JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCE

STATISTICS DEPARTMENT

MODELING TIME TO CURE OF TUBERCULOSIS PATIENTS UNDER DOTS USING MIXTURE CURE MODELS (CASE STUDY IN JIMMA UNIVERSITY SPECIALIZED HOSPITAL)

BY: TIZAZU YADETA

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

Jimma, Ethiopia

June, 2016

MODELING TIME TO CURE OF TUBERCULOSIS PATIENTS UNDER DOTS USING MIXTURE CURE MODEL (CASE STUDY INJIMMA UNIVRSITY SPECIALIZED HOSPITAL)

BY: TIZAZU YADETA

Advisor: GEREMEW MULETA (Assist .Prof.MSc)

Co. Advisor: MERSHA FILATE (MSc.)

Jimma,Ethiopia

June, 2016

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. 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. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission 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.

TizazuYadeta Gefela

Date: _____

Signature: _____

June, 2016

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES JIMMA UNIVERSITY

As thesis research advisors, we here by certify that we have read the thesis prepared by Tizazu Yadeta under our guidance, which is entitled **"Modeling Time-to- cure of tuberculosis patients under DOTS using mixture cure model"**, in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Geremew Muleta (Assist.Prof.,MSc.) _____ _____

Advisor Signature Date

Mersha Filate(MSc.) _____ _____

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.

Tesfaye Getachew_(MSc.)__ _____ _____

Name of Chairman Signature Date

Geremew Muleta (Assist.Prof.,MSc.) _____ _____

Name of Advisor Signature Date

Mersha Filate(MSc.)_____ _____ _____

Name of Co-Advisor Signature Date

Yasin Negash (MSc.)___ _____ _____

Name of Internal Examiner Signature Date

Adem Kedir (PHD) _____ _____

Name of External Examiner Signature Date

ACKNOWLEDGMENTS

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study at Department of Statistics, Jimma University.

I would like to gratefully and sincerely thank Mr. Geremew Muleta, my thesis advisor and instructor, for his invaluable comments, suggestions and patience during the entire time of the study. Next I want to thank Mr. Mersha Filate , my thesis co-advisor for his comment ,suggestion and advise.

My special thanks also go to Wollega University for sponsoring me during my study years.

I would also like to express the great debt I owed to all staff member of Jimma University Statistics Department for sharing me their individual experience specially Mr. Belay Brilie for his professional help with patience for providing me help and moral support in one way or another to accomplish this study.

Finally, yet importantly, my heart-felt thanks go to my Sister Meserat Yadeta for providing me basic need in home, love and respect when I stay in Jimma University and all my family members for all the encouragement and support that they gave me.

ACRONMY

TB	Tuberculosis
WHO	world health organization
MTB	Mycobacterium Tuberculosis
MDR-TB	multidrug-resistant tuberculosis
DOT	Directly observed treatment
MDGs	Millennium Development Goals
K-M	Kaplan-Meier
MCMs	Mixture cure models
AIC	Akakie information criterion
EM	Expectation maximization
USAID	United States Agency for international developments
IUATLD	International Union Against Tuberculosis and Lung Diseases

ABSTRUCT

Background: Tuberculosis (TB) is a global health concern; nearly one-third of the global population is infected with Mycobacterium tuberculosis and at risk of developing the disease (WHO, 2011). The objective of this study is to Model time –to- cure of tuberculosis patients using mixture cure model and identify the risk factors for the cure of tuberculosis patients under DOTS.

Methods: longitudinal retrospective cohort (1st July, 2014 and 1st January, 2016) follow up (retrospective cohort design) of tuberculosis patients' data were obtained from Jimma University Specialized Hospital TB Patient Clinic located in Jimma town. Univariate and multivariate analyses were performed using a logistic Cox PH mixture cure model and for uncured group standard Cox regression model.

Results: Of all 501 tuberculosis patients 439(87.62%) susceptible group 328(74.15%) were cured. The median cure time from TB was 6 months and 62(12.37%) non susceptible those multi-drug resistance TB none of them cured with in the follow up period. In the two population susceptible and non-susceptible or multidrug resistances TB the most of the MDR-TB patients are pulmonary negative 55 of them. From these patients of TB type pulmonary negative was risk to develop MDR-TB. An increased incidence of TB was reported for smear result (pulmonary positive, pulmonary negative and extra pulmonary), weight at initiation of treatment and HIV (HIV-/HIV+) were the risk factors predicting time to cure from tuberculosis diseases. HR and 95%CI were 2.3 [1.808 :3.121], 0.9853 [0.9766 :0.9941] and 0.7564 [0.522 , 1.095] respectively. Then both direction selection methods are applied using the software and AIC=3455.42 were small with the following end covariates.

Conclusions: body weight at initiation of treatment and smear result are the risk factor for time to cure in tuberculosis patients in multivariable Cox proportional regression model during anti tuberculosis treatment period for uncured and regiment in addition two was the risk for cured population in multivariable logistic Cox PH mixture cure model.

Contents

LIST OF TABLE **page**

List of figures page

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Tuberculosis (TB) is a global health concern; nearly one-third of the global population is infected with Mycobacterium tuberculosis and at risk of developing the disease (WHO, 2011). More than 90% of global TB cases and deaths occur in the developing world, where 75% of cases are in the most economically productive age group (WHO, 2010). Ethiopia ranks seventh among the world's 22 high-burden tuberculosis (TB) countries (Chaisson RE, Harrington M, 2009)

Worldwide, tuberculosis (TB) kills more young and middle-aged adults than any other infectious disease (WHO, 1999). Though it is curable and preventable, more than 5,000 people die of TB *every day* (2 to 3 million people per year) (WHO, 1999). TB often strikes the most vulnerable members of society and, if left untreated, causes its victims to lose weight, weaken, and eventually waste away (Ryan, 1993). TB disproportionately affects the indigent and other marginalized groups of society in whom unequal susceptibility patterns have long been recognized.

Adding further urgency to controlling TB, multidrug-resistant TB (MDR-TB) has emerged as a serious problem in many parts of the world, including Russia, Latvia, Estonia, Argentina, the Dominican Republic, and the Ivory Coast (WHO, 1999). Up to 50 million people worldwide may be infected with MDR-TB (WHO, 1999). In low-prevalence countries, drug resistance is generally more common in foreign-born populations, most likely reflecting inadequate treatment programs and sporadic drug availability in high-prevalence countries (Broekmans, 2000).

Many complex biological and social factors impact TB transmission, progression to disease, and treatment. By identifying and examining the interrelation of these factors, anthropologists and other social scientists have made important contributions toward the control of TB. Anthropological methods and approaches have been especially valuable in understanding and addressing the broad range of socio-cultural, behavioral, and structural issues pertinent to TB

control. While limited in breadth and scope, the research highlighted here has many implications and applications for TB control practice.

Tuberculosis treatment success rate (% of new cases) in Ethiopia was 91.00 as of 2012. Its highest value over the past 18 years was 91.00 in 2012, while its lowest value was 61.00 in 1995. Tuberculosis treatment success rate is the percentage of all new tuberculosis cases (or new and relapse cases for some countries) registered under a national tuberculosis control program in a given year that successfully completed treatment, with or without bacteriological evidence of success ("cured" and "treatment completed" respectively) (WHO, 2014).

To build on the achievements of DOTS and address the remaining challenges, the STOP TB strategy was launched by WHO in 2006 to help achieve the millennium development goals for TB in 2015. Ethiopia also adopted this strategy to achieve the national TB targets.

Although subtle, one needs to distinguish between the concepts of censoring and cure: censoring refers to a subject who does not fail within the monitoring time window of a particular subject, while cure refers to one who will not fail within any reasonable monitoring time window. Indeed the latter is an abstraction as we never “observe” a cure (due to a finite monitoring time). Recently much attention has been devoted to formulating parametric survival models incorporating a cured fraction – a non-zero tail probability of the survival function. These have focused upon cancer-relapse trials including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer, where due to recent advances in therapy and treatment, a significant proportion of patients are expected to be “cured”, that is to remain disease-free even after really long follow-ups. Cure rate models incorporating a cured fraction, defined as a non-zero tail probability of the survival function, adjust for this feature of the data and date back to the mixture model by Berkson and Gage (1952) (mixture-cure model, in short) and has been extensively discussed by several authors, including Farewell (1982), Farewell (1986), Maller and Zhou (1996), Ewell and Ibrahim (1997).

The mixture cure model (ScCheng at el., 1998, Yingwei, Peng and Keith B.G.Dear, 2000,J.P.Sy and J.M.G.Taylor, 2000) is particularly designed to handle the dataset with a cure fraction. Unlike the standard survival model, the mixture cure model has two components in order to model the cure probability and the survival probability of uncured patients. As discussed in

Maller and Zhou (1996), the population contains" long term survivors" (or "cured", "immune" individuals) if the cdf F(t) of failure time is improper, i.e. $F(\infty) = P(T^* < \infty) < 1$.

1.2. Statement of the Problem

The survivor function for the entire population one unstated assumption of the standard PH model is that all individuals under study are susceptible to the adverse event of interest, and they would experience the event eventually if there was no censoring. However, more and more patients will be cured now a day due to the advances in recent medical research, that is, those patients may never experience the event even after a sufficient follow-up period. With rapid developments in medical and health sciences, we now encounter more survival studies where some patients are expected to be *cured*. Survival models that account for cure are important for understanding prognosis in potentially terminal diseases. Traditional parametric survival models such as Weibull or Gamma (Cox and Oakes, 1984) do not account for the probability of cure.

Therefore, this work is different from other work by: -as stated above the standard Cox PH regression does not consider the cured patients, but this study will consider the mixture cure model that consider the cured and uncured individuals simultaneously. And again the target of the world is to stop and control tuberculosis by applying the DOTs strategies to achieve millennium development goal and it can define tuberculosis is curable disease but the survival study take their event as death which consider the survival probability of individual at time of follow up increases is zero, but the drug objective is to make patients free from disease by treating them in appropriate manner.

In 2009 TB is the 5th cause of admission, the 3rd leading cause of death with 67%cure rate for registered patient in the same year in Ethiopia (FMOH, 2010), so we say in one direction TB is curable and the data consider the patient under treatment, but using the death as event of interest by itself is contradicting the idea because of the population of tuberculosis patient is two, those MDR-TB or long term-survivors and susceptible those cured in short period. Therefore this study will use the event as cure to overcome this contradiction.

The mixture cure model is a special type of survival models and it assumes that the studied population is a mixture of susceptible individuals who may experience the event of interest, and

cure/non-susceptible individuals who will never experience the event. For such data, standard survival models are usually not appropriate because they do not account for the possibility of cure. Based on this the paper should depend on the following research question:

1. What is the incidence prevalence of tuberculosis in study area?
2. What are the cofactors affecting the susceptible and non-susceptible patients?
3. What is the appropriate baseline distribution for the susceptible patients?
4. Does the cure model appropriate for the data?

1.3 Objective of the study

1.3.1 General objectives

The general objective of this study is to Model time –to- cure of tuberculosis patients using mixture cure model and identify the risk factors for the cure of tuberculosis patients under DOTS at the study area.

1.3.2 Specific objectives

i. To fit Cure model for the MDR-TB group under the significant factors under DOTS.
ii. To determine failure time distribution of uncured patients.
iii. To identify factors associated with time to cure of TB patients.
iv. To compare the probability of cure between the tuberculosis types in terms of their excess hazards rates.

1.4. Significance of the Study

The result of this study was providing information on modeling time to cure of tuberculosis patients under DOTs by considering the mixture cure models. This results give some knowledge about the determinants or risk factors of time to cure of tuberculosis patients, provides an input for further study in the area, provides an evidence for the government and concerned bodies to evaluate their plan and used for the professional to determine the incidence prevalence of tuberculosis.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. The literature of curability of tuberculosis

Worldwide, tuberculosis (TB) kills more young and middle-aged adults than any other infectious disease (WHO, 1999). Though it is curable and preventable, more than 5,000 people die of TB *every day* (2 to 3 million people per year) (WHO, 1999).

Much literature confirms what current rates for TB diagnosis and treatment completion demonstrate: TB control programs do a better job holding rather than finding cases. Although both are pressing tasks, bringing people in to get diagnosed seems a more formidable challenge than ensuring that patients get and complete treatment once they are in the health system. As long as monitoring and supervision are adequate, the majority of patients are likely to complete treatment. Certainly, both diagnosis and treatment effectiveness require that the infrastructure of TB control programs had a minimally acceptable functioning. That is, drug supply is reliable and sustainable, laboratory facilities are available and sufficiently equipped to perform tests, monitoring systems are in place, and human capacity for different tasks (e.g. supervision, laboratory work) is adequate.

The study demonstrated that patients who died, treatment failures and defaulters were the main reason why we don't archive the 85% target treatment success set by the WHO. The youngest age group (0-14years), females, new patients and HIV negative TB patients, had higher proportions of treatment outcome. HIV negative TB patients were found to be significantly associated with better treatment outcome (Bong Ngeasham Collins, 2011).

Tuberculosis (TB) remains one of the world's deadliest communicable diseases. In 2013, an estimated 9.0 million people developed TB and 1.5 million died from the disease, 360 000 of whom were HIV-positive. TB is slowly declining each year and it is estimated that 37 million lives were saved between 2000 and 2013 through effective diagnosis and treatment. However, given that most deaths from TB are preventable, the death toll from the disease is still

unacceptably high and efforts to combat it must be accelerated if 2015 global targets, set within the context of the Millennium Development Goals (MDGs), are to be met. TB is present in all regions of the world and the Global Tuberculosis Report 2014 includes data compiled from 202 countries and territories. This year's report shows higher global totals for new TB cases and deaths in 2013 than previously, reflecting use of increased and improved national data. A special supplement to the 2014 report highlights the progress that has been made in surveillance of drug-resistant TB over the last two decades, and the response at global and national levels in recent years. Worldwide, the proportion of new cases with multidrug-resistant TB (MDR-TB) was 3.5% in 2013 and has not changed compared with recent years. However, much higher levels of resistance and poor treatment outcomes are of major concern in some parts of the world. The supplement, Drug Resistant TB: Surveillance and Response, defines priority actions needed, from prevention to cure (WHO, 2014)

The emergence and geographical spread of strains of TB that are resistant to treatment by the standard anti-TB drugs is a second major concern. Resistance to any one drug is the result of naturally occurring genetic mutation within the *Mycobacterium*, but resistance to multiple drugs emerges through incorrect, interrupted or frequently repeated treatment. The WHO estimates that there are now almost 490,000 new cases of multidrug-resistant (MDR) TB every year, and that the amount of drug resistance has been ever-increasing in countries as diverse as Peru, South Africa, China and India.(Corbett EL, Watt CJ, Walker N, *et al,*2003) An estimated 120,000 people die annually from multidrug-resistant TB.(WHO/IUATLD,2008)

In 1995 the DOTS (Directly Observed Treatment, Short course) strategy was established as the key intervention to achieve TB control worldwide. Between 1995 (when reliable records began) and 2006, a total of 31.8 million new and relapse cases, and 15.5 million new smear-positive cases were notified by DOTS programme (WHO, 2008). Later on, in 2006, WHO launched the stop TB strategy as an approach to TB care and control (WHO, 2013).

Ethiopia is one among the 27 high Multi Drug Resistant Tuberculosis (MDR-TB) burden countries and the rate of TB treatment success in the country was 83% in 2011 (WHO, 2006).

The research done on time to default in the south of Benin states that: it was challenging to know who was not susceptible and vice versa before the end of the follow up. Therefore it is

important to consider early the HIV/AIDS status and the age of patients as well as their TB history as underlying factors for compliance with anti-TB treatment in future. For instance, prospective research integrating the impact of the malnutrition, patients' immunological state and the genomics of TB may be interesting (Sylvère TA (2015)).

2.2 Determinant factors of time to cure of tuberculosis patients

Many studies suggested that tuberculosis is curable disease if it treated well. But there are many factors that determine the time of cure. Low awareness, low individual and social risk perception, high stigma, and external constraints (e.g. distance, transportation, economic limitations) account for delayed care-seeking and diagnosis. Also, conventional care-seeking behaviors (e.g. consulting traditional healers, pharmacists, and other private providers) also explain delay, particularly among women. Gender differences are crucial to understand diagnostic delay. Females in poor communities in developing countries are more likely to be less educated, be stigmatized for seeking TB care, and suffer more from external constraints (Ryan, 1993).Because only a few studies have actually correlated those variables, it is impossible to reach conclusions with general validity. Many factors that predict delay seem to be mutually associated. It is impossible to single out one factor that, regardless of other conditions, explains diagnosis delay and treatment default.

Moreover, even multivariate analyses that offer an explanation for TB detection and treatment adherence in one community do not provide conclusive results that are applicable to different settings. Although the literature has plenty of case studies with individual findings that can be tested, the dearth of comparative studies does not allow us to draw general conclusions. The inability of health systems to screen people before they develop active TB largely accounts for system delay. From the studies reviewed, a number of factors account for why health systems miss patients.

The lack of diagnostic tools in health clinics, bad implementation, overall under-utilization of health care services, and oversight by healthcare providers (due to insufficient knowledge and neglect) are plausible explanations. Once people come in contact with health clinics, then, the poor quality of services explains why a large number of patients are not diagnosed. Problems with the quality of services range from poor interpersonal communication of healthcare providers

to infrastructural deficits (Ryan, 1993).The survival status was significantly different between patient age, weight at initiation of anti-TB treatment, patient category, year of enrolment, and treatment centre(B. Getahun,*et al*,2011). In the MDR-TB treatment program patients with HIV infection and low weight had higher hazards of death. Overall treatment outcomes were poor. Efforts to improve treatment for MDR-TB are urgently needed (Jason E. Farley,*et.al* ,2011). Studies by (B. Getahun, *et al*, 2011, M. D. F. P. M. de Albuquerque,*et al*,2009, M. Vasantha, *et al,* 2008) also reported that body weight at initiation of treatment was a risk factor for death from TB and is associated with survival of patients who begin treatment for tuberculosis.

Types of tuberculosis (positive pulmonary, negative pulmonary, and extra pulmonary tuberculosis) were identified as the significant factors for mortality of tuberculosis patients (M. P. Domingos,*et al.*,2008, M. Muñoz-Sellart,*et al.*,2010, M. W. Borgdorff, *et al.,*1998). Also other literatures states that TB patient category (TBC) was statistically associated with death of patients with tuberculosis(M. D. F. P. M. de Albuquerque,*et al.*,2009, M. P. Domingos,*et al*,2008, M. Vasantha, *et a.l,* 2008, T. A. Mathew, T. N. Ovsyanikova, S. S. Shin *et al*.,2006)

TB often strikes the most vulnerable members of society and, if left untreated, causes its victims to lose weight, weaken, and eventually waste away (Ryan, 1993). TB disproportionately affects the indigent and other marginalized groups of society in whom unequal susceptibility patterns have long been recognized (Dubos, R and Dubos,J, 1952).

Age between 30-39 years compared to 40 years and above and being farmers were a negatively risk factors. Living in the same house with 5 or more family members, HIV positive, having history of previous self & family member TB infection and missing TB drug were also the risk factors for TB treatment failure (Girmatsion Fisseha *et al.,* 2014).

After the introduction of directly observed therapy (DOT), the numbers of people being cured from TB are increasing. But millions will remain ill because they lack access to high-quality care (UN, 2010). And the treatment success rate among smear positive pulmonary TB reached 87% globally in 2009, but in the Africa region it is 80% which is lower than WHO target. Among the 22 high burden countries 15 of the reached the target and Ethiopia is one of the 7 countries from which treatment successes rate lower than the target (WHO, 2010).

Even though, Ethiopia adopted and fully implemented the DOT, TB continues to be a major public health problem which puts the countries at 7thrank from the 22 TB high burden countries by estimated number of case (WHO, 2009). In 2009 TB is the 5th cause of admission, the 3rd leading cause of death with 67%cure rate for registered patient in the same year in Ethiopia (FMOH, 2010). In addition to these, for all new TB cases, 20% of them have a chance to be multidrug-resistant (MDR) TB (USAID, 2009).

2.3 Mixture cure models

'Cure models' are part of the mixture models family (Bohning& Seidel, 2003). In 'mixture cure models', it is considered that the studied population is a mixture of susceptible (i.e. that may undergo the event of interest) and non-susceptible individuals (i.e. that will never undergo the event of interest) (Farewell, 1982). Unlike classical survival analysis, they allow a separate estimation of covariate effects on incidence and incubation length. Cure models also allow estimation of the proportion of healthy (or conversely infected) individuals in a population, including individuals that did not last the total length of the study (Lam et al., 2005).

The motivation behind mixture cure modeling is the desire to address the situations in which there are cured proportions of individuals and the resulting consequence that those individuals will never experience the event of interest (Taylor, 2000). This led to exploration into cure rate estimation and development of the first mixture cure models by (Boag, 1949, Berkson and Gage, 1952, and Haybittle, 1965) from those initially developed models, various studies have proposed and assessed parametric and semi-parametric mixture cure models. Several authors have studied the parametric approach to mixture cure models (Farewell 1982, Peng et al.,1998, Yamaguchi K,1992) however, semi-parametric models, are often of greater interest than parametric models since the parametric assumption can be hard to meet. Therefore, many studies more recently have explored modeling and estimation with semi-parametric mixture cure models. (Li, C., and J. Taylor, 2002, Peng, Y., 2003,Sy J.P., and J.M.G. Taylor, 2000, Taylor, J, 1995, Zhang, J., and Y. Peng, 2007)

The modeling strategy for the mixture cure model involves separately modeling the cure proportion and the survival distribution of the uncured patients. In recent years the development of new drugs and treatment regimens has resulted in patients living longer with diseases such as

cancer and heart disease. In cohorts of patients with certain types of cancer, it is observed that some patients are cured permanently, i.e. show no recurrence of the disease. The patients who are cured are called immunes or long-term survivors, while the remaining patients who develop a recurrence of the disease are termed susceptible. The population of interest may thus be regarded as a mixture of these two types of patients. Standard parametric and non-parametric survival models are in appropriate for analyzing such data because they ignore the distinction between the immunes and susceptible in the population (Boag, J. W, 1949) proposed a two-component mixture model for analyzing breast cancer data. The model he proposed is referred to as the \cure rate" model, and is formulated in terms of a mixture model.

The model introduces a component representing the proportion of immunes in the population and a distribution representing the survival experience of the susceptible, called the latency distribution. Many researcher searches in this area include articles by (Yu etal., 2004, Farewell, 1982, Gamel et al., 1999, Yamaguchi k, 1992, and Taylor, 1995). Various parametric and semi-parametric approaches have been proposed for mixture cure models (Peng&Carriere, 2002; Lam et al., 2005). For modeling the influence of exploratory variables on the incidence, a logistic regression model is usually chosen (Kuk& Chen, 1992; Peng& Dear, 2000):

(Yu et al, 2004) used the mixture cure rate model for grouped survival data and observed that the estimate of the cure fraction can be quite sensitive to the length of follow up time and the choice of latency distribution. They investigated the effect of various distributions such as the lognormal, loglogisitc, Weibull and generalized gamma, and conclude that the estimate of the cure fraction was robust with the generalized gamma distribution. (Yu et al, 2004) also investigated the identifiablity of mixture models, and noted that the overall survival function and the survival function for the latency distribution can become unidentifiable if the follow up time is short. They suggest that a longer follow up time with respect to the median survival time and homogeneity of the observations affect the accuracy of the estimate of the cure fraction.

(Yamaguchi k, 1992) proposed an accelerated failure-time regression model with an additional regression model for the cure fraction to study inter-firm job mobility in Japan. He used the generalized gamma to model the latency distribution and the logistic function to model the cure

fraction in terms of covariates. This model helps to estimate simultaneously the effect of covariates on the acceleration (deceleration) of an event as well as the surviving fraction.

(Farewell, 1982) also used covariates to model the cure fraction. (Chen et al., 1999) proposed a new Bayesian model for survival data with a surviving fraction. This model has a proportional hazards structure, with the cure rate depending naturally on covariates. One key difference between the Bayesian approach and the cure rate approach is that the former models the entire population as a proportional hazards model while the later models only the non-cured group with a proportional hazards model. Both the models can be obtained from one another. The authors suggest that the Bayesian model is computationally attractive.

In survival analysis, it is usually assumed that if complete follow-up were possible for all individuals, each would eventually experience the event of interest. Sometimes' however, the data come from a population where a substantial proportion of the individuals do not experience the event at the end of the observation period. In some situations, some of these survivors are actually cured in the sense that, even after an extended follow-up, no further events are observed. An example is patients with tonsil cancer treated using radiation therapy (Withers et al., 1995), in which cure occurs if the radiation kills all the cancer cells.

For such data, Kaplan-Meier (K-M) estimates of the survival function for time to local recurrence level off to nonzero proportions. Furthermore, there is a time window within which most or all of the recurrences are expected to occur, and so a patient without any recurrence beyond this window can usually be considered **as** being cured. This type of data is typical of diseases where the biology of the disease suggests the possibility of cure. A K-M survival curve that shows a long and stable plateau with heavy censoring at the tail may be taken **as** empirical evidence of a cured fraction. The use of standard survival analysis for such data may be inappropriate since not all the individuals are susceptible.

In a cure model, the population is a mixture of susceptible and non-susceptible (cured) individuals. Various parametric and nonparametric methods have been proposed for the cure model. (Farewell, 1982) used logistic regression for the mixture proportion and the logistic Cox PH mixture cure model was used to account for the cure fraction of the sample. It assumes a binary distribution to model the incidence probability and a parametric failure time distribution

to model the latency (Kuk and Chen, 1992) and proposed a semi-parametric generalization of Farewell's model using a Cox proportional hazard (PH) model in the susceptible group; we call this the PH cure model. They used an estimation method involving Monte Carlo simulation. For modeling the influence of exploratory variables on the incidence, a logistic regression model is usually chosen (Kuk& Chen, 1992; Peng& Dear, 2000)

The cure model should not be used indiscriminately (Farewell, 1986). There must be good empirical and biological evidence of a non-susceptible population. The model generally requires long-term follow-up and large samples and censoring from loss to follow-up during the period when events can occur must not be excessive. Otherwise, identifiability problems between the incidence and latency parameters can occur. In medical and epidemiological studies, some individuals in the population will not experience the event of interest. i.e., those individuals are cured (immune).

Historically, the univariate mixture cure models (MCMs) (Boag, 1949 and Farewell, 1982) have been used to model the survival data with a cure fraction. Additionally, MCMs allow that both the cure fraction and the survival function of the uncured patients (latency distribution) depend on covariates. In our case these covariates are age of patient, weight of patient, HIV status of patient, regiment, TB type, and gender of the patients.

CHAPTER THREE

3. METHODOLOGY

3.1 Study area and Period

For this study, longitudinal retrospective cohort follow up (retrospective cohort design) of tuberculosis patients data obtained from Jimma University Specialized Hospital TB Patient Clinic located in jimma town. The follow-up time was the time span in month between the treatment initiation date and the date the treatment final outcome was evaluated. Treatment initiation date was between 1st July, 2014 and 1st January, 2016 when each patient started the anti-TB therapy after diagnosis.

3.2. Data source

All information are obtained from Jimma University Specialized Hospital and Shanen Gibe Jimma hospital and the data extracted from the patient's chart which contains epidemiological, laboratory and clinical information of all TB patients under directly observed treatment. Those follow their treatment at that hospital up to finish their treatment were extracted and those MDR-TB group are transferred to shanen Gibe Jimma hospital after they were tested and confirmed by laboratory. Then those are collected from this hospital in order to get the relevant information.

3.3. Study Population

TB Case

TB Case was a patient in whom TB has been bacteriological confirmed or diagnosed by a clinician (Normann B, et al., 2000). TB history and treatment outcomes were classified using the standard World health organization (WHO) definitions (Corbière F, Joly P ,2007). As with HIV/AIDS treatment, TB therapy requires high (> 90%) compliance to facilitate cure 50. TB history was divided into new, treatment after relapse, treatment after failure, treatment after default or transfer in. Final TB treatment outcome was one of the following: successful, failure, death, default and transferred out. A successful treatment outcome included "cure", or "treatment complete".

The end point

Time to cure was the end point. Cure is the success of the treatment for the tuberculosis patients. Time to cure was defined using the WHO guidelines for TB surveillance strategy. Accordingly, a cure was a patient whose treatment was success with in6-12 months or more. Thus, time to cure status was coded 1 if given TB patients "cured from TB" and 0 if a TB patient treatment outcome was set to "not cured" for both group of populations those susceptible and non-susceptible(MDR-TB in our cases).

Generally there are two group of population. These are who develop MDR-TB and who are free from bacteria with in short period of time. So those who develop MDR-TB are considered as cure (immune) since they are treated a minimum of two year it possible to say they never experience the event of interest. And those susceptible groups were uncured in this study because they are free from the bacteria within 6-12 month. These were identified based on the patient's regiment. Based on this way the total number of patients within this period of follow up were **501** patients among those **62** of them are MDR-TB group.

3.4. Variables in the Study

The variable that was included in these studies are response and the predictor or explanatory variables that can be directly or indirectly affect the outcome of the patients.

3.4.1. Dependent variable

The response (outcome) variable in this study is time to cure of tuberculosis patients, which is the probability of cure from tuberculosis. These response variables are the results of the patients though their follow up by the clinicians directly. These response were cured, treatment complete, and death.

3.4.2. Independent (Predictor) Variables

Since several variables were expected to relate with determinants of time to cure of tuberculosis, the predictor variables in this study are the possible socio-economic, demographic and environmental factors. This study considers only the predictors that clinicians and doctors were record while they follow the patient. These predictors are specifically:-

4. 1: *list of explanatory variables*

Predictors	Categories
Age in year	
Weight in kilograms at the time of admission	
Gender of the patient	0=female
	1=male
Types of TB	1=pulmonary positive
	2=pulmonary negative
	3=Extra pulmonary TB
HIV co- infection	0=HIV negative
	1=HIV positive
Regiment	0=new patient
	1=previously treated
Work place	0=non health worker
	1=health worker

3.5. Methods of Data Analysis

3.5.1 Survival Analysis

The use of survival analysis, as opposed to the use of other statistical method, is most important when some subjects are lost to follow up or when the period of observation is finite certain patients may not experience the event of interest over the study period. In this latter case one cannot have complete information for such individuals. These incomplete observations are referred to as being censored. Most survival analyses consider a key analytical problem of censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly. In reality such event can occur due to the following reasons:

A person does not experience the event before the study ends

A person is lost to follow-up during the study period and

A person withdraws from the study for unknown/known reasons

Censoring can be occurred in the following way:

i) Right censoring: Survival time is said to be right censored when an individual cannot experience an event up to the end of the study time. This type of censoring is commonly recognized survival analysis and also considered in this study as cured.

ii) Left censoring: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study.

iii) Interval censoring: Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known. The prospect of censoring complicates research design and complicates statistical analysis. Thus, censoring creates some unusual problem in the analysis of data because such data cannot be handled properly by standard statistical methods. Researchers used different techniques to respond to the complication due to censoring unsatisfactorily. New developments in statistical theory accompanied by new development in statistical computing have changed how researchers can study such data. The response variable in survival analysis is survival time and is no longer limited to only time to death. It is a non-negative random variable used loosely for the time period from a starting time point to the occurrence of any event.

In this study context, survival time will the length of time of tuberculosis patients.

Models for survival analysis typically assume that everybody in the study population is susceptible to the event of interest and will eventually experience this event if the follow up is sufficiently long. In recent years, there has been an increasing interest in modeling survival data with long term survivors. Such data may arise from clinical trials, in which, even after an extended follow-up, no further events of interest are observed. Some people in the population may be considered as cured or non-susceptible. Failing to account for such cured subjects would lead to incorrect inferences.

Moreover, researchers may be interested in estimating the cured fraction. Differences between subcategories were compared using the chi-square statistic and Fisher's exact test was used if expected value of a cell was less than 5%. The cured fraction was computed by summing the simulated individual posterior probability of being cured (Allison PD, 1995). In univariate analysis, cured time was analyzed with the Kaplan–Meier estimator (KME) (Kruk ME, 2008, Nguyen PT, 2010, and Kuk AYC, Chen C-H, 1992) and the log-rank test was applied to compare the cure time function between subcategories.

Mixture cure model:-is firstly introduced by Boag (1949) and Berkson and Gage (1952), is one of the popular models to estimate the cure rate of the treatment and the survival rate of uncure patients at the same time.

Mixture cure models assume that the studied populations are a mixture of susceptible (uncured) individuals, that may experience the event of interest, and non-susceptible (cured) individuals that will never experience it (Farewell, 1982). Therefore in this study those TB patients free from the bacteria in the study period is susceptible and those develop default is non-susceptible (MDR-TB). This approach allows estimating simultaneously whether the event of interest will occur, which is called incidence, and when it will occur, given that it can occur, which is called latency.

3.5.2 Survival functions

3.5.2.1 Univariate analysis

3.5.2.1.1. The Kaplan-Meier Product Limit Method

The Kaplan-Meier (KM) estimator is the standard non parametric estimator of the survival function, S (t), proposed by Kaplan and Meier (1958). It is also called the Product-Limit estimator. KM 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. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included.

Let ordered survival times are given by 0≤ t1≤t2 ≤tj ≤ ∞, then:

$$\hat{S}(t) = \begin{cases} 1, & if t < t1 \\ \prod_{j, tj \le t} \left(1 - \frac{d_j}{r_j}\right), & if t \ge t1 \end{cases} \text{ ---------- (1)}$$

Where; d_j is the observed number of events at time t_j and r_j is the number of individuals at riskat timet_j.

3.5.2.2 Multivariate analysis

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.

3.5.2.2.1 Standard Cox regression

Standard Cox regression was fitted under the PH assumption. That is:

$h(t; X1, X2, \ldots, Xk) = \lambda 0\,(t) exp(\lambda 1 X1 + \lambda 2\, X2 + \ldots + \lambda k Xk)$ ---- (2). Where $\lambda 0(t)$ is the baseline hazard of patient cured at time t and x1, xk are the k independent covariates. Here, t is the time to cure. Here the model can consider only the patients cured or complete the treatment as event and the rest as censored.

Assumptions of the Cox proportional hazards model are;

- ✓ The ratio of the hazard function for two individuals with different sets of covariates does not depend on time.
- ✓ Time is measured on a continuous scale and
- ✓ Censoring occurs randomly.

Use of the standard Cox PH model is linked to the hypothesis that, if complete follow-up were possible for all patients, each would have eventually experienced the cure from treatment. This hypothesis, however, did not hold for all data set.

3.4.2.2.2. Proportional Hazards Mixture Model

The PH assumption provides a way to introduce covariates into models and to separate the effect of the covariates and the shape of a baseline hazard function.

The model can be formulated as follows the survival model is assumed that the failure time can be decomposed

$$T^{*} = \eta T^{S} + (1 - \eta)\infty \text{ ---------- } 3$$

Where $T^{S} < \infty$ denotes the failure time of susceptible subject, and η is indicated by value 1 or 0 whether the sampled subject is susceptible or not.

If we assume that the proportion of susceptible $p(\eta = 1) = \pi$ where $\pi \epsilon (0,1]$, then the distribution function of T^{*} is given by

$$F(x) = P(T^{S} \leq x)p(\eta = 1) + P(\infty \leq x)p(\eta = 0) = \pi F_{0}(x) \text{ -------- } 4$$

Where $F_0(.)$ is the latent distribution for T^S (no-cure /susceptible) group. The effects of some independent covariates on both the incidence probability P and the survival function $S_o(t)$ for the susceptible group can be modeled. Therefore

The incidence model can be typically given by

$$\pi(x) = \frac{\exp(x^T\gamma 1)}{1 + \exp(x^T\gamma 1)} \text{ ---------- 5}$$

Where x is a vector for covariates $\gamma 1$ is a parameter to be estimated T denotes transpose.

Baseline survival distributions are assumed to follow parametric distribution

It has been successfully employed in Cox's PH regression model for survival data.

To model the effect of covariates x, on the failure time distribution of uncured patients in the mixture model, we also employ the PH assumption, i.e., the effect of **x,** on the Cox PH hazard function distribution for T^* is modeled by

$$h_u(ti \mid xi,) = huo(ti)exp(\beta' xi) \text{ ---------- (6)}$$

Where$h_u o(t) = \frac{\pi f_o(t)}{1 - \pi F_o(t)}$is the baseline hazard function, which can be any arbitrary unspecified hazard function but not a function of x, this assumption implies

$Su(ti|\,\boldsymbol{xi}) = SuO(ti)^{exp(\beta' xi)}$,where$Suo(t) = exp\{-\int_0^t huo(w)dw\}$. Then equation (6) can be maximized after the baseline hazard $huo(t)$is specified.

3.5.2.2.3. The logistic Cox PH mixture cure model

The logistic Cox PH mixture cure model was used to account for the cure fraction of the sample. It assumes a binary distribution to model the incidence probability and a parametric failure time distribution to model the latency.

Moreover, the distribution of survival time is characterized by three functions:

Let *U* be the indicator denoting an individual is susceptible (U = 1) or non-susceptible (U = 0) to the event of interest and T is a non-negative random variable denoting the failure time of interest, defined only when U = 1. The mixture cure model is given by

$$S(t \mid \boldsymbol{x}, \boldsymbol{z}) = \pi(\boldsymbol{z})S(t \mid U = 1, \boldsymbol{x}) + 1 - \pi(\boldsymbol{z}) \text{ ---------- (7)}$$

where$S(t|\boldsymbol{x}, \boldsymbol{z})$ is the unconditional survival function of *T* for the entire populations.

This can be written in words as follow

Usually, $\pi(\boldsymbol{z})$is refer to as “incidence" and $S(t|\boldsymbol{U}=\mathbf{1},\boldsymbol{x})$ is refer to as “latency If the PH model is used to model the latency part, the cure model is called the PH mixture cure model.”.

$$S(t \mid U = 1, \boldsymbol{x}) = P(T > t \mid U = 1, \boldsymbol{x}) \text{ ---------- (8)}$$

The survival functions for susceptible individuals given a covariate vector$\boldsymbol{x} = (x1, \ldots, xp)'$, and $\pi(\boldsymbol{z}) = P(U = 1|\boldsymbol{z})$ is the probabilityof being susceptible given a covariate vector $\boldsymbol{z} = (z1, \ldots, zq)'$,Which may include the same covariates as **x**. Starting with the incidence portion of the model, the effects of the covariate vector **z** on the cure proportion is typically modeled using a $logit$ link function

Where **b** is a vector of unknown parameters associated with the covariate vector **z.**

The latency portion of the model can be defined to be the proportional hazards (PH) model or the accelerated failure time (AFT) model. Let $S_o(t)$ be thebaseline survival function for uncured (susceptible) individuals. When$S(t \mid U = 1, \boldsymbol{x}) = s_o(t)^{e^{\beta x}}$, the proportional hazards mixture cure (PHMC) model is selected and when$S(t \mid U = 1, \boldsymbol{x}) = s_o(te^{\beta x})$, the accelerated failure time mixture cure (AFTMC) model is selected.

Characteristics of survival function

1. The survival function of cured patients can be set to one for all finite values of *t* because they will never experience the event of interest. $S(t|\boldsymbol{x}, \boldsymbol{z}) \rightarrow 1 - \pi(\boldsymbol{z})\ ast \rightarrow \infty$.
   When $\pi(\boldsymbol{zi}) = 1$ for all $\boldsymbol{zi}$,i.e. when there is no cured fraction, the mixture cures modelreduces to the standard survival model.
2. There should be nonzero probability of cure.
3. The population being studied is the mixture of two subpopulations. I.e. the proportion π cured and 1-π uncured. As modeled in equation (5) above.

There are many parametric and nonparametric models are proposed to analyze this type of distributions.

3.5.3. Computational methods and theory

3.5.3.1. Parametric and semi-parametric mixture cure models

The Cox PH model is a semi-parametric model where the baseline hazard$\beta_0(t)$is allowed to vary with time.

The effect of **z** on the probability of $\pi(\mathbf{z})$ can be modeled bythe use of binary regression models, with $logit$ link

$$logit(\pi(\boldsymbol{z})) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_q Z_q = \boldsymbol{\beta}' \boldsymbol{Z} \text{ ---------- } (\mathbf{9})$$

Whereβ_0 is the intercept and $\beta$is the vector of regressionparameters associated to **z**. Other regression models include the $probit$ link

$$\phi^{-1}(\pi(\boldsymbol{z})) = \boldsymbol{\beta}' \boldsymbol{Z} \text{ ---------- } (\mathbf{10})$$

Where$\phi$is the distribution function of a standard normal distribution, and the complementary log–log link.

$$log(-log(1-\pi(\boldsymbol{z}))) = \boldsymbol{\beta}' \boldsymbol{Z} \text{ ---------- } (11)$$

The conditional latency distribution $S(t|U=1)$ can take theform of parametric or semi-parametric distributions. Amongthe parametric models, exponential (EXP), $Weibull$ (WB), lognormal (LN) and $loglogistic$(LG) are commonly used to model survival data. After $etrization$, these distributionscan be expressed as

$$S(t \mid U = 1) = \begin{cases} \exp[-\exp(\log t - \mu)], \text{exponential} \\ \exp\left[-\exp\left(\frac{logt - \mu}{\sigma}\right)\right], Weibull \\ 1 - \phi\left(\frac{logt - \mu}{\sigma}\right), lognormal \\ [1 + \exp(\frac{logt - \mu}{\sigma})]^{-1}, loglogistic \end{cases} \text{ ---------- } (12)$$

Covariates can be included by parameterizing μsuch as$\mu = \gamma^{'}\mathbf{x}$, where γ represents the vector of unknown regression parameters. These models are also known as parametric accelerated failure time (AFT) mixture cure models. Since **x** acts multiplicatively on the scale parameter μ itaccelerates or decelerates the failure time of susceptible individuals.

In proportional hazards (PH) models, the conditional distribution of T is modeled by$S$

$$(t \mid U = 1, \boldsymbol{x}) = s_o(t \mid u = 1)^{\exp\left(\gamma^{'}x\right)}$$

$$= \exp\left(- \exp\left(\gamma^{'}x\right)\int_0^t \lambda_o(v \mid u = 1)dv\right) \text{---------- (13)}$$

Where$s_o(t|u = 1)$ and $\lambda_o(v|u = 1)$ are the baseline conditionalsurvival and hazard functions, respectively. The conditionalcumulative hazard function is

$$\lambda(t \mid u = 1) = \lambda_o(t \mid u = 1)\exp\left(\gamma^{'}x\right) \text{---------- (14)}$$

Where$\lambda_o(t \mid u = 1) = \int_0^t \lambda_o(v|u = 1)dv$

If $s_o(t|u = 1)$is left arbitrary, the model is defined as the Cox's proportionalhazards mixture cure model. Note that the Weibull (and exponential) models are both AFT and PH models.

Through the vectors of regression parameters β andγ parametric and semi-parametric mixture cure models are able to separate the covariate effects on the incidence and the latency.

3.5.3.2. Likelihood

Suppose data are form of $(ti, \delta i, \boldsymbol{x}i, \boldsymbol{z}i), i = 1, \ldots, n,$ where δi is the censoring indicator with $\delta i = 1$ if ti is uncensored and $\delta i = 0$ otherwise. The likelihood contribution for individual i is

$$\pi_i(zi)f(ti \mid U = 1,xi),for\delta i = 1\ and\left(1 - \pi_i(zi)\right) + \pi_i(zi)S(ti \mid U = 1,xi),for\delta i = 0$$

Where$f(\cdot) = S(\cdot)\lambda(\cdot)$is the conditional probability densityfunction of T

The observed full likelihood is then given by

$$L(\gamma,\beta) = \prod_{I=1}^{n}[\pi_i(zi)f(ti \mid U = 1,xi)]^{\delta i} * \left[\left(1 - \pi_i(zi)\right) + \pi_i(zi)S(ti \mid U = 1,xi)\right]^{1 - \delta i} \text{-- (15)}$$

When no cured fraction is assumed, i.e.$\pi(\boldsymbol{z}i) = 1$ for all$\boldsymbol{z}i$, the likelihood function (14) reduces to the likelihood of the standard survival model.

3.5.3.3. Estimation procedures

The procedures that will be applied to estimate the unknown parameters were based on maximization of the likelihood functions that exists on equation (15).

To estimate$\boldsymbol{\gamma}$, we must specify the failure time distributionof uncured subjects. For parametric mixture cure models, $f(\cdot|U=1)$ and $S(\cdot|U=1)$ can be defined by a few unknownparameters in (14). Therefore, maximum likelihood estimates are obtained via usual optimization methods as the Newton–Raphson method. Asymptotic standard errors are obtained by inverting the Fisher's information matrix of second order derivatives of log (L).

Unlike in the standard Cox's proportional hazard, where little information is lost by eliminating$S0(t)$, one cannot eliminate$S0(t|U=1)$ in the Cox' PH mixture cure model without losing information aboutβ. The EM algorithm provides a simpleand efficientway to estimate separately β,γ and $S0(t\,|\,U=1,\boldsymbol{x}i)$. from the introduction, U is the random variable denoting an individual is susceptible (U = 1) or non-susceptible (U = 0).

It follows that, if $\delta i=1$then$ui=1$, and if $\delta i=0$ then $ui$is not observed, where u_i is the value taken by the random variableUi. Given the $ui's$, the complete-data full log–likelihood (15) is the sum of two independent components, l_I, which depends only on β, and l_S, which depends only on γ and $\lambda_O$where

$$l_I(\beta,U)=log\prod_{i=1}^{n}\pi(zi)^{ui}*[1-\pi(zi)]^{1-ui} \text{ ---------- (16)}$$

$$l_S(\gamma,\lambda_O,U)=log\prod_{i=1}^{n}\lambda(t|U=1,xi)^{\delta iui}*S(t\,|\,U=1,xi)^{ui} \text{ ------- (17)}$$

where U is the vector of ui value.

The EM algorithm starts with initial values$\beta^{(0)}$, $\gamma^{(0)}$and$S_0{}^0(t|U=1)$. The *E* step in the $(r)t$h iteration calculates the expectation of the complete log–likelihood function with respect to **u**,

conditional on the observed data and$\beta^{(r)}$, $\gamma^{(r)}$and$S_0{}^{(r)}(t|U=1)$, the estimates of $\beta,\gamma,and\ S(t|U=1)$ at the r^{th} iteration. This is given by the following conditional expectation

$$yi^{(r)} = E\left\{\beta^{(r)},\gamma^{(r)},S_0{}^{(r)}(t|U=1,xi)\right\}$$

$$= \delta i + (1 - \delta i) * \frac{\pi^{(r)}(zi)S^{(r)}(t \mid U=1,xi)}{1 - \pi^{(r)}(zi) + \pi^{(r)}(zi)S^{(r)}(t \mid U=1,xi)} \text{ ------- } (18)$$

which is the r^{th}estimator of the probability of the i^{th}individualbeing susceptible. Given he*M* stepyi^{r} in the$(r+1)^{th}$ iteration maximizes the expected complete log–likelihood function with respect to $\beta$and γ to obtainβ^{r+1}, $\gamma^{r+1}$and$so^{r+1}(t|u=1)$. The algorithm iterates until convergence on estimates of β, γ and $S_o(t|u=1)$.

As we see from equation (16) above conditional baseline is accommodated in. so estimating the conditional baseline survival function in semi parametric models are possible. Accordingly two non-parametric methods are discussed by peng and Dear, andSy and Taylor, in 2000. The two methods are:-

- Profile likelihood estimate of$S_o(t|u=1)$, similar to theBreslow's likelihood for the standard Cox's PH model.
- product limit estimator (PLE)after Kalbfleisch and Prentice

In order to obtain good estimators for γ and $\boldsymbol{\beta}$, it is important for $S0(t|U=1)$ to approach 0 as t→∞. However, the estimates from the Breslow or the PLE methods do not approach zero as t→∞ when there are censored survival times after tk, where tk is the last observed failure time. Setting$S0(t|U=1)=0$ for all $t \geq tk$ allows for a proper distribution function for susceptible individuals and avoid identifiability problems (Taylor, 1995 and Li*, et.al.,* 2001)

3.5.4. Model selection

3.5.4.1. TheAkaike information criterion

From the list of the parametric and semi parametric models that is assumed for the baseline distribution for the latency parts there should be the criterions to select the best fit. Therefore the AIC is the advisable criterion for the mixture cure models.

For estimation of mixture cure models, Cai et al. (2012) explain the use of the expectation maximization (EM) algorithm to deal with the latent *Y* values. The log-likelihood for the data triplets ($Ti;\delta i;\ Ui$) could be used in the AIC.

$$AIC = \ - logL_{Ti,\delta i,Ui}(\hat{\Theta},Ti;\delta i;\ U) + 2d \text{ ---------- } (19)$$

Where d counts the number of parameters in the model, and $\hat{\theta}$is the maximum likelihood estimator of the parameter vector. Variable selection was performed using stepwise backward elimination with a significance threshold of 5%. For backward elimination selection, a full model was initially fitted with all candidate covariates included. The covariates were then tested for statistical significance and the largest p-values were removed one by one until the remaining covariates had p-values of less than 5%.

3.5.5. Model diagnosis

3.5.5.1. Evaluation of the semi-Parametric Baselines

For the mixture cure model the baseline distribution is diagnosed by using graphical methods.

The Kaplan-Meier (KM) plot:-is the plot of the survival function estimate versus time. The KM survival function estimate levels off at the right tail and exhibits a long and stable plateau, which ensures the applicability of the mixture cure model approach.

If the KM survival estimates were plotted against the model-based survival estimates, a well-calibrated model would produce estimates that fall on a 45^0 line through the origin.

3.5.5.2. Schoenfeld residuals

There are defined at each observed failure time as:

$$r^{s}_{ij} = Z_{ij}(t_i) - \overline{Z_j}(t_i) \text{ --------- } ---------- (20)$$

Where $r_{ij}$is the i^{th}observation under j^{th}cofactor and $t_i$is the observed failure time

- Represent the difference between the observed covariate and the average over the risk set at that time
- Calculated for each covariate
- Useful for assessing time trend or lack or proportionality based on plotting versus event time

CHAPTER FOUR

4. STATISTICAL ANALYSIS AND RESULTS

4.1. Descriptive summaries.

Of all 501 tuberculosis patients 439(87.62%) susceptible group 328(74.15%) were cured (Table 4.1). The median cure time from TB was 6 months and 62(12.37%) non susceptible those multi-drug resistance TB none of them cured with in the follow up period. proportion of male (65.45%) and female (34.53%) were involved of which 38.7% of females and 61.29 % of males were non susceptible group were none of them cured within given time and 66.06% of males and 33.94% of females were susceptible group were 75.2% of males and 73.8% of females was cured. The median cure time of susceptible group was 6 months.

From those the total patients 10.2% of them are HIV positive while 89.8% of them were none infected. From the co-infected 11.29% of them was the non-susceptible group while 88.7% of them were from susceptible group. In general there are two population susceptible and non-susceptible or multidrug resistances TB from the summaries of table 4.1 below the most of the MDR-TB patients are pulmonary negative 55 of them.

From these patients of TB type pulmonary negative was risk to develop MDR-TB. And in the case of HIV co-Infection most of the co-infection was from MDR-TB group. The proportion of the HIV co-infection in non-susceptible group was 11.3% while the proportion of susceptible group was 10%. The median age 24 year ranged between 1- 90 year. And mean age was 27year. The median weight and mean weight of the patients were 52kg and 50.58kg respectively with range 6-89kg.From table 4.1 below time to cure of MDR-T B is more than 18 months and the median time to cure for the susceptible patients is 6 months, but from confidence interval even the median survival time is equal under each factor, regiment, smear test result and work place have wide confidence interval comparatively. This imply under thesefactor there are variability on the survival time of the disease it elongate or decrease the survival time.Pulmonary positive TB increase the time of treatment by one month compared to the other two. Unlike pulmonary positiveTB previously treated and working in health center decrease the treatment by one month.

4.2：Time to cure analysis of tuberculosis patients based on the categories of the predictors.

Characteristics		No. Patients	Non -susceptible	Cured	Median	Standard error	95%CI of median
Sex	Male	328	38	218(75.2%)	6	.04339	[6 , 6]
	Female	173	24	110(73.8%)	6	.0576628	[6, 6]
HIV co-infection	Infected	51	7	31(70.45%)	6	.1234161	[6, 6]
	Not infected	450	55	297(75.2%)	6	.0358633	[6, 6]
Smear. Test	Pulmonary +ve	195	1	145(74.7%)	6	.0644125	[6, 7]
	Pulmonary–Ve	170	55	83(72.2%)	6	.0440645	[6, 6]
	Extra pulmonary	136	6	100(76.9%)	6	.0638126	[6, 6]
Regiment	New	441	61	282(74.2%)	6	.0346754	[6 ,6]
	Treated	60	1	46(77.97%)	6	.1755743	[5, 6]
Work place	Health worker	53	9	31(70.45%)	6	.1755743	[5 , 6]
	Non health worker	448	53	297(75.2%)	6	.0346754	[6, 6]

4.1.1. Test of equalities of probabilities across the different groups of categorical variables.

Before proceeding to more complicated models, we make a descriptive analysis that will use as initiation to our subsequent findings. Here we start with the test of equality of probabilities across the different groups of a categorical variable using Log Rank test. The null hypothesis to be tested is that there is no difference between the probabilities of an event occurring at any time point for each population. The Stata and R results have been summarized in Table 3 below. The table shows that the different groups of work place of patient, regiment of patient, sex and HIV status is statistically not equal in experiencing the cure event, whereas smear test result is statistically the same in experiencing the event cure. The Log–Rank test results suggests that smear test result is significant covariate whose different levels have an impact in the survival longevity of TB patients; while work place of patient, regiment of patient, sex and HIV status does not have an impact.

4. 3:Log rank test for equality of survival experience among the different groups of covariates

Factor	*Chi-square*	*DF*	*P-value*
Regiment	3.5	1	0.0618
Smear test result	55.9	2	7.36e-13
HIV test result	2.5	1	0.111
Work place	1.3	1	0.252
Sex	1.5	1	0.223

This indicate there is the significance difference between the categories of smear test result with p-value=7.36e-13 on the time to cure of tubeculosis patients.

4.2. Graphical presentation of data

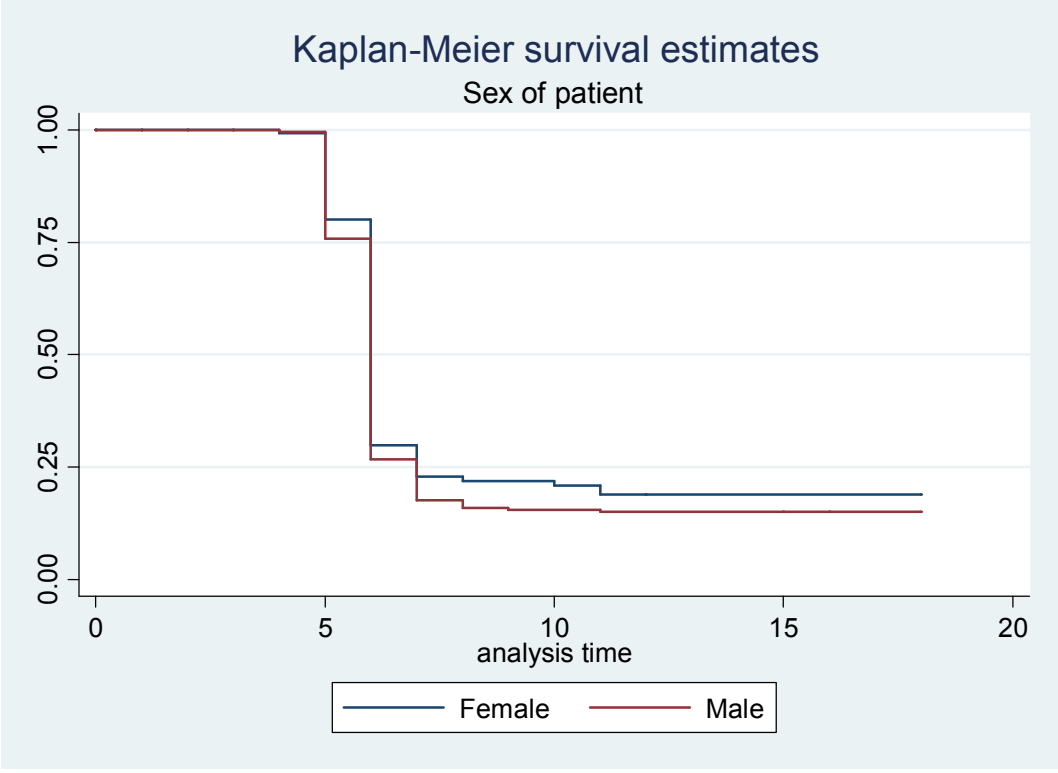

Figure 4. 1: ***Kaplan Meier survival plot by covariate sex for modeling time to cure of tuberculosis patientsat Jimma University specialized hospital.***

Of the group of patient based on their gender the survival time of the diseases is greater for the female than for the male even if there is no so much gaps between them as supported by the logrank test indicates there is no difference on time to cure of the patient based on the sex categorie (p=0.223).

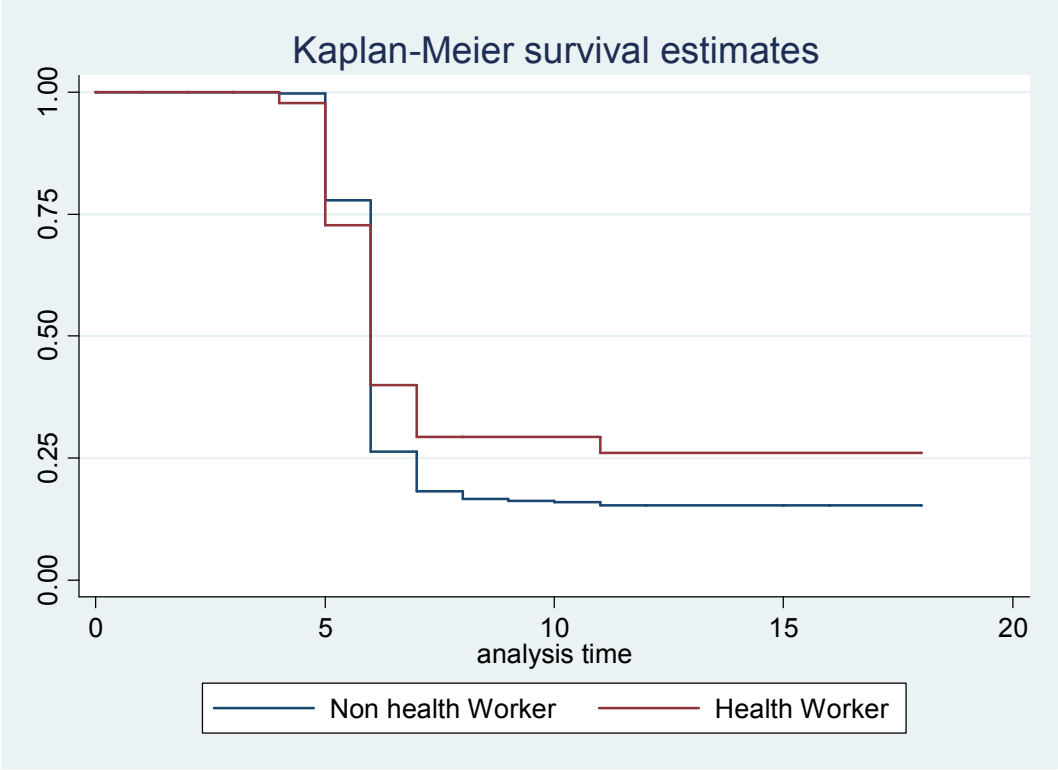

Figure 4. 2:*Kaplan Meier curve for the covariate working place for modeling time to cure of tuberculosis patients at Jimma University specialized hospital.*

Figure 4.2 indicates that those patients whose work was not related with health centre were low survival time with diseases than those health workers. Working at health centre is risk for TB if he\she may affected.Directly the curve is not easy to say there is a difference at all. Because,the curve is above and below based on time this is supported by logrank test (p=0.252) work place was no significant effect on the time to cure of the patient.

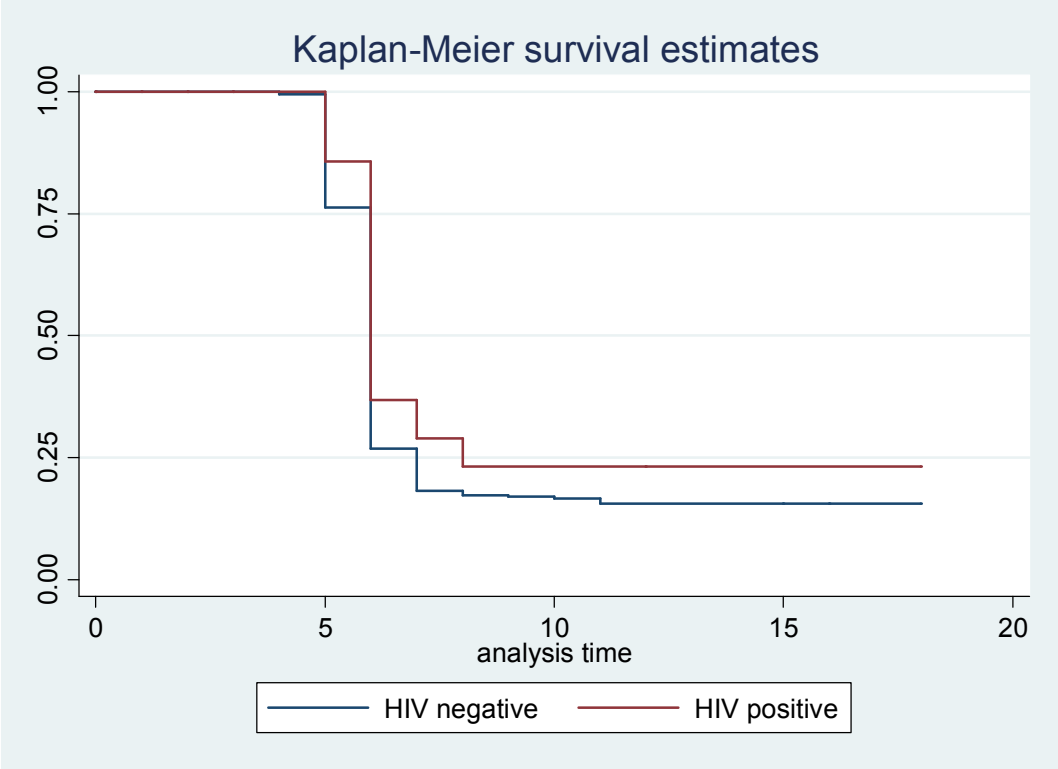

Figure 4. 3:*Kaplan Meier curve by covariate HIV .test for modeling time to cure of tuberculosis patientsat Jimma university specialized hospital*

Based on figure 4.3 it seems patients of hiv co-infected were long surviving with diseases compared to those not ifected. This implies the patients co-infected were need long period of time of folow up than not infected under the use of drug by observing righ tail of the curve, but through all the curves it can't show this trend. So, it is supported by logrank test (p=0.111) there is no significant effect of hiv on time to cure of the patient.

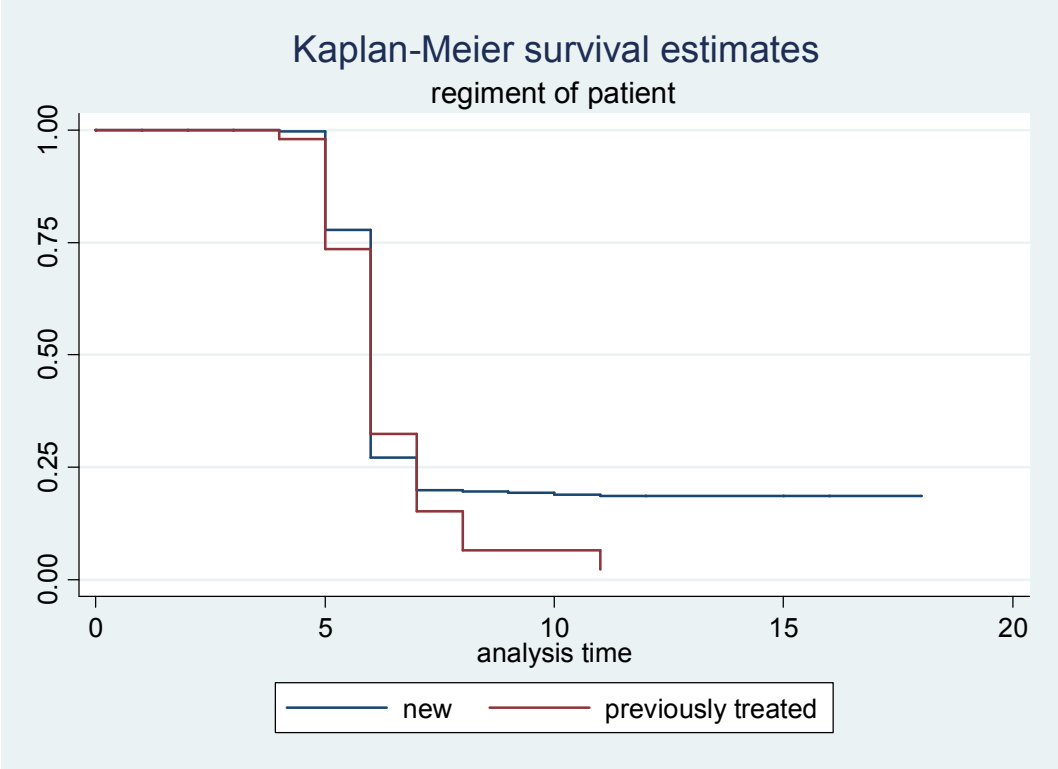

Figure 4. 4:*Kaplan Meier curve by covariate regiment for modeling time to cure of tuberculosis patients at Jimma University specialized hospital*.

This seems, these patiens who previously treated could be cured in short time than those newly comerafter 6 months. But, the survival of the two group is almost equal. Similarly the logranktest shows there is no group variation on time to cure of the patients (p=0.0618) at 5% level of significance.

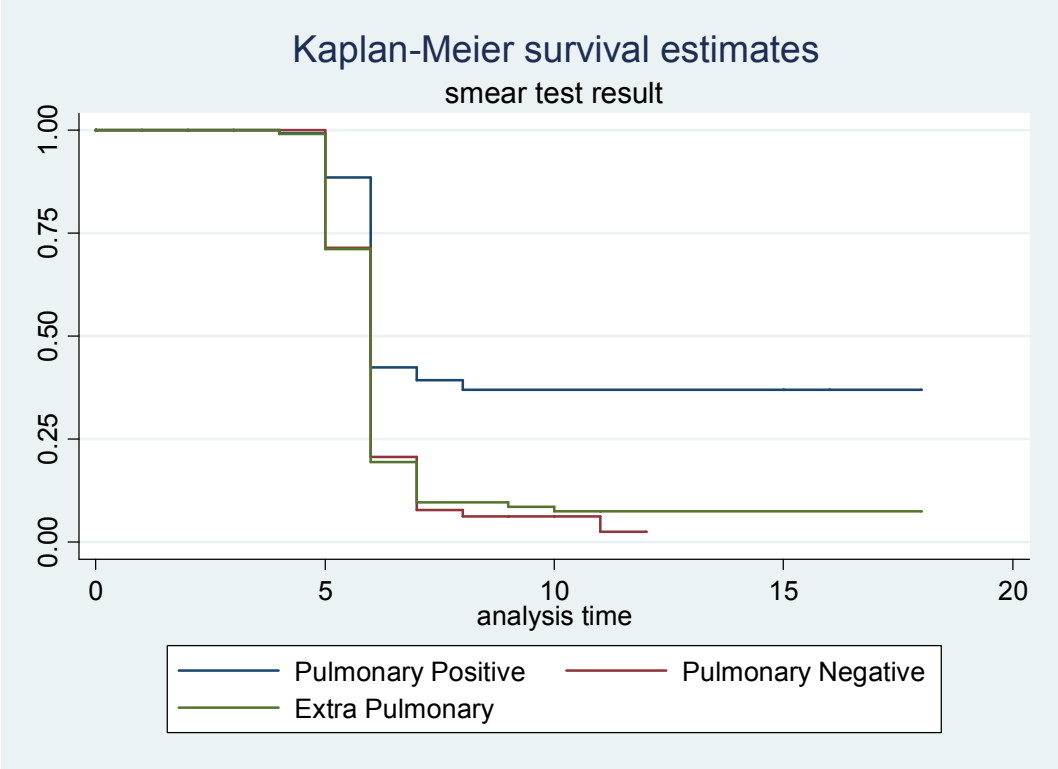

Figure 4. 5: *Kaplan Meier curve for the covariate smear result of the patient. Source: Jimma University specialized Hospital*

Patients of plumonary positive were stay with diseases longer than the other type of TB and extra plumonary TB were also have a long survival probability of the diseases compared to plumonary negative. Similarly logrank test shows us p=7.36e-13 which is smear result categories have the significant effect on the time to cure of the tuberculosis patients.

4.3. Effect of covariates using the mixture cure model

Data was designed such that each of the selected covariates has an effect on both the cured fraction and the survival of the uncured individuals or patients who cured within the follow up period. The Use of the standard Cox PH model is linked to the hypothesis that, if complete follow-up were possible for all patients, each would have eventually cured from tuberculosis. This hypothesis, however, did not hold for the dataset at hand. Some individuals were cured or immune against the event, resulting in the fact that cure time distribution was improper as it has total mass less than 1.

Indeed, from the figure 4.1-4.5, the KME curve levels off at nonzero proportion (around 20%) at the right tail and exhibits a relatively long and stable plateau (Figure 4.1-4.5). Combined with the fact that the last cure time was censored, this supported the applicability of the mixture cure model (McLachlan GJ, Chrishnan T., 1996, Kruk ME, Schwalbe NR, Aguiar CA, 2008). Another evidence of the presence of immune individuals in the 2015 cohort of TB patients was based on the value of the largest event time (Kruk ME, Schwalbe NR, Aguiar CA, 2008). The largest cure time was censored. This led to the rejection of the hypothesis of no immune patient in the source population of the cohort which establishing the evidence of sufficient follow-up. Added to this, we found 12.37% of non-susceptible, satisfying the cut-off criterion of at least 5% of event needed to apply the mixture cure model (McLachlan GJ, Chrishnan T., 1996).

To estimate the probability of cured the logistic cox PH mixture cure model is assumed according to (Kuk& Chen, 1992; Peng& Dear, 2000) reference For modeling the influence of exploratory variables on the incidence, a logistic regression model is usually chosen and the factors affecting the cure probability and fluire time distributions were identified as listed in table 4.2 below in univariable analysis using semi parametric mixture cure model.

4. 4: ***The univariate estimates in modeling time to cure of tuberculosis patients for logistic PH mixture cure model***

Cure probability model				
Covariates	Estimate	Std.Error	Z value	Pr(>|Z|)
Weight	-0.05117782	0.01801626	-2.840646	4.502222e-03
Sex	0.2048506	0.3110912	0.6584906	5.102230e-01
Age	--0.01315393	0.008909613	-1.476375	1.398433e-01
Work place	-0.6596622	0.4028949	-1.637306	0.1015666
HIV.test	-0.522814	0.397798	-1.31427	0.1887553
smear.result	1.765821	0.4569642	3.864244	0.0001114338
Regment	16.637733	2.0278454	8.204636	2.220446e-16
Failure time distribution model				
Weight	-0.002096773	0.003111946	-0.6737819	0.50045
Sex1(male)	0.0864443	0.06167498	1.40161	0.1610316
Age	-0.001261908	0.002690223	-0.4690719	0.6390182
Workplace1(health worker)	0.007841045	0.1551715	0.05053149	0.9596989
HIV.test1(HIV+)	-0.1039459	0.1169365	-0.8889089	0.3740521
Smear.result2(pulmonary negative)	0.002775936	0.08304287	0.03342775	0.9733335
Smear.result3(Etra pulmonary)	0.045624579	0.08533037	0.53468158	0.5928701
Regment1(previously treated)	-0.249114	0.1219655	-2.042496	0.04110235

The Cox PH mixture cure models showed that from the selected covariates regiment is the only variable has effects on the time to cure of the uncured individuals and cured individual. Those covariates are selected by univariable analysis and those has the significant effect based on their p-value in the univariate is included in multivariable fit. Based on their p-value and their standard error with logit link the signifiant factors are proposed to next analysed in multivariable

analysis. Accordingly weight,age, smear.result and regiment are the significant factor on the cure portion and similrly at the same time regiment affect the survival of the diseases at 5% level of significant. Based on those significant factors with the same distribution and link function the multi-variable analysis are given in table 4.3 below.

4. 5: The multivariable estimates in modeling time to cure of tuberculosis patients for logistic-mixture cure model

Cure probability model(cure portion)				
Covariates	Estimate	Std.Error	Z value	Pr(>|Z|)
Intercept	2.83734016	1.4729411	1.926309	5.406576e-02
Weight	-0.07638137	0.02782684	-2.744881	6.053289e-03
smear.result	1.72854423	0.43911592	3.936419	8.270652e-05
Regment	17.14157081	2.23464153	7.670837	1.709743e-14
Failure time distribution model(survival)				
Weight	-0.001432372	0.003320818	-0.43133125	0.66622753
Smear.result2(pulmonary negative)	-0.005579360	0.077912513	-0.07161058	0.94291182
Smear.result3(Extra pulmonary)	0.013690014	0.075980771	0.18017735	0.85701334
Regment1(previously treated)	-0.246115637	0.115980256	-2.12204772	0.03383373

From the parameter estimates in the cure probability model, the cure rate can be calculated for each of cofactors. The cure rate for the weight group is calculated to be

$1 - \pi(weight) = 1 - \frac{e^{weight}}{1 + e^{weight}} = 1 - \frac{e^{2.83734016 - 0.07638137}}{1 + e^{2.83734016 - 0.07638137}} = 0.05896,$ Suggesting that 5.9% of patients are cured under the effect of weight. However, the p-value for the weight variable is significant at α = 0.05 level of significance (p = 6.053289e-03). Therefore, weight has a significant effect on the cure rate of the TB patients, and similarly the cure rate at the effect of smear result and regiment are 15.5%, and 5.03%, respectively. And even though their p-value cannot indicate significances since it have the significant effect in cure rate the excess hazard

rates for the covariates are calculated under the assumption of Cox PH. Therefore the hazard of pulmonary negative is $e^{-0.00558}$ times the hazards of pulmonary positive it implies that pulmonary negative decrease the risk by 1% compared to pulmonary positive and the hazards of extra pulmonary TB group was $e^{0.0137}$ times hazards of pulmonary Positive TB group. And a hazard of pulmonary negative is $e^{-0.00558-0.0137} = e^{-0.02}$this implies that the risk of plumonaery negative TB type is greater than the risk of extra plumonary TB type.

4.4. Assesing the accuracy of the model

The accuracy of the model is checked by ploting the predicted survival probabilty versus time taken to cure. If the plot shows plateau as time goes to infinity at probabilty above zero the model is accurate.

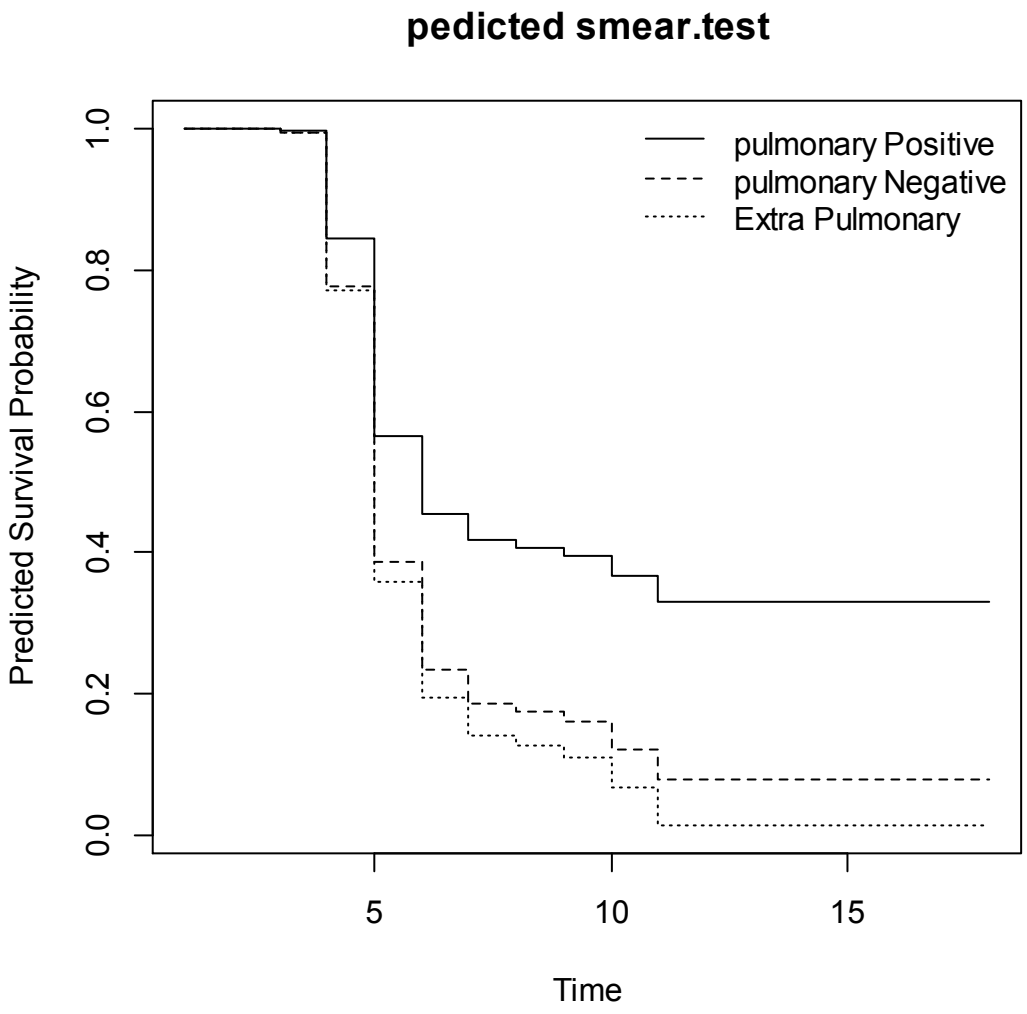

Figure 4. 6*:Predicted survival probability using logistic Cox PH mixture cure model for the covariate smear result of the patient.*

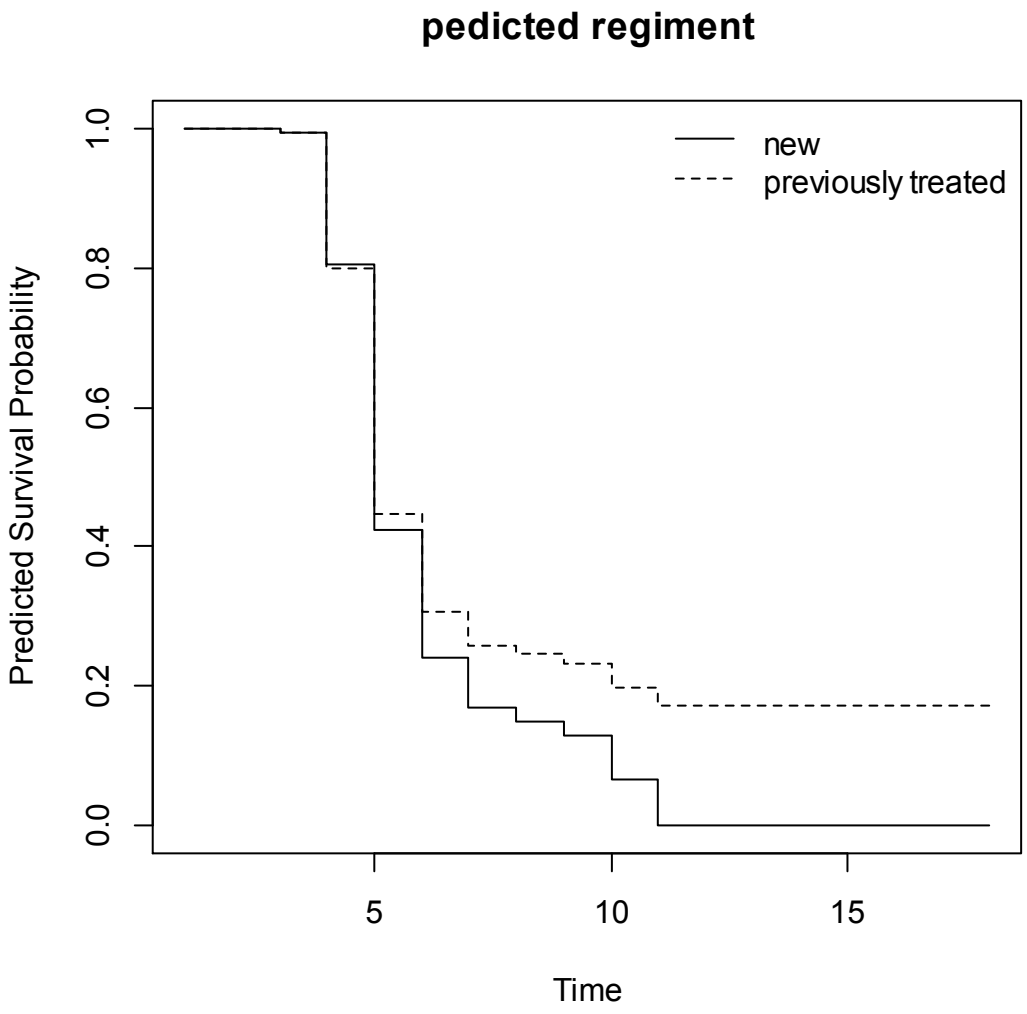

Figure 4. 7: *Predicted survival probability using logistic Cox PH mixture cure model for the covariate regiment of the patient.*

From the predicted cure model the plot pulmonary positive is above the two types while Extra pulmonary tuberculosis diseases is the bottom line and the middle line was the pulmonary negative plot. Based on this it is clearly seen that being attacked by pulmonary positive type of tuberculosis patient take more time to be free from the diseases under the TB drug. And pulmonary negative is second TB categories which need long time compared to extra pulmonary TB categories. From the Predicted survival probability curves of regiment the survival probability of the previously treated patients are above the newly come patients. So from this the patients of TB if they are treated before the time taken to be free from the diseases is longer than those newly affected patients.

4.5. Semi-parametric survival analysis of uncured group

4.5.1. Parametric baseline hazard distribution

Under the all distribution regiment of the patient, TB categories and weight of the patients were having the significant effect on the time to be free from the tuberculosis disease in all distribution. But, the smallest AIC was the good model; accordingly, log normal distribution is the appropriate baseline distribution for the Cox PH regression.

4. 6***: Baseline distribution comparison based on their AIC for Standard Cox regression***

Distribution	**AIC**
Exponential	2108.622
Weibull	1801.363
Logistic	1995.563
Log normal	1667.14

4.5.2.Univariable Analysis using standard Cox PH regression

From the univariable analysis at 5% level of significance only three variables are significant. These were given in table 4.5 below. The hazard of pulmonary negative and extra pulmonary were almost equal 2.376 times the hazards of pulmonary positives. Also from 95% confidence interval for the hazards ratio the confidence interval were not contain 1 which mean that the estimates are different from zero. That means those cofactors were influence the curing time of the patients from the diseases. From the coefficient of the estimates age and weight have a negative impact on the survival of the diseases compared to below median age and below median weight respectively. Similarly pulmonary negative and extra pulmonary TB have positive coefficients indicating that they have the positive impact on the survival probability of diseases compared to the pulmonary positive TB.

4. 7: ***The parameter estimates of the Cox PH fit for the uncured patients in univariate.***

Cofactor	Estimates	Standard errors	95%CI of HR	Hazard ratio	p-value
Age	*-0.006315*	*0.004835*	[0.5198 ,0.8082]	*0.993705*	*0.192*
Weight	*-0.014815*	*0.004544*	[0.9766 ,0.9941]	*0.985294*	*0.00111* **
Sex1(male)	0.1287	0.1170	[0.9043 , 1.43]	1.137	0.271
Workplace1 (health worker)	-0.2622	0.1890	[0.5312 ,1.114]	0.7693	0.165
HIV.test1 (HIV positive)	-0.2792	0.1888	[0.5224 ,1.095]	0.7564	0.139
Smear test2 (pulmonary negative)	*0.8654*	0.1392	[1.808 , 3.121]	2.376	5.14e-10 ***
Smear test3 (etra pulmonary)	0.8356	0.1492	[1.721 , 3.090]	2.306	2.15e-08 ***
Regiment1 (previously treated)	0.1964	0.1592	[0.8909 ,1.663]	1.217	0.217

4.5.2. Multivariate analysis of time to cure using the Cox's PH model

For the classical Cox' PH analysis, covariates included were those with p-value less than or equal to 15%. When adjusted for all these potential risk factors at hand, smear result (pulmonary positive, pulmonary negative and extra pulmonary) and weight at initiation of treatment are the risk factors predicting time to cure from tuberculosis diseases (Table 4.4). HR and 95%CI were 2.3 [1.808 :3.121] for pulmonary negative and 0.9853 [0.9766 :0.9941] for extra pulmonary. Then both direction selection methods are applied using the software and AIC=3455.42 were small with the following end covariates. The analysis indicates that at 5% level of significance smear test and weight has the significant effect on the curing time of the TB patients. And weight

has the negative slope which means at baseline weight, a kilogram increase in weight accelerate the time to cure by0.9902.

4. 8：*The parameter estimates of the Cox PH fit for the uncured patients in multivariabe case.*

Cofactor	Estimates	Standard errors	95%CI of HR	Hazard ratio	p-value
Weight	-0.009880	0.004397	[0.9817 , 0.9987]	0.9902	0.0246 *
Smear.test2(pulmonary negative)	0.838193	0.141933	[1.7507 , 3.0538]	2.3122	3.51e-09 ***
Smear test3 (Etra pulmonary)	0.798534	0.149960	[1.6564 , 2.9816]	2.1533	1.01e-07 ***
HIV.test1 (HIV positive)	-0.350887	0.191307	0.4839 1.0244	0.7041	0.0666

The variables significant for the cure portion and survival of the individual case in addition to weight and smear result regiment of the patient has an impact on the time to cure of the patients but in uncured part here regiment is not statistically significant.

4.5. Model diagnosis

4.5.1. Plot of log of cumulative hazard vs log time

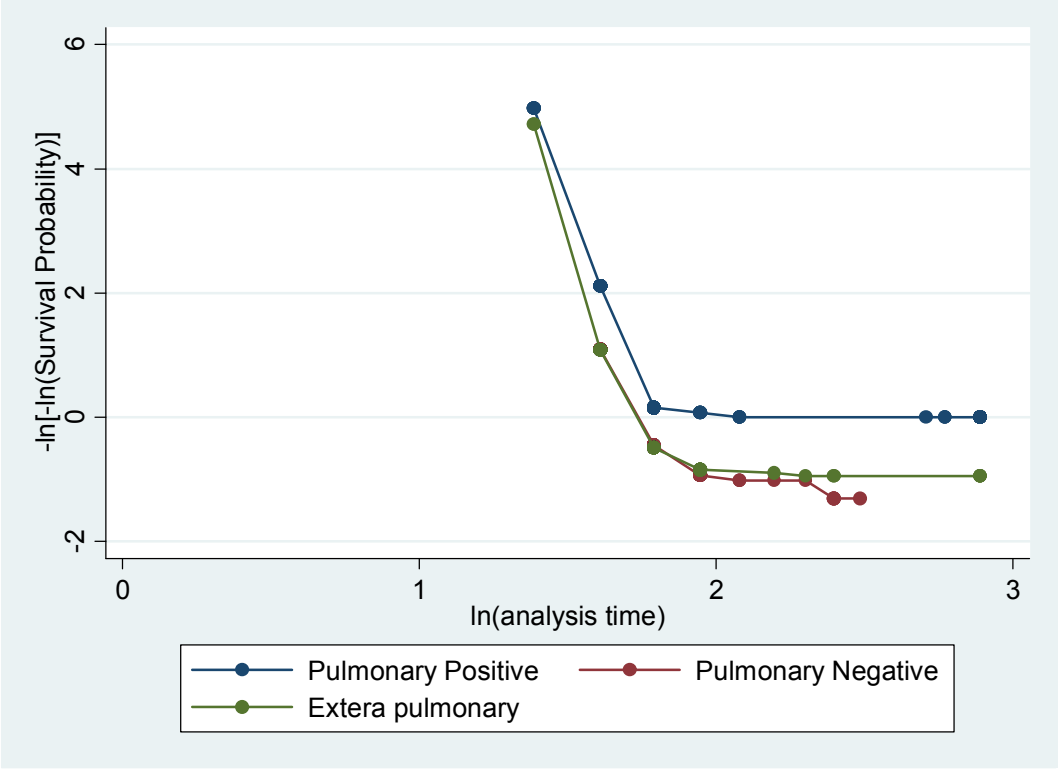

Figure 4. 8: *The log cumulative hazard vs log time plot for smear result factor.*

For testing the PH assumption where smear=1 pulmonary positive, smear=2 pulmonary negative and smear=3 is extra pulmonary TB for Proportional Hazards Mixture Model

It is easy to see that the two curves are not crossover which indicates that the PH assumption is appropriate for this data set.

4.5.2. SchoenfeldResiduals.

This schoenfeld residual is one method of assessing the goodness of fit and outlier. If the plot shows a flat shape the model is a good fit. And similarly this plot seems flat and the outliers are not as much influential as observed below.

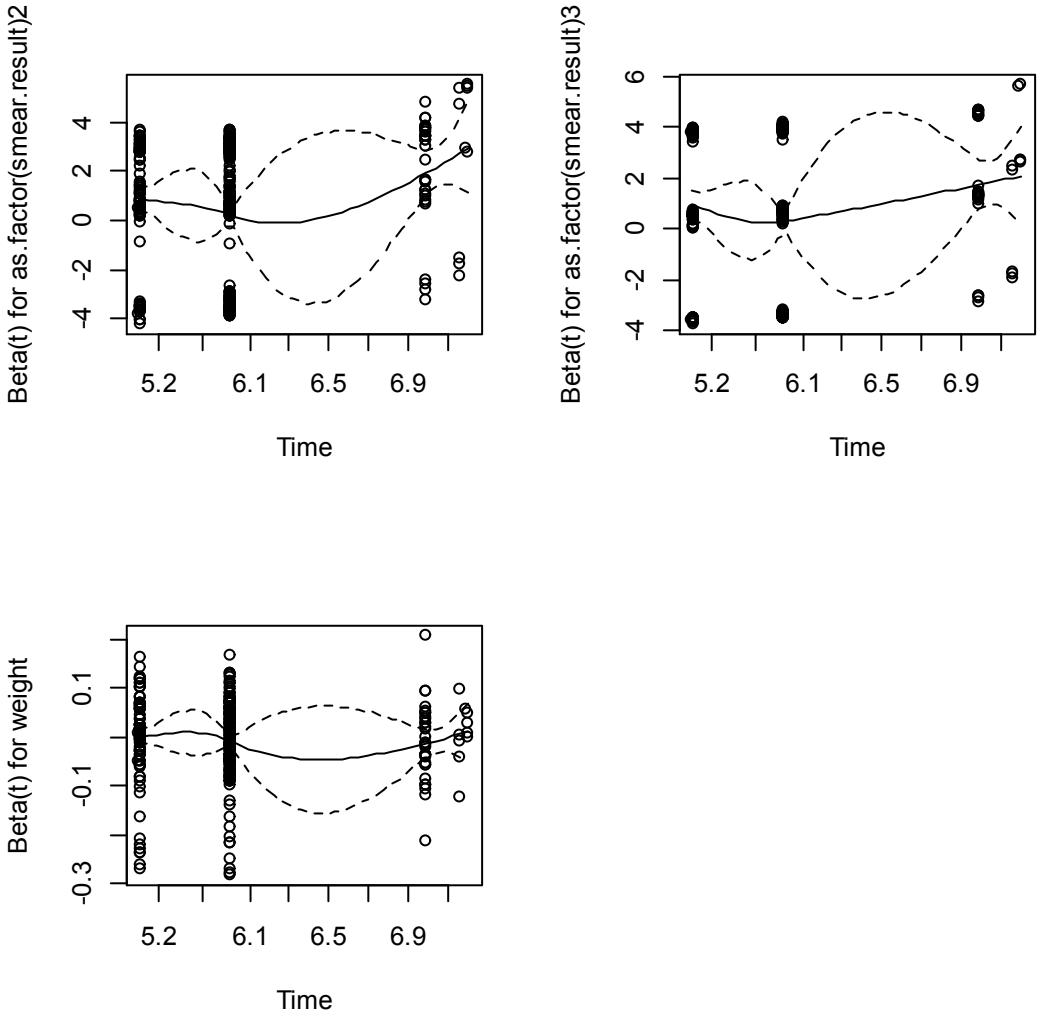

Figure 4. 9: *schoenfeld residual plot.*

From figure 4.9 there were no observable patterns, which mean that PH-assumption is not violated. If this can show increasing or decreasing trend with time the assumption of proportional hazard ratio is violated. Also it is used to examine fit and detect outlying covariate values

CHAPTER FIVE

5. DISCUSSION AND CONCLUSION

5.1 Discussions

The research was targeted on a positive event which is curing from tuberculosis diseases. Adding further urgency to controlling TB, multidrug-resistant TB (MDR-TB) has emerged as a serious problem in many parts of the world, including Russia, Latvia, Estonia, Argentina, the Dominican Republic, and the Ivory Coast (WHO, 1999). Up to 50 million people worldwide may be infected with MDR-TB (WHO, 1999). In low-prevalence countries, drug resistance is generally more common in foreign-born populations, most likely reflecting inadequate treatment programs and sporadic drug availability in high-prevalence countries (Broekmans, 2000).

The most infectious patients are those with smear-positive pulmonary disease. Delay in the diagnoses and treatment of such patients increases the risk of transmission in the community and is associated with increased mortality (Rylance J, *et al*, 2010). The aims of tuberculosis control programs are to cure the patients, to reduce transmission, and to prevent the emergence of drug resistance (WHO, 2010). Treating drug-resistant TB is notoriously arduous for patients and presents huge difficulties for health programmes.

Most of the second-line TB drugs used to treat drug-resistant TB are known for their relative ineffectiveness against the bacilli, meaning a lengthy treatment of up to two years. Patients must receive daily injections for up to six months and take a handful of different drugs once or twice a day for a further eighteen months or more. Treatment is also fraught with numerous side effects which require additional medical management (NathansonE,*et al*,2004) such intense treatment places very big demands on patients. Many have to give up work in order to see their treatment through. Some patients who are hospitalized for periods of their treatment are isolated from their families, which can again give rise to psychological problems and major loss of income. The toxicity of the drugs is perhaps the most striking feature of drug-resistant TB treatment.

Indeed, the severity of the side effects has been compared to cancer chemotherapy, with the difference that MDR-TB therapy is not administered in cycles, but in a continuum over two years. With a long list of commonly experienced side effects added to such a lengthy treatment, it is not surprising that many patients give up, some considering the treatment worse than the disease((Bong Ngeasham Collins, 2011).

Therefore, in reality analyzing those two groups within 18 month follow up data is knowingly making those groups censored. But those censored are the right censored data due to the end of the study. This results wrong conclusion in order to account this mixture cure model is best survival models. The main aim of the study was to model time to cure of tuberculosis patients using mixture cure model by considering the population are two groups. Those susceptible (uncured) and non-susceptible (cured or MDR-TB). Then based on the data 12.4% of the patients are the patients of MDR-TB and 87.6% are the patients of susceptible one. The median cure time for those susceptible groups was 6 months while no one patient is cured with in follow up period for the non-susceptible group. Almost seven explanatory variables are included to see the effect of them on the cure time of the patient based on the literature and information of the clinicians. Those covariates are: Age, Sex, Weight, HIV co-infection, Work place, Regiment and Smear test.

The effect of those covariates was seen in two different ways. The studies of (Kuk& Chen, 1992; Peng& Dear, 2000) states for modeling the influence of exploratory variables on the incidence, a logistic regression model is usually chosen. According to, studies by [B. Getahun, *et al*, 2011, M. D. F. P. M. de Albuquerque,*et al*,2009, M. Vasantha, *et al,* 2008] reported that body weight at initiation of treatment was a risk factor for death from TB and is associated with survival of patients who begin treatment for tuberculosis. Similarly the studies by [M. P. Domingos,*et al.*,2008, M. Muñoz-Sellart,*et al.*,2010, M. W. Borgdorff, *et al.,*1998] indicates types of tuberculosis (positive pulmonary, negative pulmonary, and extra pulmonary tuberculosis) were identified as the significant factors for mortality of tuberculosis patients where time to event is death using Cox ph. regression with Weibull baseline distribution and again the studies by (M. D. F. P. M. de Albuquerque,*et al.*,2009, M. P. Domingos,*et al*,2008, M. Vasantha, *et a.l,* 2008, T. A. Mathew, T. N. Ovsyanikova, S. S. Shin *et al*.,2006) also states TB patient category (TBC) was statistically associated with death of patients with tuberculosis with the

same distribution. My study also found that baseline weight, regiment and smear result (TB patient categories) are statistically significant factors for time to cure of TB patients by logistic Cox PH mixture cure model and weight at initiation and smear result were significant factor by standard Cox regression with lognormal baseline distribution. In addition to these, regiment of patients was affecting the cure and survival probability of the patient at the same time so it needs a great attention. The lognormal distribution has been used for both the non-mixture and mixture cure models (Koti,2001; Sposto 2002) identifies in population based cancer studies age has the significant effect on the curing time of cancer with the same model and distribution.

The effect on the probability of cure by using R-3.3.0 version with package smcure was used to analysis the data. smcure fit and effect of them on the survival probability of uncured patients was analyzed with Cox.ph fit. For the smcure fit the logistic Cox PH mixture cure model was assumed in order to fit and started from univariable analysis then regiment, weight and smear test are identified based on their p-values. And then after the full model is fitted similarly they show significant effect on the cure probability (table 4.2). Predicted survival probabilities are plotted by using plotpredictsmcure function. And the plot shows that the importance's of mixture cure model. These predictions also measure the accuracy of the model.

For the uncured populations similarly univariable analysis was applied with Cox ph. Model and regiment, smear test and weight were identified at 5% level of significance and when the level of significance increases to 15% HIV test also significant then it is also proposed to be included in the multivariable analysis. Those significant in univariable model by p-value were included in multi-variable analysis based on AIC criterion weight and smear test (TB types) was having the significant effect on the survival probability of the uncured population (Table4.3) at 5% level of significant and HIV. Test is nearly significant at 5% level of significant (p=0.066). Hazard of pulmonary negative was 3.2699 times the hazards of pulmonary positive, indicating that patients of HIV positive decelerate time to cure by 30% compared to HIV negative. And hazards of extra pulmonary were 3.3549 times hazards of pulmonary positive that means extra pulmonary was 29.8% faster to heal than pulmonary positive. For the regiment of the patient's hazards of previously treated patient was 0.9851 times hazards of newly come patients. Previously treated increase the hazards of being cure. From the Parametric baseline distribution Exponential,

Weibull, Logistic and lognormal were tested and based on the AIC of the fit Log normal was best fit with (AIC=1679.178) for the Cox ph. fit.

5.2. CONCLUSION

The study identified factors or variables which influence the survival of tuberculosis under DOTS at Jimma University specialized hospital. The result showed that baseline weight, regiment and smear test result have the significant influence on the cure rate of the MDR-TB group and regiment have the significance influence on the survival of the TB patients when multivariable logistic Cox PH mixture cure model is used. But in standard Cox regression baseline weight and smear test result have the significant influence on the survival of the uncured group by the lognormal distribution. The predicted survival probability under the risk factors indicates plateau as time of follow up increases, indicating that mixture cure model is better than standard survival analysis for the given data. The time to cure for the uncured patients is in average 6 months.

5.3. RECOMMENDATION

From the result the maximum time of follow up was 18 months for MDR-TB, but with this time period none of patients are cured which indicates that also it need great attention for the concerned stakeholders in order to prevent and reduce the time to cure from the diseases since TB is curable diseases. In addition to this from the analysis most of the patients 55 of them are those extra pulmonary TB types. Therefore, it is better if concerned bodies observe the strategies clearly and work on it. Generally In order to mitigate the consequences of TB; concerned stakeholders need to design successful mechanisms to improve awareness of the community on TB and to reduce time to cure for the MDR-TB patients from seeking care after experiencing TB

5.4. LIMITATION OF THE STUDY

This study had some limitations: the first is that the study used data from single hospital. Thus, the findings of this study should be interpreted very carefully when they are inferred to the national level. The second limitation is lack of published literature on our country related to the time to cure of tuberculosis using mixture cure model; the references are more of other countries outcome. Finally as different literature pointed out, there are different factors that are assumed to have impacts on time to cure of tuberculosis patients, such as educational level and life style of the patient. However, data on these variables could not be available in hospital records, so these variables were not integrated in this study.

REFERANCES

Allison PD, 1995. Survival Analysis Using SAS: A Practical Guide. Cary, NC: SAS Institute Inc. SAS Institute Inc., Cary, NC, USA.

Berkson J., and R. Gage, 1952. "Survival curve for cancer patients following treatment." Journal of the American Statistical Association.

Boag, J, 1949. "Maximum likelihood estimates of the proportion of patients cured by cancer therapy." Journal of the Royal Statistical Society: Series B (Methodological).

Bohning, D. & Seidel, W. (2003). Editorial: recent developments in mixture models. Comput Stat Data Anal.

Bong Ngeasham Collins, 2011, Assessing the Outcome of Tuberculosis Treatment in the Cameroon Baptist Convention Health Board Tuberculosis Treatment Centers, Umeå University, Sweden

C.S. Li, J.M.G. Taylor, J.P. Sy, 2001, Identifiability of cure models, Stat. Probab. Lett. 54 Caminero, 2003. A tuberculosis guide for specialist physicians. *International Union Against Tuberculosis and Lung Diseases*.

Chaisson RE, Harrington M (2009) how research can help control tuberculosis. Int J Tuberc Lung Dis13[PubMed]

Chen et al., 1999, "A new Bayesian model for survival data with a surviving fraction", Journal of the American Statistical Association, vol. 94.

Corbière F, Joly P, 2007. A SAS macro for parametric and semi parametric mixture cure models. Comput Methods Programs Biomed 85(2).

Dubos, R., & Dubos, J., 1952. *The white plague: Tuberculosis, man, and society*. New Brunswick, NJ: Rutgers University Press.

Ewell, M. and Ibrahim, J.G., 1997. The large sample distribution of the weighted log-rank statistic under general local alternatives. *LifetimeData Analysis*, **3**.

Farewell, 1982. "The use of mixture models for the analysis of survival data with long-term survivors." Biometrics.

FMOH, 2010. Health and health related indicator. Addis Ababa Ethiopian federal ministry of health.

Freireich et al, 1963. “The effect of 6-Mercaptopurine on the duration of steroid-induced remissions in acute leukemia; a model for evaluation of other potentially useful therapy” Blood, vol. 21.

Gamel et al., 1999. “Proportion cured and mean log-survival time as functions of tumor size", Statistics in Medicine, vol. 9.

Getahun B et al., 2013, Treatment outcome of tuberculosis patients under directly observed treatment in Addis Ababa, Ethiopia.

Girmatsion Fisseha,2014,Determinant Factors of Treatment Failure among Tuberculosis

Haybittle, 1965. “A two-parameter model for the survival curve of treated cancer patients.” American Statistical Association.

J.D. Kalbfleisch, R.L. Prentice, 1980, the Statistical Analysis of Failure Time Data, John Wiley, New York.

Keshavjee S, Seung K., 2008 stemming the tide of multi-drug resistant tuberculosis: major barriers to Addressing the growing epidemic. Harvard Medical School, Partners in Health, Francois-Xavier Bagnoud Center for Health and Human Rights, Brigham and Women’s hospital

KOTI, K. M. (2001). Failure-time mixture models: yet another way to establish efficacy. *Drug Information Journal* **35**.

Kruk ME, Schwalbe NR, Aguiar CA (2008) Timing of default from tuberculosis treatment: a systematic review. Trop Med Int Health 13(5)

Kuk and Chen, 1992. A mixture model combining logistic regression with proportional hazards regression. Biometrika 79(3).

Lam, K. F., Fong, D. Y. & Tang, O. Y. (2005). Estimating the proportion of cured patients in a censored sample. Stat Med 24,

Lemos et al., 2004. Risk of tuberculosis among household contacts in Salvado, Bahia. *Brazil Journal of Infectious Diseases* .

Li, C., and J. Taylor, 2002. “A semi-parametric accelerated failure time cure model.”Statistics in Medicine,.

Maller and Zhou, 1996. Survival analysis with long-term survivors, John Wiley & Sons, Inc., New York.

McLachlan GJ, Chrishnan T.,1996 The EM Algorithm and Extensions. Second Edition ed. WILEY- INTERSCIENCE.

N.E. Breslow, 1972 Contribution to the discussion of D.R. Cox, J. R.Stat. Soc. Ser. B 34
Nelson, W., 1982. Applied life data analysis, John Wiley & Sons, New York.
Nathanson E, Gupta R, Huamani P, et al.,2008 Adverse events in the treatment of multidrug-resistant tuberculosis: results from the DOTS-Plus initiative International Journal of Tuberculosis and Lung Disease
Nguyen PT, Henrard S, Coche E, Goffin E, Devuyst O, et al. (2010) Coronary artery calcification: a strong predictor of cardiovascular events in renal transplant recipients. Nephrol Dial Transplant 25(11).
Normann B, et al., 2000. Shortages in Swedish tuberculosis care. Good results only in 71 percent of cases after 12-month treatment as shown in a current study. Lakartidningen97
Peng, et al., 1998. "A generalized F mixture model for cure rate estimation." Statistics in Medicine.
Peng, Y, 2003. Fitting semi-parametric cure models." Computational Statistics &Data Analysis.
Peng, Y. &Carriere, K. C. (2002). An empirical comparison of parametric and semi-parametric cure models. Biom J 44.
Peng, Y. & Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. Biometrics 56.
Ryan, F.,1993. *The forgotten plague: How the battle against tuberculosis was won—and lost*. (1st ed.). Boston: Little, Brown.
Rylance J, Pai M, Lienhardt C, Garner P (2010) Priorities for tuberculosis research: a systematic review. Lancet Infect Dis10[PMC free article][PubMed]
Struthers, C. A. and Farewell, V. T. (1989), "A mixture model for time to AIDS data with left truncation and an uncertain origin", Biometrika, vol. 76.
Sy J.P., and J.M.G. Taylor, 2000. "Estimation in a Cox Proportional Hazards Cure Model." Biometrics.
Sylvère TA, 2015. Default Time from Tuberculosis Treatment in the Southern Republic of Benin Using Mixture Cure Model for Survival Analysis. BiomBiostatInt J 2(5): 00039. DOI: 10.15406/bbij.2015.02.00039
Taylor, J, 1995. "Semi-parametric estimation in failure time mixture models."
UN, 2010. The Millennium Development Goals Report. New York.
USAID, 2009. Ethiopian Tuberculosis Profile.

V.T. Farewell, 1988, D.A. Sprott, The use of a mixture model in the analysis of count data, Biometrics.

WHO, 2006 The Stop TB Strategy, Building on and enhancing DOTS to meet the TB-related Millennium Development Goals. Geneva, Switzerland.

WHO, 2008 Global tuberculosis control - surveillance, planning, financing. Geneva, Switzerland.

WHO, 2009. Global Tuberculosis Control. A short update to the 2009 report, Geneva.

WHO, 2010 Global Plan to Stop Tuberculosis 2011–2015. Geneva: World Health Organization.

WHO, 2014, Global tuberculosis report. 20 Avenue Appia, 1211 Geneva 27, Switzerland

WHO, 2011 Global tuberculosis control. Geneva: World Health Organization.

WHO, 2013 Global Tuberculosis Report 2013. Geneva, Switzerland.

WHO. 1999. Guidelines for prevention of tuberculosis in health care facilities in resource limited settings. *World Health Organization , Geneva*:13

WHO. 2006(a). Tuberculosis and air travel: Guidelines for prevention and control. *World Health Organization, Geneva*: 1-30.

WHO. 2010 Global Tuberculosis Control Geneva World Health Organization.

WHO/IUATLD, 2008. Anti-Tuberculosis Drug Resistance in the World. Report no.4. Geneva: World Health Organization and International Union Against Tuberculosis and Lung Diseases

Y. Peng, K.B. Dear, 2000, A nonparametric mixture model for cure rate estimation.

Yamaguchi K, 1992. "Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'Permanent Employment' in Japan." American Statistical Association.

Yu,et al.,2004. Cure fraction estimation from the mixture cure models for grouped survival data, Stat. Med. 23.

Zhang, J., and Y. Peng, 2007. "A new estimation method for the semi-parametric accelerated failure time mixture cure model." Statistics in Medicine