rank test in table 4.2 above demonstrated that season of diagnosis had no significant difference (p=0.263) at 5% level of significance. According to the log rank test in Table 4.2 above, treatment types taken by patients were statistically significant (p=0.017) in estimating the survival time of pneumonia patients. Similarly KM plots in Figure 4.11 appendix A, showed patients who took treatment type Penicillin and Ampicillin had better survival rate when compared to the remaining treatment types.

4.3 Semi-Parametric Cox PH

4.3.1. Univariable Analysis

In any data analysis it is always a great idea to do some univariable analysis before proceeding to more complicated models. Single covariate Cox proportional hazards model analysis is an appropriate procedure that is used to screen out potentially important variables before directly included in the multivariable model. The relationship between each covariates and survival time of under-five pneumonia patients are presented in table 4.3 below. As shown from this table, survival of the patients is significantly associated with sex, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patient refer status and patient nurse ratio (PNR) at 5% level of significance.

Table 4.3: Univariable Cox PH Analysis with time to event of Pneumonia Patients

Covariates	Categories	$\widehat{\beta}$	HR	SE	Wald	Sign	95%CI for HR
Sex	Female	Ref					
	Male	0.7088	2.0315	0.3144	2.254	.0242*	[1.097, 3.762]
Age	1-11	Ref					
	12-23	0.05267	1.05408	0.37343	0.141	0.888	[0.5070, 2.191]
	24-35	0.23989	1.27111	0.44263	0.542	0.588	[0.5338, 3.027]
	36-47	0.28436	1.32891	0.50249	0.566	0.571	[0.4963 ,3.558]
	48-59	0.47460	1.60737	0.62175	0.763	0.445	[0.4752 ,5.437]
Weight		-0.0123	0.98777	0.04432	-0.278	0.781	[0.9056 ,1.077]
Residence	Rural	Ref					
	Urban	-1.1343	0.3216	0.3539	-3.205	0.001 ***	[0.1608,0.6436]
Season of	Autumn	Ref					
Diagnosis	Spring	0.8307	2.2949	0.4545	1.828	0.0676.	[0.9416 , 5.593]
	Summer	0.9439	2.5701	0.4471	2.111	0.0347*	[1.0701 , 6.173]
	Winter	1.0797	2.9438	0.5625	1.920	0.0549.	[0.9775 , 8.865]
Co-	No	Ref					
morbidity	Yes	1.443	4.233	0.399	3.617	0.0003***	[1.9370, 9.2520]
SAM	No	Ref					
	Yes	1.154	3.170	0.326	3.539	0.0004***	[1.6703 , 6.0070]
	Penicillin	Ref					
Treatment	Ceftriaxone	0.4809	1.6176	0.4894	0.983	0.326	[0.6198 , 4.222]
types	Ampicillin	0.2261	1.2538	0.4515	0.501	0.616	[0.5175, 3.038]
	Combined	0.4155	1.5151	0.4343	0.957	0.339	[0.6468, 3.549]
Pat refer	No	Ref					
status	Yes	-1.0627	0.3455	0.4126	-2.575	0.01*	[0.1539, 0.7757]
BOR		-2.2601	0.1043	1.7066	-1.324	0.185	[0.0037, 2.959]
PPhR		-0.0851	0.91843	0.05346	-1.592	0.111	[0.8271, 1.02]
PNR		-0.3064	0.7361	0.1395	-2.196	0.0281 *	[0.56, 0.9675]

SE=Standard Error, HR= Hazard Ratio, CI=Confidence Interval, Ref. Reference, * significant (P-value < 0.05)

4.3.2. Multivariable Analysis

Results presented in Table 4.4 below indicates that the parameter estimates of coefficients 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. According to Table 4.3 above the predictor's such as sex of children, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patient refer status and patient nurse ratio (PNR) was passed the first filtration of variables for multiple covariates analysis and then forward variable selection method was used to select the important variables to be included in Cox proportional hazards 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 0.05 cut point or 5% significance level are considered as important variables and hence, are included in the final model. In the cox PH final model the survival time of under-five aged children who had suffered pneumonia was statistically significantly associated with sex, residence, season of diagnosis, co-morbidity, severe acute malnutrition (SAM), patient refer status and patient nurse ratio (PNR).

Table 4.4: Multivariable analysis of Cox PH with time to death of under-five pneumonia Patients.

Covariate	Categories	$\widehat{\beta}$	HR	SE	Wald	Sign	95%CI for HRs
Sex	Female	Ref	-------	-------	-------	-------	-------
	Male	0.7978	2.2207	0.3387	2.356	0.0184*	[1.1435,4.3128]
Residence	Rural	Ref	----------	--------	--------	-------	-------
	Urban	-0.9967	0.3691	0.4257	-2.341	0.0192*	[0.160, 0.8502]
Season of Diagnosis	Autumn	Ref	-------	-------	-------	-------	-------
	Winter	0.8301	2.2935	0.5990	1.386	0.1658	[0.709, 7.420]
	Spring	1.4276	4.1688	0.5237	2.726	0.0064**	[1.494,11.636]
	Summer	1.4668	4.3353	0.4812	3.048	0.0023**	[1.688, 11.134]
Co-morbidity	No	Ref	-------	-------	-------	-------	-------
	Yes	1.5450	4.6879	0.4445	3.476	0.0005***	[1.962, 11.202]
SAM	No	Ref	-------	-------	-------	-------	-------
	Yes	1.2665	3.5484	0.3874	3.270	0.0011**	[1.095, 5.0107]
Pat. refer status	No	Ref	-------	-------	-------	-------	-------
	Yes	-1.2030	0.3003	0.4404	-2.732	0.0063**	[0.127, 0.7118]
PNR		-0.5010	0.6059	0.1921	-2.608	0.0091 *	[0.416, 0.8829]

*SE=Standard Error, HR= Hazard Ratio, CI=Confidence Interval, Ref. Reference, * 5% significance level*

4.3.3. Statistical Tests of Proportional Hazards Model Assumptions

Goodness of fit testing approach is appealing because it provides a test statistic and p-value for assessing the PH assumption for a given covariates of interest. rho tells the relation between time and residuals. When the test of correlation (rho) is insignificant that indicates proportional hazards assumption is fulfilled. Table 4.5 below provided rho, chi-square test statistic and p-values for goodness-of-fit tests for each variable in the fitted model; based on the p-values in the table below variables such as sex, residence, co-morbidity, severe acute malnutrition (SAM) and patient refer status were satisfy the PH assumption. But variables such as season of diagnosis and patient nurse ratio (PNR) were not satisfying the PH assumption. Moreover it is also possible to see its global test and if it is greater than 0.05 the assumption have satisfied by the covariates in the model. In this study the global test is less than 0.05 the assumptions do not satisfied by the covariate in the model.

Table 4.5: Test of proportional hazards assumption

Covariates	rho	Chi-sq	DF	Sign
Sex	0.033	0.06	1	0.8049
Residence of patient	0.030	0.06	1	0.8134
Season of Diagnosis	-0.094	0.49	1	0.0484
Co-morbidity	-0.090	0.48	1	0.4907
SAM	0.066	0.24	1	0.6271
Patient refer status	0.062	0.17	1	0.6772
PNR	0.018	0.02	1	0.0361
Global test		7.98	7	0.0021

Chi-sq= chi-squared, DF=degree of freedom

The scatter plots of Scaled Schoenfeld residuals in Appendix A also used to check PH assumption. If the PH assumption is met, Schoenfeld residuals should look horizontal since the scaled Schoenfeld residuals would be independent of survival time. The plot of season of Diagnosis against survival time was slightly downward (not horizontal) and patient nurse ratio against survival time was also upward (not horizontal) in figure 4.3 below. These also revealed that there is a violation of the proportional hazard assumption for the covariates season of

Diagnosis and patient nurse ratio (PNR). Thus, the researcher doubt about the accuracy of the PH assumption, therefore the researcher was considered the AFT model for this data set.

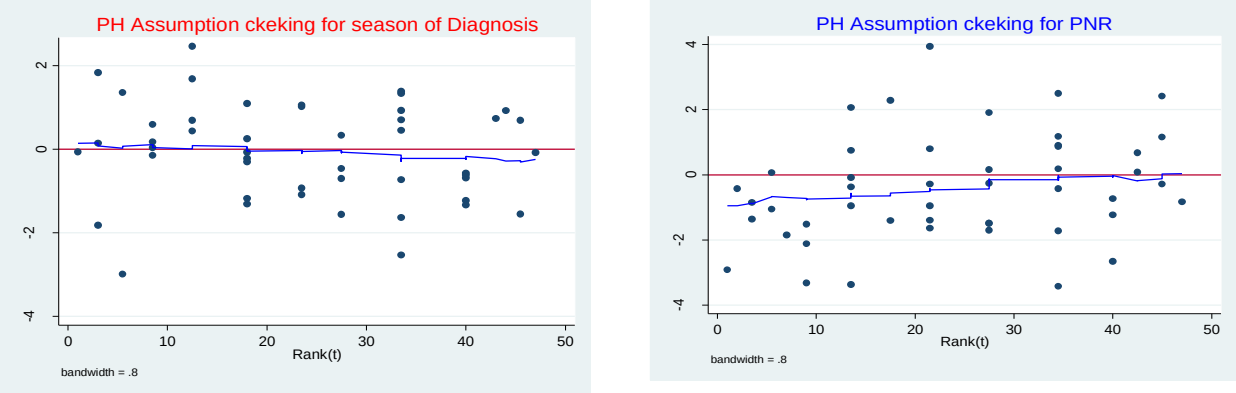


Figure 4.3: The plot of Scaled Schoenfeld residuals for season of Diagnosis and PNR to check PH assumption

4.4 Accelerated Failure Time Model

When Cox PH assumptions were not satisfied, alternatively parametric AFT models were used instead of the Cox PH model. The p-values of the goodness of fit for covariates were significant for the variables season of Diagnosis and patient nurse ratio (PNR). Due to these the researcher used AFT models to fit the under-five pneumonia data set in Tercha General Hospital.

4.4.1. Univariable AFT Analysis

Univariable analysis is used to see the effect of each covariate on survival time before proceeding to the multivariable analysis. The univariate analysis was fitted for each covariate by AFT models using different baseline distributions i.e. Weibull, lognormal and log-logistic distributions. In all univariable analysis of AFT model sex, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patients refer status and patient nurse ratio (PNR) were significantly associated with survival time of patients at 5% level of significance. The summary of univariable analysis is given in table 4.6 below. Hence, based on the univariable analysis, all explanatory variables filtrated in Cox PH are also candidate predictors for further analysis in AFT models.

Table 4.6: Univariable Weibull AFT model of under-five Pneumonia Patient.

Covariates	Categories	$\widehat{\boldsymbol{\beta}}$	SE	z	Sign	95%CI
Sex	Female	*Ref*				
	Male	-0.196	0.0718	-2.73	0.006*	[-0.3370, -0.0555]
Age	1-11	*Ref*				
	12-23	-0.0212	0.0969	-0.219	0.827	[-0.2112, 0 .1688]
	24-35	-0.0385	0.1146	-0.336	0.737	[-0.2631, 0.1860]
	36-47	-0.0926	0.1299	-0.713	0.476	[-0.3471, 0 .1619]
	48-59	-0.1425	0.1608	-0.886	0.375	[-0.4577, 0.1726]
Weight		0.0052	0.01165	0.442	0.658	[-0.0177, 0 .0280]
Residence	Rural	*Ref*				
	Urban	0.2495	0.0782	3.19	0.001*	[0.0962, 0.4028]
Season of	Autumn	*Ref*				
Diagnosis	Winter	-0.309	0.145	-2.13	0.033	[-0.5931, -0.0246]
	Spring	-0.221	0.119	-1.85	0.064	[-0.4552, 0.0128]
	Summer	-0.250	0.118	-2.11	0.035	[-0.4817, -0.0183]
Co-	No	*Ref*				
morbidity	Yes	-0.2865	0.0754	-3.80	0.000*	[-0.4343, -0.1388]
SAM	No	*Ref*				
	Yes	-0.277	0.0751	-3.69	0.000*	[-0.4245, -0.1299]
Treatment	Penicillin	*Ref*				
types	ceftriaxone	-0.1414	0.1258	-1.124	0.261	[-0.3880, 0.1052]
	Ampicillin	-0.0431	0.1149	-0.375	0.708	[-0.2683, 0.1821]
	Combined	-0.1179	0.1122	-1.051	0.293	[-0.3377, 0.1019]
Pt refer	No	*Ref*				
status	Yes	0.264	0.1057	2.5	0.012*	[0.0565, 0.4707]
BOR		0.6247	0.4524	1.38	0.167	[-0.2620, 1.511]
PPhR		0.0222	0.0139	1.60	0.109	[-0.0050, 0.0494]
PNR		0. 0745	0.0366	2.04	0.042*	[0.0028, 0.1462]

SE=Standard Error, HR= Hazard Ratio, CI=Confidence Interval, Ref. Reference, * significant (P-value < 0.05).

4.4.2. Multivariable AFT Analysis

For survival time of under-five pneumonia patients data, multivariable AFT models of Weibull, log-logistic and log-normal distribution were fitted by including all the covariates those are significant in the univariable analysis at 5% level of significance. To compare the efficiency of different models, the AIC was used. It is the most common applicable criterion to select model. Based on AIC, a model having the minimum AIC value was preferred. Accordingly, from the table 4.7 below Weibull AFT model has (AIC=334.128) found to be good for the survival time of pneumonia patients data set from the given alternatives when include all the covariates those are significant in the univariable analysis. All covariates significant in the univariable become significant in the multivariable analysis model. Finally, the effect of interactions terms were also tested and found to be statistically insignificant in multivariable Weibull AFT model at 5% level of significance. The final model covariates are sex, residence, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patients refer status and patient nurse ratio (PNR). All AFT models and the corresponding AIC and BIC values were displayed in Table 4.7 below to compare classical AFT models with different baseline distributions.

Table 4.7: Classical AFT models Comparison

Model types	Log-Likelihood	AIC	BIC
Weibull	-146.0641	334.1282	324.8574
Lognormal	-160.5716	363.1433	335.6928
Log-logistic	-149.9426	341.8852	331.0583

AIC=Akaike Information criteria, BIC=Bayesian Information criteria

Assessment of model Adequacy for Weibull accelerated failure time Regression Model

From the likelihood ratio test in table 4.8 below, it implies that the model is significant and log likelihood values of the null model and the full model showed that the model has a significant improvement after the covariates were added in the model.

Table 4.8: Assessment of model adequacy for Weibull AFT model

Loglik(intercept only)	Loglik(model)	Chi-sq	DF	Sign
-176.8	-146.1	61.39	19	0.000

4.4.3. Interpretation and presentation of the Classical final AFT model

The output of the final Weibull AFT model is presented in Table 4.9 below. Thus the acceleration factors in the Weibull model were interpreted as follows:

Table 4.9: Final Multivariable Analysis for classical Weibull AFT model.

Covariates	**Categories**	$\widehat{\boldsymbol{\beta}}$	**SE**	**z**	**sign**	$\boldsymbol{\gamma}$	**[95% CI]** $\boldsymbol{\gamma}$
Sex	*female*	*Ref*	-------	-------	-------	-------	-------
	male	-0.130	0.060	-2.18	0.029*	0.878	[0.782, 0.987]
Residence	*rural*	*Ref*	-------	-------	-------	-------	-------
	urban	0.147	0.070	2.11	0.035*	1.158	[1.010, 1.328]
Season of	*Autumn*	*Ref*	-------	-------	-------	-------	-------
Diagnosis	Winter	-0.060	0.115	-0.52	0.602	0.942	[0.751, 1.181]
	Spring	-0.169	0.088	-2.07	0.003*	0.845	[0.720, 0.991]
	Summer	-0.207	0.088	-2.35	0.001*	0.813	[0.683, 0.966]
Co-Morbidity	*No*	*Ref*	-------	-------	-------	-------	-------
	Yes	-0.127	0.068	-1.88	0.007*	0.881	[0.640, 0.916]
SAM	*No*	*Ref*	-------	-------	-------	-------	-------
	Yes	-0.139	0.066	-2.09	0.036*	0.870	[0.764, 0.901]
Patient refer	*No*	*Ref*	-------	-------	-------	-------	-------
status	Yes	0.174	0.070	2.48	0.013*	1.190	[1.038, 1.366]
PNR		0.091	0.037	2.44	0.015*	1.095	[1.018, 1.177]
Intercept		2.316	0.295	7.86	0.000*		[1.738, 2.893]
Scale= 0.189							

*γ Indicates Acceleration factor; *indicate: Significance at 5% level; 95%CIγ: 95% confidence interval for acceleration factor; SE: standard error for estimates; Ref. Reference*

Interpretation of Classical Accelerated Failure Time Model Parameters

Based on the above table 4.9 the final model were interpreted using acceleration factor, 95% confidence interval of acceleration factor and p-value of the estimate of accelerated failure time model. Under the Weibull AFT model, when the effect of other factor keep fixed, the estimated acceleration factor for male patient is estimated to be 0.878 with [95% CI: 0.782, 0.987]. The confidence interval for the acceleration factor did not include one and p-value is small (p=0.029). This indicates that male patients have less survival time than female patients or in the other way female patients survived 12.2% longer that male patients. The acceleration factors for patients whose residence was urban were estimated to be 1.158 with [95% CI: 1.010, 1.328]. The confidence interval for the acceleration factor did not include one and p-value is small (p=0.035). This indicates that patients whose residence was urban had prolonged survival time than patients from rural residence at 5% level of significance.

As shown in table 4.9 above the estimated acceleration factor for patients diagnosed at spring season and summer season were 0.845 and 0.813 with [95% CI: 0.720, 0.991 and 0.683, 0.966] respectively. The confidence intervals for both acceleration factor did not include one and p-values were small (p=0.003 and 0.001) respectively. This implies that patients who were diagnosed at spring and summer season had less survival time than patients who was diagnosed at autumn season. But patients who were diagnosed at winter season survival time were not significantly different from patients who were diagnosed at autumn season at 5% level significance. The acceleration factor for patients who were suffered co-morbidity was estimated to be 0.881 with [95% CI: 0.640, 0.916]. The confidence interval did not includes one and p-values is small (p=0.007). This implies that patients who were not suffered co-morbidity had longer survival time than patients who were suffered co-morbidity.

The acceleration factor for patients who were suffered severe acute malnutrition (SAM) was estimated to be 0.870 with [95% CI: 0.764, 0.991]. The confidence interval did not includes one and p-values is small (p=0.036). This indicates that patients who were not suffered severe acute malnutrition (SAM) had longer survival time than patients who were suffered severe acute malnutrition (SAM) or in the other ways patients who were not suffered severe acute malnutrition (SAM) survived 13% longer than that of patients who were suffered severe acute malnutrition (SAM) at 5% level of significance. The acceleration factor for patients who were referred from other health centers was estimated to be 1.190 with [95% CI: 1.038, 1.366]. The confidence interval for the acceleration factor did not include one and p-value is small (p=0.013). This indicates that patients who were referred from other health center had prolonged survival time than patients who were not referred from other health center at 5% level of significance. Acceleration factor for patient nurse ratio was estimated to be 1.095 with [95% CI: 1.018, 1.177]. The confidence interval did not include one and p-value is small (p=0.015). This indicates that patient nurse ratio had significant effect on the survival time of patients at 5% level of significance.

4.4.4. Model Diagnostics

After the model has been fitted, it is desirable to determine whether a fitted parametric model adequately describes the data or not.

Checking Adequacy of Parametric Baselines using Graphical Methods

To check the adequacy of our baseline hazard the Weibull, lognormal and log-logistic are plotted. If the plot is linear, the given baseline distribution is appropriate for the given dataset. Accordingly, their respective plots are given in figure 4.4 below; the plot for the Weibull baseline distribution make approximately straight line which is better than Log-logistic and lognormal baseline distribution. This evidence also strengthens the decision made by AIC value that Weibull baseline distribution is appropriate for the given dataset.

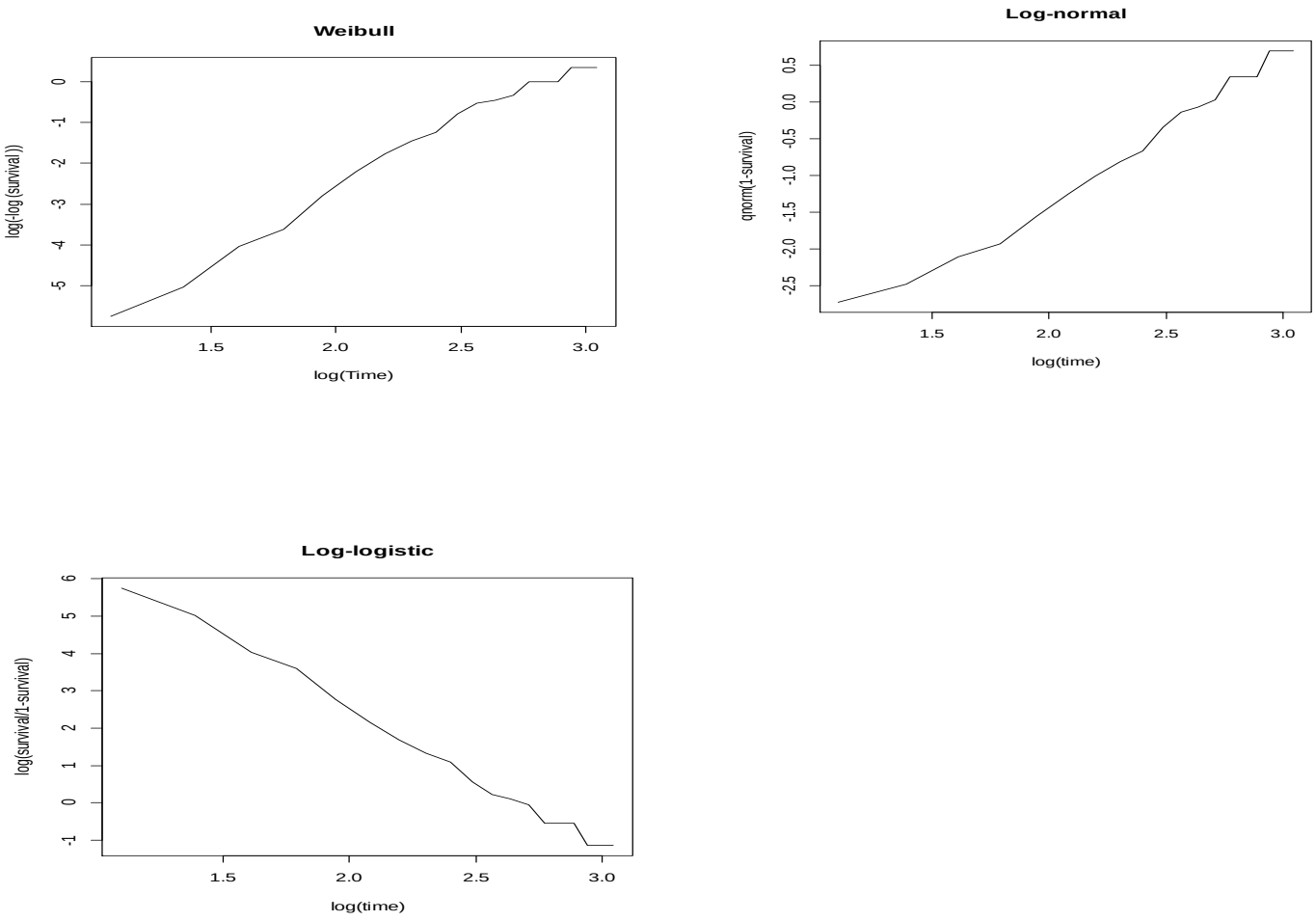


Figure 4.4: Graphs of Weibull, Log-normal and Log-logistic baseline distributions for survival time of pneumonia patients' data set.

Cox- Snell residuals plots

The Cox-Snell residuals are one way to investigate how well the model fits the data. From the figures 4.5 below the Weibull baseline distribution plot makes approximately the straight line through the origin than log-logistic and lognormal baseline distributions, so this plot suggests that Weibull AFT model is an appropriate model to fit the survival time of under-five pneumonia patients' data set in Tercha General Hospital.

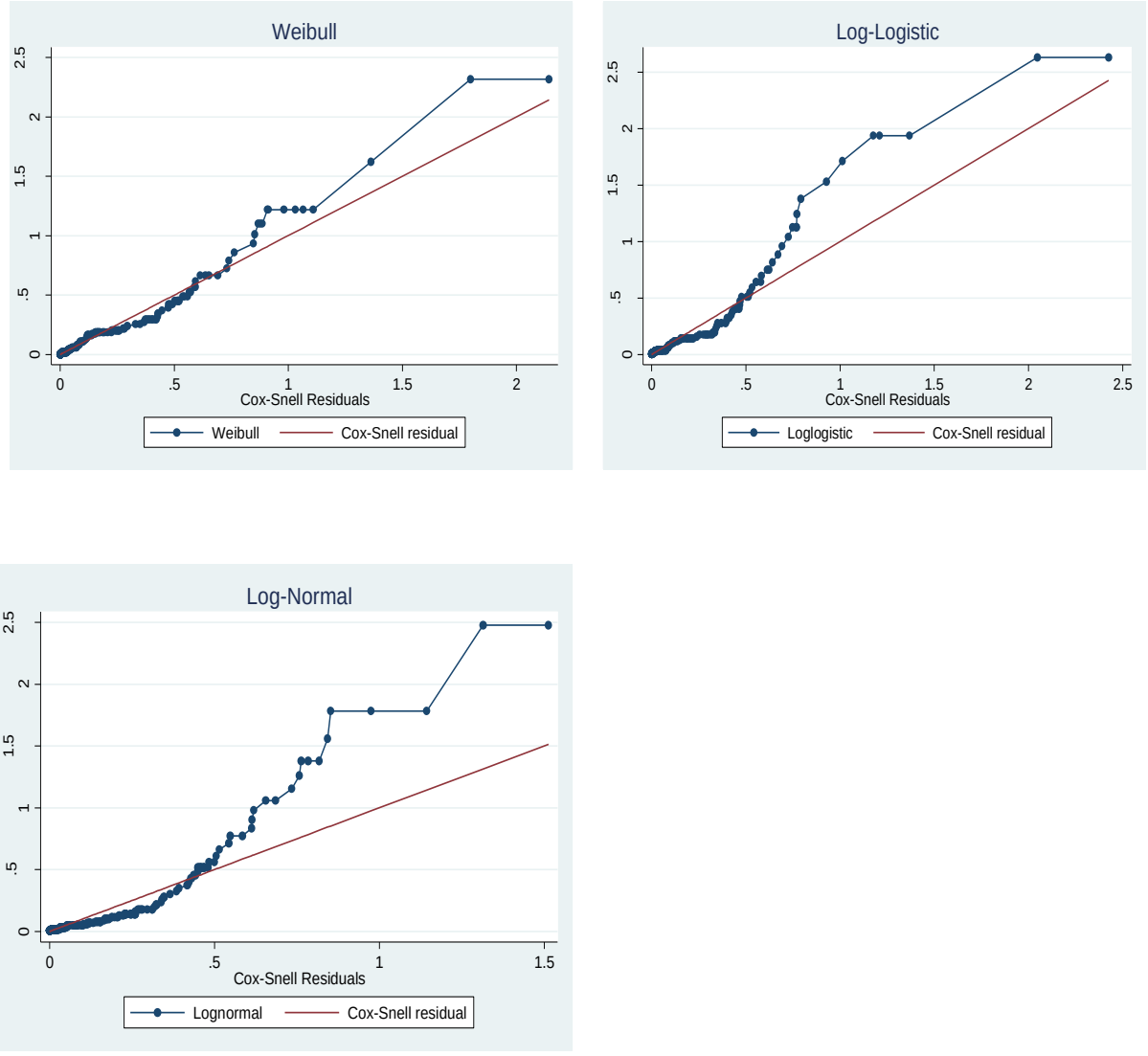


Figure: 4.5: Cox- Snell residual plots for Weibull, log-logistic and lognormal baseline distribution for survival time of under-five pneumonia patients' data set.

4.5 Bayesian method of Survival data Analysis

Bayesian Survival analysis procedure was used to make inference about the parameters of a Survival model. The Gibbs sampler algorithm was implemented with 40,000 iterations in three different chains, 15,000 burn-in terms discarded, as to obtain 75,000 samples for full posterior distribution. In Bayesian approach of Parametric AFT model, seven out of twelve predictor variables were found statistically significant. The Gibbs sampler with more than one chain simultaneously provide Gelman Rubin statistic plot**,** autocorrelation and time series plots of each chain in different colors that help us to check convergence. If all the chains appear to be overlapping, we are confident that convergence has been attained.

4.5.1. MCMC Estimation Method

The researcher used non-informative normal prior distribution with mean zero and variance 1000 and Inverse gamma distribution for sigma with scale =0.01, shape=0.01 parameters (Ghassan H., 2013). In the iteration of these Bayesian inference using MCMC the researcher was used 40,000 Markov Chain samples by fixing the burn-in state at 15,000, thinning 1 and three chains. This implies the parameters of the covariates were estimated by 25,000 Markov chain sample values, simply using the Markov Chain samples after the burn-in state.

Bayesian Accelerated Failure Time Model Comparison

Table 4.10: Bayesian AFT model comparison

Model	Dbar	Dhat	PD	DIC
Weibull	1400.830	1380.550	20.283	1421.110
Log-normal	1430.900	1410.530	20.368	1451.270
Log-logistic	1462.480	1441.670	20.816	1483.300

Based on the above table 4.10 Weibull distribution has smallest Dbar, Dhat, PD and DIC values. The distribution with small Dbar, Dhat, PD and DIC values is the best distribution that fit the data well, Because of this Bayesian Weibull Accelerating failure time model is selected to be the preferable model to analyze the data in Bayesian approach.

4.5.2. Posterior summary for Bayesian Accelerated Failure Time Model

Table: 4.11. Posterior summary for Bayesian Weibull AFT model parameter Estimates.

node	Variables	mean	Sd	MCerror	median	95% CrI	start	sample
beta[1]	constant	3.008	0.2859	0.00302	3.078	(1.494, 4.554)*	15001	75000
Sex	*Female*		*Ref.*	-------	-------	-------	-------	
beta[2]	male	-0.125	0.1278	0.00113	-0.1248	(-0.267, -0.0345)*	15001	75000
Residence	*Rural*		*Ref.*	-------	-------	-------	-------	
beta[8]	urban	0.165	0.1397	9.08E-4	0.1647	(0.1025, 0.1712)*	15001	75000
Season D	*Autumn*		*Ref.*	-------	-------	-------	-------	
beta[9]	Winter	-0.148	0.2373	0.00295	-0.1488	(-0.3183, 0.6090)	15001	75000
beta[10]	Spring	-0.023	0.1142	0.0036	-0.0245	(-0.319,-0.0797)*	15001	75000
beta[11]	Summer	-0.140	0.2282	0.0097	-0.138	(-0.411, -0.1042)*	15001	75000
Comorbidity	*No*		*Ref.*	-------		-------	-------	
beta[12]	yes	-0.119	0.1438	0.00190	-0.1181	(-1.1, -0.5357)*	15001	75000
SAM	*No*		*Ref.*	-------	-------	-------	-------	
beta[13]	yes	-0.286	0.1362	0.00523	-0.2879	(-0.2482, -0.105)*	15001	75000
prefer status	*No*		*Ref.*	-------	-------	-------	-------	
beta[17]	yes	0.372	0.1454	0.00126	0.371	(0.0916,0.4623)*	15001	75000
beta[20]	PNR	0.106	0.1157	0.0052	0.1056	(0.0865, 0.1552)*	15001	75000
Sigma ($\boldsymbol{\sigma}$)		3.002	0.1427	0.00391	3.0	(2.721, 3.288)*	15001	75000

Sd=standard deviation, MC error=Mont Carlo error, 95%CrI = 95% credible Intervals.

Table 4.11 above shows that the Posterior summary for Bayesian Weibull AFT model parameter Estimates include Monte Carlo error (MC-error), sample standard deviation (SD), median and the 95% credible intervals for all parameters. It can be seen that for all parameter estimates the Monte Carlo error (MC-error) is less than 5% of standard deviation. So researcher can use those parameter estimates for inferential purpose.

Interpretation of Bayesian Accelerated failure time model parameters

Based on the above table 4.11 the final model were interpreted using acceleration factor, 95% credible interval of Bayesian accelerated failure time estimated values.

Under the Bayesian Weibull AFT model, when the effect of other factors keep fixed, the estimated acceleration factor for male patient was estimated to be $e^{\beta} = e^{-0.125} = 0.882$ with [95% CrI: -0.2673, -0.0345]. The credible interval for the Bayesian acceleration failure time didn't include zero or on other hand researcher can say that the credible interval for the Bayesian acceleration factor did not include one by exponentiation of the Bayesian acceleration failure time credible interval that is [95% CrI: $e^{-0.2673}, e^{-0.0345}$: 0.765, 0.9661] . This indicates that male patients have less survival time than female patients or in the other way female patients survived 11.8% longer than male patients. The acceleration factors for patients whose residence was urban were estimated to be 1.179 with [95% CrI: 0.1025, 0.1712]. The credible interval did not include zero. This indicates that patients whose residence was urban had prolonged survival time than patients from rural residence at 5% level of significance.

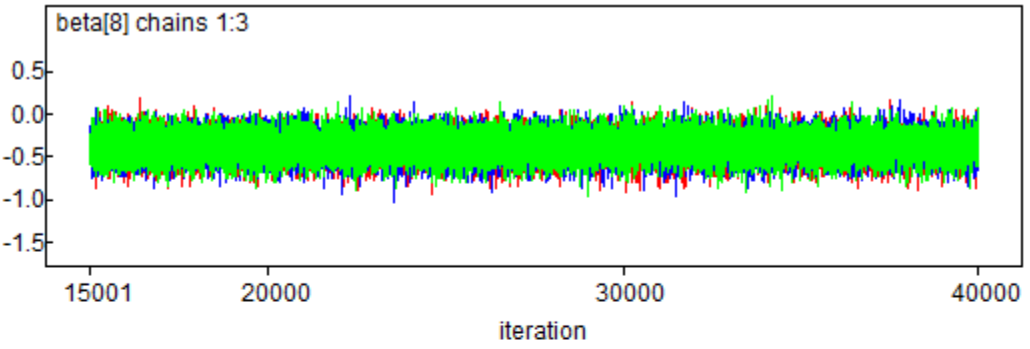
As shown in table 4.11 the estimated acceleration factor for patients diagnosed at spring season and summer season were 0.977 and 0.869 with [95% CrI: -0.3194, -0.0797 and -0.4108,-0.1042]. The credible intervals for both estimated values did not include zero respectively. This implies that patients who were diagnosed at spring and summer season had less survival time than patients who was diagnosed at autumn season. But patients who were diagnosed at winter season survival time were not significantly different from patients who were diagnosed at autumn season at 5% level significance. The acceleration factor for patients who were suffered other extra disease or comorbidity was estimated to be 0.888 with [95% CrI: -1.1, -0.5357].The credible interval did not include zero. This implies that patients who were not suffered other extra disease had longer survival time than patients who were suffered other extra disease or co-morbidity.

The acceleration factor for patients who were suffered severe acute malnutrition (SAM) was estimated to be 0.751 with [95% CrI: -0.2482, -0.105]. The credible interval did not include zero. This indicates that patients who were not suffered severe acute malnutrition (SAM) had longer survival time than patients who were suffered severe acute malnutrition (SAM) or in the other ways patients who were not suffered severe acute malnutrition (SAM) survived 24.9% longer than that of patients who were suffered severe acute malnutrition (SAM) at 5% level of significance. The acceleration factor for patients who were referred from other health centers was estimated to be 1.451 with [95% CrI: 0.0916, 0.4623]. The credible interval did not include

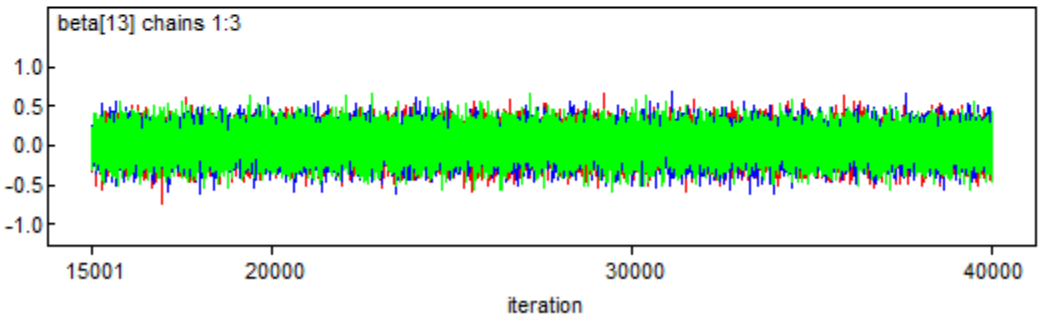
zero. This indicates that patients who were referred from other health center had prolonged survival time than patients who were not referred from other health center at 5% level of significance. Acceleration factor for patient nurse ratio was estimated to be 1.112 with [95% CrI: 0.0865, 0.1552]. The credible interval did not include zero. This indicates that patient nurse ratio had significant effect on the survival time of patients at 5% level of significance.

4.5.3. Assessment of Convergence

Time Series (History) Plots: are commonly used to assess convergence of the parameter estimates in Bayesian analysis. The WinBUGS package gives the plot with number of iterations on the x-axis and parameter values on the y-axis for each significant parameter. If the plot looks like a horizontal band, with no long upward or downward trends, then researcher have evidence that the chain has converged. For all simulated parameters, time series plot indicates a good convergence since three independent generated chains are mix together or overlapped, all time series plots are available at appendix B figure 4.14.



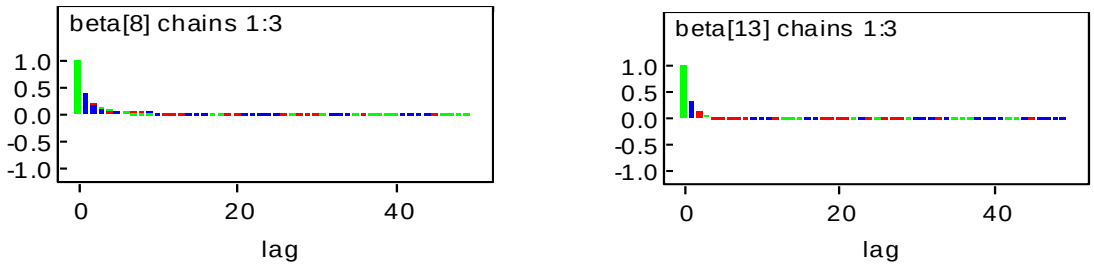
(a). Convergence checking for Residence



(b). Convergence checking for Severe Acute Malnutrition (SAM)

Figure 4.6: Time series plot to check convergence of covariates

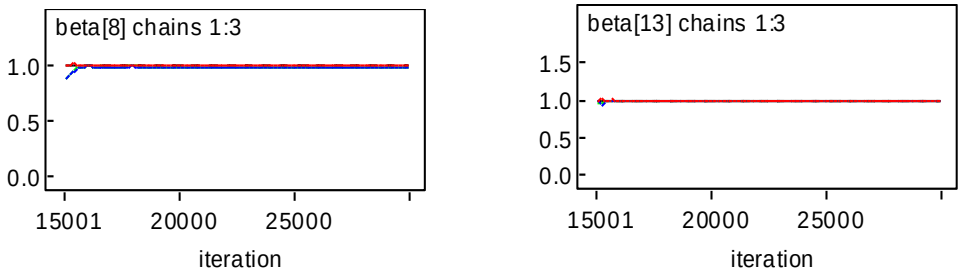
Autocorrelation Plot: It is a test used for convergence of Bayesian analysis. High autocorrelations in parameter chains often signify a model that is slow to converge. For all simulated parameters, the plot of the first 40 lags of three independently generated chains demonstrated good chain mixture indication of convergence; all Autocorrelation plots are available at appendix B figure 4.15.



(a). Convergence checking for Residence (b). Convergence checking for (SAM)

Figure 4.7: Autocorrelation plots to check convergence of covariates

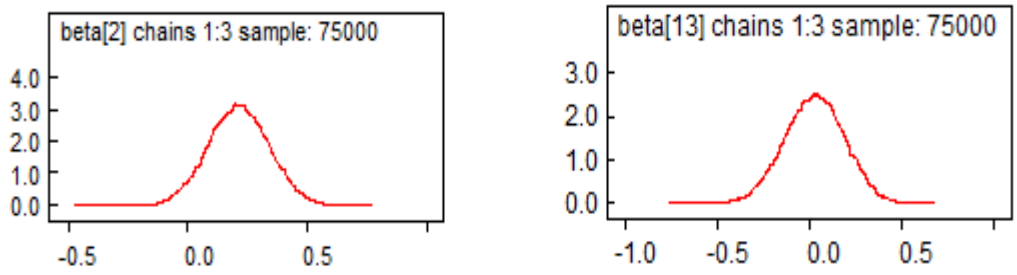
Gelman-Rubin Statistics: It is another way of assessing convergence for Bayesian analysis. It can also be applied only when multiple chains are used. For a given parameter, this statistic assesses the variability within parallel chains as compared to variability between parallel chains. The model is judged to have converged if the ratio of between to within variability is close to 1. The green line represents the between variability, the blue line represents the within variability, and the red line represents the ratio. Evidence for convergence comes from the red line being close to 1 on the y-axis and from the blue and green lines being stable (horizontal) across the width of the plot. Hence the Gelman-Rubin statistic of this study emphasis that one should be concerned convergence of ratio close to one, all Gelman-Rubin Statistics plots are available at appendix B figure 4.16.



(a). Convergence checking for Residence (b). Convergence checking for (SAM)

Figure 4.8: Gelman-Rubin statistic plots of covariates

Density Plot: This is also the statistical techniques to recognize convergence in Bayesian analysis. When coefficients of independent covariates were normal distributed. Then, it indicates that the Markov chain has attained its posterior distribution, all Kernel Density plots are available at appendix B figure 4.17



(a). Convergence checking for Sex (b). Convergence checking for SAM

Figure 4.9: *Kernel Density plots of covariates*

4.5.4. Assessing Accuracy of the Bayesian Survival Analysis

The posterior summary estimates by the MCMC algorithm (Gibbs sampler), like posterior mean, standard deviation, Monte Carlo error and credible interval were estimated using winBUGS software. To assess the accuracy of Bayesian survival analysis, researchers were used Monte Carlo error for each parameter. If the MC error value is less than 5% of its posterior standard deviation, then the posterior density is estimated with accuracy. In this study, MC error for each significant variable is less than 5% of its standard deviation. This indicates that convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics.

Both method give almost consistent results, but most of the parameters in Bayesian approach analysis have smaller standard error compared to the corresponding classical AFT model. Therefore, Bayesian AFT model gives better fit than the classical Weibull AFT model.

4.5.5. Model Comparison of Classical and Bayesian Approach of Survival analysis

In this section the researcher compared the model which is fitted in both Classical and Bayesian approach analysis to address good model. Based on the model comparison criteria in both approaches the Weibull AFT model is selected as good model. Therefore, the researcher compared the Classical Weibull AFT model and Bayesian Weibull AFT model using parameters' numerical value of standard error in both approaches. The model with the smaller standard error is the better model for fitting the data. The comparisons of standard errors of both approaches have been presented in the Table 4.12 below. The results in the Table 4.12 below shows that all estimated coefficients' standard errors in Bayesian Weibull AFT model are smaller than the Classical Weibull AFT model.

Table 4.12: Model Comparison of Classical and Bayesian Approaches of survival analysis

Covariates	Categories	CWAFT		BWAFT		SE Comparison
		β	SE_C	Mean(β)	SE_B	
Sex	female	Ref	-------	------	----	
	male	-0.130	0.060	-0.125	0.0011	$SE_B < SE_C$
Residence	rural	Ref	-------	-----	------	
	urban	0.147	0.070	0.165	9.08E-4	$SE_B < SE_C$
Season of Diagnosis	Autumn	Ref	-------	------	------	
	Spring	-0.169	0.088	-0.023	0.0036	
	Summer	-0.207	0.098	-0.140	0.0097	$SE_B < SE_C$
	Winter	-0.060	0.115	-0.148	0.00295	
Co-Morbidity	No	Ref	-------	--------	-------	
	Yes	-0.127	0.068	-0.119	0.00190	$SE_B < SE_C$
SAM	No	Ref	-------	------	------	
	Yes	-0.139	0.066	-0.286	0.00523	$SE_B < SE_C$
Patient refer status	No	Ref	-------	-------	-------	
	Yes	0.174	0.070	0.372	0.00126	$SE_B < SE_C$
PNR		0.091	0.037	0.106	0.0052	$SE_B < SE_C$

NOTE: CWAFT = Classical Weibull Accelerated Time Model, BWAFT= Bayesian Weibull Accelerated Time Model, SE_C=Standard Error of Classical Survival Analysis, SE_B=Standard Error of Bayesian Survival Analysis.

4.6 Discussion of the Results

The objective of this study was to identify the risk factors of mortality of under-five pneumonia patients in TGH using both classical and Bayesian survival analysis. For determining the risk factors of the mortality of under-five pneumonia patients; a total of 281 patients were included in the study out of which 16.7% were died and this study is agree with study conducted at Bushulo Major Health Center that is 18.79% by (Zinabu T. et al., 2014), with study conducted at Hawassa city that is 16.4% by (Tariku T., 2017) and with the Global, regional and national causes of child mortality report that is 14.1% by (Li Liu, 2012).

Descriptive statistics revealed that male patients were more exposed to pneumonia than female patients; this result is in line with other study conducted in JUSH by (Firaol B., 2017) and in Pakistan by (Christa L, 2013) and also children in the age group 1-11 months were more exposed than other age groups and this study agrees with study conducted in Sidama Zone Wondo Genet District by (Teshome A., 2017) similarly patients in the rural residence were more exposed than urban residence agrees with study conducted in JUSH by (Firaol B., 2017) and in China by (Feng X., 2012) and the prevalence of pneumonia in season of spring and summer were higher than other seasons and the results are in line with studies in Hawassa city by (Tariku T., 2017), in Malawi by (Ellubey R., 2004) and in Southern Israel by (Lieberman D and Porath A, 2005).

This study focused on classical and Bayesian approaches of survival analysis next to descriptive statistics. In classical parts it included nonparametric, semi-parametric and parametric survival analysis. Nonparametric methods used to compare the difference between each categorical covariate based on Kaplan-Meier estimation method and Log-rank test. The semi-parametric method of analysis using Cox model is applied starting from univariable analysis and the covariates which are significant in univariable analysis with p-value =0.05 were included in the multivariable analysis. The assumptions were checked for each covariate and for overall model fitted and assumptions were violated for Cox PH model. Then the researcher introduced an alternative model for Cox PH model which is parametric AFT survival model to fit the pneumonia data in TGH based by (Kalbfleisch J.and Prentice R., 2002).

The researcher used different types of the baseline distributions to fit AFT models for pneumonia dataset in TGH. The baseline distributions used in this study were Weibull, Lognormal and Log-logistic. The Weibull AFT model was selected as good AFT model than lognormal and log-logistic models in classical survival analysis based on comparison criteria with smaller AIC value (Munda M., 2012) and the Bayesian parts of analysis were centered by classical survival analysis in these study, therefore; the Bayesian analysis were applied on parametric AFT models and the comparison made using DIC values. The Bayesian Weibull AFT model is also selected as the best model in the Bayesian survival analysis based on smaller DIC value (Spiegelhalter D., 2004). Based on the classical Weibull AFT model and Bayesian Weibull AFT model the study showed that the survival of under-five pneumonia patients was significantly and strongly associated with Sex of children, Residence of children, Season of Diagnosis patients were admitted to hospital, Comorbidity, Severe Acute Malnutrition(SAM), Patient refer status from other health center and Patient to Nurse ratio(PNR). The current study is consistent with other findings by (Firaol B., 2017; Tariku T., 2017).

The findings of this study was revealed that female patients and patients whose residence was urban had prolonged the timing death of pneumonia while male and patients whose residence was rural had shorten timing death of pneumonia, the study agrees with study conducted at Pakistan (Christa L, 2013) and study conducted at JUSH (Firaol B., 2017) and also with the report of Integrated community case management of childhood illness in Ethiopia (Miller N., 2014). In this study it was found that Patients who were admitted in summer and spring season have shorter survival time and had high risk of dying from CAP as compared with autumn and winter seasons; this result agrees with study conducted in Hawassa city by (Tariku T., 2017) and in southern Israel Hospital by (Lieberman D and Porath A, 2005) as reported that there is high incidence of CAP during spring and summer seasons.

The findings of this study also showed that the patients who were suffered comorbidity or any other disease had shorter survival time than patients without comorbidity and also patients suffered Severe acute malnutrition (SAM) had shorter survival time than that of patients without severe acute malnutrition (SAM) the studies that support this results were conducted in Pakistan by (Duke T, 2002) , in Malawi by (Ellubey R., 2004), in JUSH by (Firaol B., 2017), in southern Israel Hospital by (Lieberman D and Porath A , 2005) and the Child Health Epidemiology

Reference Group (CHERG) report by (Fischer W., 2013). Patients who admitted during patients to nurse ratio (PNR) was high had high risk of dying from pneumonia. Since patients in hospital are nurtured by nurses and this has a positive impact on the recovery from their illness. Fortunately, patients who admitted during ratio of patient to nurse is high has less chance to survive as it is compared to others patients the study agrees with study conducted at Hawassa city by (Tariku T., 2017) and in Europe by (Andrea D., 2017; Penyoyer D., 2010).

All variables that were statistical significant in classical AFT model were also become statistical significant in the Bayesian AFT model. One of the objectives of the findings was to compare the classical and Bayesian approach analysis based on the survival time of under-five pneumonia patients. The Bayesian survival analysis is started from MCMC simulation of 40,000 samples with burn-in state of 15,000 and using the 25,000 sample for posterior inference using Win BUGS software for iteration and the convergence of the parameters were checked. After 40,000 sample generated the data was converged and the 25,000 sample were used for posterior inference in Bayesian survival analysis. The MCMC iterations were generated by setting the initial values and burn-in state without any criteria, since there is no established method for determining an appropriate number of iterations and burn-in size. Rather, the researcher use a trial-and-error process in which the ultimate goal is to obtain stable parameter estimates that minimize simulation error. This statement confirms with study conducted in USA by (Ghassan H., 2013). The MCMC simulation helped to increase the accuracy of the results by narrowing the credible interval and minimizing the standard error, but did not changed the direction of the results this agrees with studies by (Hakim E., 2009; Geir Storvik, 2014).

Bayesian survival analysis of this study was showed that smaller standard error and narrow credible interval for all significant parameters than that of classical or frequentist survival analysis models. This study is consistent with studies conducted in United Arab Emirates on Overview of Frequentist and Bayesian Approach to Survival Analysis by (Cluj-Napoca and Romania., 2016). As observed from two approaches Bayesian Weibull AFT model had narrow credible interval and smaller standard error (MCSE) than Classical Weibull AFT model. This implies that Bayesian survival analysis is good compared to Classical survival analysis; the current study is consistent with the studies conducted in Beirut Lebanon by (Pascale S., 2014) and (Wong W. et al , 2005)

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

This study employed the classical and Bayesian approach accelerated failure time model to determine risk factors associated with under-five pneumonia patients in Tercha General Hospital. Increasing the survival time of the children is the major goal of every country. To do that it is better to identify the risk factors that are related with time to under-five death due to pneumonia. From different types of AFT models fitted using different baseline distributions, Weibull accelerated failure time model is selected as the good model than Lognormal and Log-logistic in both classical and Bayesian approach analysis to fit our dataset in these study. From both approaches employed in this study, the Bayesian approach analysis is selected as an appropriate method to fit under-five pneumonia data in Tercha General Hospital.

The results of both classical and Bayesian approach Weibull accelerated failure time model showed that sex of the children, residence of the children, season of diagnosis, comorbidity, severe acute malnutrition (SAM), patients refer status and patient nurse ratio (PNR) were found to be significant predictors for survival time of patients in Tercha General Hospital. Of which patients whose residence was urban and patients who admitted during the patient to nurse ratio was low is prolong timing death of pneumonia patients in Tercha General Hospital. Similarly being male patient, season of diagnosis was Spring and summer, patients with comorbidity and patients with severe acute malnutrition (SAM) were statistically significantly shorten timing of death of under-five pneumonia in Tercha General Hospital.

Finally, the model adequacy checking in classical approach was applied by using baseline distribution graphical method and the Cox-Snell residual, and the result revealed that Weibull distribution seems better to describe the data. In the Bayesian approach analysis the model diagnosis were checked using Time series or History plot, Autocorrelation plot, Gelman-Rubin statistic plot and kernel density plots, and the convergence was attained.

5.2. RECOMMENDATIONS

Based on the results obtained from the findings the following recommendations are made for Federal ministry of health, Dawro Zone Department of Health, TGH and researcher.

1. The risk of dying due to pneumonia is higher in rural individuals than the urban dwellers; therefore the Federal ministry of health should work on awareness by giving health promotions on appropriate and effective treatment earlier.
2. Federal ministry of health should prepare well designed pediatric registration charts for all hospitals and health centers by including all risk factors.
3. Dawro Zone Department of Health should improve public and professional awareness by early detection and prompt treatment using feasible, effective regimens and detailed patient characteristics in the pediatric registry data collaboration with hospitals and health centers.
4. Tercha Genera Hospital need to improve Health facilities based on quality variables of hospitals like bed occupancy rate, patient to nurse ratio and patient to physician ratio in hospitals should be handled with effective management to minimize in-patient mortality.
5. The physicians, clinician and health extension workers should give attention to prevent the morbidity and mortality of pneumonia by giving health promotion to the community.
6. The researchers who are interested to investigate on the same area are recommended to introduce frailty modeling to account the correlation which comes from the cluster and to accounts unobservable random effect using classical and Bayesian analysis.

5.3. LIMITATION OF THE STUDY

As the data was gathered from the pediatric registration charts and cards of patients in the study there were a lot of patients with insufficient information. Lack of published literature on the country related to the survival time of under-five pneumonia patients in both classical and Bayesian approach. Lack of important risk factors related to survival time of patients in the pediatric registration chart and cards like age of mother, mother education, father education, mother occupation, father occupation, number of under-five children, and number of family size, monthly family income and vaccination status of children.

REFERENCES

Amare D. Tessema, F. & Girma, B (2007).Determinants of under-five mortality in Gilgel Gibe Field Research Center, Southwest Ethiopia,. Ethiop.J.Health Dev., 21(2):117-124.

Andrea D. Grant, M. Carroll, D., Dalton, S., & Astin, F. (2017). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. European Journal of Cardiovascular Nursing, 17(1), 6-22.

Antonietta M. (2013).Efficient computational strategies for doubly intractable problems with applications to Bayesian social networks. Statistics and Computing, 25(1), 113-125.

Betuel S. & Anna R . (2009). Severe Pneumonia in Mozambican Young Children: Clinical and Radiological Characteristics and Factors. Journal of Tropical Pedi Vol.55, No. 6:379-387

Black E., Jones G., Steketee, R., (2003). Child Survival Study Group. How many child deaths can we prevent this year? The lancet, 362 (9377): 65-71.

Brown R., Morris, S. S., & Bryce, J. (2004). Where and Why Are 10 Million Children Dying Every Year? Lancet, 361, 2226-2234

Campbell H and Nair H. (2015). Child pneumonia at a time of epidemiological transition. The Lancet Global Health., 3(2), e65-e66.

Christa L, Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z. A. & Black, R. E. (2013) Global burden of childhood pneumonia and diarrhoea .The Lancet , 381: 1405–1416.

Christensen R. Johnson, W., Branscum, A., Hanson, T. E., (2011) Bayesian Ideas and Data Analysis, an Introduction for Scientists and Statisticians. CRC Press.

Cluj-Napoca and Romania. (2016) Overview of Frequentist and Bayesian Approach to Survival Analysis. Applied Medical Informatics, Vol.38, No. 1, 25-38.

Cochran G. (1977). Sampling Technique. John Wiley and Sons. New York.

Collett D. (2003). Modelling survival data in medical research ,third edition.

Cox D. (1972). Regression models and life tables (with discussions). Journal of the Royal Statistical Society., 34:187-220.

Crighton E., et al . (2007). A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992–2001). Soc Sci Med , 64:1636–1650.

CSA. (2012). Central Statistical Agency ,Addis Ababa, Ethiopia. Measure DHS, ICF Macro Calverton, Maryland, USA. Ethiopia Demographic and Health Survey.

Datwyler C. and Stucki T. (2011). Parametric Survival Model, Handout.

Dezfuli H., Kelly, D., Smith, C., Vedros, K., & Galyean, W. (2009) Bayesian inference for NASA probabilistic risk and reliability analysis.

Duke T, Poka H, Dale F, Michael A, Mgone J, Wal T.(2002). Chloramphenicol versus benzylpenicillin and gentamicin for the treatment of severe pneumonia in children in Papua New Guinea: a randomized trial. The Lancet, 359(9305), 474-480.

Ellubey R. (2004). Pneumonia Case Fatality Rate In Children Under-Five: Understanding Variations In District Hospitals In Malawi (Master's thesis).

Fatmi K. and Franklin W. (2002). A comparison of 'cough and cold' and pneumonia: risk factors for pneumonia in children under 5 years revisited. International Journal of Infectious Diseases, 6 (4):295-301.

Feng X.,Theodoratou E, Liu L, Chan KY, et al. (2012). Social, economic, political and health system and program determinants of child mortality reduction in China between 1990 and 2006: A systematic analysis. Journal of global health, 2(1).

Firaol B.,Sinaga, M., Quadri, J., Kumar, A., Shariff, A., & Malik, T. (2017) Factors associated with outcomes of severe pneumonia in children aged 2 months to 59 months at jimma university specialized hospital. Current Pediatric Research, 21(3), 447-454.

Fischer W., R. I. (2013). Global burden of childhood pneumonia and diarrhea. The Lancet, 381, 1405-1416.

Ganjali M. and Baghfalaki T. (2012). Bayesian Analysis of Unemployment Duration Data in the Presence of Right and Interval Censoring. Journal of Reliability and Statistical Studies, Vol. 5, Issue 1: 17-32.

Geir Storvik, H. L. (2014). Bayesian methods . Statistical Machine Learning.

Gelfand, A. E., & Mallick, B. K. (2005). Bayesian analysis of proportional hazards models built from monotone functions. Biometrics, 51(3), 843–852.

Gelman A. and Brooks A. (1998). General methods for monitoring convergence of iterative simulations. Journal of computational and graphical,7(4), 434-455

Ghassan H., Richard M. and David R.(2013) Markov Chain Monte Carlo: an introduction for Epidemiologists, International Journal of Epidemiology , 42:627–634 doi:10.1093/ije/dyt043

Hakim E., U. M. (2009). Squamous cell carcinoma and keratoa canthoma of the lower lip associated with “Goza” and “Shisha” smoking. Int J Dermatol, 108-110.

Hosmer and Lemeshow. (1999). Applied Survival Analysis. New York: John Wiley and Sons,

Ibrahim J. et al. (2001). Bayesian Survival Analysis. Spring Science & Business Media.

Jamison D.T, B. J. (2006). Disease control priorities in developing countries. World Bank.

Kalbfleisch J.and Prentice R. (2002). The statistical Analysis of Failure Time Data. New York: 2nd ed.Wiley.

Khan A., H. H. (2009). High incidence of childhood pneumonia at high altitudes in Pakistan: a longitudinal cohort study. pakistan: Bulletin of the World Health Organization.

Klein, D. G. (2005). Survival Analysis, third edition.

Klein, J. and Moeschberger, M. . (2003). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

Kristina L., M. K. (2017). Morbidity and mortality of children aged 2–59 months admitted in the Tanzania Lake Zone’s public hospitals: cross-sectional study. BMC Res Notes,10(1),502.

Lehmann D, H. P. (2005). Nutrition and morbidity: Acute lower respiratory tract infections, diarrhoea and malaria. Papua New Guinea Medical Journal, 48(1/2):87.

Leung, D. T. (2016) Prevention and Control of Childhood Pneumonia and Diarrhoea. . Paediatric clinics, 63(1), 67-79.

Li Liu, H. L. (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet, 379: 2151–2161.

Lieberman D and Porath A . (2005). Seasonal variation in community-acquired pneumonia. Eur Respir J , 9(12):2630–2634.

Liu, L. Oza, S., Hogan D., Chu Y. and Perin, J. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. The Lancet, 388(10063), 3027-3035.

M.L. Calle et al. (2006). Bayesian survival analysis modeling applied to sensory shelf life of foods. Food Quality and Preference 17, 307–312.

Mantel, N. a. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the national cancer institute, 22(4), 719-748.

Marrie T. and Wu L. (2005). Factors influencing in-hospital mortality in community-acquired pneumonia: a prospective study of patients not initially admitted to the ICU. doi:10.1378/chest.127.4.1260, Chest 127(4), 1260 – 1270.

Merke P., N. .. (2005). Measurement of the tt production cross section in pp collisions.

Miller, N., Amouzou, A., Tafesse, M. amd Hazel E. (2014) Integrated Community Case management of common childhood illness in Ethiopia: implementation strength and quality of care. The American journal of tropical medicine and hygiene, 91(2), 424-434.

Monir R.., Aemmi, S. & Emami Moghadam, Z. (2015). Factors Affecting the Rate of Pediatric Pneumonia in Developing Countries: a Review and Literature Study. International Journal of Pediatrics, 3(6.2), 1173-1181.

Monto A. (2004). Studies of the community and family: acute respiratory illness and infection. Epidemiological Review, 16, 351-373.

Muluneh S. et al . (2011). Factors Influencing the Intention Not to Use Contraceptives among Sexually Active Women in Ethiopia. Journals of statistical Society, vol 20.

Pascale S.,Waked, M., Khayat, G. & Dramaix, M. (2014). Bayesian and Frequentist Comparison for Epidemiologists: A Non Mathematical Application on Logistic Regressions. The Open Epidemiology Journal, 7, 17-26, Beirut Lebanon.

Penyoyer D. (2010). Nurse staffing and patient outcomes in critical care: a concise review. Crit Care Med, 38(7): 1521–1528.

Ramazani M., A. F. (2006). The effects of a designed care on the weight of infants suffering from pneumonia. . Payesh Journal of The Iranian Institute For Health Sciences Research, 5(1): 5-13.

RCN. (2010). Royal College of Nursing. Guidance on safe nurse staffing levels in the UK.

Rudan I, Boschi P., Biloglav Z. & Campbell, H (2008). Epidemiology and etiology of childhood pneumonia. Bulletin of the World Health Organization,86, 408-416.

Schoenfeld D. (1982). partial residuals for the proportional hazard regression model.Biometrika 69(1), 239-241

Soofi S., A. S. (2012). Effectiveness of community case management of severe pneumonia with oral amoxicillin in children aged 2-59 months in Matiari district, rural Pakistan: a cluster-randomised controlled trial. The Lancet, 379(9817), 729-737.

Spiegelhalter D., Abrams, K. R.,& Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation. Vol.13, John Wiley & Sons.

Suwanjutha S. (2005). Risk factors associated with morbidity and mortality of pneumonia in Thai children under 5 years . Southeast Asian J Trop Med Public Health, 25(1):60-66.

Tariku T. (2017). Modelling Under-Five Mortality among Hospitalized Pneumonia Patients in Hawassa City: A Cross-Classified Multilevel Analysis. Annals of Data Science, 1-22

Teshome A. (2017). Prevalence of pneumonia and factors associated among children 2-59 months old in Wondo Genet district, SNNPR, Ethiopia . Curr Pediatr Res , 21 (1): 19-25.

UNICEF (2014). Pneumonia and Diarrhea Progress Report. International Vaccine Access Center (IVAC).Johns Hopkins Bloomberg, School of Public Health, www.jhsph.edu/ivac.

UNICEF(2016). United Nations Children's Fund United Nations Children's Fund.

WHO. (2012). Improving health and saving lives by ensuring access to priority medicines.

WHO.(2016).Pneumonia.[online]Available at: www.who.int/mediacentre/factsheets/fs331/en/.

Williams, B. G.P. (2002). Estimates of world-wide distribution of child deaths from acute respiratory infections. pp. Lancet Infect. Dis. 2:25–32.

Wong M. et al . (2005). Bayesian analysis of clustered interval-censored data. Journal of Dental Research, 84(9), 817–821.

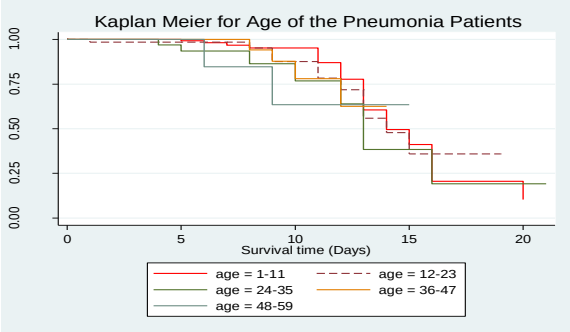
World health statistics. (2007). Available from: http://www.who.int/whosis/whostat2007.

Zafar F., W. F. (2002). A comparison of cough and cold and pneumonia: risk factors for pneumonia in children under 5 years revisited. Int J Infect Dis, 6:294-301.

Zinabu Teka et al. (2014): Analysis of Risk Factors for Mortality of In-Hospital Pneumonia Patients in Bushulo Major Health Center, Hawassa, Southern Ethiopia, Science Journal of Public Health; 2(5): 373-377

APPENDIX A: CLASSICAL SURVIVAL ANALYSIS

Kaplan Meier survival time plot for Pneumonia with different covariates



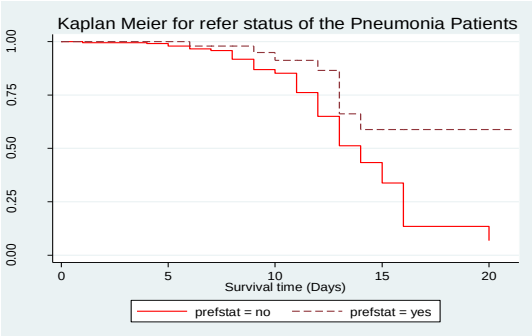
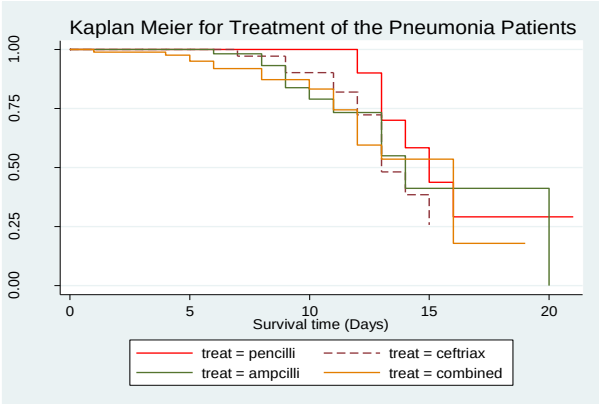


Figure 4.10: K-M plot for Age and patient refer status of pneumonia patients



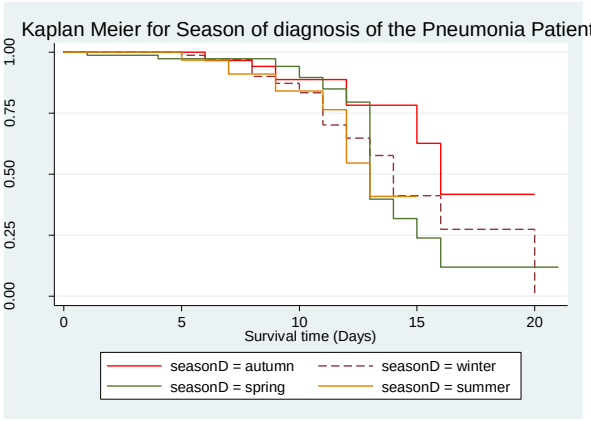
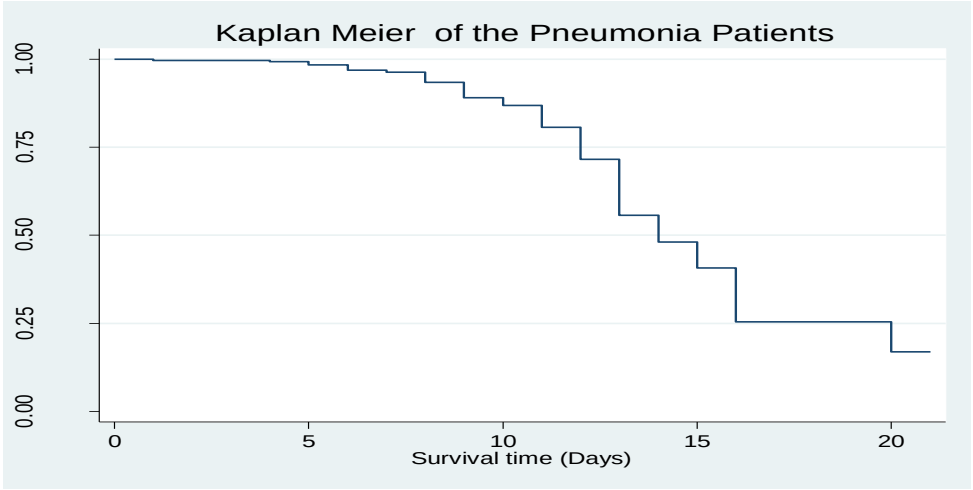
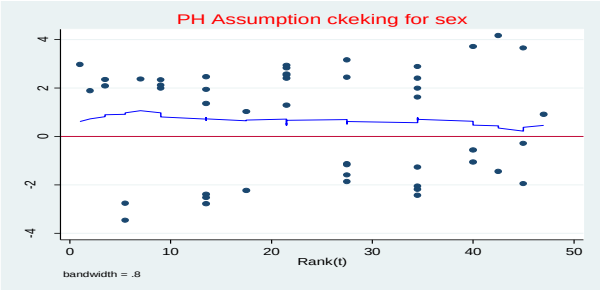


Figure 4.11: K-M plot for Treatment and season of Diagnosis of pneumonia patients



Test of proportional hazards assumption by Scaled Schoenfeld residuals



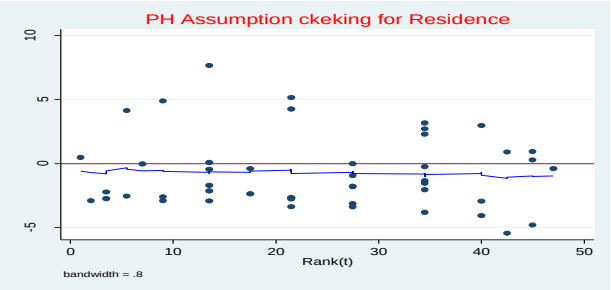
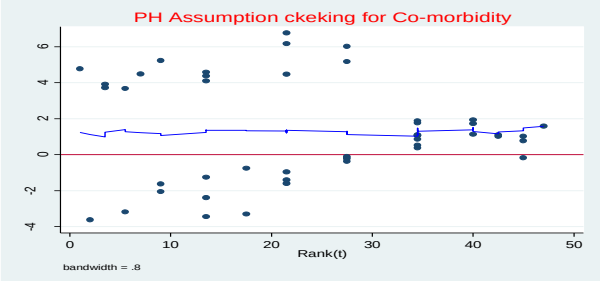
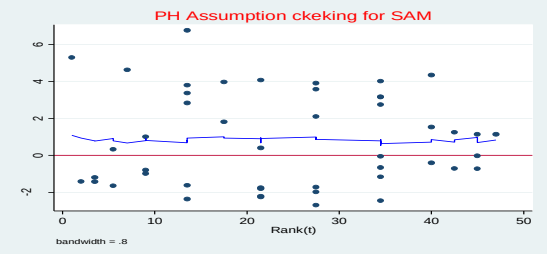


Figure 4.12: The plot of Scaled Schoenfeld residuals for Sex and Residence to check PH assumption





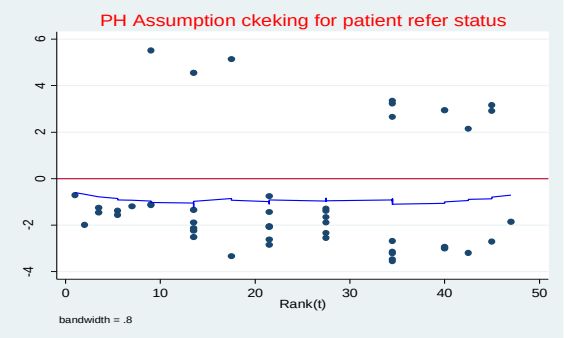
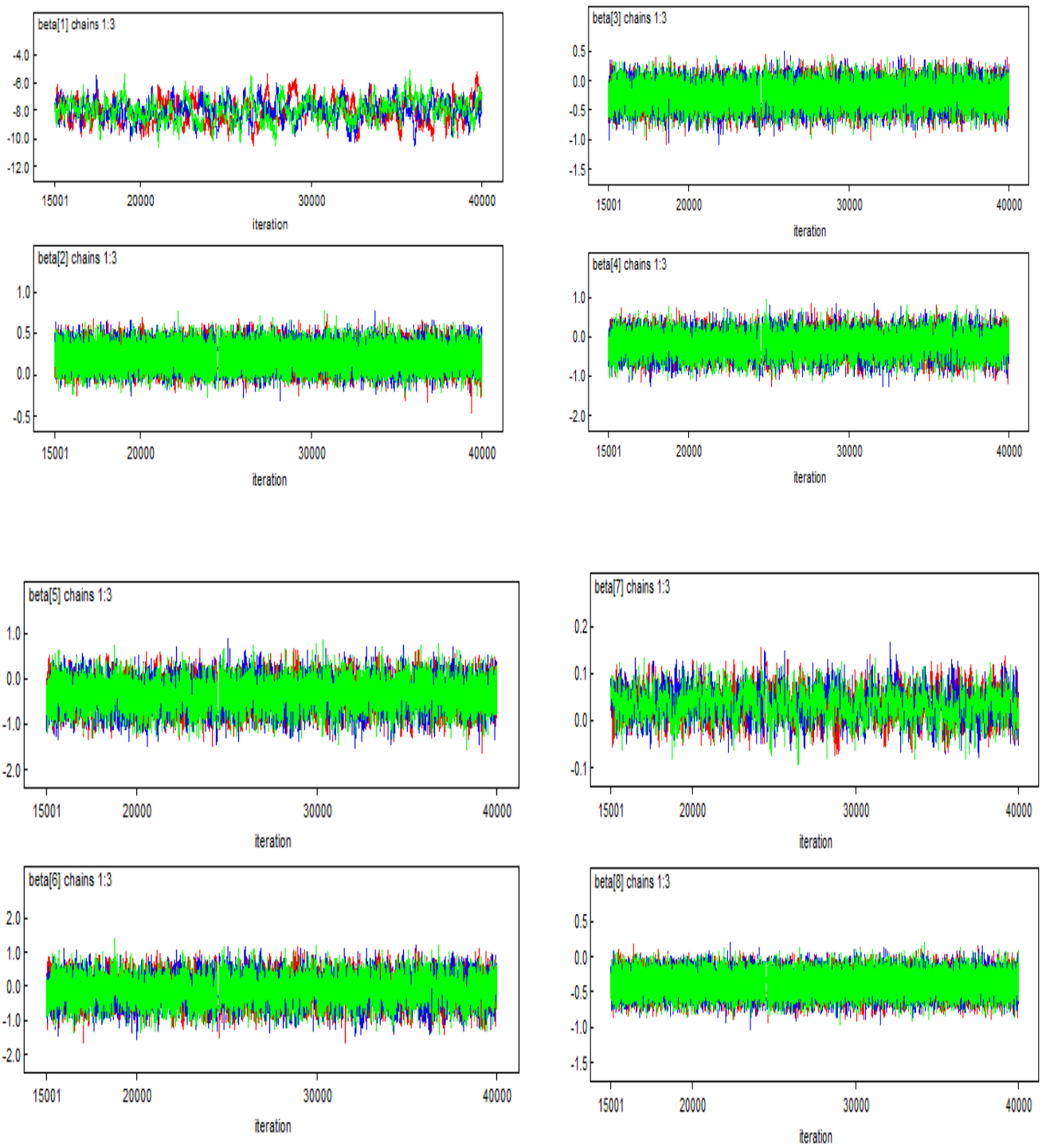
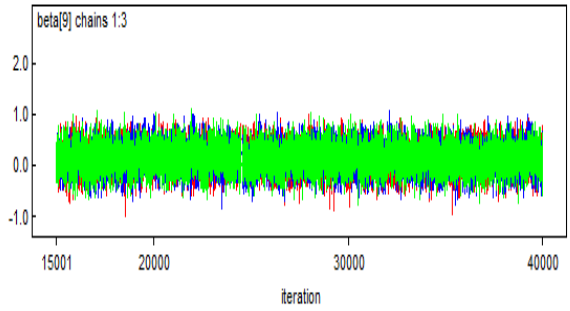


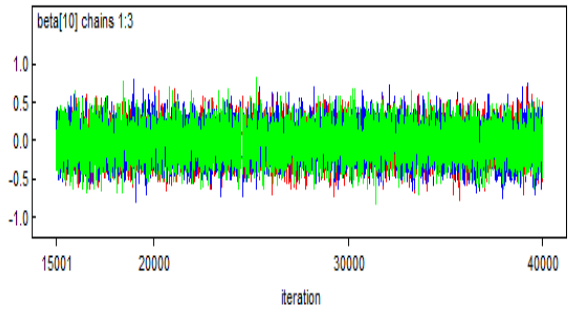
Figure 4.13: The plot of Scaled Schoenfeld residuals for Co-morbidity, SAM and patient refer status to check PH assumption

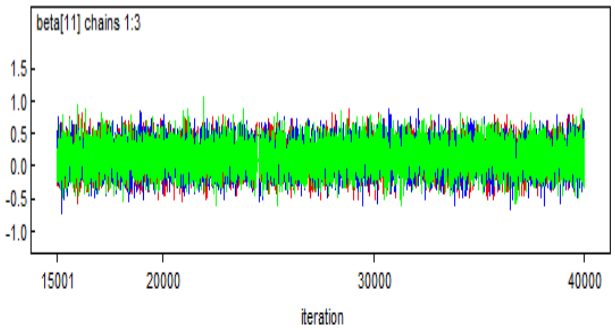
APPENDIX B: BAYESIAN CONVERGENCE CHECKING PLOTS

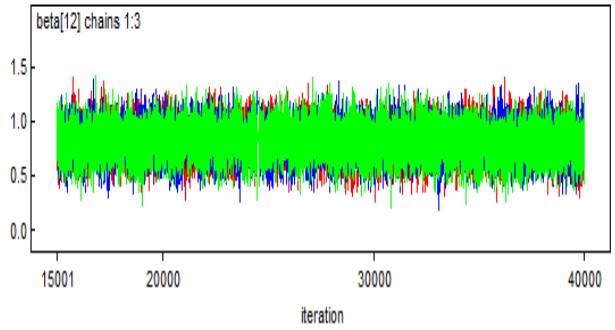
Figure 4.14: Time series (History) plots for all covariates

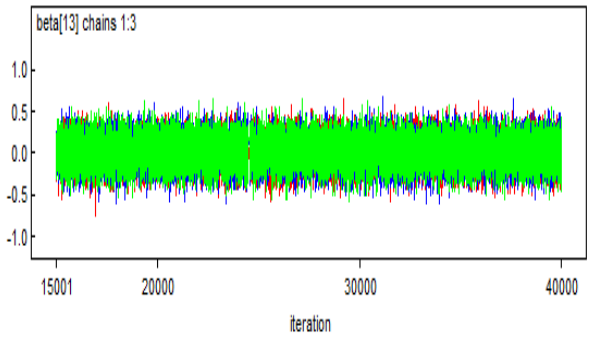


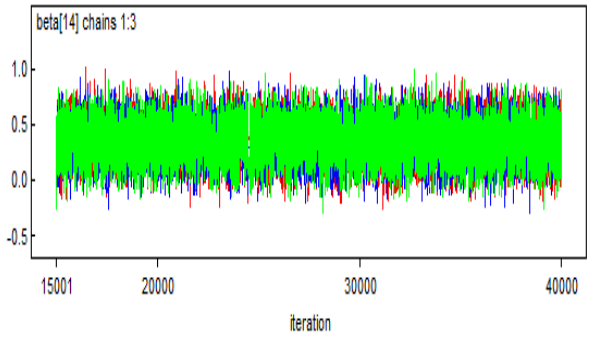

beta[9] chains 1:3
2.0
1.0
0.0
-1.0
15001
20000
30000
40000
iteration

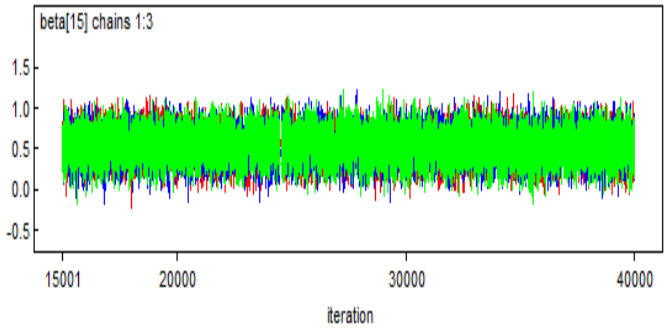

beta[10] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
15001
20000
30000
40000
iteration

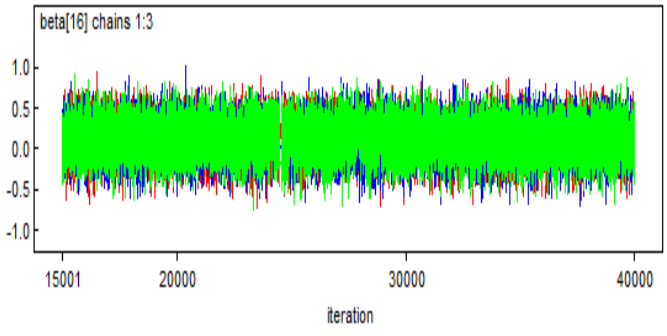

beta[11] chains 1:3
1.5
1.0
0.5
0.0
-0.5
-1.0
15001
20000
30000
40000
iteration


beta[12] chains 1:3
1.5
1.0
0.5
0.0
15001
20000
30000
40000
iteration


beta[13] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
15001
20000
30000
40000
iteration


beta[14] chains 1:3
1.0
0.5
0.0
-0.5
15001
20000
30000
40000
iteration


beta[15] chains 1:3
1.5
1.0
0.5
0.0
-0.5
15001
20000
30000
40000
iteration


beta[16] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
15001
20000
30000
40000
iteration

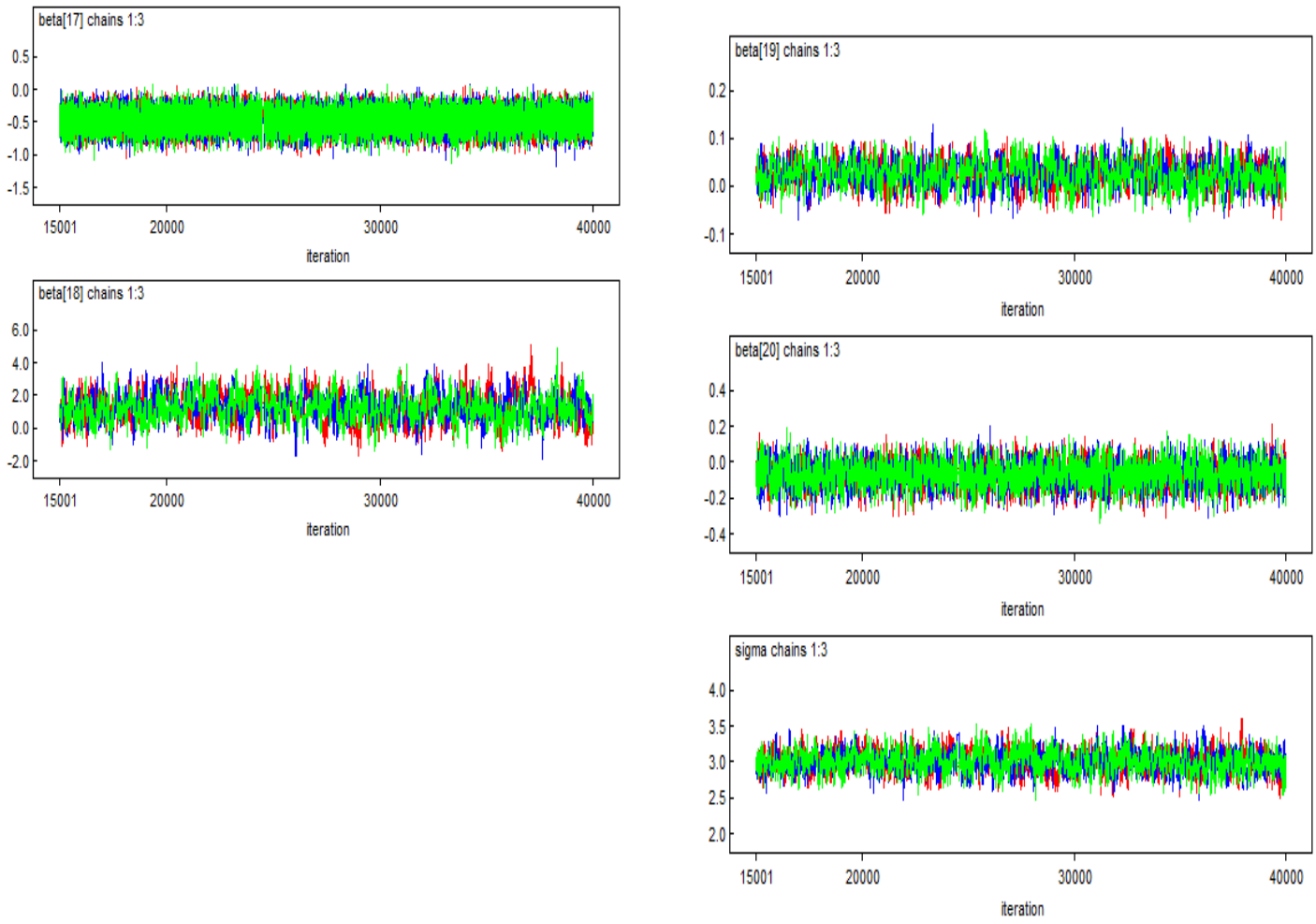
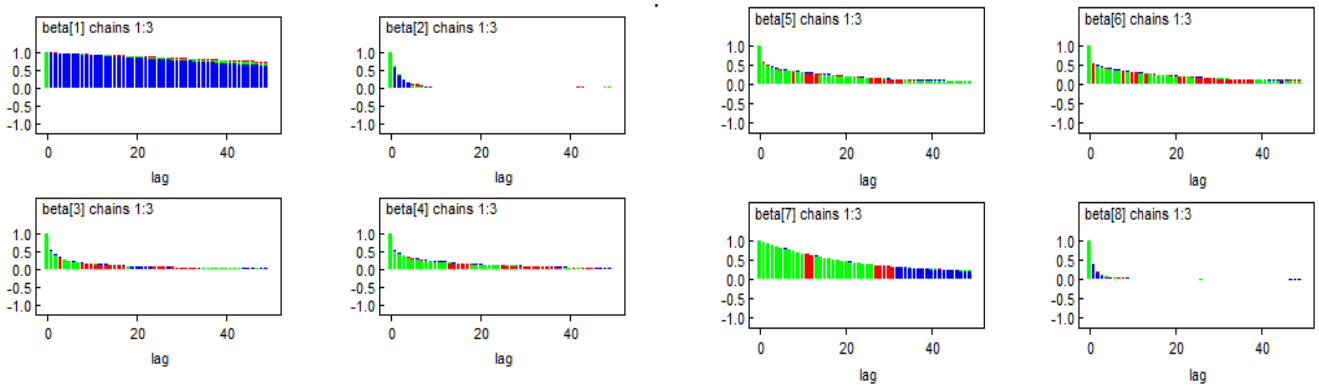
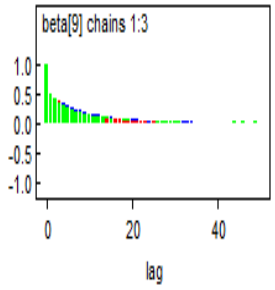
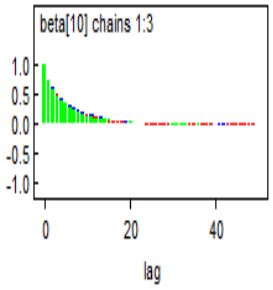
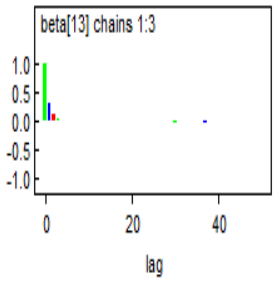


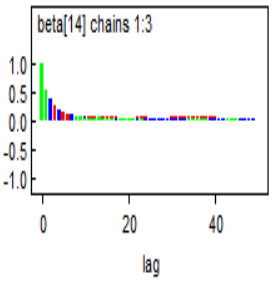
Figure 4.15: Autocorrelation plots for all covariates

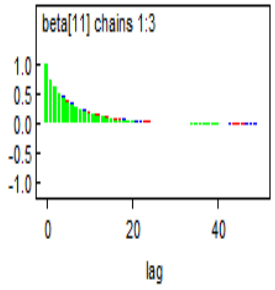


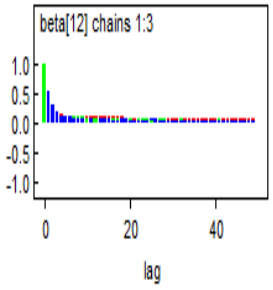

beta[9] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

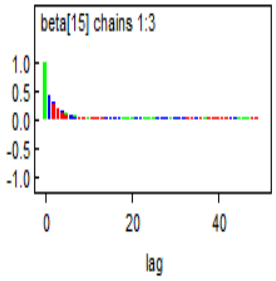

beta[10] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

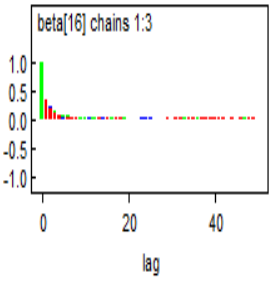

beta[13] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

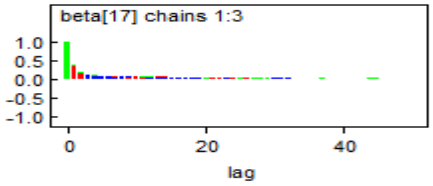

beta[14] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

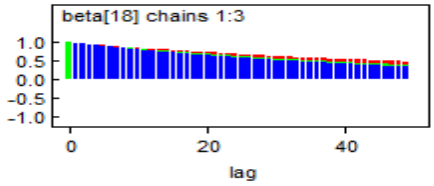

beta[11] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

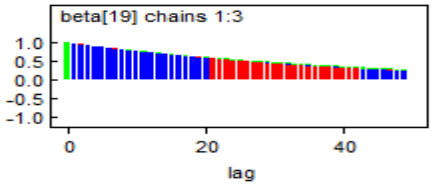

beta[12] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

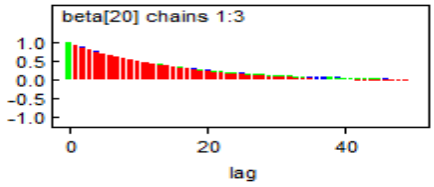

beta[15] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[16] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[17] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[18] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[19] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[20] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

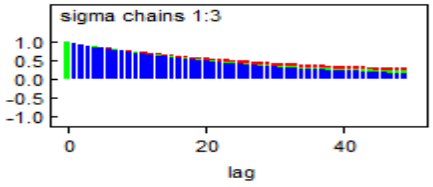

sigma chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

Figure 4.16: Gelman Rubin statistic plot for all covariates

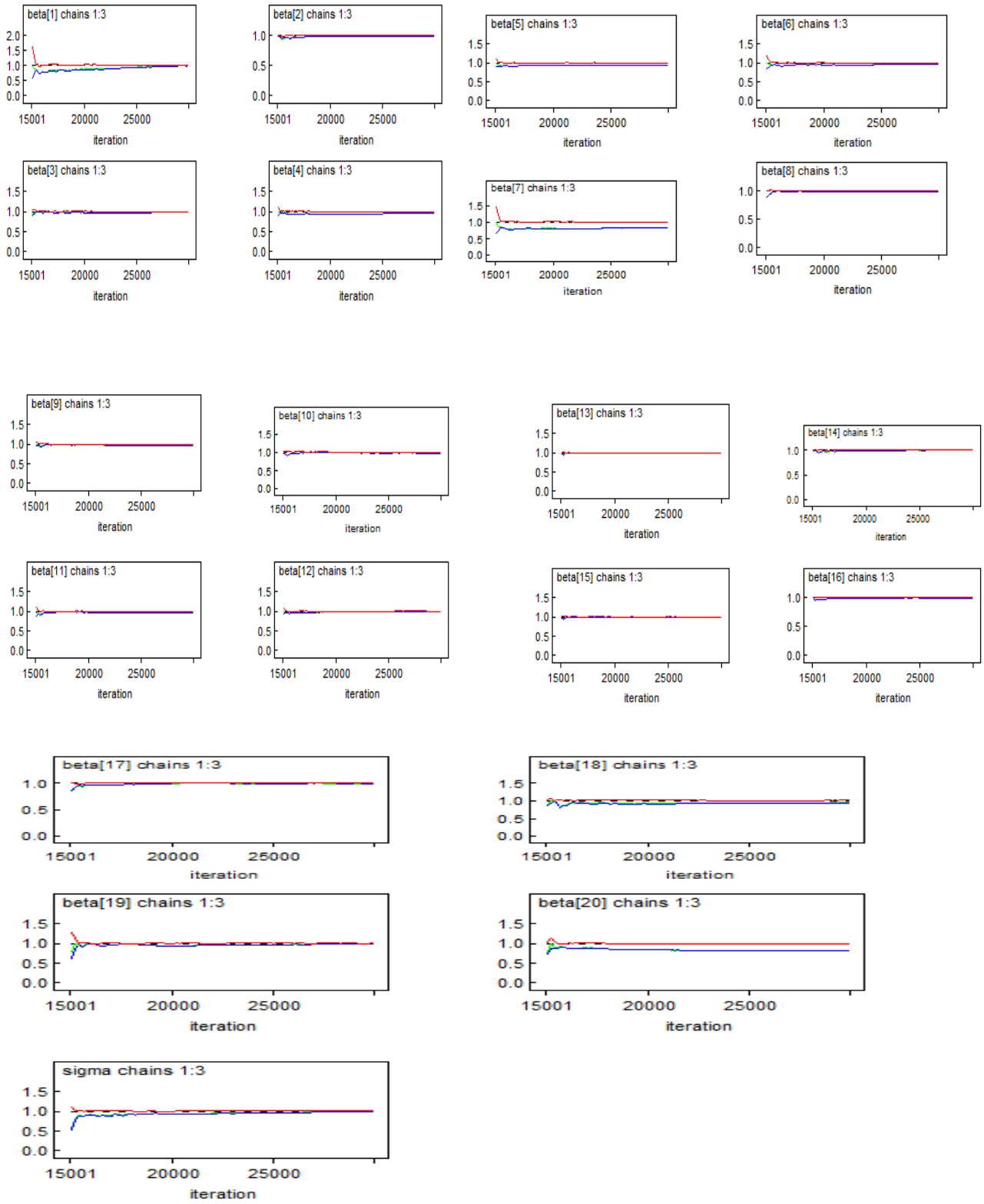


Figure 4.17: Kernel Density plot for all covariates

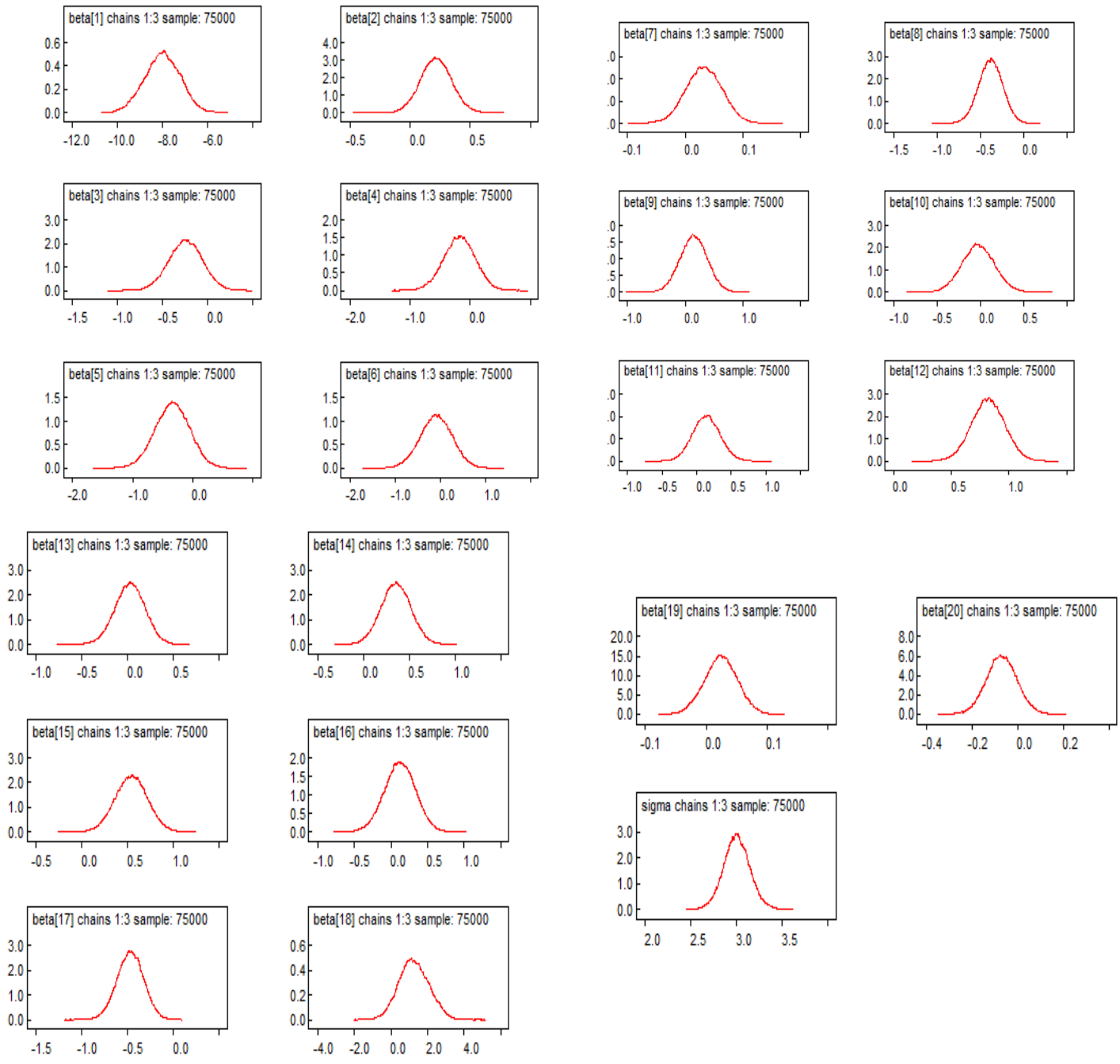


Table 4.13: General Posterior summary for Bayesian Weibull AFT model.

node	mean	Sd	MC error	5%*Sd	2.5%	median	97.5%	start	sample	Variable
beta[1]*	3.008	0.2859	0.00302	0.0143	1.494	3.078	4.554	15001	75000	Constant
Sex (female**)**	**Ref**									
beta[2]*	-0.125	0.1278	0.00113	0.0064	-0.2673	-0.1248	-0.0345	15001	75000	male
Age (1-11**)**	**Ref**									
beta[3]	-0.244	0.1858	0.00234	0.0093	-0.6097	-0.2436	0.1183	15001	75000	12-23
beta[4]	-0.179	0.2611	0.00376	0.0131	-0.7007	-0.1765	0.3216	15001	75000	24-35
beta[5]	-0.338	0.2883	0.00494	0.0144	-0.9082	-0.3357	0.2206	15001	75000	36-47
beta[6]	-0.090	0.357	0.00626	0.0179	-0.8022	-0.0882	0.6013	15001	75000	48-59
beta[7]	0.033	0.0313	8.09E-4	0.0016	-0.0289	0.0326	0.0934	15001	75000	weight
Residence (Rural**)**	**Ref**									
beta[8]*	0.165	0.1397	9.08E-4	0.0070	0.1025	0.1647	0.1712	15001	75000	urban
Season D (Autumn**)**	**Ref**									
beta[9]	-0.148	0.2373	0.00295	0.0119	-0.3183	-0.1488	0.6090	15001	75000	Winter
beta[10]*	-0.023	0.1142	0.0036	0.0057	-0.3194	-0.0245	-0.0797	15001	75000	Spring
beta[11]*	-0.140	0.2282	0.0097	0.0114	-0.4108	-0.138	-0.1042	15001	75000	Summer
Co-morbidity (No**)**	**Ref**									
beta[12]*	-0.119	0.1438	0.00190	0.0072	-1.1	-0.1181	-0.5357	15001	75000	yes
SAM (No**)**	**Ref**									
beta[13]*	-0.286	0.1362	0.00523	0.0068	-0.2482	-0.2879	-0.105	15001	75000	yes
Treatment (penicillin**)**	**Ref**									
beta[14]	0.355	0.1615	0.00191	0.0081	-0.0368	0.3545	0.6686	15001	75000	ceftrixon
beta[15]	0.540	0.1755	0.00182	0.0088	-0.1939	0.5411	0.8813	15001	75000	Ampicln
beta[16]	0.122	0.2124	0.00191	0.0106	-0.3011	0.1236	0.5321	15001	75000	combn
Patient refer status (No**)**	**Ref**									
beta[17]*	0.372	0.1454	0.00126	0.0073	0.0916	0.371	0.4623	15001	75000	yes
beta[18]	1.261	0.8203	0.02598	0.0410	-0.2971	1.236	2.895	15001	75000	BOR
beta[19]	0.024	0.0266	7.24E-4	0.0013	-0.0285	0.0244	0.0760	15001	75000	PPhR
beta[20]*	0.106	0.1157	0.0052	0.0058	0.0865	0.1056	0.1552	15001	75000	PNR
Sigma*	3.002	0.1427	0.00391	0.007	2.721	3.0	3.288	15001	75000	σ

*Sd=standard deviation, MC error= Mont Carlo error, *indicates significant covariates, Ref.=reference categories.*