MODELING TIME TO RECOVERY FROM TUBERCULOSIS IN SOUTH WEST ETHIOPIA: A CASE OF JIMMA UNIVERSITY SPECIAL HOSPITAL

By:

Alemu Bekele

Advisor: Dr.Yehenew Getachew

Co-advisor: Mr. Akalu Banbeta

A Thesis Submitted to the Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University in Partial Fulfillment for the Requirements of Masters of Science (MSc) Degree in Biostatistics

December, 2014

Jimma, Ethiopia

MODELING TIME TO RECOVERY FROM TUBERCULOSIS IN SOUTH WEST ETHIOPIA: A CASE OF JIMMA UNIVERSITY SPECIAL HOSPITAL

By:

Alemu Bekele

Advisor: Dr. Yehenew Getachew

Co-advisor: Mr. Akalu Banbeta

A Thesis Submitted to the Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University in Partial Fulfillment for the Requirements of Masters of Science (MSc) Degree in Biostatistics

October, 2014

Jimma, Ethiopia

DECLARATION

I declare that this thesis submitted in partial fulfillment of the Master of Science in Biostatistics degree requirement is my own work in accordance with the University of Jimma academic regulations. It has not been submitted for any degree or examination in any other University, and all the sources I have used or quoted have been indicated and acknowledged by complete references.

Alemu Bekele

Date: _____

Signature: _____

Jimma University, Jimma

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we hereby certify that we have read and evaluated the thesis prepared by Alemu Bekele under our guidance, which is entitled: Modeling time to recovery from Tuberculosis in South West Ethiopia: A case of Jimma University Specialize Hospital. We recommend that the thesis be submitted as it fulfills the requirements for the degree of Master of Science (MSc).

Dr. Yehenew Getachow ---------- ----------

Main advisor Signature Date

Mr. Akalu Banbeta ---------- ----------

Co-advisor Signature Date

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

---------- ---------- ----------

Name of Chairman Signature Date

---------- ---------- ----------

Name of Main Advisor Signature Date

---------- ---------- ----------

Name of Co-Advisor Signature Date

---------- ---------- ----------

Name of Internal Examiner Signature Date

---------- ---------- ----------

Name of External Examiner Signature Date

DEDICATION

This thesis is dedicated to my family members, particularly my beloved father who have tirelessly made provision for me both morally and materially throughout my life as a student and who was suddenly assassinated without seeing of my success.

Table of Contents **pages**

List of tables

List of figures

Acronyms

AFB	Acid-Fast Bacilli
AFT	Accelerated Failure Time
AIDS	Acquired Immune Deficiency Virus
CDC	Centers for Disease Control and Prevention
COXPH	Cox proportional hazard model
DOT	Directly Observed Treatment
DOTS	Directly Observed Therapy ShortCourse
EHRZS	Ethambutol (E) Isoniazid (H), Rifampicin (R), Pyrazinamide (Z), Streptomycin (S)
EPTB	Extra Pulmonary Tuberculosis
FMOH	Federal Ministry Of Health
NTCP	National Tuberculosis Control Programme
GHM	Gompertz Hazard Model
HIV	Human Immunodeficiency Virus
HC	Health Center
JUSH	Jimma University Specialized Hospital
LNTB	Lymph Node Tuberculosis
LTB	Latent Tuberculosis
LTBI	Latent Tuberculosis Infection
MDRTB	Multi Drug Resistant Tuberculosis
MTB	Mycobacterium Tuberculosis
NTB	National Tuberculosis Program
PTB	Pulmonary Tuberculosis
TB	Tuberculosis
WHO	World Health Organization

Abstract

Background : Tuberculosis is a chronic infectious disease that has long been one of the major health problem. Survival analysis is a set of methods used for analysis of the data which exist until the occurrence of an event. Many researches on tuberculosis treatment have reported varying recovery times. This research gets the average recovery time from TB in South West Ethiopia by using parametric survival models.

Objective: The objective of this study is to identify the best predictors for the time to recovery from tuberculosis and to compare different parametric survival models on the time to recovery from TB.

Method: The study has been used the retrospective data collected from 384 tuberculosis patients' selected randomly from last three years in JUSH. Survival analysis was used as the population under study is changing, we only consider the individual risk to recovery for those who are still ill, but this means that many standard statistical approaches cannot be applied. Parametric survival models are statistically more powerful than non-parametric or semi-parametric models.

Result: Among the potential parametric models fitted, Gompertz model was the best model to study TB data. The covariates: age, body weight, dose level of drug, type of TB, residence and HIV status were statistically significant covariates that affect the time to recovery from TB. The average time to recovery from TB was 172 days while 53.65% (206) of patients' were recovered from TB.

Conclusion: The Gompertz model is the best model to study time to recovery from TB data. From the Gompertz model result we conclude that being old, rural residence, having Extra-pulmonary TB, having HIV, lower doses and body weight at baseline prolonged the recovery time. Covariates significant in the Gompertz model are also significant in the Weibull model.

Key words: TB, Gompertz, Hazard, Weibull, Deviance residuals

Acknowledgement

First of all, I would like to thanks God, the Almighty, for having made everything possible by giving me strength and courage to write this paper. My Lord and God, praise you for your great deeds, you accomplished as you have spoken.

Secondly, I would like to thank my supervisor Dr. Yehenew Getachew, for his valuable suggestions and comments. My deepest thank to my co-advisor Mr. Akalu Banbeta for his critical suggestions which enrich the entire thesis.

Of course, I am grateful to my parents and my friends for their patience and love throughout the years of my study and without them this work would never have come into existence.

Finally, thanks to Mr. Negera Akuma for his endless support at the academic as well as the professional level.

CHAPTER ONE

1. INTRODUCTION

1.1. Back Ground

Tuberculosis is contagious and airborne disease and it is one of the most dangerous and killer diseases. Tuberculosis is a bacterial disease caused mainly by Mycobacterium tuberculosis. The mode of transmission of Mycobacterium species from person to person is well established. Virtually new infections with MBT are acquired via airborne transmission. The sources of infections are persons with tuberculosis of the lung who are coughing and sneezing. Coughing and sneezing produces air droplets containing bacilli. Persons in the same household, or who are in frequent contact with an infectious patient have the greatest risk of being exposed to the bacilli (Murray and Lopez, 1996).

Detection of the bacterium by direct microscope from sputum has been the main way for the diagnosis of tuberculosis. However smear positive rates among HIV positive tuberculosis patients have markedly decreased and thus the most expensive and time taking culture technique was used for the diagnosis of tuberculosis. Important challenges for TB control were HIV co-infection and multidrug resistance. HIV co-infection is the strongest known risk factor for progression of latent TB infection to tuberculosis disease. Although HIV co-infection has been shown not to affect the failure rate of TB treatment, high mortality has been reported among HIV-infected TB patients in sub-Saharan Africa (FMOH, 2008).

Ethiopia ranks seventh among the world's 22 countries with a high tuberculosis burden. Based on the WHO estimates, the incidence of TB of all forms in Ethiopia was 341 per 100,000 population while smear positive TB was 152 per 100,000 population. The prevalence of all forms estimated to be 546 per 100,000 populations and mortality of tuberculosis 73 per 100,000 during that year (WHO, 2012).

Since 1995, 56 million people were successfully treated for TB in countries that had adopted the WHO TB strategy, saving 22 million lives (WHO, 2012). In developed countries where TB drug supplies are not a problem, it is assumed that once the diagnosis of TB is made and the treatment instituted, the outcome will be a cure. However, even in developed countries with low incidence

of TB, treatment failure can lead to rising TB rates and consequently multidrug resistance was observed in New York City in 1988 (Daniel, 2011).

The main objectives of TB treatments are: to recovery the patient's from TB by rapidly eliminating most of the bacilli; to prevent death by its late effects; to prevent relapse by eliminating the dormant bacilli; to prevent development of drug resistance by using a combination of drugs; and to reduce transmission to others.

The first line anti-TB drugs are a multidrug which contains usually a mixture of four antibiotics. These are: Isoniazid, Rifampicin, Pyrazinamide & Ethambutol (WHO, 2009; Bong Ngeasham Collins, 2012). Although a TB patient may feel better, if he/she don't finish treatment the TB bacteria are still in their body. This become seriously ill and develop drug-resistant TB. Treatment of multi drug resistant TB (MDR TB) is more difficult than the treatment of drug susceptible TB, and it requires the use of "second line" or reserve drugs that are more costly, cause more side effects, and the drugs must be taken for up to two years. Cure rates for MDR TB are lower, typically ranging from around 50% to 70%. Finishing treatment is the only way to cure tuberculosis completely (WHO, 2011). The TB patient will need to have TB treatment for at least six months, to make sure all the TB bacteria are killed.

The primary objective of the TB treatment is to recovery the patients while the time to cure is the second issue to be considered. People generally want immediate relief and may switch off a medicine mid-course if they perceive it isn't working or if the medicine makes them sick. With any medication, it is possible to have side effects. There is a heavy stigma associated with having TB and the consequent isolation. In rural Ethiopia, neighbors isolate patients with TB. People do not want to tell others if they are diagnosed with having TB, even if they are successfully treated and recover (WHO, 2011).

Many researches on tuberculosis treatment have been reported varying recovery times. It was reported that if TB is detected early, it can be treated in six months whiles it takes between six to ten years to treat for late detection (Gavrilenko, 2001). It is reported that the addition of the antibiotic moxifloxacin to the usual TB drugs reduces recovery time from six to four months (Dioggban and Michael, 2012).

Tuberculosis is a curable disease so we have to model time to recovery from TB rather than time to death. To model the time to recovery from TB in South West Ethiopia, 384 TB patients' data were collected from JUSH. The time of treatment was determined by calculating the difference between initial time and outcome of treatment time.

In Ethiopia, barriers to modern medical treatment for TB include lack of money, transportation, and clinic availability. The most common experience is that people with TB can't get medicine and they die. Many times the treatment of illness involves a combination of modern and traditional remedies. People reported various causes of TB:-God, Draft, Wind, Clash of cold and hot air, Strong Heat, Sunlight, Bad sanitation, Progression from other illnesses, like pneumonia or a common cold, *Mich* (*mich* is sickness due to the clash of cold and hot air), Contagion by sweat, human touch, spit in the air, fault, bad behavior, vices, curse, taboo behavior. Other factors were reported as contributing to causing TB, including: Malnourishment, Weak immune system and Stress (WHO, 2011). The risk factors in this study include: residence, age, and sex, HIV status, body weight. The factors of TB treatment include the TB type and dose level of drugs with demographic and socio-economic factors.

The statistical analysis of lifetime data (time-to-event) plays an important role in medicine because we are observing something that develops dynamically over time. There are two points related to this development of survival data. First, Survival times are usually a mixture of discrete and continuous data that lend themselves to a different type of analysis. The Kaplan-Meier estimator of the survival function is a major step in the development of suitable models for such kind of data. Second, most evaluations are made conditionally on what is known at the time of the analysis, and changes over time (Kaplan & Meier, 1958). Usually, as the population under study is changing, we only consider the individual risk to recovery for those who are still ill, but this means that many standard statistical approaches cannot be applied. While non-parametric methods work well for homogeneous samples, they do not determine whether or not certain variables are related to the survival times. This need leads to the application of regression methods for analyzing survival data. The standard multiple linear regression model is not well suited to survival data for several reasons. Firstly, survival times are rarely normally distributed. Secondly, censored data result in missing values for the dependent variable (survival time) (Klembaum, 1996).

Models based on the hazard function have dominated survival analysis since the construction of the proportional hazards model by Cox in 1972. The Cox proportional hazards (PH) model is now the most widely used for the analysis of survival data in the presence of covariates or prognostic factors. One of the reasons this model is so popular is because of the ease with which technical difficulties such as censoring and truncation are handled. This is due to the appealing interpretation of the hazard function as a risk that changes over time. However the Cox PH model may not be appropriate in many situations and other modifications such as stratified Cox model or Cox model with time-dependent variables can be used for the analysis of survival data. (Klembaum, 1996).

The accelerated failure time (AFT) model is another alternative method for the analysis of survival data. The second important regression model in survival analysis is the accelerated failure time model AFT (Lawless, 1982). We use parametric survival models when the assumptions of Cox-PH should not fulfilled. Parametric survival models are statistically more powerful than non-parametric or semi-parametric models (Elvan Akturk Hayat. et al., 2010). The purpose of this thesis is to compare the performance of the different AFT models. This will be studied by means of real dataset which is collected from tuberculosis (TB) patients in JUSH. A survival analysis was conducted by using Exponential, Weibull, Gompertz, Gamma, Loglogistic and Lognormal distribution models.

1.2. Statements of the problem

Tuberculosis is an important global health problem. About one third of the world population has latent TB which means, people have been infected by TB bacteria but are not yet ill with disease and cannot transmitted the disease. Without proper treatment up to two third of people ill with tuberculosis will be die (WHO, 2012). It ranks as the second leading cause of death from a single infectious agent, after the HIV/AIDS. Tuberculosis imposes a high burden of human suffering and loss. In the Worldwide, 8.6 million new cases and 1.3 million deaths reported in 2012. About 3 million people who developed TB in 2012 were missed by national systems from treatment. WHO estimates the toll of TB mounts annually to 3 million deaths till 2020 and the global burden of TB infection will reach more than one billion. Surprisingly more than 80% of the disease burden comes from the poor resourced countries where the rate of reemergence is faster due to poor TB control and spending extremely in adequate treatment (WHO, 2012). According to WHO report in 2013, Ethiopia was **7**th among the 22 highest TB-burdened countries and 15th among the 27 highest MDRTB burdened countries. This study is necessary to model the time to recovery from TB in South West Ethiopia.

Most researches have been done on the time to death from TB rather than time to recovery from tuberculosis. Participants reported that many Ethiopians fully aware which illnesses are curable and which are not, including TB. Recovery from TB is recognized in someone's weight and appetite gain. Of the thesis done on the time to recovery from TB they report different recovery times from TB (Dioggban and Michael, 2012). Accelerated failure time model performs better than the proportional hazard model in applications where the effects of treatment are to accelerate or delay the event of interest (Kay and Kinnersley, 2002). Hence this study addressed the following research questions:-

1. What is the average time to recovery from TB for the case of South West of Ethiopia?
2. Which type of TB prolonged time to recovery from it?
3. What are the covariates influencing the time to recovery from TB; and
4. Which model is best to predict the recovery time from TB?

1.3. Objectives

1.3.1. General objective

The general objective of this study is to identify the best predictor factors for the time to recovery from tuberculosis and to compare parametric survival models on the time to recovery from TB in South West Ethiopia.

1.3.2. Specific objectives

The specific objectives of this study were:

- To identify the TB Type that prolonged duration of recovery ;
- To investigate important covariates that significantly associated with time to recovery from TB;
- To compare different parametric survival models for time to recovery from TB ;

1.4. Significance of the study

The TB treatment has the advantage of recovery the patients and indirectly preventing the transmission of the diseases from the patient to the normal persons. The purpose of this study was to develop evidence based policies, strategies and standards for TB prevention, care, and control. That means this study results may be used in the monitoring and providing leadership on matters critical to TB. The study can be used for public health and policy makers to make decisions and plan for future treatment control. The study is important to give care to reduce complications, costs, TB related illness and premature death. Finally it may use as references material for researchers that will be study on the TB.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Definition and general overview of tuberculosis

According to the standard definitions of the National Tuberculosis and Leprosy Control Program guideline (NLCP) adopted from WHO (MOH, 2008) the following clinical case and treatment outcome definitions were used:

2.1.1. Pulmonary TB (PTB)

A patient with at least two sputum specimens which were positive for AFB by microscopy, or a patient with only one sputum specimen which was positive for AFB by microscopy, and chest radiographic abnormalities consistent with active pulmonary TB smear positive. A patient with symptoms suggestive of TB, with at least two sputum specimens which were negative for AFB by microscopy and with chest radiographic abnormalities consistent with active pulmonary TB or a patient with two sets of at least two sputum specimens taken at least two weeks apart, and which were negative for AFB by microscopy, and radiographic abnormalities consistent with pulmonary TB smear-negative and lack of clinical response to one week of broad spectrum antibiotic therapy(MOH, 2008; Teklu B.,1993).

2.1.2. Extra- pulmonary TB (EPTB)

This included tuberculosis of organs other than the lungs, such as lymph nodes, abdomen, genitourinary tract, skin, joints and bones, meninges, etc. Diagnosis of EPTB was based on fine needle aspiration cytology or biochemical analyses of cerebrospinal/pleural/ascitic fluid or histo pathological examination or strong clinical evidence consistent with active extra-pulmonary tuberculosis, followed by a decision of a clinician to treat with a full course of anti-tuberculosis chemotherapy. In all the cases of EPTB, sputum examinations and chest radiographs were used to investigate the involvement of lung parenchyma (MOH, 2008; Belay T. et al., 2009).

2.1.3. Tuberculosis treatment outcome

The TB treatment outcome is divided into seven categories according to NLCP and WHO guideline. These categories were: Recovery (finished treatment with negative bacteriology result at the end of treatment), Completed treatment (finished treatment, but without bacteriology result at the end of treatment), Failure (remaining smear positive at five months despite correct intake of medication), Defaulted treatment (patients who interrupted their treatment for two consecutive months or more after registration), Died (patients who died from any cause during the course of treatment), Transferred out (patients whose treatment results are unknown due to transfer to another health facility) and Successfully treated (a patient who was recovered or completed treatment). In this study we categorized the above TB treatment out comes in to event recovered and censor (non-cured or failed) (WHO, 2012).

2.1.4. Cause and transmission of TB

Tuberculosis is a chronic infectious disease mainly caused by mycobacterium tuberculosis (MTB). The main source of infection is untreated smear-positive pulmonary tuberculosis (PTB) patient discharging the bacilli. It mainly spreads by airborne route when the infectious patient expels droplets containing the bacilli. It is also transmitted by consumption of raw milk containing Mycobacterium (Teklu B., 1993).

TB affects many parts of the body, it mainly affects the lung. Its clinical presentation, therefore, depends on the site of infection, the organ affected and its severity. Patients with PTB present with pulmonary symptoms (like productive cough, haemoptysis, chest pain and shortness of breath), constitutional symptoms (like fever, poor appetite, Weight loss, night sweats and anorexia) and other symptoms depending on the site of the infection (MOH, 2008).

Understanding of the symptoms is important to inform the community about the symptoms to seek medical advice and to inform health workers in order to increase the index of suspicion to easily pick suspects and detect tuberculosis cases presenting to health institutions. Early detection of the cases and prompt treatment are crucial for TB control. TB diagnosis mainly depends on the clinical presentation of the disease and identification of the offending bacilli. Many TB diagnostic tests are available although no single diagnostic test for TB exists that can be performed rapidly, simply, inexpensively, and accurately as a stand-alone-test. Thus, the diagnosis of active TB is a clinical

exercise; and sputum microscopy remains the mainstay of diagnosis because of its availability, operational feasibility and ability to identify the highly infectious forms of TB, the smear-positive PTB cases (MOH, 2008).

2.1.5. Risk factors of Tuberculosis

Studies done in Hawasa city showed that age, weight, smear negative pulmonary TB, dose of anti-TB drugs, and HIV status were all factors associated with death of TB patients during the period of DOT (Debebe and Alemayehu, 2012). The common risk factors for the TB were given as follows:

Age: The risk of TB infection increases with age from infancy to early adult life, probably, because of increasing number and frequency of contacts. TB is mainly a disease of adults in the age group of 15 - 49 years. In a population where the transmission has been stable or increasing, the incidence rate is higher in children mostly because of recent infection or re infection. As transmission falls, the case load shifts to older adults mainly because of reactivation of LTBI at later Ages (Rajagopalan, 2001).

Sex: Many Reports showed that men account for high proportion of notified TB cases than women. This was explained by sex (biological determinant progression from TB infection to disease is likely to be faster for women compared with men in their reproductive years) and gender (socio-cultural determinants influencing access to TB care leading to differential access to health care (like economic problem, inability to make decisions, poor health seeking behavior and stigma) that compromise the women's ability to utilize the available health service. In addition, higher risk of HIV infection among women makes them susceptible to develop active TB. Tuberculosis is the leading infectious cause of death in young women in developing countries. This could be worse in settings with health services insensitive to gender-specific needs. Studies that consider the interplay of biological, socio-cultural and health system determinants of sex and gender-based differences are needed to understand how and why women are affected (Baltussen. et al., 2005; Wikstrom, 2011).

Residence: More TB patients were reported from urban than rural areas because of overcrowding, poverty and HIV infection. In contrast, the presumed lower risk of TB infection in rural settings could be misleading and should be cautiously taken in high burden countries. In the rural settings,

access to the health service is limited; health seeking behavior is poor and the living condition favour disease transmission. As a result, understanding the burden of TB in rural areas will have a wider implication for TB control in such settings (Berhane. et al., 2008).

Tuberculosis and HIV co-infection: A complex interaction exists between TB and HIV infection. HIV increases the risk of infection, as it reactivates LTBI and increases the progression to active disease. TB-HIV co-infection has fatal consequences as TB becomes the leading cause of death in HIV infected individuals and patients with acquired immunodeficiency syndrome (AIDS). HIV lowers the host's immune response to MTB. The lifetime risk of developing active TB in HIV infected individuals is 10% per year compared with lifetime risk of 5 - 10% in individuals without HIV. As a result, the TB case notification rate (CNR) has increased four to six fold in sub-Saharan Africa (Shane, 2005; Amare Deribew. et al., 2009).

Socio-economic conditions: TB has been associated with factors linked to socioeconomic deprivation: poverty, overcrowding and malnutrition. The magnitude of TB is high among the poor, displaced, homeless, drug addicts, elderly, malnourished and women (Reichman and Hershfield's, 2006).

2.2. Global tuberculosis control

History of TB control started from attempts of treating unidentified cause to treating cases infected with the bacilli, from no remedy to effective treatment, from compulsory isolation to chemical isolation (treating infectious cases with anti-TB drugs), and from vertical to integrated approach where the service delivery was progressively decentralized to peripheral health institutions in the communities (Raviglione. et al., 2001).

Robert Koch's identification of the bacilli and the proposal to isolate patients was followed by compulsory isolation of the patients as the main principle of TB control. This included social support and contact examination in TB clinics (that were accessible and open at convenient time for the patients).

Global efforts to control TB were strengthened in 1991, when a World Health Assembly resolution recognized TB as a major global public health problem. Two targets for TB control such as 70% of case detection rate and 85% of recovery rate were established as the part of this resolution.

Eventually these two targets were embedded within the DOTS strategy launched by WHO in 1994, and subsequently endorsed by the WHO STOP TB Strategy in 2006 (WHO, 2006).

2.3. Tuberculosis control in Ethiopia

The NTLCP was started in 1992 and Ethiopia adopted the WHO recommended DOTS strategy in 1995. According to WHO report in 2012, 21.6 per 1000 mortality from TB in Ethiopia was reported. Tuberculosis is among the leading causes of morbidity and mortality in Ethiopia (FMOH, 2007). According to the national routine report and the global TB report 2012, more than 155,000 new TB cases of all forms are annually notified. Drug resistant TB has also become a major challenge of the TB control program in Ethiopia. According to the first national drug resistance survey completed in 2005, the rate of MDRTB is 1.6% and 11.8% among new and previously treated cases, respectively.

The study done on 4224 patients in the Gambella region showed that, 2579 (61.1 %) of the patients completed their treatment (Getahun. et al., 2010). The Cox-PH analysis of time to delay from the treatment of TB in Jimma zone showed that residence is the main factor of delaying time and the median of delaying is 10 weeks. Among the total of 565, 245 (43.4%) smear positive pulmonary, 91(16.1%) were smear negative pulmonary while the left are extra-pulmonary (Ayalew & Meseret, 2009). The majority of papers done on the TB are considered the time to death, rather than time to recovery. This thesis is considered time to recovery from TB and identify the covariates associated to TB.

A study done on 422 patients in Addis Ababa city showed that generally over all knowledge and attitude of TB patients about TB and its treatment were low in Addis Ababa city the creation of awareness on TB was recommended (Senait, 2011).

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1. Data source and study area

The study had been carried out at Jimma University Specialized Hospital, Jimma Zone, South West Ethiopia. Jimma town is the main city of Jimma zone and located 350 kilometers away from Addis Ababa in the South West direction. The Zone is found in Oromiya Region with a total population of 2,788,390 according to 2007 national census report. It covers an area of 199316.18 Kilometre squares and an average altitude of about 2180m above sea level. The zone has warm and humid climate with a mean annual maximum temperature of 33°C and a mean annual minimum temperature of 10°C. It lies in the climatic zone locally known as 'Woyna Daga' which is considered ideal for agriculture as well as human settlement.

Jimma town has two governmental hospitals and three health centres's which provides health care services for Jimma. Jimma University Specialized Hospital (JUSH) is one of the oldest public hospitals in the country. It was established in 1930 E.C during the Italian occupation for the service of their soldiers. It became the only teaching and referral hospital in the southwestern part of the country. It provides services for approximately 9,000 inpatient and 80,000 outpatient attendances. It has a bed capacity of 450 and a total of more than 750 staffs of both supportive and professionals. The data was collected from Jimma University Specialized Hospital (JUSH) only because it is referral, oldest and had more information than the left health centers.

The National health policy of Ethiopia emphasizes the development and provision of equitable and acceptable health service to the people. Jimma University Specialized Hospital is the hospital that give DOTS program in South West Ethiopia.

3.2. Study design

A retrospective data was collected from the profile and treatment outcome of all tuberculosis patients' registered from September 1, 2004 to June 30, 2006 at DOTS Clinic. The registration documents reviewed contain basic information such as patient's age, sex, address, tuberculosis type, and treatment outcome.

3.3. Sample size determination

Among all TB patients who placed under TB follow up in between September, 2004 to June 2006 in JUSH, 1769 patients those fulfilled the inclusion criteria had been taken as study population. Therefore, among the total of 1769 TB patients fulfilled the inclusion criteria during the study period, 384 patients' information selected using random sampling technique.

3.4. Study variables

For any type of statistical modeling we have two types of variables which are known as response variable and the independent covariates.

Dependent variable: Time to recovery from TB was the dependent variable in the study. The patient was said to be recovered from TB if the patient finished treatment with negative bacteriology result at the end of treatment. Patients who died, who are lost to follow up or got treatment failure, or whose recovery time exceeded the eight month in the course of the treatment were considered as censored. The pulmonary TB treatment needs an average of six months while the treatment of extra pulmonary TB needs more than six months. The eight month is selected since the study include both TB types.

Independent variables: The independent variables included in this study were demographic and clinical factors. Two of these covariates are continuous while five of them are categorical covariates (Table3.1). The dose level is categorized according to WHO and CDC programs (WHO, 2009; CDC, 2014).

Table 3.1: The covariates used in the time to recovery from TB model

No.	Variable	Description	Values/ codes
1	Age	Age	Years
2	Sex	Sex	0=Female, 1=male
3	TBT	Type of TB	0=Extra pulmonary ,1= pulmonary
4	Resident	Residence	0=Urban , 1=Rural
5	BW	Body weight	Weight in kilograms
6	HIV	HIV status	0=Positive, 1=negative
7	Dose	Dose of drugs	0=II, 1=III, 2=IV.

3.5. Inclusion and exclusion criteria

Inclusion criteria: All TB patients whose age more than five years were included in the study.

Exclusion criteria: Exclude patients that had ever received any TB treatment before the study period.

3.6. Methods of data analysis

3.6.1. Censoring

The background of this report is the analysis of time-to-event. That is, data are related with the individual time elapse in certain situation or state. The main characteristic of these data is the issue of censoring which occurs when the periods of time for some individuals cannot be completely observed. The presence of censoring makes these data unsuitable to analyze with traditional regression method and hence, calls for appropriate or specific techniques and analyses, usually called Survival Analysis (Hosmer and Lemeshow, 1999). Survival analysis consists of a set of specialized statistical techniques used to study response time data. In analyzing such data, the main objects are to determine the length of time interval spent in a state and the transition probabilities from the current state to the entered state (Gharibvand. et al., 2009).

The interest of this statistical tool is mainly focused on two distinguishing features of time data. Firstly, duration (times) is non-negative values usually exhibiting highly skewed distribution and therefore assumption of normality may be violated. Secondly, the true duration is not always observed or known. According to Hosmer and Lemeshow a censored observation is one whose value is incomplete due to random factors for each subject (Hosmer and Lemeshow, 1999).

Right censored: An observation is said to be right censoring if it is recorded from its beginning until a well defined time before its end time. In other words, an observation is said to be right censored if it begins at time initial and terminate before the outcome of interest is observed (Elisa T. Lee & John Wenyu Wang, 2003).

3.6.2. Survival analysis

Survival analysis is the axiom used to describe the analysis of data in the form of times from a well defined time origin until the occurrence of some particular event or end points. The main

feature of survival data that renders standard methods inappropriate is that survival times are frequently censored (Hosmer & Lemeshow, 1999; Collett, 2003). The term survival data is now used for all kind of events. In medical studies, often the main emphasis is the timing of event, time to recovery from disease, time to death and time to finish the treatment.

Suppose there are *n* subjects followed over a certain time interval $[0,\tau)$. The i^{th} subject at times {ti, i=1,2,…,n} and a (possibly censored) survival time ti to a certain endpoint. Let T_i denotes the response for the i^{th} subject (time to event), C_i denote the censoring time for the i^{th} subject, δ_i denote the event indicator

$$\delta i = \begin{Bmatrix} 1\ if\ the\ event\ was\ observed\ (Ti \geq Ci) \\ 0\ if\ the\ event\ was\ cenored\ (Ti < Ci) \end{Bmatrix}$$

The observed response yi=min[Ti,Ci]. The covariates of interest are denoted by Xi.

Let T denote a nonnegative random variable, representing time taken for recovery to occur. Let f(t) and F(t) be the respective density and cumulative distribution functions of T. The distribution of survival times is characterized by the survival and the hazard functions.

Survival Function

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

$$S(t) = P(T \geq t), t \geq 0 \text{ ---------- } 3.1$$

Hazard Function

The hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t, i.e.

$$h(t) = \lim_{\nabla t \to 0} P \frac{(t \leq T < \nabla t / T \geq t)}{\nabla t}, t \geq 0$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt} \text{ ---------- } 3.2$$

The hazard function is the probability that an individual will experience an event. The survival function is most useful for comparing the survival progress of two or more groups. The hazard function gives a more useful description of the risk of failure at any time point.

3.6.3. Non parametric survival modeling

The Kaplan-Meier Estimator is a nonparametric estimator of the survival function which is not based on the actual observed event and censoring times, but rather on the order in which events occur will be applied to estimate the survival probability of the infected patients. This principle of nonparametric estimation of the survival function is to assign probability to and only to uncensored failure times. The log-rank tests are also applied to test the survival function by covariate values (groups).

Let $t_1 < t_2 < \ldots\ldots < t_k$ be the ordered recovery times across two groups. Suppose that dj failures occur at tj and that rj subjects are at risk just prior to t_j, j, = 1, 2,3,..., k Let d ij and rij be the corresponding numbers in group i(i= 1, 2). The log-rank test compares the observed number of recovery with the expected number of recovers for group i. Consider the null hypothesis: $S_{1(t)} = S_{2(t)}$ i.e. there is no difference between survival curves in two groups. Given rj and d j the random variable d_1j has the hyper geometric distribution.

$$\frac{\binom{\mathrm{dj}}{\mathrm{d1j}}\binom{\mathrm{rj-dj}}{\mathrm{r1j-d1j}}}{\binom{\mathrm{rj}}{\mathrm{r1j}}}$$

Under the null hypo thesis, the probability of recovery at t (j) does not depend on the group, the probability of recovery at t(j) is $\frac{\mathrm{dj}}{\mathrm{rj}}$.

$$\chi^2 \log rank = \left[\frac{\sum_{j=1}^{k} (d_{1j} - r_{1j} * dj/rj)}{\sum_{j=1}^{k} \frac{r_{2j} r_{1j} dj(rj - dj)}{\left| rj^2 (rj - 1) \right|}} \right]$$, this statistic approximate χ^2 with 1 degree of freedom

There are two approaches to estimate the survival hazard function: parametric approach and empirical approach. The parametric approach specifies the parametric model while non-parametric

or semi-parametric models popular in medical science. In this study we focused on parametric models since the proportional assumption of Cox-PH is not full filled.

3.7. Parametric survival analysis

The Cox model is the most widely used survival model in the health sciences, but it is not the only model available (Kleinbaum & Mitchel, 2012). In the parametric model, survival time is assumed follow a known distribution. The commonly used parametric survival time models are: Exponential, Weibull, Loglogistic, Lognormal, Gamma and Gompertz distributions (Xian, 2012).

3.7.1. Exponential distribution

The exponential model X ~ Exp (λ) is the simplest parametric model because it has only a scale parameter and assumes a constant risk over time. The probability to cure within a particular time interval depends only on the length but not on the location of this interval. Probability density function $f(x) = \lambda e^{-\lambda x}$ is characterized by parameter λ is positive integer. This model has: $S(t,\lambda) = \exp(-\lambda t)$, survival function, $f(t,\lambda) = \lambda \exp(-\lambda t)$, probability density function and $h(t,\lambda) = \lambda$ hazard function. In and of itself, not reasonable in many applications: hazards usually not constant overtime.

3.7.2. Weibull distribution

The Weibull model introduced by Waloddi Weibull in 1939.This is an important generalization of the exponential model with two positive parameters. The shape parameter in the model allows great flexibility of the model and different shapes of the hazard function. The convenience of the Weibull model for empirical work stems on the one hand from this flexibility and on the other from the simplicity of the hazard and survival function. Characterized by a scale parameter λ > 0 and a shape parameter $\gamma > 0$

$$S(t,\lambda,\gamma) = \exp\left(-\lambda t^{\gamma}\right)$$
$$f(t,\lambda,\gamma) = \gamma\lambda t^{\gamma-1}\exp\left(-\lambda t^{\gamma}\right)$$
$$h(t,\lambda,\gamma) = \gamma\lambda t^{\gamma-1}$$

Hazard decreases monotonically with time if $\gamma < 1$, hazard increases monotonically with time if $\gamma > 1$ and hazard is constant over time if γ value is 1. The Weibull distribution is inappropriate when the hazard rate is indicated to be unimodal or bathtub-shaped (Jiezhi, 2009).

3.7.3. Gompertz distribution

British actuary Benjamin Gompertz made simple but important observation that a law of geometrical progression pervades large portions of different tables of mortality for humans in 1825. A random variable follows a Gompertz distribution with parameters $\lambda > 0$ and $\gamma > 0$ (T ~ Gompertz (λ , γ)), if the following relations hold:

$$S(t,\lambda,\gamma) = \exp\left(-\lambda\gamma^{-1}\exp(\gamma t)^{-1}\right)$$
$$f(t,\lambda,\gamma) = \exp(\gamma t)\exp\left(-\lambda\gamma^{-1}\exp(\gamma t)-1\right)$$
$$h(t,\lambda,\gamma) = \lambda\exp(\gamma t)$$

Gompertz distribution can also be presented with $\gamma \in R$ for $\gamma < 0$ the hazard is decreasing and the cumulative hazard is not going to ∞ when $t \to \infty \to \lambda$ part of the population will never experience the event (Kleinbaum. et al., 2012). The Gompertz survival model is the best model applicable for recovery data (Lee and Wang, 2003; Ohishi, Okamura & Dohi, 2009; Willemse & Koppelaar, 2000)

3.7.4. Loglogistic distribution

An alternative model to the Weibull distribution is the Loglogistic distribution. The Loglogistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases (David Hosmer, et al., 2012). The distribution imposes the following functional forms on the density, survival and hazard: A random variable T has a Loglogistic distribution if log T has a logistic distribution and Characterized by two parameters λ and $\gamma > 0$

$$S(t) = \frac{1}{1+(t\lambda)^{\gamma}}$$
$$f(t) = \frac{\gamma t^{\gamma-1}\lambda^{\gamma}}{\left[1+(t\lambda)^{\gamma}\right]^2}$$
$$h(t) = \frac{\gamma t^{\gamma-1}}{1+\lambda t^{\gamma}}$$

The median event time is only a function of the parameter λ Median (T) = exp (1/λ). This model is suitable for half censored data (Gupta, et al., 1999).

3.7.5. Lognormal distribution

Resembles the Loglogistic distribution but is mathematically less tractable. A random variable T has a Lognormal distribution if log T has a normal distribution. Used for peaked hazard functions (Tai. et al., 2007). Characterized by two parameters μ and $\gamma > 0$

$$S_0(t) = 1 - \Phi\left(\frac{\log(tj) - \mu j}{\sqrt{\gamma}}\right)$$

$$f(t) = \frac{1}{t\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2\gamma}(\log(t) - \mu)^2\right)$$

$\mu j = X_j\beta$ is the parameterization and Φ is the standard normal distribution. The median event is only the function of parameter μ, median (T) = exp (μ).

3.7.6. Gamma distribution

Gamma distribution, for its compliance with the patient data, is a suitable distribution to use in survival distribution models (Ponnuraja & Venkatesan, 2010). A lifetime T has a gamma distribution if for $\lambda > 0$ and $\beta > 0$,

$$f(t, \lambda, \beta) = \frac{\lambda^\beta t^{\beta-1}}{\Gamma(\beta)} \exp(-\lambda t) \ , \ t > 0$$

$$S(t, \lambda, \beta) = 1 - \frac{1}{\Gamma(\beta)} \int_0^{\lambda t} u^{\beta-1} \exp(-u) du \ , \ t > 0$$

Where λ is a scale parameter while β is a shape parameter.

3.8. Model estimation

Before any distribution can be fit to the data, the parameter values need to be estimated. The method of choice for parameter estimation is maximum likelihood estimation (MLE). For survival data, the likelihood function is:-

$$L(\theta)=\prod_{i=1}^{n}[f(ti,\theta)]^{\delta}[S(ti,\theta)]^{1-\delta} \quad \text{----------3.3}$$

Where θ is parameter of interest, f(ti) is the probability density function, S(ti) is survival function, and ti is failure time. The values of the parameters that maximize the likelihood function also maximize the log likelihood function because of f(ti)=h(ti)S(ti).

$$\begin{aligned} Log(L(\theta)) &= \log(\prod_{i=1}^{n}[f(ti,\theta)]^{\delta}[S(ti,\theta)]^{1-\delta}) \\ &= \log(\prod_{i=1}^{n}[h(ti,\theta)S(ti,\theta)]^{\delta i}[S(ti,\theta)]^{1-\delta i}) \\ &= \sum_{i=1}^{n}\delta i \log[h(ti,\theta)S(ti,\theta)]+\sum_{i=1}^{n}(1-\delta i)\log[S(ti,\theta)] \\ &= \sum_{i=1}^{n}\delta i \log[h(ti,\theta)]-H(ti,\theta)-\sum_{i=1}^{n}(1-\delta i)\log[H(ti,\theta)], \end{aligned}$$

Where log[S (ti,θ)] =-log [H (ti,θ)], which is negative of cumulative distribution.

A likelihood ratio test is a statistical test used to compare the fit of two models, one of which (the null models) is a special case of other (the alternative model. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value, or compared to a critical value to decide whether to reject the null model in favor of the alternative model. The use of likelihood ratio in statistical inference is common (Royall, 1997).

3.9. Model selection

3.9.1. Akaike's information criterion (AIC)

The exponential model, the Weibull model and lognormal model are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) can be used (Akaike, 1974) and it is defined as:-

$$AIC=-2(LL)+2(k+c) \quad \text{----------3.4}$$

Where LL is loglikelihood, k is the number of covariates in the model and c is number of shape parameters in model. The addition of 2(k+c) can be thought of as a penalty if non predictive parameters are added to the model. Although the best fitting model is the one with the largest log

likelihood, the preferred model is the one with the smallest AIC value. But there is a difficulty in using the AIC in that there are no formal statistical tests to compare different AIC values. When two models have very similar AIC values, the choice of model may be hard and external model checking or previous results may be required to judge the relative plausibility of the models rather than relying on AIC values alone (Pan, 2001).

3.10. Model checking and diagnosis

The use of diagnostic procedures for model checking is an essential part of the modeling process to check whether the fitted model is correct or not. There are different commonly used model checking to evaluate whether the appropriate functional form for a covariate is used in the model to assess the fitted model.

3.10.1. Cox-snell residuals

The residual that is most widely used in the analysis of survival data is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (Cox & Sell 1968; Cox & Oakes, 1984). The Cox-Snell residual for the i[th] individual, $i = 1,2,\dots,n$, is given by properties and features of residuals, when survival outcome are modeled, have been extensively studied in the literature. The Cox -Snell residuals are commonly used for a direct assessment of excess events (i.e., to reveal subjects that are poorly fit by the model), and for evaluating whether the appropriate functional form for a covariate is used in the model.

$S(t:X) = [S_o(t)]^{\exp(\beta X)}$Or, in terms of hazards: h (t; X) =$h_o(t)$ exp($\boldsymbol{\beta X}$). So, for each person with covariates x_i, S (t: x_i) =$[S_o(t)]^{\exp(\beta x_i)}$ then we can calculate

$\hat{h}_i = -\log[\hat{S}(T_i;\ x_i)]$ Or first predict survival probability at the actual survival time for individual, then log-transform it.

The residuals in right censored data constitute a censored sample of the unit exponential distribution

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

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

Then the modified Cox-Snell residual is given by

$$rci' = 1 - \delta i + rci \quad \text{----------} \quad 3.5$$

- Plotting –log($\hat{S}(T_i)$) vs t should yield a straight line
- Plotting log [–log ($\hat{S}(T_i)$)] vs log (t) should yield a straight line through the origin with slope=1.

If the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

3.10.2. Martingale residuals

The modified residuals $r_{Ci}{}'$ defined in equation (3.5) have mean of unity for uncensored observations. Accordingly, these residuals might be further refined by relocating the $r_{Ci}{}'$ so that they have a mean of zero when an observation is uncensored. If in addition the resulting values are multiplied by -1, we obtain the residuals

$$rmi = \delta i - rci \quad \text{----------} \quad 3.6$$

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

3.10.3.Deviance residuals

Although martingale residuals share many of the properties possessed by residuals encountered in other situations such as in linear regression analysis, they are not symmetrically distributed about zero, even when the fitted model is correct. This skewness makes plots based on the residuals difficult to interpret.

The deviance residuals, which were introduced by (Therneau et al., 2000), are much more symmetrically distributed about zero. They are defined by

$$rDi = sign(rmi)\left[-2\{rmi + \delta i \log(\delta i - rmi)\}\right]^{1/2} \quad \text{----------3.8}$$

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

3.11. Ethical consideration

The Ethical clearance was obtained from Jimma University, department of Statistics. The official ethical clearance was also obtained from Jimma University Specialized Hospital medical director. Careful recruitment and training for data collectors was undertaken. To maintain the confidentiality, the data collector and the supervisor will extract the necessary data from the patient baseline and follow up card. The data obtained had been coded carefully for the analysis.

CHAPTER FOUR

4. DATA ANALYSIS AND RESULTS

4.1. Descriptive statistics

The study was conducted retrospectively on 384 TB patients admitted into JUSH in the South West Ethiopia for the last three years. Of the 384 patients studied over the specified period, 252 representing 65.63% were males and 132 representing 34.37% of the patient population used were females. Among 384 TB patients registered for the DOT program 240(62.5%) of the patients were came from Urban while 144(37.5%) were came from Rural areas. The percentage of patients in dose levels II, III, and IV were 54.17%, 32.81%, and 13.02% respectively. This means 208(54.17%) of the patients used dose II, 126(32.81%) used dose III and 50(13.02%) of the patients had used dose IV of the anti-TB drug. About 289(75.26%) of the patients had Pulmonary TB, while 95(24.74%) were diagnosed with extra-Pulmonary TB. About 66(17.19%) of the patients had HIV/ADS virus, while 318(82.81%) were had no HIV/AIDS virus.

From this study about 54 percent of the patients treated recovered from TB while 46 percent treatment failures. 166(69.2%) of urban residents were recovered from TB during the study while 40(27.8%) of the patients were recovered from the rural patients. A recovery proportion seems lower for females 61(46.21%) than for males 145(57.54%). Proportion of recovery for HIV negative groups 191(60.06%) was greater than proportion of recovery of HIV positive patients 15(22.73%). Pulmonary TB patients' recovery was 190(65.31%) which was larger than Extra-pulmonary TB patient's recovery which was 16(16.84%). Proportion of recovery from TB by the dose levels of the drug were almost the same. Hence, the proportion patients recovered from TB whose used dose II, dose III and dose IV were 108(51.92%, 70(55.56%) and 28(56.0%) respectively.

Table 4.1: Frequencies and percentages for categorical covariates together with status of the events in time to recovery from TB

Demographic and health variables			Status of the events			
Variables	Categories	Total	Recovery	Censor	Censor (%)	Recovery (%)
Type of TB	Extra pulmonary	95	16	79	83.16	16.84
	Pulmonary	289	190	99	34.26	65.74
Residence	Urban	240	166	74	30.8	69.2
	Rural	144	40	104	72.2	27.8
Sex	Female	132	61	71	53.79	46.21
	Male	252	145	107	42.46	57.54
HIV	Positive	66	15	51	77.27	22.73
	Negative	318	191	127	39.94	60.06
Dose level	II	208	108	100	48.08	51.92
	III	126	70	56	44.44	55.56
	IV	50	28	22	44	56
Total		384	206	178	46.35	53.65

From table 4.2, the patients were followed up for a median period of 166.5 days. The minimum follow up time was 18 days and the maximum time was 232 days. The average Age and Weight of the patients' in the study were 27 years and 50kgs respectively. The minimum Age of patients was 5 years and the maximum Age of patient was 70 years while the minimum and maximum body Weights of the patients were 15kgs and 74kgs respectively.

Table 4.2: The descriptive statistics of continuous covariates used in time to recovery from TB

Variable	Total	Mean	Std. Dev.	Min	Max	Median	Q1	Q3
Age	384	27.09375	10.80405	5	70	22.50	20	30
Weight	384	50.13227	8.50031	15.5	74	50.00	45	55
Time	384	158.362	29.79804	18	232	166.5	142	173

4.2. Univariate survival analysis

In order to select variables in the model, first univariate analysis is used to check all the covariates associated with recovery time. To explore whether there are significant differences among different groups of categorical predictors we employ the log-rank statistical test. The null hypothesis being tested is that there is no difference between the survival curves.

The log-rank test Table 4.3, indicated that statistically there is a significant difference of survival experience among groups of TB type, HIV status and dose of drug used while sex is not statistically significantly different. Controlling for sex, the log rank test statistic (1.3091, Pr=0.2526) showed that there was no significant difference in recovery times for both sexes.

Table 4.3: The log-lank test among the categorical variables of TB data in JUSH

Variables	DF	Chi-square	P-value
Sex	1	1.3091	0.2526
HIV	1	22.2618	<0.0001
TBT	1	38.4050	<0.0001
Dose	2	29.6486	<0.0001

The pulmonary TB is more accelerated to cure than Extra-pulmonary TB. The plot of the survivor function for this is presented in Fig. 4.1. It could be observed that the two survivor curves were the same until the 50^{th} days when the curve for the pulmonary TB started a gradual fall. The curves are distinguishable and serve as a confirmation of the earlier result that there was statistically significant difference in recovery times between pulmonary TB and extra-pulmonary TB.

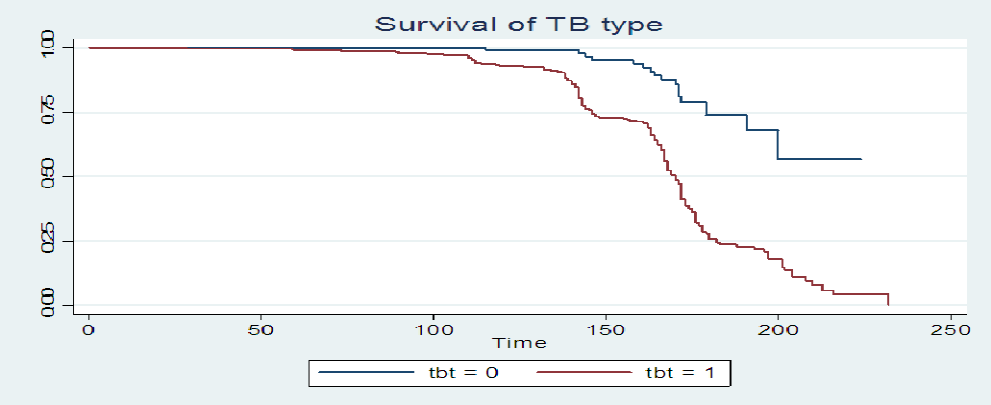

Figure 4.1: The survival plot on type of TB of TB data in JUSH

Similarly on the plots appendix B figure1, the residents of urban were accelerated to cure from TB than the residents from the rural and the patients those had no HIV were accelerated to recovery from TB than that of TB/HIV co infected patients.

From table 4.4, we have identified the covariates statistically significantly included in the time to recovery from TB. Accordingly TB category, HIV status, residence, age, dose of anti-TB drugs and body weight are included while gender is removed by using forward selection of variables.

Table 4.4: The forward selection of the variables in the study

Variables	Coef.	Std. Err.	z	[95% Conf.interval]	
TB type	1.363236	0.261993	5.2	(0.849738	1.876733)*
HIV	1.005833	0.286335	3.51	(0.444628	1.567039)*
Residence	-0.69201	0.178502	-3.88	(-1.04187	-0.34215)*
Age	-0.02613	0.00819	-3.19	(-0.04218	-0.01008)*
Dose	1.226565	0.126667	1.98	(1.001815	1.501737)*
Weight	0.022562	0.009423	2.39	(0.004093	0.04103)*

*= pvalue less than 0.05

4.3. Parametric survival model selection

Table 4.5 shows Akaike's Information Criterion (AIC) for the different considered models. According to this criterion, among the desired models, a model that has the lowest AIC is the best and the most efficient one. Therefore, in our study the Gompertz model was the best fitted model to the time to recovery from TB data among other parametric models. Residence of patients, types of TB, body weight of patients, age and dose level IV of the drugs were significant factors for the duration of recovery from TB under Weibull, and Gompertz model (Appendix A, Table 3). The model using Gompertz model fitted the data better than other models because it had smaller AIC value which is 50.99097.

Table 4.5: The parametric survival model comparisons of time to recovery from TB data in JUSH

Models	Exponential	Weibull	Gompertz	Loglogistic	Gamma	Log normal
AIC	607.9297	58.5392	50.9909	86.0368	58.25142	159.5004
BIC	639.5348	94.0950	86.5467	121.5927	97.75784	195.0562
-2LL	591.9300	40.5392	32.9910	68.0368	38.251	141.5004

Therefore we performed the time to recovery from TB using the Gompertz model which is our best model as shown below in Table 4.6. The slope coefficient gamma (γ) value for time to recovery from TB was 0.042985 which was greater than zero. The model estimated gamma was positive, significant, and this implies increasing risk (hazard) of recovery with time. The Gompertz model showed that Residence of the patients, TB type, and HIV status, body weight of patients, age, and dose level of drug used were statistically significant factors to time to recovery from TB in the data of JUSH.

The hazard ratio value for age of patients' was 0.974842, and it implies that as age increases the risk of recovery from TB decreases by 0.025158. The hazard ratio of the dose level of anti-TB drug of dose level III and dose level IV were 1.088487 of CI 0.802116 - 1.477099, and 1.66499 of CI 1.084629-2.555889 respectively while dose level II kept constant. The dose IV of the anti-TB drug

was statistically significant to the risk recovery from TB at 5% level of significance; that means dose level four was 1.66499 times more effective than dose two while dose III kept constant. This shows that there was a statistical significant difference between the three dose levels of the drug to recovery from TB.

The hazard ratio for TB type was 4.13976 and it was found between 2.475238 and 6.923635. The patients' those who had pulmonary TB were 4.13976 times more recovery from TB than those had Extra-pulmonary TB.

HIV status had statistically significant influence on the survival time to recovery from TB. The estimated hazard ratio was 3.277343 and found between 1.895266 and 5.667264. The patients' those had no HIV/ AIDS were 3.277343 times more recovery from TB than those had HIV/ AIDS virus.

If the weight increased, the risk of recovery from TB was also increases. For one kg increments of weight of the patient the risk of recovery from TB increased by 1.026 than the base weight. Suppose we take an increment of Weight by 5 kg to make comparisons. Then, the estimated hazard ratio for a five kg increase in initial Weight will be 1.141576=exp (5*0.026482). This means the risk of recovery was 14% higher for a patient whose Weight is 5kg higher than another patient controlling for other covariates in the model.

The hazard ratio for Residence was 0.4779 and it was found between 0.334493 and 0.682829. The patients' those who came from urban were 0.5221 times more recovered from TB than those came from rural. The Gompertz hazard model is given by:

$$h(t,x) = \lambda \exp(\gamma t)$$

$$h(t,x) = 4.1397\,TBT\,(pulm.) \times 3.2773\,HIV\,(neg.) \times 0.4779\,\mathrm{Re}\,si(rural) \times 1.0269\,BW \times 0.9749\,Age \times 1.0849\,Dose(III) \times 1.6961 Dose(IV) \times \exp(0.0429\,time)$$

Table 4.6: The Gompertz hazard model of time to recovery from TB data in JUSH

Variables	HR	SE($\widehat{\boldsymbol{\beta}}$)	[95% conf. interval]	
$Resi_{(rural)}$	0.477914	0.087005	(0.334493	0.682829)*
$Tbt_{(pulmonary)}$	4.10287	1.076362	(2.453468	6.86112)*
$Hiv_{(neg'v)}$	3.270064	0.923905	(1.879588	5.689181)*
Weight	1.027582	0.01011	(1.007958	1.047589)*
Age	0.974842	0.008155	(0.958988	0.990957)*
Dose(III)	1.084958	0.169065	(0.799418	1.472489)
Dose (IV)	1.69616	0.371734	(1.103864	2.606262)*
Gamma(γ)	0.042985	0.002282	(0.038513	0.047457)*

Wald test =119.99 p-value = 0.0001

* = pvalue less than 0.05, HR= is hazard ratio

Gamma (γ) was the coefficient of the slope,

Afterwards, we evaluated and compared the considered models (Exponential, Weibull, Lognormal, Gamma, Loglogistic and Gompertz) using the Cox-Snell residuals. For each model, we calculated the Cox-Snell residuals, estimated their survival functions using Kaplan-Meier method and, then, calculated the cumulative hazard functions for these estimations. Finally, according to Cox-Snell residuals, the hazard function graphs were drawn. Considering that the closer the graph to the bisector the better fitted model to the data, we saw that in our study the Gompertz model was the best fitted model to the studied TB data of all (Appendix B, Figure: 4).

4.4. Model diagnostics

Before fitting the covariates into the model, proportional hazard assumption were checked by plotting “Schoenfeld residuals, regressing Schoenfeld residuals against time to test for independence between time and residuals and by examining log (-log (time) plots see (Appendix

B, Figure. 3 & 5). Sex which violated the assumptions was not included in the model by forward selection of variables (Table: 4.4).

From fig.4.2 below data is not symmetrically distributed about zero.

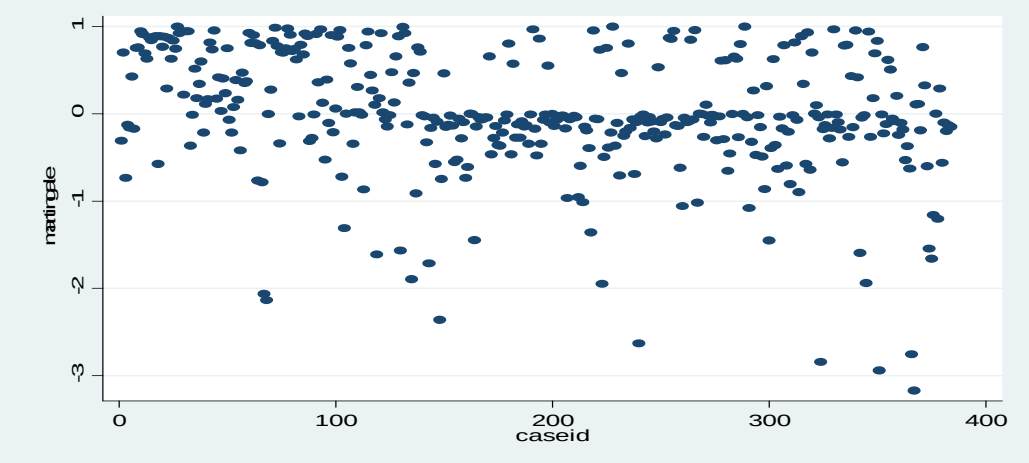

Figure 4.2: Martingale plot of TB data in JUSH

From figure 4.3, the Cox-Snells plotted points lie on a line that has a unit slope and zero intercept. So there is no reason to doubt the suitability of this fitted Gompertz model. We conclude that the Gompertz model is the best fitting TB data based on AIC and residuals plot.

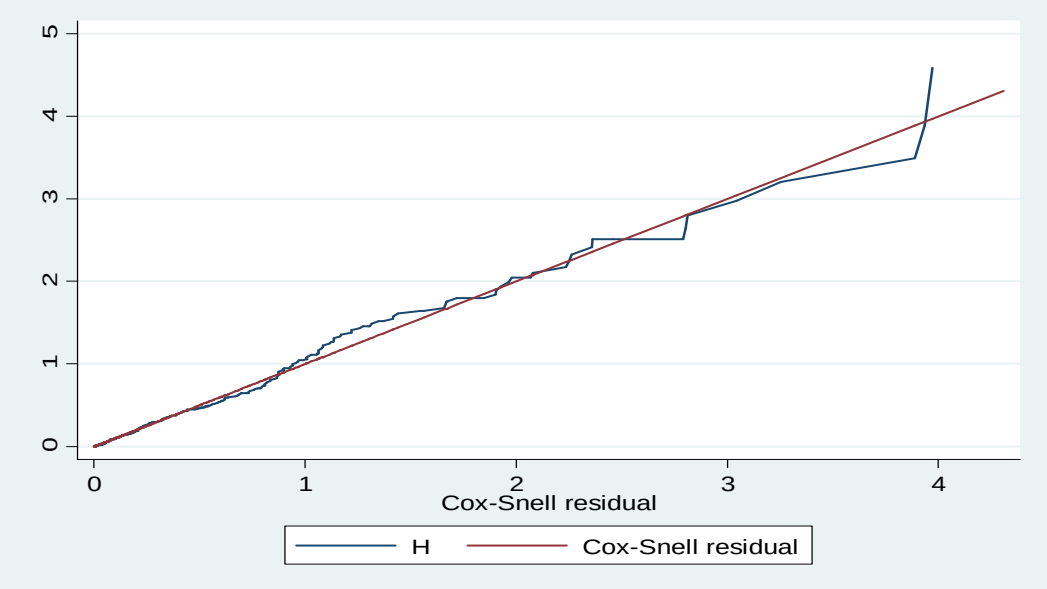

Figure 4.3: The Cox-Snell's plot of the Gompertz model of TB data in JUSH.

From Figure 4.4, the deviance residuals to be relatively well behaved and it is almost symmetric and the model is good fit. The deviance residuals are a rescaling of the martingale-like residuals so that they are symmetric about zero and thus more like residuals obtained from linear regression.

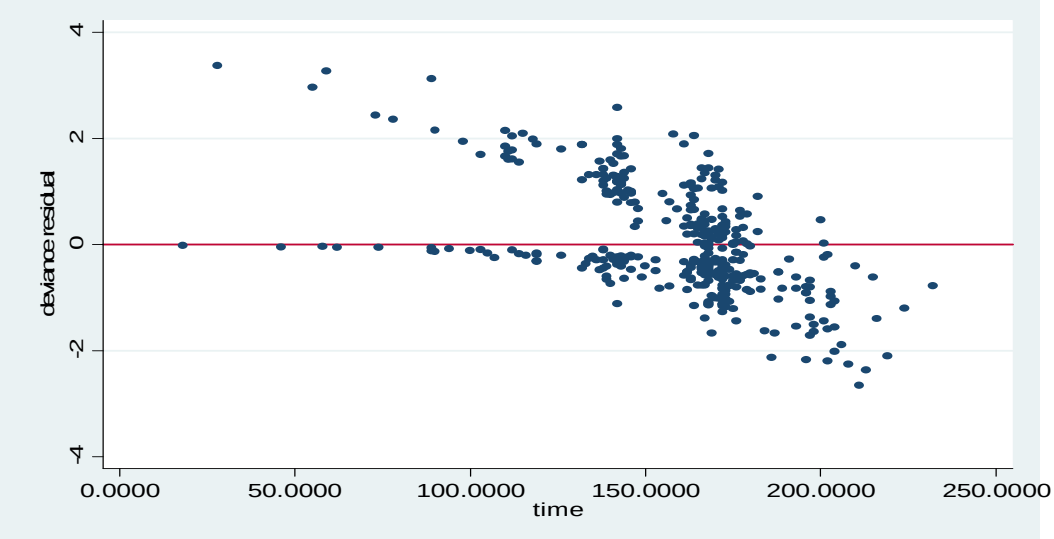

Figure 4.4: Deviance residuals plot used to evaluate model fit of GHM on TB data in JUSH

CHAPTER FIVE

5. DISCUSSION AND CONCLUSISON

5.1. Discussion

The purpose of this study was to identify the best predictor factors for the time to recovery from tuberculosis and to compare different parametric survival models on the time to recovery from TB in JUSH. The study used the secondary data collected from JUSH. Accordingly, descriptive survival analysis, Exponential, Weibull, Loglogestic, Gamma, Lognormal and Gompertz models were compared. The results, obtained have been discussed as follows.

Emily L. Butler showed that, the Gompertz model is better than Exponential and Weibull models when the data has different failure rates (Emily L. Butler, 2011). We confirmed Emily L. Butler, because of the cure rate of TB patients almost constant up to hundred days of treatment and exponentially increase after these days (see appendix B: Fig.4).

The Gompertz model is the best model to study the time to recovery from TB for data from JUSH by using AIC. While Ponnuraja and Venkatesan, applied likelihood based criteria for model selection indicated that the Gamma model was the best fitting parametric model for tuberculosis clinical trial data (Ponnuraja & Venkatesan, 2010).

Jiezhi Qi viewed that, after comparison of all the models and assessment of goodness of fit, found that the loglogistic AFT model fits better, for randomized placebo controlled trial to prevent Tuberculosis in Uganda adults infected with HIV (Jiezhi Qi, 2009).

A study conducted by Nakhaee and Law showed that, survival following a diagnosis of HIV infection was modeled by applying parametric survival models on people who were only diagnosed with HIV or HIV and AIDS registered in the national surveillance system from 1997 to 2003. Likelihood based criteria for model selection indicated that the Weibull model was the best fitting parametric model for predicting survival following both HIV and AIDS diagnoses (Nakhaee and Law, 2011).

The study showed that the types of TB, age, doses of anti-TB drug, weight, HIV status and residence of the patients were the covariates significant to time to recovery from TB. A case study done in the Hawasa city reported that the age, weight, doses and residences were significant on the time to death from AIDS/TB co-infection. Even though the above study had done on the time death of AIDS /TB, the covariates included in the study were similar with this study (Fikadu, 2012).

The cure rate of patients in JUSH during this study was about 53.65% while 46.35% of the patients were failed from their disease (TB). The cure rate in this study was much lower than the NTP and WHO reports. The NTP and WHO reported that even though not all health facilities have adequate there was 70% of case TB detection rate and 85% of cure rate by DOTS (WHO, 2006). The study in Jimma Zone Kersa district showed that, the cure rate of smear positive PTB was 64.6%. (Desalegn, 2012).

This study showed that the larger dose level is significant to recovery from the TB. From Tuberculosis register review of American society 5.8% of patients under intensive phase received below the recommended dose and 2.9% above the recommended dose. Similarly, 40.6% patients under continuation phase of treatment received below the recommended dose and 4.7% above the recommended dose. In addition to this 21 children received the adult dose of drug due to the absence of pediatric dose in all HCs. This contrasts the NTP as well as WHO standards; and also could cause drug side effect hepatotoxicity (Jussi et al., 2006).

According to this study Pulmonary TB was more curable than that of TB on the other organs of the patients. British healthy center reported that Pulmonary TB is treated using a six month course of a combination of antibiotics while the extra-pulmonary is 12 months (BHIV, 2014). This study confirmed the time to recovery proposed by NTP and WHO, and British health center reports. The study result shows that, the Extra-pulmonary TB needs more time to recovery than pulmonary TB (WHO, 2013).

Taking medication for six months is the most effective method of ensuring that the TB bacteria were killed. If the patient stops taking their antibiotics before they complete the course or if you skip a dose, the TB infection may become resistant to the antibiotics. This is potentially serious, as it can be difficult to treat and will require a longer course of treatment. If treatment is completed correctly, the patient should not need any further checks by a TB specialist afterwards. However,

they may be given advice about spotting signs that the illness has returned, although this is rare. In rare cases, TB can be fatal even with treatment. Death can occur if the lungs become too damaged to work properly (CDC, 2014).

This study showed that the urban residents had the quick time to recovery from TB than rural residents. The median time of urban patients' to recovery from TB was 169 days while the median time of recovery of rural TB patients' was 201 days. From the Gompertz model there was statistically significant difference between residences of patients to recovery from TB. Studies reported this was been the reason of that the rural residents are economically low and far from the healthy center (Daniel, 2011).

As the age of the patient increase the risk to recovery from TB is decrease. The studies done on the South West Ethiopia showed that mortality from the TB increases by 3% in every year (Debebe and Alemayehu, 2012). This result is the same with the result we obtained that the risk of recovery decreased with increase with age because death and recovery were controversy. Most papers reported that TB attacks adults than children because of the more contacts and it's mode of transmission. The weight and risk of recovery had the direct relationship. The weight of the patient decreased when caused by TB and increase when she/ he recovered.

This finding showed that the recovery time of TB in HIV/TB co-infection was longer than that of HIV un-infected patients. In the absence of data from clinical trials, it is not known if duration of treatment of TB in HIV infected patients should be longer than in HIV un-infected patients. The few data that exist suggest that in HIV infected patients duration of treatment for tuberculosis sensitive to first line therapy should be no difference to HIV uninfected patients (BHIV, 2014).

The study reported from Ghana showed that if TB is detected early, it can be treated in six months whiles it takes between six to ten years to treat for late detection (Gavrilenko, 2001). This study confirmed the average recovery time reported from the study done on sixty one patients in Northern Region, Ghana by using Cox-PH six months (Dioggban & Michael, 2012).

The study done on the breast cancer in Turkey reported that Gompertz model is the best model to study breast cancer among potential parametric models by using AIC (Elvan Akturk Hayat. et al., 2010).

5.2. Conclusion

Based on AIC and BIC the Gompertz model provided a suitable choice in order to model time of recovery from TB for the data obtained from Jimma University Specialized Hospital as compared to other parametric models. The average time to recovery from TB was 172 days which was approximately six months. Age of patent, body weight, HIV status, residence, dose level of drug used and types of TB are significant factors for time recovery from tuberculosis. The Extra-pulmonary TB was taken longer time to recovery from TB. The patients with HIV/TB co-infections take more time to recovery from TB than that of the patients uninfected with HIV/AIDS.

From the Gompertz model result we conclude that being old, rural residence, having Extra-pulmonary TB, Having HIV, lower doses and body weight at baseline prolonged the recovery time. From this study, we conclude that the Gompertz model is the best fitting TB data based on both AIC and residuals plot.

Inspite of different models suggested and recommended by many researchers, who followed different methodologies in their experiments, the result of this study has a strong inclination for the Gompertz method as the most suitable one; better than other based on AIC for time to recovery from TB studies.

5.3. Recommendation

The following recommendations have been given concerning the result of the study:-

The average recovery time from TB is 172 days. This recovery time depends on first line TB treatment. I have recommend to further study on the second line TB treatments and MDR-TB duration.

The policy makers and health service providers should give special need for rural residence, those having Extra-pulmonary and TB/HIV co-infection since they need prolong time to recovery from TB.

Since different models suggested and recommended by many researchers, who followed different methodologies in their experiments, the result of this study has a strong recommendation for the Gomperz method as the most suitable one; better than other based on AIC for time to recovery from TB.

Since the study used retrospective data collected from TB patients' card, the covariates like smoking status, marital status, occupation, education level and alcoholism are not included in the study. Therefore I recommend for the future study to include these variables.

Lastly, when the proportional assumption of Cox-PH is either not satisfied or not checked the parametric models are better to model the survival data.

6. REFERENCES

Abrham Keraleme, (2009). Survival of tuberculosis treated under directly observed treatment short course in Addis Ababa Ethiopia: Addis Ababa University.

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

Amare Deribew, Markos Tesfaye, Yohannes Hailmichael, Nebiyu Negussu4, Shallo Daba, Ajeme Wogi, Tefera Belachew, Ludwig Apers and Robert Colebunders,(2009). Tuberculosis and HIV co-infection: its impact on quality of life*: Bio Med Central Ltd* 105*(***7***)*

Ayalew T, & Meseret Y, (2009). Delays in tuberculosis treatment, and associated factors in Jimma Zone, south west Ethiopia. *Ethiopia J Health sci,* 19(**1**)

Belay Tessema, Abebe Muche, Assegedech Bekele, Dieter Reissig, Frank Emmrich, and Ulrich Sack, (2009).Treatment outcome of tuberculosis patients at Gondar UniversityTeaching Hospital, Northwest Ethiopia. A five - year retrospective stud: *BMC Public Health* ***9****:371*

Berhane Y. Wall S., Fantahun M. et al., (2008). A rural Ethiopian populations undergoing epidemiological transition over a generation: Butajira from 1987 to 2004.

BHIV, (2014). British Health Report on TB/HIV. WWW. BHIVA - 4.0 Type and duration of TB treatment.htm

Bong Ngeasham Collins, (2012). Assessing the Outcome of Tuberculosis Treatment in the Cameroon Baptist Convention Health Board Tuberculosis Treatment Centers. Umea University, Sweden.

CDC. (2014). Centers for Disease Control and Prevention. Treatment of Tuberculosis, American Thoracic Society, CDC, and Infectious Diseases Society of America.

Collett, David, (2003). Modeling Survival Data in Medical Research. Chapman and Hall, London.

Cox, (1972). Regression models and life tables with discussion Journal of the Royal Statistical Society.

Cox, D.R. and Oakes, D., (1984). Analysis of survival data. London: Chapman and Hall.

Daniel Gemechu, (2011). Improving Tuberculosis Control in Ethiopia, University of Bergen, Norway: *Centre for International Health.*

David Hosmer W., (1999). Applied survival analysis: Regression model of time to event data, John Wiley & sons inc.

David W. Hosmer, Stanley Lemshow, Susanne May, (2012). Survival analysis: Regression modeling of time to event data. 2nd ed. Wiley series.

Debebe S. & Alemayehu. W., (2012). Tuberculosis treatment survival of HIV positive TB patients on directly observed treatment short course in Southern Ethiopia: A retrospective cohort study.

Desalegn Dabaro, (2012). Process evaluation of quality in the diagnosis and treatment of tuberculosis at Kersa district, Jimma zone, *Ethiopia J Health sci.*

Dioggban.j and Michael o., (2012). Survival analysis of average recovery time of tuberculosis patients in northern region, Ghana, *International journal of current research.*

Elisa T. Lee & John Wenyu Wang, (2003). Statistical method for survival data analysis, 3rd ed, John wiley inc

Elvan Akturk Hayat, Asli Suner, Burak Uyar, Omer Dursun, Mehmet N. Orman and Gul Kitapcioglu MD., (2010). Comparison of Five Survival Models: Breast Cancer Registry Data from Ege University Cancer Research Center; *Turkiy, J Med Sci*, 30(**5**), 1665-74.

Emily L. Butler, (2011). Estimating the Survival Distribution of Aluminum Processing Pots, Carnegie Mellon University; *Dietrich College honors thesis*

Extending SAS survival analysis techniques for medical research. http://WWW. Extending SAS survival analysis techniques for medical research.

Federal Ministry of Health of Ethiopia, (2007): Federal Ministry of Health of Ethiopia. Health and health related indicators.

Federal Ministry of Health of Ethiopia, (2012). TB control program annual performance report 2011/12. Addis Ababa: *Federal Ministry of Health of Ethiopia*

Fikadu Z., (2012). Survival analysis of tuberculosis patients under dots program: a case study at Hawassa city, Ethiopia.

Gavrilenko, V. S., (2001). Recovery criteria and time in patients with pulmonary tuberculosis. Tuberk, (8): 10-4.

Getahun Asebe, Haimanote Dissasa, Takel Teklu, Gebremedihin Gebreegizeabher, Gobena Ameni, (2010). Treatment outcome of tuberculosis patients at Gambella Regional Hospital, Southwest Ethiopia: A ten-year retrospective study.

Gharibvand, Lida Liu, Lei, (2009). Analysis of Survival Data with Clustered Events, SAS global forum.

Gupta R.C., Akman O. and Lvin S., (1999). "A study of log-logistic model in survival analysis", Biom J, 41(4), 431-43.

Hosmer, D.W., Jr., and Lemeshow, S. (1999), Applied Survival Analysis: Regression Modeling of Time to Event Data. New York: Wiley

Jiezhi Qi, (2009). Comparison of Proportional Hazards and Accelerated Failure Time Models, Thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Jussi J, (2006). An Official ATS Statement: Hepatotoxicity of Antituberculosis Therapy. American Thoracic Society Documents.

Kaplan, E. L., and P. Meier. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53: 457–481.

Kay R. and Kinnersley N., (2002). "On the use of the accelerated failure time model as an alternative to the proportional hazard model in the treatment of time to event data: A case study in Influenza", *Drug information Journal*, **36**, 571-79.

Kleinbaum, Mitchel Klein, (2012). Survival analysis: a self-learning text, 3nd edition, springer.

Klembaum, D. G., (1996). Survival Analysis: A Self learning text. Springer, New York

Lawless, J. F., (1982). Statistical Methods and Model for Lifetime Data. Wiley, New York.

Lee, E. T. and J. W. Wang., (2003). Statistical Methods for Survival Data Analysis. 3rd ed. New York: Wiley.

Ministry of Health of Ethiopia (MOH), (2008). Tuberculosis, Leprosy and TB/HIV Prevention and Control Programme Manual. *Addis Ababa: MOH* 4th edition.

Murray C.J., Lopez A.D., (1996). Global health statistics. Cambridge, MA, Harvard School of Public Health, (Global Burden of Disease and Injury Series, vol. II).

Nakhaee, F. and Law M., (2011). "Parametric modeling of survival following HIV and AIDS in the era of highly active anti-retroviral therapy: data from Australia", Eastern Mediterranean Health Journal, 7(**3**)

Ohishi, K., Okamura, H. and Dohi, T., (2009). Gompertz software reliability model: estimation.

Pan, W, (2001). Akaike's information criterion in generalized estimating equations

Ponnuraja, C. and Venkatesan P., (2010). "Survival models for exploring tuberculosis clinical trial data an empirical comparison", *Indian Journal of Science and Technology*, 2**(7**), 755-58.

Rajagopalan, S., (2001). Tuberculosis and aging: a global health problem. *Clin Infect Dis*; **33**: 1034–1039.

Raviglione MC, Pio A., (2002). Evolution of WHO policies for tuberculosis control, 1948-2001. *Lancet*, 359(**9308**):775-780.

Reichman and Hershfield's, (2006). Tuberculosis a comprehensive International Approach. Volume 219. New York.

Royall, R. M. (1997), "Statistical evidence: A likelihood paradigm", London. Chapman & Hall.

Senait A., (2011). Assessment of knowledge attitude and practices of tuberculosis patients towards tuberculosis and its treatment in Addis Ababa city government, Ethiopia

Shane H., (2005). Co-infection with HIV and TB: double trouble. *Int J STD AIDS*, **16(2**):95-100;

Tai P, Chapman J.A., Yu E, Jones D, Yu C, Yuvan F, (2007), "Disease-specific survival for limited-stage small cell lung cancer affected by statistical method of assessment", BMC cancer, 7:31

Teklu B., (1993). Symptoms of pulmonary tuberculosis in consecutive smear-positive cases treated in Ethiopia. *Tuber Lung Dis* 1993, 74(2):126-128.

Therneau, T.M and P.M. Grambsch, (2000). Modeling Survival Data: Extending the Cox Model. Springer Verlag, New York.

WHO, (2011). Treatment of Tuberculosis guidelines, Geneva, www.who.int/tb/en/

Wikstrom, G., (2011). Women's Perspectives on Pathway to Diagnosis of Pulmonary Tuberculosis Women Voices from Community Level in Uganda. Nordic School of Public Health.

Willemse, W. J. and Koppelaar H., (2000). "Knowledge elicitation of Gompertz' law of mortality", Scandinavian Actuarial Journal, 2, 168–79.

World Health Organization, (2006). Global tuberculosis report. *WHO 2006 report*. WHO/HTM/TB.

World Health Organization, (2009). Dosing instructions for the use of currently available fixed-dose combination TB medicines for children. WHO/HTM/TB.

World Health Organization, (2012). Global tuberculosis report. *WHO 2012 report*. WHO/HTM/TB.

World Health Organization, (2013).Global Tuberculosis report. *WHO 2013 report* WHO/HTM/TB

Xian Liu, (2012). Survival analysis: Models and applications, john Wiley and sons, ltd. Publications.

APPENDIXES

Appendix A

Tables

Table 1: The survival estimate of Kaplan-Meier

Time	Total	Fail	Lost	Survival	Std. error.	[95% Conf.	Int.]
18	384	0	1	1.0000	.	.	.
28	383	1	0	0.9974	0.0026	0.9816	0.9996
46	382	0	1	0.9974	0.0026	0.9816	0.9996
55	381	1	0	0.9948	0.0037	0.9793	0.9987
58	380	0	1	0.9948	0.0037	0.9793	0.9987
59	379	1	0	0.9921	0.0045	0.9758	0.9975
62	378	0	1	0.9921	0.0045	0.9758	0.9975
73	377	1	0	0.9895	0.0052	0.9723	0.9961
74	376	0	1	0.9895	0.0052	0.9723	0.9961
78	375	1	0	0.9869	0.0058	0.9688	0.9945
.	.	.	.	.	.	.	.
206	10	0	1	0.1709	0.0355	0.1081	0.2457
208	9	1	0	0.1519	0.0363	0.0893	0.2298
210	8	1	0	0.1329	0.0364	0.0719	0.2128
211	7	0	1	0.1329	0.0364	0.0719	0.2128
213	6	1	0	0.1108	0.0364	0.0524	0.1936
215	5	0	1	0.1108	0.0364	0.0524	0.1936
216	4	1	0	0.0831	0.0363	0.0298	0.1716
219	3	0	1	0.0831	0.0363	0.0298	0.1716
224	2	0	1	0.0831	0.0363	0.0298	0.1716
232	1	1	0	0	.	.	.

Table 2: Cox-PH model assumption checking table

Variables	rho	chisq	P
TBTplumonary	-0.0233	0.113	0.737
sexmale	-0.1584	5.499	0.019
Age	-0.0505	0.627	0.428
Bwt	0.0604	0.909	0.34
hivnegative	0.0495	0.53	0.467
dose	0.0673	0.759	0.384
residrural	-0.037	0.287	0.592
GLOBAL	NA	8.053	0.328

Table 3: The comparison results of the significant factors of the potential parametric models

Variables	Exponential Coef.(Std. Err	Weibull Coef.(Std. Err.)	Gompertz Coef.(Std. Err.)	Log logistic Coef.(Std. Err.)	Log normal Coef.(StdErr.)
$Resi_{(rural)}$	-0.714 (0.179)*	-0.723 (0.181)*	-0.723 (0.182)*	0.098 (0.027)*	0.147 (0.035)*
$Tbt_{(plumona)}$	1.308 (0.262)*	1.41 (0.262)*	1.412 (0.262)*	-0.207 (0.036)*	-0.275 (0.046)*
$Hiv_{(neg'v)}$	0.758 (0.272)*	1.069 (0.279)*	1.185 (0.283)*	-0.155 (0.041)*	-0.166 (0.050)*
weight	0.014 (0.009)	0.025 (0.009)*	0.027 (0.009)*	-0.003 (0.001)	-0.003 (0.002)
age	-0.018 (0.008)*	-0.025 (0.008)*	-0.025 (0.008)*	0.004 (0.001)*	0.003 (0.002)*
Dose III	0.092 (0.154)	0.094 (0.155)	0.082 (0.156)	-0.034 (0.026)	-0.021(0.034)
Dose IV	0.277 (0.214)	0.492 (0.218)*	0.528 (0.219)*	-0.068 (0.035)*	-0.066 (0.046)
cons	-7.555 (0.615)*	-38.769 (2.142)*	-13.984 (0.779)*	5.508 (0.094)*	5.579 (0.125)*
AIC	607.9297	58.53924	50.99097	86.03688	159.5004
BIC	639.5348	94.09502	86.54675	121.5927	195.0562
-2LL	-591.93	40.53924	32.991	68.0368	141.5004

APPENDIX B

Graphs

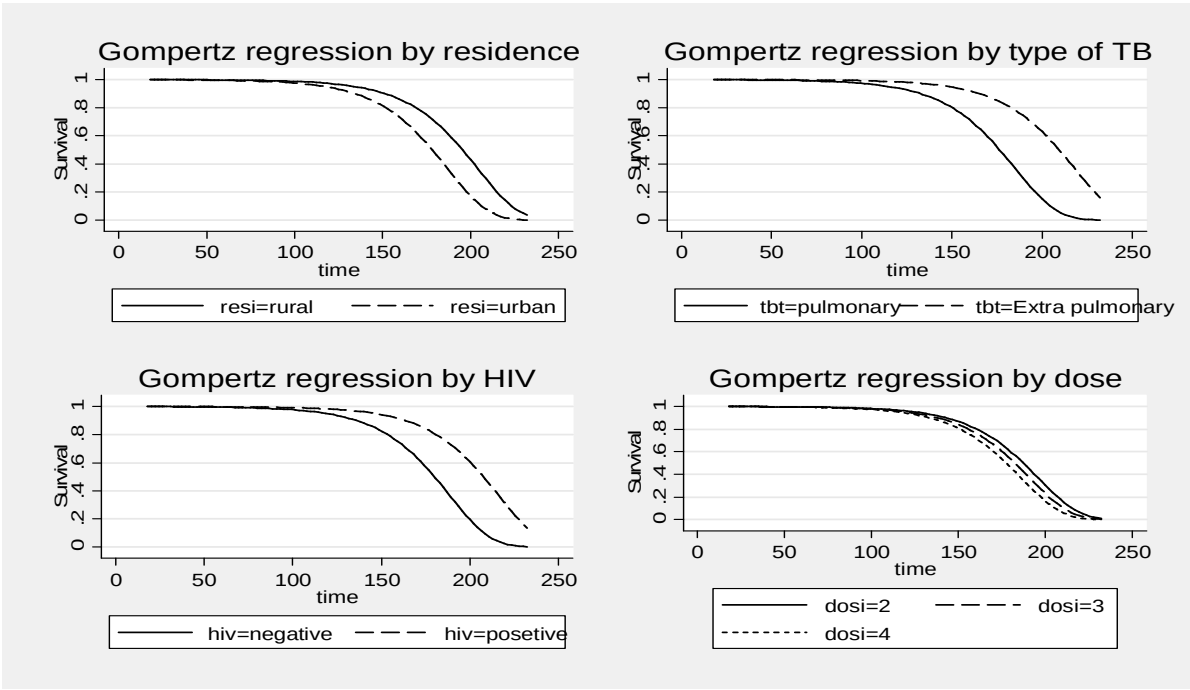

Figure 1: The Gompertz regression plot of categorical variables

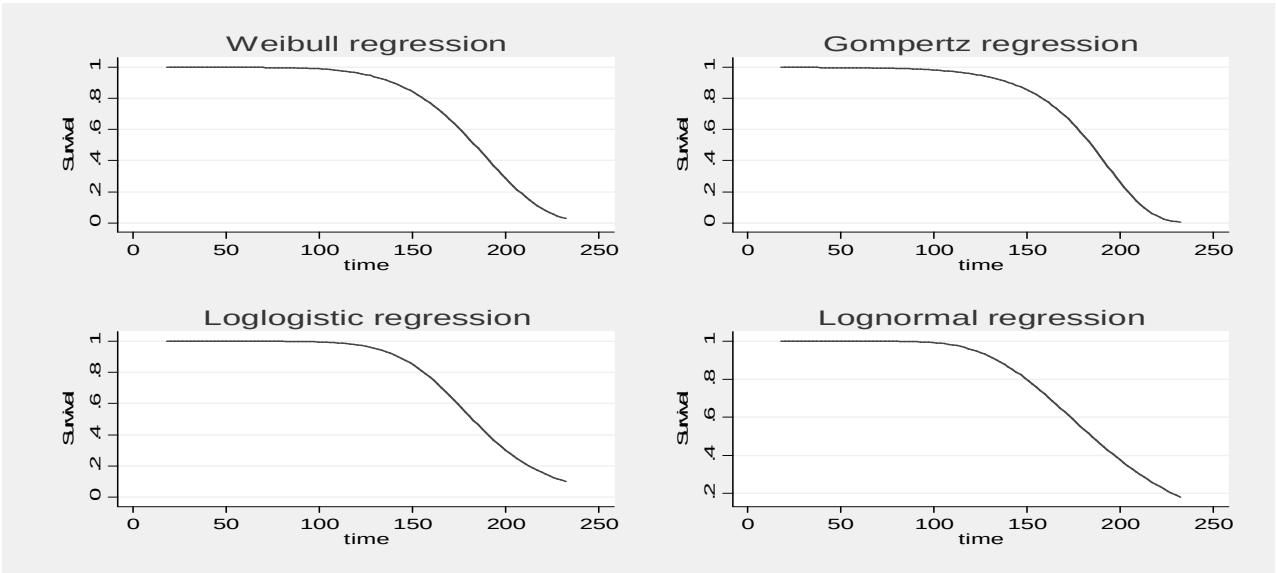

Figure 2: The survival estimate of different survival models

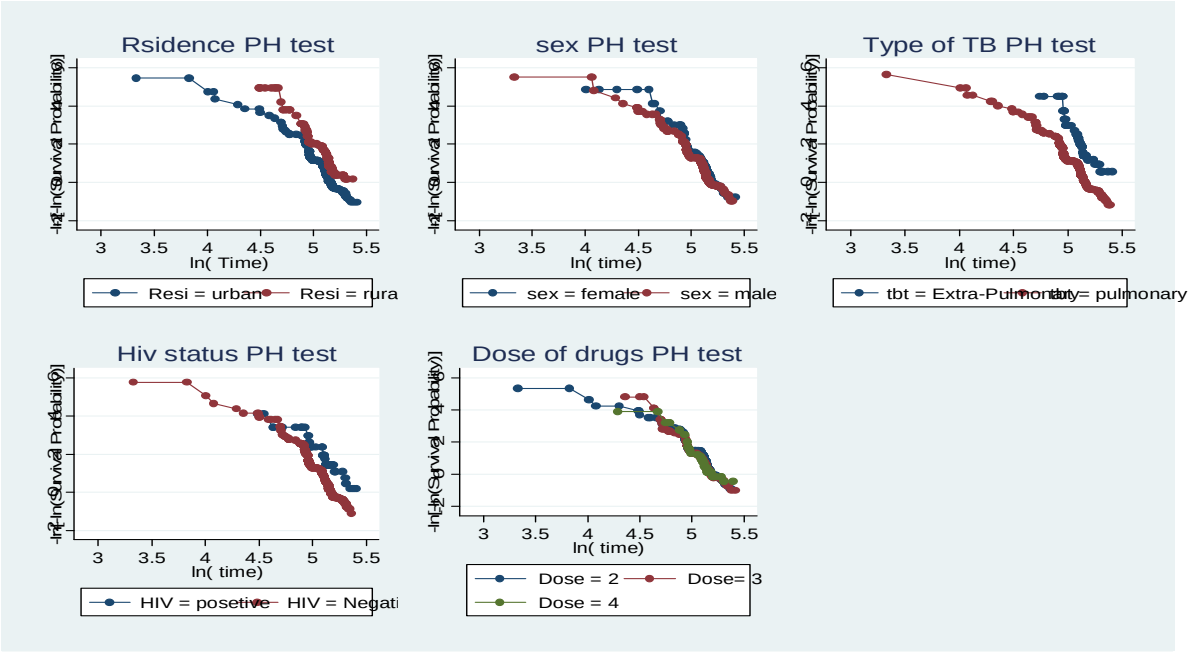

Figure 3: The proportional assumption the covariates.

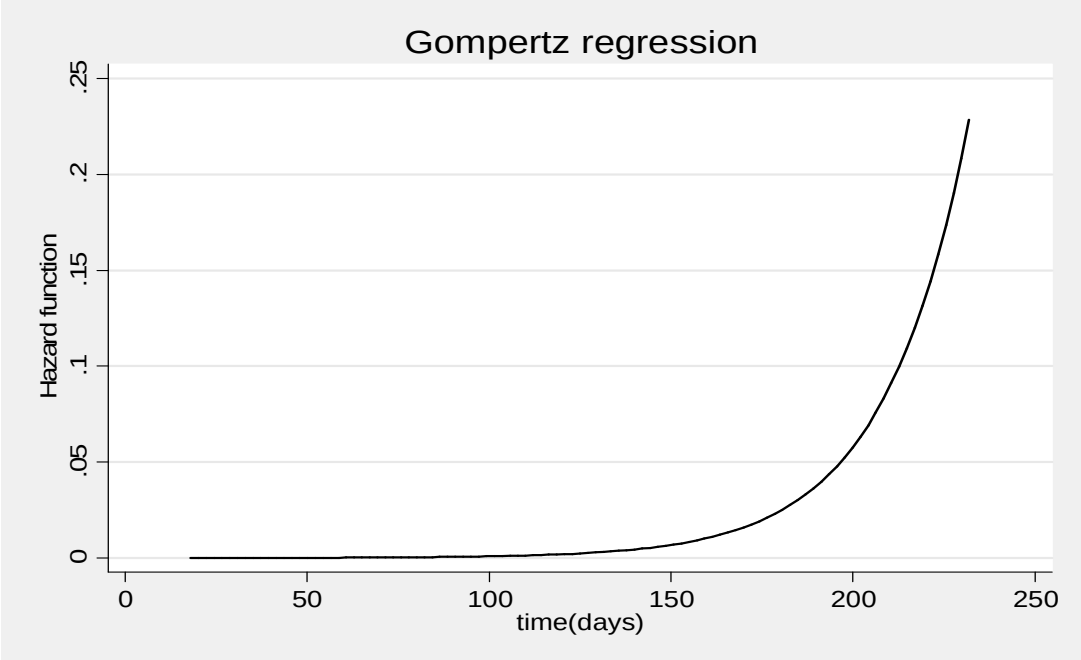

Figure 4: The Gompertz hazard model plot of TB data in JUSH

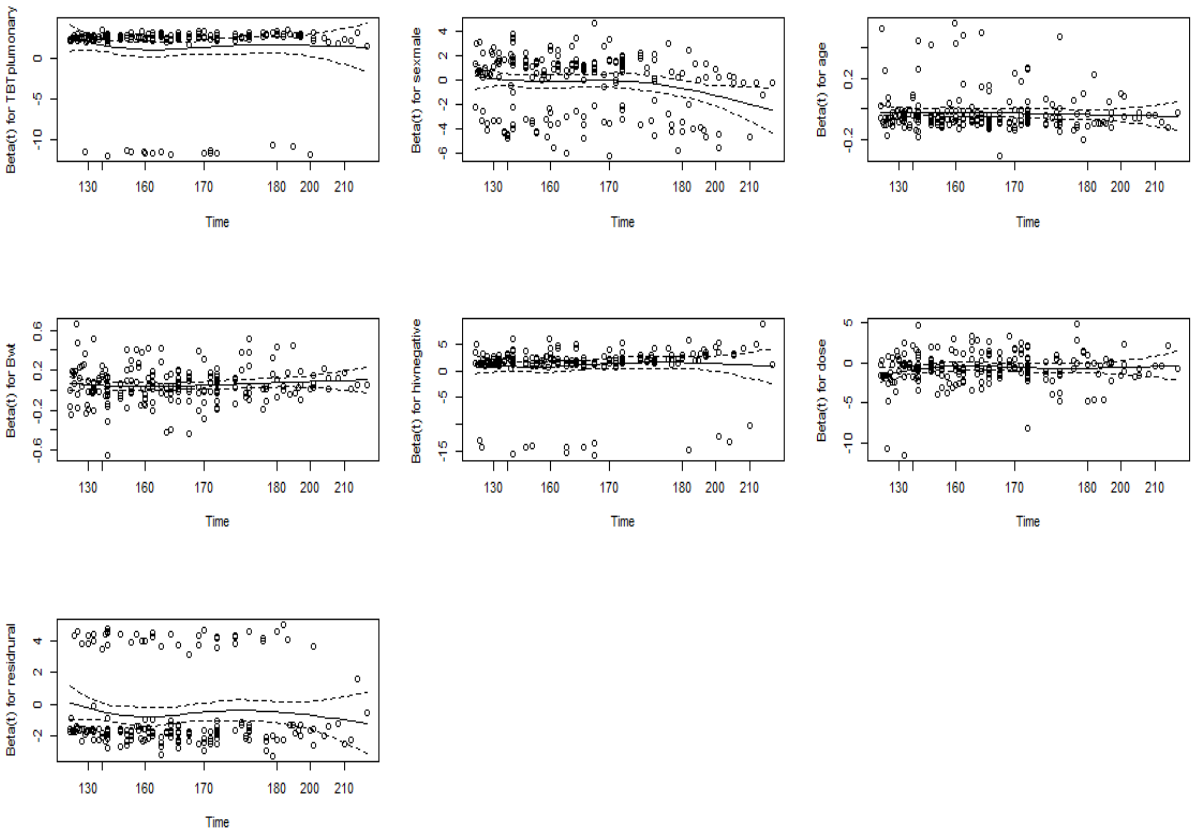

Figure 5: Plots of scaled Schoenfeld residuals against transformed time for each covariate in a model fit to the TB data.

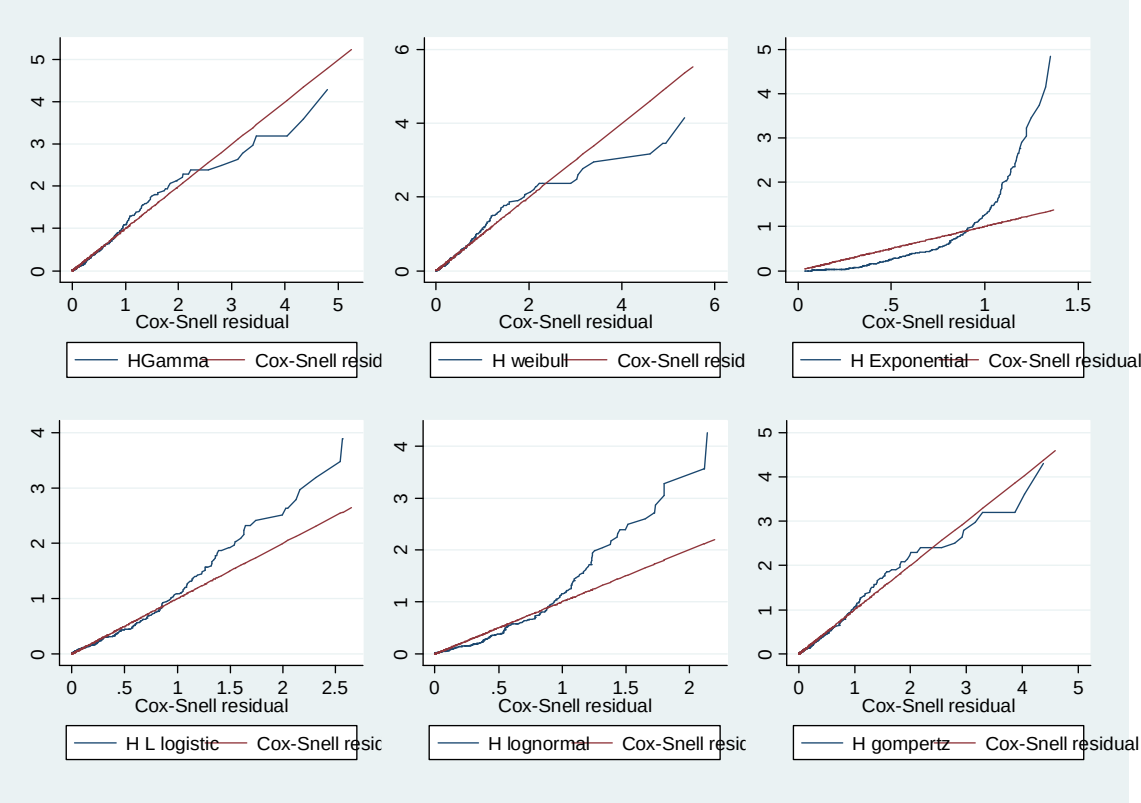

Figure 6: The cox -snell plots of parametric survival models