

Full Length Research Paper

Survival Analysis of Tuberculosis Death Under Directly Observed Treatment: A Case of Adama Hospital, Ethiopia

Zebider Assegid¹ and O. Chandra Sekhara reddy²

¹M.Sc. Statistics Student, Hawassa University, Ethiopia.

²Associate Professor, Mainefhi College of Science, Eiretra

ARTICLE INFORMATION

ABSTRACT

Corresponding Author:
OCS Reddy

Article history:

Received: 04-07-2020

Revised: 08-08-2020

Published: 30-08-2020

Key words:

Stratified Cox regression model; parametric regression model; Tuberculosis coordinates.

Tuberculosis is a major public health problem among developing world due to delay in detection and treatment of patients with active TB. It was proposed as one of the leading causes of morbidity and mortality in Ethiopia and in Adama as well. This study assessed the risk factors associated with time to death among TB patients treated under directly observed treatment program in Adama hospital, Ethiopia. Data were taken from patients' medical record card that enrolled during September 2013 to December 2018. To estimate, compare and model the survival time as well as examine the association between the survival time with different demographic and risk factors. The Kaplan Meier estimation method, stratified Cox regression model and the weibull Accelerated failure time regression model were applied. The analysis is done with the help of STATA, R statistical package and SPSS software's. The result from Kaplan-Meier estimation revealed that the survival time of patients is significantly related with sex categories, weight category, TB patient category and smoker category. The Log rank result also indicated that the survival probability of Tuberculosis patients was statistically different in experiencing the death event among groups classified by sex, age, initial weight, type of TB, patient category, smoking, smear result and HIV status. A Stratified Cox regression model result show that a significant covariates contributed to shorter survival time of a TB patient. The best-fitted weibull Accelerated failure time model is selected by using the Akaike information criterion (AIC). The AIC confirms that the Weibull AFT model is found to be the best fit of the survival of tuberculosis patients under the DOTS at Adama hospital, Ethiopia. The results of Weibull AFT model showed that sex, age categories, HIV status, patient category, Type of TB, smoking and smear result significantly contributed shorter survival time. The findings of this study showed that age categories, sex, type of TB, patient category, initial weight, smear result, HIV status and smoking were major factors related to survival time of TB patient. Survival probability of patients having older age, smoker TB patient, HIV positive TB patient and less initial weight were less survival probability of patients during treatment period.

Introduction

Tuberculosis (TB) is one of the most dangerous and killer diseases. Tuberculosis (TB) is among the leading causes of morbidity and mortality worldwide. More than 70% of the deaths of TB patients occur during the first two months of TB treatment Ashenafi *et al.*, 2014. The major risk factors that increase early death of TB patients are being positive for human immunodeficiency virus (HIV), being of old age, being underweight or undergoing re-treatment (Birlie, *et al.*, 2015). Most of the time it affects the lungs, but it can also damage other organs in the body (WHO, 2017). Tuberculosis is a serious infectious disease which can lead to disability and death. So

people with TB should be diagnosed and treated as early as possible to protect their own health and to prevent infection of people around them (Fisseha, 2018). TB transmission occurs through airborne spread, from a person that has TB of the lung during coughing, speaking and sneezing of infectious droplets. The main symptoms described by TB patients, specifically cough, night sweats, tiredness and Weight loss are among the symptoms with high predictive value. About one-quarter of the world's population has latent TB, which means people have been infected by TB bacteria but are not (yet) ill with the disease and cannot transmit the disease (WHO, 2018). Globally TB is one of the top 10 causes of death and the leading cause from a single

infectious agent (above HIV/AIDS). Millions of people continue to fall sick with TB each year. Globally estimated 10.0 million people (range 9:0 - 11:1million) developed TB disease in 2017: 5.8 million men, 3.2million women and 1.0 million children. In 2017, TB caused an estimated 1.3 million deaths (range 1.2–1.4 million) among HIV-negative people and there were an additional 300000 deaths from TB (range, 266000–335000) among HIV-positive people (GTB-R, 2018). TB occurs in every part of the world. In 2017, the largest number of new TB cases occurred in the South-East Asia and Western Pacific regions with 62% of new cases and African region with 25% of new cases. In 2017, 87% of new TB cases occurred in the 30 high TB burden countries (GTB-R, 2018). TB remains among the top five causes of death in Sub-Saharan Africa. In terms of the years of life lost to TB.

The prevalence of TB is high in low-income countries than developed nations (Tauseef, 2018). Tuberculosis patients who had middle and low income were 87% and 90% respectively to die compared with those who had high socio-economic level. The study revealed that TB patients 58.3% died within 2 months of the start of treatment while the entire patients survived in the sixth month warranted early diagnosis and the start of the appropriate treatment (Ashenafi *et al.*, 2014). Ethiopia is one of highly tuberculosis affected countries, related to low level of awareness on the disease in the population (Gelaw, 2016). Tuberculosis is the third cause of hospital admission (after deliveries and malaria), and the second cause of death in Ethiopia, after malaria. Ethiopia ranks eighth in the list of the 30 high TB burden countries. Ethiopia is among the four countries where treatment outcomes of more than 10% of TB cases were not evaluated and documented (WHO-TBR, 2017). According to 2017 Global TB Report of the World Health Organization lists Ethiopia as one of high-burden TB, TB/HIV and multi-drug resistant TB (MDR-TB) countries, with an estimated incidence of 182,000 in 2016. The rate of MDR-TB was estimated to be 2.7% of new TB cases and 14% of previously treated TB cases.

The prevalence of TB retreatment cases at Adama was higher as compared to similar studies. Similarly, their treatment success rate (TSR) was lower as compared to the national figure (Lenjisa *et al.*, 2015). TB is the 9th leading cause of death worldwide and the leading cause among infectious agents, ranking above HIV/AIDS (WHO, 2017). Ethiopia ranked at 27th among 30 high MDR-TB burdened countries in the world (WHO, 2015). Tuberculosis is a major public health problem throughout the world. According to the Ministry of Health hospital statistics data Tuberculosis is the third causes of hospital admission (after deliveries and malaria) and the second causes of death in Ethiopia after malaria. Tuberculosis is an obstacle to socio-economic development; 75% of people affected by TB are within economically productive age group of 15-54 years (Anteneh and Kebede, 2017). Although efforts to control the epidemic have reduced mortality and incidence, there are several predisposing factors that should be modified in order to reduce the burden of the disease. Tuberculosis can also lead to complications in the course and management of other diseases, such as diabetes. It is therefore important to identify these co morbidities in tuberculosis patients in order to ensure adequate management of both conditions. TB treatment success rate in Ethiopia was below the threshold suggested by WHO (90%). There was also a

discrepancy in TB treatment success rate among different regions of Ethiopia (Seid *et al.*, 2018). However, the direction to where Tuberculosis in Ethiopia is heading hasn't been well analyzed and unpackaged by epidemiologically relevant factors and appropriate statistical models. A higher death rate was noted in patients who started MDR-TB treatment with initial low body weight, HIV positive (Fantaw, 2018) at Adama and Bishoftu General Hospitals.

Although the Government of Ethiopia and different stakeholders were working on reduction of TB, it is still burning public health concern in Ethiopia. The mean treatment success rate was good but case detection rate was decrease through time so the hospital needs to consider active case detection and improve defaulter tracing mechanism. Significant number of hospital staffs is being infected with TB so concerned health authorities including professionals need to give emphasis on the identified factors and on infection prevention control. Further study is recommended in retrospective design to include other predictors (Fisseha, 2018). Hence, this study aimed to assess the survival time and risk factors for mortality in tuberculosis patients at Adama hospitals, Oromia, Ethiopia.

This study focus on survival of tuberculosis patient under directly observed treatment at Adama hospital in Ethiopia by considering the following research questions.

1. What are the factors that significantly affect the death rate of tuberculosis patients?
2. Which statistical model is the most appropriate for analyzing the predictor of death rate of TB?
3. Is there survival difference of the TB patients with respect to different risk factors?
4. How much is the median survival time of tuberculosis patients?

This study was undertaken with the following objectives:

To identify the significant factors for death of TB patients during treatment period; To identify a statistical model that predicts the survival probability of TB patient's; To identify weather there is a survival difference of the TB patients with respect to different risk factors; To estimate median survival time of tuberculosis patient during treatment period.

Materials and Methods

Description of the Study Area

This study was conducted in Adama Hospitals located in Oromia Regional State, Ethiopia, that provides survival of tuberculosis patient under directly observed treatment. Adama Hospital is located about 100 km south east of Addis Ababa, capital city of Ethiopia. Adama is a city in central Ethiopia; it located at 8.54°N 39.27°E at an elevation of 1712 meters. The city sites between the base of an escarpment to the west, and the Great Rift Valley to the east. The city name Adama has been derived from the Oromo word *adaamii*, which means a cactus or a cactus-like tree.

Sample Population

The target population in this study is the full set of individuals could be included in the retrospective study. It consists of all TB patients' male and female who could have follow-up time in the Adama hospital, Ethiopia.

Data description

In this study we used secondary data and collect from patient follow-up records. Information for this study was extracted from documents of all TB cases registered from September 2013 daily followed until December 2018 in adama hospital. Each patient was identified with a medical register card code and the necessary information for this study extracted from their follow up history (medical document) at adama hospital.

Study design

A retrospective five year cohort study was design to assess the outcome death of tuberculosis patients who follow up treatments based on TB patients that were registered in unit TB registers in the health facilities providing direct observed treatments strategy (DOTS) in adama hospital.

Study variables

In this study, several variables that are supposed to be associated with mortality of TB patients were considered. The variables considered in the study were selected based on the findings of previous study and those are expected to be determining factors for TB diseases.

The Response Variable

The dependent variable in this study is “The survival time of Tuberculosis patients” or it can be defined as “The time to death of Tuberculosis patient. It is duration of time from date of treatment until date of death or censor. Tuberculosis patients, who are alive during the study time or dropped before death, are considered as censored. Right censoring is applicable in current study and the total number of population were 601 tuberculosis patients during follow-up only 106(17.6%) observed and the remaining 495(82.4%) censored at Adama hospital Ethiopia.

The survival analysis is required if there is censoring and skewed data our study was analogous with literature with (Tolosie, *et al.*, 2014: Enquoselassie, 2015: Abayneh Birlie, *et al.*, 2015: Yeshiwork Beyene, *et al.*, 2016 and Fantaw, *et al.*, 2018) the percentage of censoring is 87.3%, 85.6%, 92.6%, 90.7% and 81.7% respectively.

Predictor Variables

In this study the explanatory variables expected to be the risk factors of Tuberculosis diseases related mortality were categorical. The risk factors for the death of Tuberculosis patients are several, but in current study only some variables are listed as risk factors. All risk factors did not exist in the patient’s card.

Method for data analysis

Survival analysis

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

The use of survival analysis, as opposed to the use of other statistical methods, is most important when some subjects are lost to follow up or when the period of observation is finite and

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. There are generally three reasons why censoring may occur.

- a person does not experience the event before the study ends;
- a person is lost to follow-up during the study period;
- A person withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., adverse drug reaction or other competing risk)

There are three categories of censoring, (Klein, 1992).

- Right censoring: Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study.
- 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; this is not common in survival studies.
- 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.

Mathematics of Survival Analysis

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. In summarizing survival data, the two common functions applied are the survivor function and the hazard function (Hosmer and Lemeshow, 1999).

Survival function

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

- (a) The survivorship function,
- (b) The probability density function and
- (c) The hazard function

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

$$F(t) = P(T > t) = \int_0^t f(u) du, \quad t > 0 \dots \dots \dots (1)$$

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

$$S(t) = p(T \geq t) = 1 - F(t), \quad t \geq 0 \dots \dots \dots (2)$$

And density function is given by

$$f(t) = \frac{d}{dt}(F) = \frac{d}{dt}(1-s(t)), t \geq 0 \dots \dots \dots (3)$$

Theoretically, as t ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. The characteristics of Survivor functions are.

They are non-increasing.

- at time t = 0, S(t) = S(0) = 1; that is, at the start of the study, since no one has experienced the event yet, the probability of surviving past time 0 is one and
- As time t → ∞, S(t) → 0; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually converge to zero.

Median Survival Time

In analysis of the survival data we use the median survival time because of the existence of censored and positively skewed nature of survival time. This is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value t(50) which is such that S{t(50)}=0.5.

Median survival time is defined as that value for which S(t0.5) = 0.5. If S(t) is strictly decreasing (0.5) is the smallest number such that S(t0.5) ≤ 0.5 or t0.5 = s⁻¹(0.5).....(4)

Hazard Function

The hazard function is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual dies somewhere between t and t + Δt, divided by the probability that the individual survived beyond time t. The hazard function λ(t) can be formulated as:

$$\lambda(t) = \max_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$\lambda(t) = \frac{f(t)}{s(t)} \dots \dots \dots (5)$$

The survival and cumulative hazard functions can be given in terms of the hazard function as:

$$\Lambda(t) = \int_0^t \lambda(u) du \text{ and } S(t) = \exp(-\Lambda(t)) \dots \dots (6)$$

Using the above expressions the hazard function λ(t) can also be given as:

$$\lambda(t) = \frac{d \log s(t)}{dt} = \frac{d \Lambda(t)}{dt} \dots \dots \dots (7)$$

Estimation of survivor function

In practice, when using actual data, we usually obtain estimated survivor function and obtain curves that are step functions, rather than smooth curves.

Kaplan-Meier estimator

The Kaplan-Meier (KM) estimator of the survivorship function (Kaplan and Meier, 1958) also called the product limit estimator,

is mostly used to estimate the survivor and hazard functions. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. Suppose we have a sample of independent observations, their survival times denoted by:

t₁, t₂, ..., t_n And indicators of censoring denoting by δ₁, δ₂, ..., δ_n

$$\text{Where } \delta_i = \begin{cases} 1 & \text{if the event occur} \\ 0 & \text{otherwise} \end{cases} \dots \dots \dots (8)$$

The main quantity of interest is the probability that an event were not occur by time t:

S(t) = P(T > t). Kaplan and Meier (1958) develop an estimator for the survival function.

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \left(\frac{n_i - d_i}{n_i} \right) \delta_i$$

Where

- d_i = number of TB patients died at t_(i)
- n_i = number of TB patients at risk before t_(i)

The variance of the Kaplan-Meier estimators which is referred to as Greenwood's formula is given as.

$$\hat{V}(\hat{S}_{KM}) = \hat{S}_{KM}(t)^2 \sum_{t(i) \leq t} \frac{d_i}{n_i(n_i - d_i)} \dots \dots \dots (9)$$

An alternative estimator for the Kaplan-Meier estimator is the Nelson-Aalen estimator.

Here by we used the link between the survival function and the cumulative hazard Function,

S(t) = exp(-Λ(t)). Estimating first the cumulative hazard function, we then get another estimator for the survival function.

$$\hat{F}(t) = \sum_{t(i) \leq t} \frac{d_i}{n_i} \text{ implies } \hat{S}_{NA}(t) = \exp[-\hat{F}(t)] = \prod_{t(i) \leq t} \exp\left\{-\frac{d_i}{n_i}\right\} \dots \dots \dots (10)$$

It is merely in the case of small samples that the Nelson-Aalen estimate of the survivor function prevails over the KM estimate (Hosmer and Lemeshow, 1999). Moreover, the Kaplan-Meier estimate of the survivor function can be regarded as an approximate to the Nelsen-Aalan estimate.

$$\hat{S}_{KM}(t) \approx \hat{S}_{(t)} = \prod_{t(i) \leq t} \exp\left\{-\frac{d_i}{n_i}\right\} \dots \dots \dots (11)$$

Log-Rank Test

When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. In survival analysis, the Kaplan-Meier estimators of the survivor function for each group are plotted. The statistical question is whether the observed difference in the graph is significant. A number of statistical tests have been proposed to answer this question such as Log-rank, Generalized Wilcoxon, and Tarone-Ware test and so on (Hosmer, et al., 2008). The calculation of each test is based on a contingency table of groups by status at each observed survival time. The general form of these test statistics for the comparison of survival functions between two groups can be defined as follows:

$$Q = \left[\frac{\sum_{i=1}^m w_i(d_{1i} - \hat{e}_{1i})}{\sum_{i=1}^m w_i^2 \hat{v}_{1i}} \right]^2 \dots \dots \dots (12)$$

Where:

- m is the number of rank-ordered failure (death) times by TB.
- n_{1i} is the number of individuals at risk due to TB in group 1 just prior to failure time t_i

n_{2i} is the number of individuals at risk due to TB in group 2 just prior to failure time t_i

n_i is the number of individuals at risk in both groups 1 and 2 just prior to failure time t_i

d_{1i} is the observed number of failure (death) by TBs in group 1 at failure time t_i .

$\hat{e}_{1i} = \frac{n_{1i} \times d_{1i}}{n_i}$ is the expected number of failures or death due to

TB corresponding in group 1 at time t_i

$\hat{v}_{1i} = \frac{n_{1i} n_{2i} d_{1i} (n_i - d_{1i})}{n_i^2 (n_i - 1)}$ is the variance of the number of failures or death by TBs in group 1 at time t_i .

W_i is the weight for right censor adjustment at failure time t_i by TB.

The most common method used to compare survival curves is then a statistical hypothesis test called log-rank test, this test is based on weights equal to one, i.e. $w_i=1$. The null hypothesis for the log-rank test is that there is no difference between the survivals of two or more populations that are being compared (i.e. the probability of the event of interest occurring at any time point is the same for each population). The significance of Q may be tested using the chi-square distribution with one degree of freedom. We can also use the above test to compare more than two groups.

Regression Models for Survival Data

In most medical studies which give rise to survival data, supplementary information referred to as covariates or independent variables needs to be collected on each individual, so that the relationship between survival experience of individuals and various explanatory variables have to be investigated. In the analysis of survival data, interest centers on the risk of hazard of failure at any time after the time origin of the study. As a consequence, the hazard function is modeled directly in survival analysis. There are two broad reasons to model survival data. One objective of the modeling process is to determine which combinations of potential explanatory variables affect the form of the hazard function. Another reason for modeling the hazard function is to obtain an estimate of the hazard function itself for an individual from a set of explanatory variables (Klein and Moeschberger, 1997). A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches. One of the most popular types of regression models used in survival analysis is the Cox proportional hazard model (Cox, 1972).

The Cox Proportional Hazards Regression Model

A popular regression model for the analysis of survival data is the Cox Regression model. The strength of this model is that $h_0(t)$ or baseline hazard is left unspecified (unknown function).

It represents the hazard of an individual with covariates equal to zero. We call this a Semi-parametric Regression model; this model is often also called Cox Proportional Hazard model. In the other way, Cox Proportional Hazard model is a semi parametric model, which contains two parts, non-parametrically estimated baseline hazard $h_0(t)$ and parametrically estimated set of

covariates (Gill, 1984)). The distinguishing feature of Cox PH model is its ability to estimate the relationship between the hazard rate and explanatory variables without having to make any assumptions about the shape of the baseline hazard function. The Cox proportional hazards regression model relates covariates to the hazard function as follows:

$$\lambda(t/x) = h_0(t) c(\beta'x) \dots \dots \dots (13)$$

Where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero. $\beta' = (\beta_1; \beta_2; \dots; \beta_p)$ is a parameter vector of regression coefficients, $x = (x_1; x_2; \dots; x_p)$ is the value of the vector of explanatory variables for a particular individual, and $c(\cdot)$ is a fixed, known scalar function (Samore, et al., 2004). This is a semi-parametric model where the baseline hazard $h_0(t)$ is estimated non-parametrically, Where, $c(\cdot)$ take an exponential form:

$$C(\beta'x) = e^{\beta'x} = e^{\sum_{i=1}^p \beta_i x_i}$$

Which assures that the hazard is non-negative and assumes that covariate effects on the hazard are multiplicative. So

$$\lambda(t/x) = h_0(t) e^{\beta'x} = h_0(t) e^{\sum_{i=1}^p \beta_i x_i} \dots \dots \dots (14)$$

Assumption of Cox proportional hazard model

- The baseline hazard $h_0(t)$ depends on t , but not on covariates x .
- The hazard ratio, that is $e^{\beta'x}$ depends on the covariates x , but not on time t .
- The covariates x do not depend on time t .
- Time is measured on a continuous scale.
- Censoring occurs randomly.

Interpretation

The survival function of the subpopulation with covariate x is

$$S(t/x) = [S_0(t)] \exp^{-\beta'x} \dots \dots \dots (15)$$

where, $S_0(t) = e^{-\int_0^t h_0(t) dt} \dots \dots \dots (16)$

Then, the hazard ratio becomes: For any two sets of covariates

$$\frac{\lambda(t, x_1, \beta)}{\lambda(t, x_2, \beta)} = \frac{h_0(t) e^{\beta'x_1}}{h_0(t) e^{\beta'x_2}} = \frac{e^{\beta'x_1}}{e^{\beta'x_2}} = e^{\beta'(x_1 - x_2)} \dots (17)$$

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

Estimation of Parameters in proportional hazard model

In Cox proportional hazards model we can estimate the vector of parameters β without having any assumptions about the baseline hazard $\lambda_0(t)$. As a consequence, this model is more flexible and an estimate of the parameters can be obtained easily.

Cox Partial Likelihood

The partial likelihood method proposed by Cox in 1975 gives essentially the same results as the last section for the model. Let t_1, t_2, \dots, t_n be the observed survival time for n individuals, and the ordered death time of r individuals be $t(1); \dots, t(r)$ The set of

individuals who are at risk at t_j is denoted by $R(t_j)$. So that $R(t_j)$ is the group of individuals who are alive and uncensored at a time just prior to t_j . The conditional probability that the i^{th} individual dies at t_j given that one individual from the risk set on $R(t_j)$ dies at t_j is:

$$\frac{h_i(t_j)}{\sum_{k \in R(t_j)} h_k(t_j)} = \frac{h_{o(t_j)} e^{\beta'x_j}}{\sum_{k \in R(t_j)} h_{o(t_j)} e^{\beta'x_k}} = \frac{e^{\beta'x_j}}{\sum_{k \in R(t_j)} e^{\beta'x_k}} \dots\dots\dots 18$$

By taking the product of these conditional probabilities over r death times gives:

$$L(\beta) = \prod_{j=1}^r \frac{e^{\beta'x_j}}{\sum_{k \in R(t_j)} e^{\beta'x_k}} \dots\dots\dots (19)$$

Then the partial likelihood function for the Cox PH model is given by:

$$L(\beta) = \prod_{j=1}^r \left[\frac{e^{\beta'x_j}}{\sum_{k \in R(t_j)} e^{\beta'x_k}} \right] \delta_j \dots\dots\dots (20)$$

Where $R(t_j)$ is the risk set at time t_j and δ_j is the event indicator which is zero if the i^{th} survival time is right censored and unity otherwise. This is the partial likelihood defined by Cox. The Cox methodology uses the partial likelihood to yield estimates of β that are consistent and efficient regardless of the form of $h_0(t)$. The partial likelihood is valid when there are no ties in the data set.

Tests to assess the significance of the coefficient

There are three different tests to assess the significance of the coefficients: the partial likelihood ratio test, the Wald test and the score test.

The partial likelihood ratio test

The partial likelihood ratio test R^2 is computed as twice the

difference between the log partial likelihood of the model not containing the covariates the log partial likelihood of the model containing covariates and. specifically,

$$R^2 = 2\{L_P(0) - L_P(\beta)\} \dots\dots\dots (21)$$

The Wald test

The Wald statistic is defined as the ratio of the estimated coefficient to its estimated standard error.

$$Z = \frac{\hat{\beta}}{SE(\hat{\beta})} \dots\dots\dots (22)$$

Under the null that is a single parameter $H_0: \beta=0$, the Wald statistic follows a standard normal distribution (i.e., $Z \sim N(0, 1)$). Obviously, the square of Wald statistic follows a chi square distribution with 1 degree-of-freedom.

The score test

The score test which is obtained by computing the ratio of the derivative of the log partial likelihood to the square root of the observed information, all evaluated at $\beta=0$.

Model Building for Cox PH

An initial step in model building process is to identify a set of explanatory variables that have the potential for being included in the linear component of a proportional hazards model. This set were contain those covariates and factors that have been recorded for each individual, but additional terms corresponding to interactions between factors or between covariates and factors may also require. Model building in proportional hazards regression analysis requires critical decisions in selecting subsets of covariates as it is likely that more covariates are present in real life problems; selection of interaction terms to be included in the model and checking the linearity of continuous covariates and choosing the appropriate transformation for nonlinear covariates. The methods of selecting a subset of covariates in a proportional hazards regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections (Hosmer and Lemeshow, 1999).

Recommendable procedure in selecting variables in the study

Hosmer and Lemeshow (1999) and Collett (2003) recommended the following procedure in variable selection.

1. Include all variables that are significant in the univariable analysis at the 20 to 25 percent level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.
2. The variables that appear to be important from step 1 are then fitted together in a multivariable model. In the presence of certain variables others may cease to be important. Consequently, backward elimination is used to omit non-significant variables (i.e. those variables that do not significant increase the value of $-2\log \hat{\varpi}$ when they are omitted) from the model. We therefore compute the change in the value of $-2\log \hat{\varpi}$ when each variable on its own is omitted from the set. Only those that lead to a significant increase in the value of $-2\log \hat{\varpi}$ are retained in the model. Once a variable has been dropped the effect of omitting each of the remaining variables in turn should be examined.
3. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method (i.e. any that reduce $e^{-2\log \hat{\varpi}}$ significantly are retained in the model). This process may result in terms in the model determined at step 2 ceasing to be significant.
4. A final check is made to ensure that no term in the model can be omitted without significant increasing the value of $-2\log \hat{\varpi}$ and that no term not included significantly reduces $-2\log \hat{\varpi}$. At this stage the interactions between any of the main effects currently in the model can be considered for inclusion if the inclusion significantly modifies the model. In searching for possible confounders, it is useful to examine the simple relationship between the other explanatory factors and survival, adjusted for the factor of interest because, if there is obviously no relationship between a factor and survival, then, it is not likely to be a confounder. Thus, the next step is to consider the relationship between each of the other explanatory factors and survival, given that the factor stated in the basic hypothesis is already in the

model. This process is continued by exploring the relationships between each of the remaining explanatory variables and survival, given that the factor stated in the basic hypothesis and the one next most related to survival (assuming that the basic variable is in the model) are in the model. If no significant confounders are found at any step in this process, then we stop and base our inference about the primary hypothesis on the last model.

Model Assumption

The Cox proportional hazards assumption means that the hazard ratio is constant over time, or that the hazard for an individual is proportional to the hazard for any other individual. When there are covariates in the analysis, which are not satisfy proportionality assumption, since we have covariates information only at the time of the survey.

Model Adequacy Checking

After the model has been fitted, the adequacy of the fitted model needs to be assessed which is usually performed using model residuals.

Cox-Snell Residuals

The Cox-Snell residual is given by Cox and Snell, which is used for assessing the fitness of PH model (Cox, D.R. and Oakes, D., (1984)). The Cox-Snell residual for the i^{th} individual is defined as:

$$r_{ci} = \exp(\hat{\beta}'x_i) \hat{H}_0(t_i) \dots \dots \dots (23)$$

Where $\hat{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i . In practice the Nelson Aalen estimate is generally used. . If the final PH model is correct and the $\hat{\beta}$ are close to the true values of the β , then r_{ci} should resemble a censored sample from a unit exponential distribution. Therefore, a plot of the Nelson-Aalen cumulative hazard estimate of residuals $\hat{H}(r_{ci})$ versus residuals r_{ci} should be a straight line through the origin with a slope of 1, if the fitted model is correct.

Martingale residuals

To check the linearity of continuous variables we plot hazard against the midpoint of the class and using plot of martingale residuals. For this particular study the plot of martingale residuals against continuous covariates are used to check linearity. The Martingale residuals are defined as:

$$r_{Mi} = \delta_i - \hat{H}_i(t_i) = \delta_i - r_{ci} \dots \dots \dots (24)$$

Where δ_i for uncensored observations and zero otherwise, and r_{ci} are Cox-Snell residuals. The Martingale residuals take values between negative infinity and unity. They have a skewed distribution with mean zero. In large samples, the Martingale residuals are uncorrelated with one another and have an expected value of zero. However, the Martingale residuals are not symmetrically distributed about zero (Barlow and Prentice, 1988)

Proportional Hazard Assumption Checking

The main assumption of the Cox proportional hazards model is proportional hazards, which mean that the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality (Graphical

method, Scaled Schoenfeld residuals, Adding time dependent covariate) (Collett, 1994).

Graphical method

To assessing the validity of PH assumption, By plotting estimated $-\log(-\log(\text{survival}))$ versus survival time for two groups. The Researcher would see parallel curves if the hazards are proportional assumptions are satisfied (Klembaum, 1996). This method does not work well for categorical predictors with many levels because the graph becomes cluttered.

Scaled Schoenfeld Residuals

Scaled Schoenfeld residuals are defined as the product of the inverse of the estimated variance-covariance matrix of the k^{th} Schoenfeld residual and the k^{th} Schoenfeld residual (Klembaum, 1996). The scaled Schoenfeld residual can be used to assess time trends and lack of proportionality.

$$r^*_{pji} = (v^{-1}) r_{pji} \dots \dots \dots (25)$$

Where, r^*_{pji} the scaled schoenfeld residual and r_{pji} is the schoenfeld residual. Under the null hypothesis, we expect to see a constant function over time. When the proportional hazards assumption holds, straight horizontal line with zero slopes is expected. If this assumption is violated, the simple Cox PH model is invalid and more complicated analysis such as the stratified Cox regression model or the extended Cox regression model are required.

The Stratified Cox Regression Model

The stratified Cox regression model is a modification of the Cox regression model by the stratification of a covariate that does not satisfy the proportional hazards assumption. Covariates that are assumed to satisfy the proportional hazards assumption are included in the model, whereas the predictor being stratified is not included. The covariates not satisfying the proportional hazards assumption are denoted by z_1, z_2, \dots, z_k and the

covariates satisfying the proportional hazard assumption are denoted by x_1, x_2, \dots, x_p to form the stratified Cox regression model,

a new variable is defined from z variables and denoted by z^* . The

stratification variable z^* has k^* categories, where k^* is the

total number of combinations (strata) formed after categorizing each of z' s (Ata and Sözer, 2007)

Parametric Survival Models

Parametric models need some special assumptions about $h_0(t)$. But the advantage of Cox model is the fact that such assumptions can be avoided. The Parametric PH model is the parametric versions of the Cox proportional hazards model. It is given in similar form to the Cox PH models. The main difference between the two kinds of models is that the baseline hazard function is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data, while the Cox model has not such assumption. A number of different parametric PH models can be derived by choosing different hazard functions Lawless, (2003). The commonly used models are exponential, Weibull, lognormal,

log logistic and Generalized gamma models. If the proportionality assumption is not valid, the Cox proportional hazard models cannot be used in modeling rather some parametric approaches are appropriate. In a parametric model, the distribution of outcomes (time to event) is specified in term of a finite number of unknown parameters. One of the famous parametric models is accelerated failure time (AFT) model in which the time to event is assumed to be a function of explanatory variables.

Parametric models are useful when we want to predict survival rather than identify factors that influence survival. Parametric models can be expressed in: (1) proportional hazard form, where a one unit change in an explanatory variable causes a proportional change in hazard; and (2) accelerated failure time (AFT) form, where a one unit change in an explanatory variable causes a proportional change in survival time.

Parametric Proportional Hazards model

The parametric proportional hazards model is the parametric versions of the Cox proportional hazards model where a one unit change in an explanatory variable causes a proportional change in hazard. The hazard function is given by;

$$\lambda(x/t) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \dots \dots \dots (26)$$

This is the base line hazard function is distributional based. One of the most appropriate parametric models is accelerated failure time (AFT) model in which the time to event is assumed to be a function of explanatory variables. The time of survival for a single covariate, which is called, accelerated failure time, expressed as:

$$T = \exp(\beta_0 + \beta_1 x + \epsilon) \dots \dots \dots (27)$$

Where $\{t \sim \text{Exp}(\alpha)\}$ and ϵ be error component

Accelerated Failure Time Model

Although parametric PH models are very applicable to analyze survival data, there are relatively few probability distributions for the survival time that can be used with these models. In these situations, the accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models the researcher measure the direct effect of the explanatory variables on the survival time instead of hazard.

The AFT model assumes that there is a linear relationship between the logarithm of time and the covariates. It can be written as an ordinary regression model for log survival time of the form,

$$Y = \log T = X' \beta + \delta W \dots \dots \dots (28)$$

Where; $\beta = (\beta_1, \beta_2, \dots, \beta_p)^t$ are vector parameters, X's are vector covariates, δ is scale parameter and W is the error term and given as;

$$W = \frac{\log T - X' \beta}{\delta} \dots \dots \dots (29)$$

Exponential Model

The exponential model is the simplest type of parametric model it assumes that the baseline hazard is constant over time. The exponential distribution is the only distribution with a constant hazard i.e $\lambda(t, \lambda) = \lambda, \lambda > 0$. This implies that the conditional "probability" of an event is constant over time. In other words, the risk of an event occurring is flat with respect to time. The

survivorship function may be obtained by expressing in terms of time given by the formula;

$$S(t, x, \beta) = \exp\left(\frac{-t}{e^{\beta_0 + \beta_1 x}}\right) \dots \dots \dots (30)$$

And the hazard function of the exponential regression model is:

$$h(t, x, \beta) = \lambda_0(t) e^{-(\beta_0 + \beta_1 x)} \dots \dots \dots (31)$$

The exponential regression model for the p covariates and individual TB patients is expressed as:

$$(t, x_i, \beta) = \lambda_0(t) \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \dots \dots \dots$$

Weibull Distribution

If the survival time T is Weibull with parameter $\lambda > 0$ and $\alpha > 0$,

denoted $T \sim W(\lambda, \alpha)$: where λ and α the scale and the shape parameters respectively. The survival time t is a positive random variable with Weibull density function can be expressed as:

$$f(t; \lambda, \alpha) = \frac{\alpha}{\lambda} \left(\frac{t}{\lambda}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{\lambda}\right)^\alpha\right) \dots \dots \dots (33)$$

Where, $\lambda > 0$ and, $\alpha > 0$

The baseline hazard function of the distribution becomes;

$$h(t; \lambda, \alpha) = \alpha \lambda t^{\alpha-1} \dots \dots \dots (34)$$

Now incorporate covariates x in the hazard function, the Weibull regression models become:

$$h(t; x, \beta) = \lambda \alpha t^{\alpha-1} \exp(x\beta) \dots \dots \dots (35)$$

This yields the following survivor functions; $S(t) = e^{-\lambda t^\alpha}$ and the cumulative hazard function becomes: $H(t) = \lambda t^\alpha$

The Weibull model is more general and flexible than the exponential model and allows for hazard rates that are non-constant but monotonic. It is a two-parameter model (λ and α) where λ , is the location parameter and α is the shape parameter determines whether the hazard is increasing, decreasing, or constant over time. The shape parameter works in the following way:

> If $0 < \alpha < 1$, then the hazard is monotonically decreasing with time.

> If $\alpha = 1$, then the hazard is flat and we have the exponential model i.e. the Weibull model nests the exponential model. This means that we can use the Weibull model to test to see if the exponential model is appropriate.

Log-logistic Distribution

A random variable T has the log-logistic distribution with the following density, hazard and survivorship functions are given the following equations.

$$f(t, \lambda, \alpha) = \frac{\lambda \alpha t^{\alpha-1}}{1 + \lambda t^{\alpha^2}} \dots \dots \dots (36)$$

$$\lambda(t, \lambda, \alpha) = \frac{\lambda \alpha t^{\alpha-1}}{1 + \lambda t^{\alpha^2}} \dots \dots \dots (37)$$

The survival function S (t /X) at any covariate x can be expressed as:

$$S(t, \lambda, \alpha) = \frac{1}{(1 + \lambda t^{\alpha^2})} \dots \dots \dots (38)$$

Where, scale parameter $\lambda > 0$, shape parameter $\alpha > 0$.

- If $\alpha < 1$, the hazard decreases monotonically over time.
- If $\alpha = 1$, the hazard decreases monotonically from λ .
- If $\alpha > 1$, however, the hazard increases to a maximum point and then decreases over time. In this case ($\alpha > 1$), the hazard function is said to be uni-modal.

Unlike the Weibull model, a log-logistic AFT model is not a PH model. However, the log logistic AFT model is a proportional odds model. A proportional odds survival model is a model in which the odds ratio is assumed to remain constant over time. This is analogous to a proportional hazard model where the hazard ratio is assumed constant over time.

Lognormal Distribution

The log-normal model may take censored time dependent variable that allows the hazard rate to increase and decrease. The log-normal model assumes that $\varepsilon \sim N(0, 1)$. Let $h(t)$ be the hazard function of T for (3.23) when $\beta=0$ i.e. $\beta_0 = \beta_1 = \dots = \beta_p = 0$

Then, it can be shown that $h(t)$ has the following functional form:

$$h(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\log(t)}{\sigma}\right)\right] \sigma t} \dots \dots \dots (39)$$

Where $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ be the probability density function and $\int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$ is the cumulative density function. Obviously we no longer have a proportional hazard model

Method of Parameter Estimation

All parametric models may be fit by maximizing the appropriate likelihood function. In each distribution there are several parameter(s) which determined the shape of this distribution. In survival analysis, some observations are censored. Hence, estimation Method has to be adapted to censoring. In parametric modeling, maximum likelihood estimation is commonly used.

Comparison of Models

Model comparison and selection are among the most common problems of statistical practice, with numerous procedures for choosing among a set of models (Rao, *et al.*, 2001). There are several methods of model selection. The most commonly used methods include Akaike information and likelihood based criteria. A data-driven model selection method such as an adapted version of Akaike's information criterion AIC (Akaike, 1974) is used to find the truncation point of a series of models. In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Burnham, *and Anderson*, 2002). In this study, we use the AIC criterion and log likelihood to compare four of parametric models. AIC is defined as: $ALC = -2L + 2(K+C) \dots \dots \dots (40)$

Where L is the log-likelihood, k is the number of covariates in the model and c is the number of model-specific ancillary parameters. The addition of $2(k+c)$ can be thought of as a penalty if non predictive parameters are added to the model. Small values of AIC suggest a better model.

Result and discussion

Table 1. Demographic and Health factors by TB death

Patients characteristics		Summary of the Number of death and censored Values				
		Total	Death	Percent death	Censored	Percent censored
Sex	Male	259	68	26.3%	191	73.7%
	Female	342	38	11.1%	304	88.9%
Age	Less than 24	214	27	12.6%	187	87.4%
	24-44	269	47	17.5%	222	82.5%
	Above 45	118	32	27.1%	86	72.9%
Smoker	Non smoker	488	57	11.7%	431	88.3%
	Smoker	113	49	43.4%	64	56.6%
Weight	Less than 35	74	33	44.6%	41	55.4%
	Above 35	527	73	13.9%	454	86.1%
HIV	Negative	488	85	17.4%	403	82.6%
	Positive	113	21	18.6%	92	81.4%
Type of TB	Category I (pulmonary positive)	170	16	9.4%	154	90.6%
	Category II (Extra pulmonary)	269	69	25.7%	200	74.3%
	Category III (pulmonary negative)	162	21	13%	141	87%
TB patient	Non-New case	85	28	32.9%	57	67.1%
	New case	516	78	15.1%	438	84.9%
Smear result	Negative	327	42	12.8%	285	87.2%
	Positive	274	64	23.4%	210	76.6%
Family size of TB patient	Patient have no family	266	26	9.8%	240	90.2%
	Less than 3	194	37	19.1%	157	80.9%
	Above 3	141	43	30.5%	98	69.5%

MDR.TB	No	488	85	17.4%	403	82.6%
	Yes	113	21	18.6%	92	81.4%
Work place	NHF	559	97	17.4%	462	82.6%
	HF	42	9	21.4%	33	78.6%
Anti-TB drug	RHZE	599	105	17.5%	494	82.5%
	Others	2	1	50%	1	50.0%
Source of initial referral	HEW	13	1	7.7%	12	92.3%
	Self	542	101	18.6%	441	81.4%
Dose	Other	46	4	8.7%	42	91.3%
	Tablets1	8	5	62.5%	3	37.5%
	Tablets2	56	8	14.3%	48	85.7%
	Tablets3	335	59	17.6%	276	82.4%
	Tablets4	200	33	16.5%	167	83.5%
	Tablets5	2	1	50%	1	50.0%

Out of the total 601 registered TB patients 106(68 male and 38 female) or (17.6%) died during the study period and 495(82.4%) were censored. A death proportion seems lower for females (11.1%) than for males (26.3%).The age group (≥ 45 years) showed the highest percentage (27.1%) with respect to death proportions than the other two age groups. A smoker TB patient had higher percentage (43.4) with respect to death proportion than non smoker. Patients with body weight at initiation of treatment less than 35 had higher percentage (44.6) risk group for death. HIV positive TB patients are highest risk group for death i.e. (18.6%). In TB patient category non-new case had higher percentage (32.9%) of death than new case. The percentage of death 9.4%, 13% and 25.7% occurred in the patients with pulmonary positive, pulmonary negative and extra pulmonary types of TB, respectively (Category II TB patients have the highest death proportion (25.7%) as compared to the other two groups while Category I TB patients show the lowest death

rate).The Family size of TB patient had above 3 showed the highest percentage (16.87%) with respect to death proportions than the other two groups. Positive smear results of TB patients have the highest death proportion (23.4%) as compared to Negative smear result (12.8) of TB patients. The death proportions were highest for those patients who work in healthy facility which is 21.4% compared to Non healthy facility worker with 17.4%.Regarding MDR.TB the death proportion were highest for those patients who had MDR.TB which is 18.6%,and lowest 17.4% for patients who had No MDR.TB. From this, the death proportion were highest for those patients who were use others drug is 50% , while patients who were use RHZE had the lowest proportion of death 17.5%.

Survival analysis

Non-parametric survival analysis

Table 2: Summary statistics of TB patient and days of follow up time.

Median survival time of TB patient in day	Minimum	Maximum	Prob
664	1	982	0.272

From the Kaplan Meier survival estimate, we have seen the minimum, maximum and median survival time of TB patients in Adama hospital was give respectively (1, 982 and 664). This median value indicated that the half of the TB patients died with probability 0.272. In this case the estimate of the median survival is 664 days as we have seen from summary. Then the

survival curve reaches the 27% line before the end of the study by the probability 0.272 which is approximation 0.5. Median survival time is less than or equal to 0.5, therefore 0.272 is less than 0.5. A similar study was conducted by (Fantaw, *et al.*, 2018) at Bishoftu General Hospitals, Oromia, Ethiopia the median survival time was 400.5 day

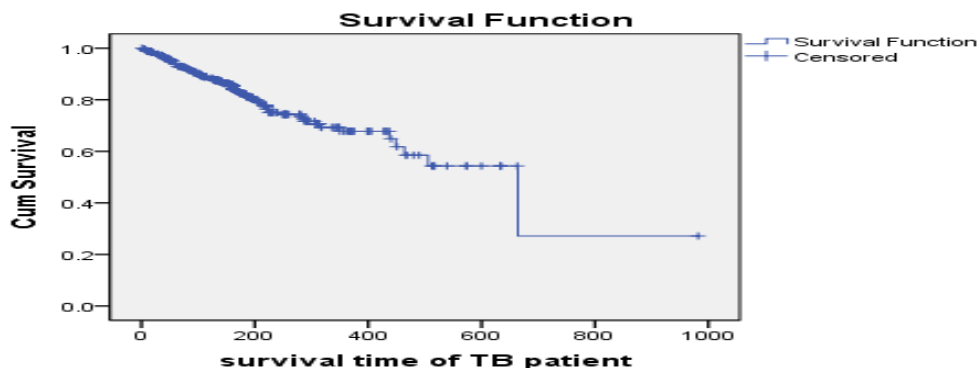
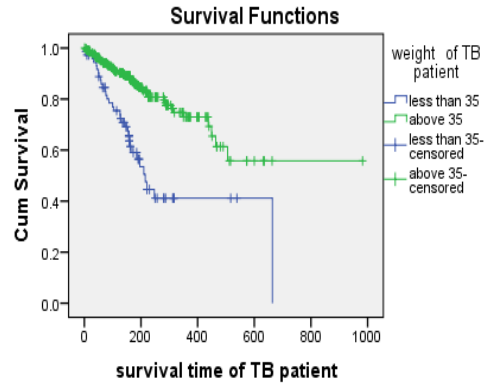
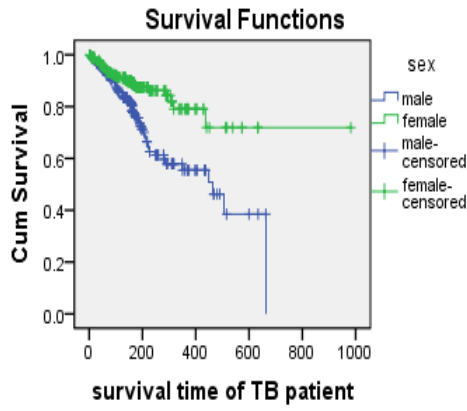


Fig 1: The plot of the overall estimate of Kaplan-Meier survival curves of TB under DOTS in Adama hospital.

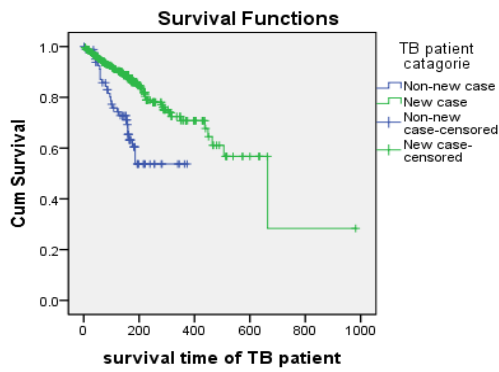
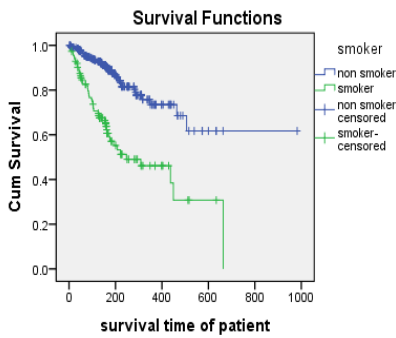
The graph of the estimate of overall Kaplan-Meier survival curve shows the cumulative proportion of TB patients over time and

International Journal of Basic and Applied Sciences

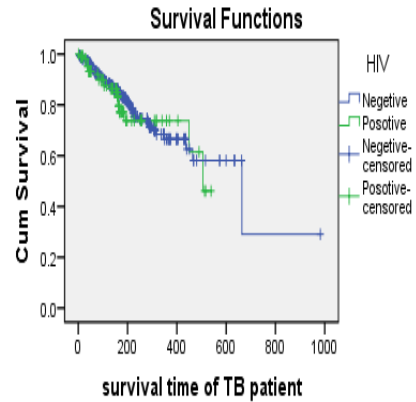
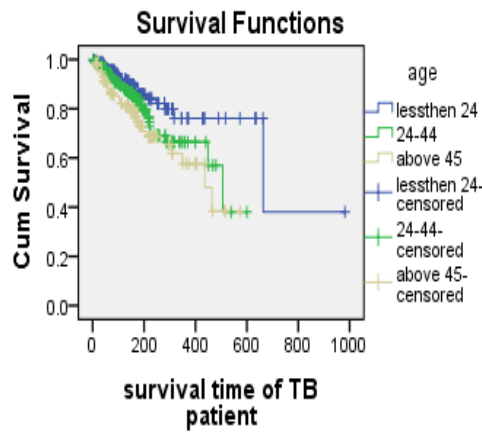
the plot shows that most of the event/deaths occurred in the earlier months of TB treatment initiation.



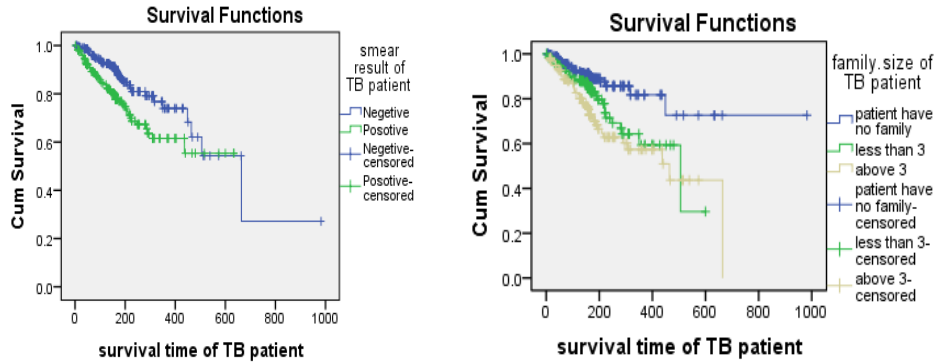
Comparison of Kaplan-Meier survival curves, by sex category and initial weight



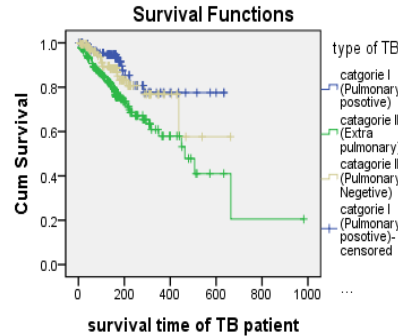
Comparison of Kaplan-Meier survival curves of smoking category and TB patient category



Comparison of Kaplan-Meier survival curves, by age category and HIV



Comparison of Kaplan-Meier survival curves, smear result category and family size



Comparison of Kaplan Meier survival curves of type of TB

Fig 2: The plot of Kaplan-Meier survivor curve of TB patients under DOTS in Adama hospital.

Log rank

We use log-rank test and Kaplan-Meier survival estimates to look into the significance of the difference in survival experience among different factors. The results of the log-rank test for the

equality of survivor functions are presented in Table 2: The null hypothesis to be tested has been no difference between the probabilities of an event occurring at any time point for each population.

Table 3: Logrank test for equality of Survival function among the different groups of covariates for the Death of TB Patients.

Test of Equality over strata				
Variable	Chisq	Df	Pr>	Chi-Square
Sex	17.5	1		0.0000284
Age	9.8	2		0.00742
Family size	19.6	2		0.0000555
Weight	34	1		0.00000000553
Patient category	22.5	1		0.00000206
HIV	20.8	1		0.00000523
Smoker	51.8	1		0.00000000000063
Smear result	11.4	1		0.000723
Type of TB	17.6	2		0.000152
Dose	16.8	4		0.00216
MDR.TB	0.4	1		0.542
Drug	1.4	1		0.237
Source	3.3	2		0.187
Work. Place	0.3	1		0.591

The log-rank test results in Table 3 shows that the different groups of Sex, age, family size, weight, patient category, HIV, smoker, smear result, type of TB, are a significant covariates at 5% level of significance, are those significant covariates whose different levels have an impact in the survival time of TB patients.

Univariate Cox PH Regression Analyses

In order to explore the relationship between the survival experience of a patient and covariate variables, we applied a Cox proportional hazard modeling approach for the analysis of survival data. To determine the variables to be included in the final model, the single covariate Cox PH regression analysis is applied first to identify the impact of individual variable on time to event before proceeding more complicated model selection. Variables are identified as significant using a 0.05 significance level in the single covariate analysis.

Table 4: Results of univariate Cox PH Analysis

Covariates	Categories	Coef	SE(coef)	p-value	exp(coef)	Lower CI 95%	Upper CI 95%
Age	Age0@				1		
	age1	0.4590	0.2452	0.06117	1.5825	0.9787	2.559
	age2	0.8121	0.2643	0.00212	2.2526	1.3420	3.781
Family size	Family size0@				1		
	Family size1	0.6876	0.2562	0.00729	1.9889	1.204	3.287
	family size2	1.0636	0.2490	1.95e-05	2.8968	1.778	4.720
Weight	Weight 0@				1		
	weight1	-1.1660	0.2111	3.32e-08	0.3116	0.206	0.4713
Patient category	Patient.cat0@				1		
	Patient cat1	-1.0178	0.2235	5.28e-06	0.3614	0.2332	0.5601
HIV	HIV 0@				1		
	HIV1	0.8614	0.1949	9.88e-06	2.3665	1.615	3.467
Smoker	Smoker 0@				1		
	smoker1	1.3142	0.1957	1.9e-11	3.7218	2.536	5.462
Type of TB	Type of TB0@				1		
	type of TB1	1.0202	0.2779	0.000242	2.774	1.6088	4.782
	type of TB2	0.4039	0.3320	0.223849	1.498	0.7812	2.871
Sex	Sex0@				1		
	sex1	-0.8253	0.2028	4.7e-05	0.4381	0.2944	0.6519
Smear result	smear result1	0.6646	0.2001	0.000898	1.9438	1.313	2.877
	Dose	dose1@			1		
	dose2	-1.8943	0.5725	0.000937	0.1504	0.04898	0.4620
	dose3		0.4677	0.000792	0.2081	0.08322	0.5206
		-1.5696					
	dose4	-1.6347	0.4818	0.000692	0.1950	0.07584	0.5206
	dose5	-0.7469	1.1043	0.498815	0.4738	0.05440	4.1269

According to the univariate Cox PH analysis (Table 4), that the covariates Age, Sex, smear result ,weight, family size ,patient category, type of TB , HIV status ,smoker and dose are statistically significant at at 20%-25% level of significance and selected as significant risk factors for the death of TB patients.

Multivariable Cox PH Regression Analysis

The multivariable Cox PH regression analysis conducted by including all the potential risk factors that had P-value less than 20%- 25% significant level in univariate Cox PH regression

analysis (by using stepwise selection process). To select the most appropriate subgroup of covariates in our model, the approach of stepwise was applied. The stepwise selection process consists of a series of alternating forward selection and backward elimination steps. It means only covariates with P-value less than or equal to 0.05 will be tested in the model, and to keep it in the model, its P-value should be less than or equal to 0.05. The results of the stepwise proportional hazard regression analysis are presented in the following Table 5.

Table 5: Results of the Multivariable Cox PH Analysis

Covariates	Categories	coef	se(coef)	p-value	exp(coef)	Lower CI 95%	Upper CI 95%
Age	age 0@				1		
	age1	0.2801	0.2999	0.350355	1.3232	0.73514	2.3817
	age2	0.7538	0.3564	0.034422	2.1250	1.05684	4.2727
Weight	Weight 0@				1		
	weight1	-0.6358	0.2666	0.017082	0.5295	0.31402	0.8929
Sex	Sex 0@				1		
	sex1	-0.5491	0.2186	0.011987	0.5775	0.37626	0.8862
Type of TB	Type of TB 0@				1		
	type of TB1	1.0490	0.3041	0.000562	2.8547	1.57294	5.1809
	Type of TB2	0.8250	0.3545	0.019966	2.2818	1.13896	4.5714
Smoker	Smoker 0@				1		
	smoker1	0.9747	0.2138	5.15e-06	2.6502	1.74297	4.0298
HIV	HIV0@				1		

Smear result	HIV1	0.6168	0.2113	0.003514	1.8529	1.22458	2.8036
	Smear result 0®				1		
Family size	smear result1	0.7960	0.2198	0.000293	2.2166	1.44076	3.4102
	Family size0®				1		
	family size1	0.2502	0.3105	0.420323	1.2843	0.69885	2.3601
Dose	family size2	0.1548	0.3332	0.642189	1.1674	0.60761	2.2430
	Dose 1®				1		
	dose2	-1.2133	0.6047	0.044795	0.2972	0.09085	0.9722
	dose3	-0.7480	0.5313	0.159197	0.4733	0.16707	1.3410
	dose4	-0.9674	0.5487	0.077871	0.3801	0.12967	1.1140
Patient cat	dose5	0.3550	1.1883	0.765105	1.4262	0.13890	14.6446
	Patient cat 0®				1		
	Patient cat1	-0.7414	0.2478	0.002766	0.4764	0.29316	0.7743

To further optimize the Cox model, the variable with the highest P-value and over threshold of significance are removed from the predictive model one by one until all the rest variables are shown

significant impact on the prediction of hazard rate. From Table 5 the variables family size is the variables with highest p-value, so it is removed. The result is shown as Table 6.

Table 6: Elimination of Variables with High p-value by backward Elimination Process

Covariate	Categories	coef	se(coef)	p-value	Exp(coef)	Lower CI 95%	Upper CI 95%
Age	age0®				1		
	age1	0.3621	0.2610	0.165334	1.4363	0.8612	2.3955
	age2	0.7938	0.2814	0.004792	2.2118	1.2741	3.8396
Weight	Weight 0®				1		
	weight1	-0.6155	0.2454	0.012136	0.5403	0.3340	0.8741
Sex	sex0®				1		
	sex1	-0.5707	0.2142	0.007702	0.5651	0.3714	0.8599
Type of TB	Type of TB0®				1		
	Type of TB1	1.0753	0.2999	0.000337	2.9310	1.6282	5.2763
	Type of TB2	0.8241	0.3528	0.019504	2.2798	1.1418	4.5520
Patient category	patient.cat 0®				1		
	Patient cat1	-0.7606	0.2418	0.001660	0.4674	0.2910	0.7508
Smoker	Smoker 0®				1		
	Smoker1	1.0106	0.2051	8.39e-07	2.7471	1.8376	4.1068
HIV	HIV0®				1		
	HIV1	0.6029	0.2042	0.003149	1.8274	1.2247	2.7268
Smear result	Smear result 0®				1		
	Smear result1	0.8682	0.2169	6.29e-05	2.3825	1.5573	3.6450

Table 6: As we have seen the final model is generated by including the variables age, weight, sex, type of TB, patient category, smoker, HIV, smear result of TB disease are significant covariates at 5% level of significance. So we interpreted as these covariates are more risk factors for the death of TB patients. After fitting the final model the multivariable Cox proportional hazard model is given as follows.

$h(t) = h_0(t) \exp(0.7938\text{age} - 0.6155\text{Weight} - 0.5707\text{sex} + 0.6029\text{HIV} - 0.7606\text{patient category} + 1.0753\text{type of TB} + 0.8682\text{smear result} + 1.0106\text{smoker})$. The final multivariable Cox PH model concluded that: The variables weight, Age, sex, HIV, patient category, type of TB, smear result, smoking are the more risk factor for the death of TB patients, those variables at 5% of significance level affects the survival of TB patients based on the data from Adama hospital, Ethiopia.

Model checking

Adequacy of a fitted model needs to be assessed after a model has been constructed. It is desirable to determine whether a fitted

Cox PH regression model adequately describes the data or not. This includes a test for violation of the assumption of proportional hazards and measuring the overall goodness of fit of the model.

Assessing the goodness of fit of the model

In our survival regression analysis assessment of model adequacy we must. i). test the assumption of proportional hazards and ii). We also need to check the goodness of fit of the proportional hazards model based on the empirical data. The statistic used to determine the overall significance of a Cox model is called the likelihood ratio test. The likelihood ratio test compares the likelihood of the full model (the model with covariates) with the likelihood of the null model (a model which contains only the intercept or empty model). Therefore, the model fitted in this study the Likelihood Ratio, score and Wald tests are used to compare (at 5% significance level) the goodness of fit of the model.

The PH Assumption Checking

The final model is based on a major assumption that the hazards between groups are proportional. To test the assumption of

proportionality, the scaled Schoenfeld residuals and log-log survival plot have been used.

Scaled Schoenfeld residuals

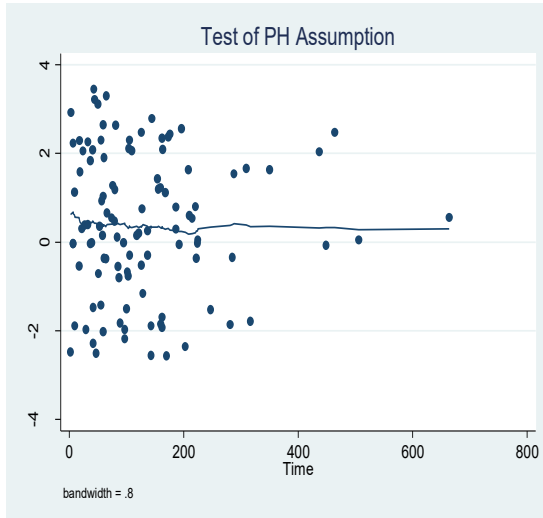
Table 7: rho statistical results for All Fitted Covariates of Cox PH Model.

Covariates	Rho	Chisq	P
Age	-0.01133	0.0168	0.897
Weight	-0.01573	0.0322	0.858
Sex	-0.20904	4.7086	0.030
Type of TB	-0.00705	0.0049	0.944
patient.cat	-0.04441	0.2092	0.647
Smoker	-0.07301	0.5596	0.454
HIV	0.17529	3.4308	0.064
smear.R	-0.12534	1.9583	0.162

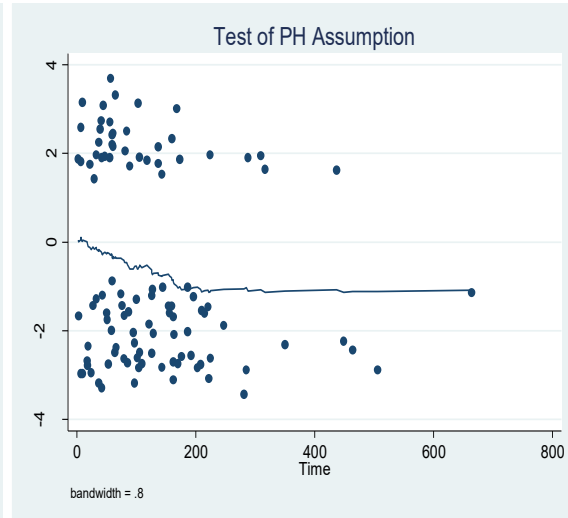
Table 7: shows that the p-value corresponding to covariates sex are less than 0.05 indicating that p-value of rho statistic is less than 5% for this covariates indicates the rejection of null hypothesis of the proportionality of Cox proportional hazard model, the assumption of proportional hazard is not satisfied for this variables. Conversely, the p-values were greater than 0.05

for all the remaining covariates like weight, type of TB, patient category, smoker , HIV, smear result Therefore, there is no enough evidence to reject the null hypothesis and satisfy the assumption of proportional hazard weight, type of TB, patient category, smoker , HIV and smear result satisfy the assumption of proportional hazard.

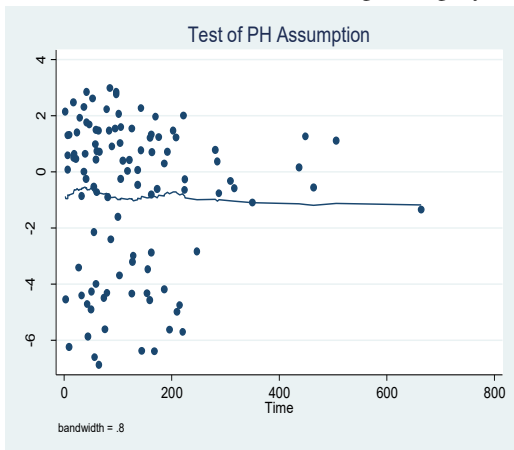
Graphical assessment of the proportional hazard assumptions



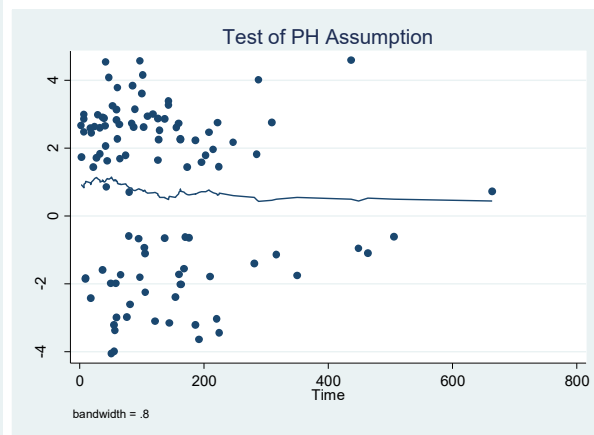
(a) Scaled schoenfeld Residual for age category



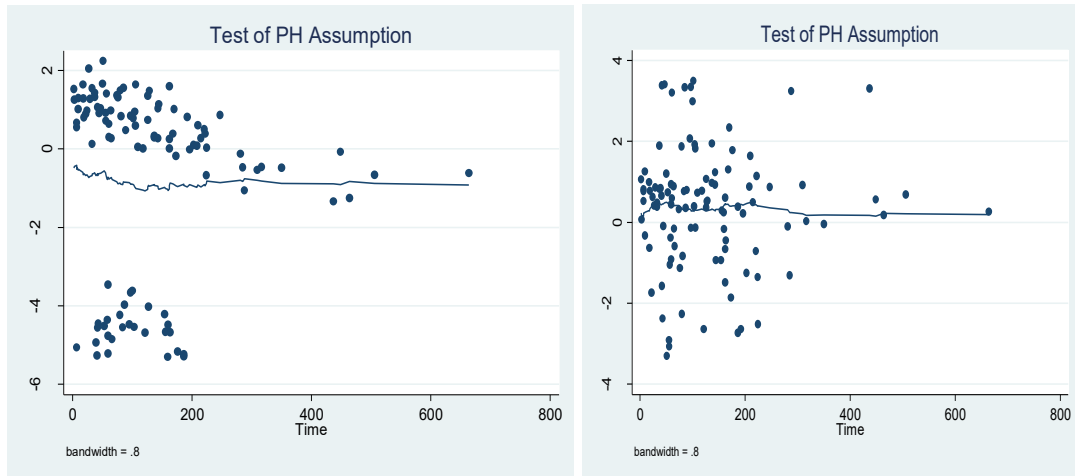
(b) Scaled schoenfeld Residual for sex category



(c) Scaled schoenfeld Residual for weight category

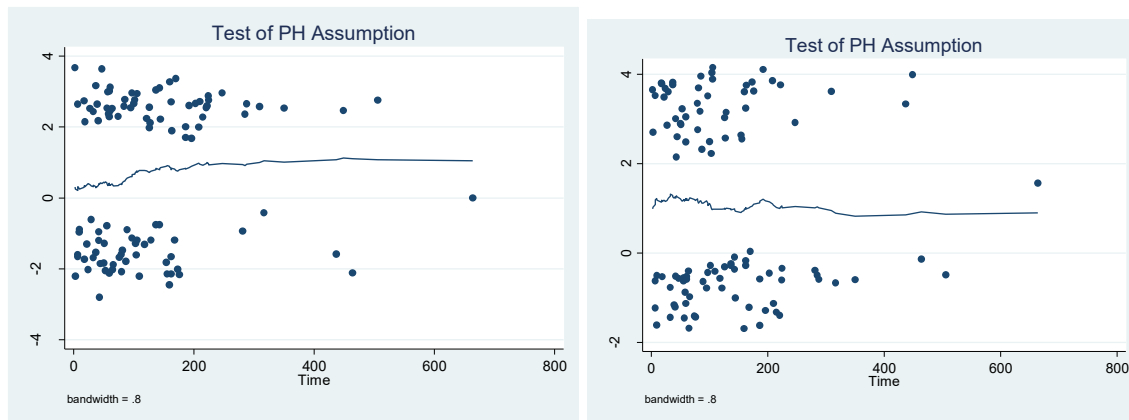


(d) Scaled schoenfeld Residual for smear result category



e). Scaled schoenfeld Residual for patient category

f) Scaled schoenfeld Residual for type of TB category



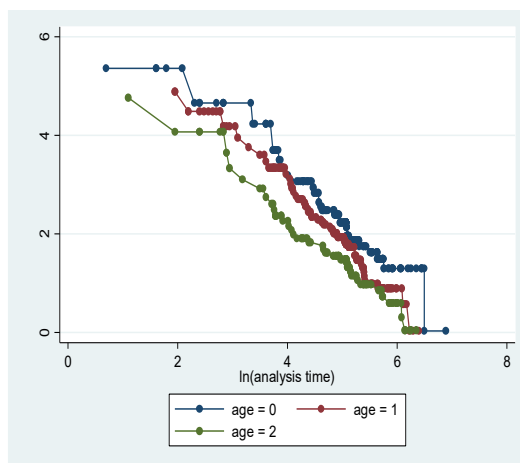
g) Scaled Schoenfeld Residual for HIV category

h) Scaled Schoenfeld Residual for smoker category

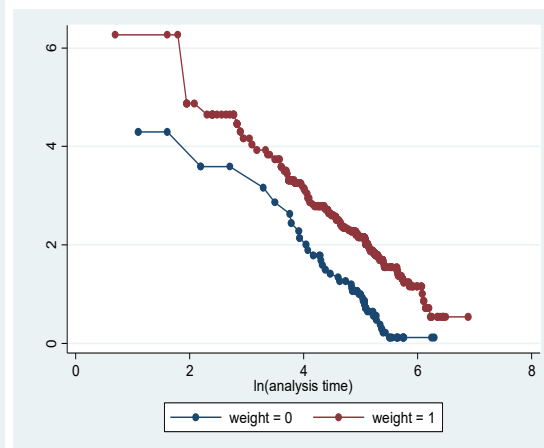
Fig 3: Plots of Scaled Schoenfeld Residuals for each covariates.

The graphical display shows plots of the scaled Schoenfeld residuals against the survival time for each covariate namely Age, sex, Weight, smear result, patient Category, Type of TB, HIV status and smoker of a patient. In Figure 3 (a, c, d, e, f, g and h) and plots of scaled Schoenfeld residuals show randomness. Moreover, the smoothed curve is an approximate

horizontal line; so the above seven covariates satisfied the assumption of proportional hazards. On the other hand Figure 3 (b) sex category of TB patients does not manifest randomness in the residuals over time and the smooth curve is not a horizontal straight line. This indicates that the sex category does not satisfy the proportional hazards assumption.



a) Log-Log Survival Plot (age category)



b) Log-Log Survival Plot (weight)

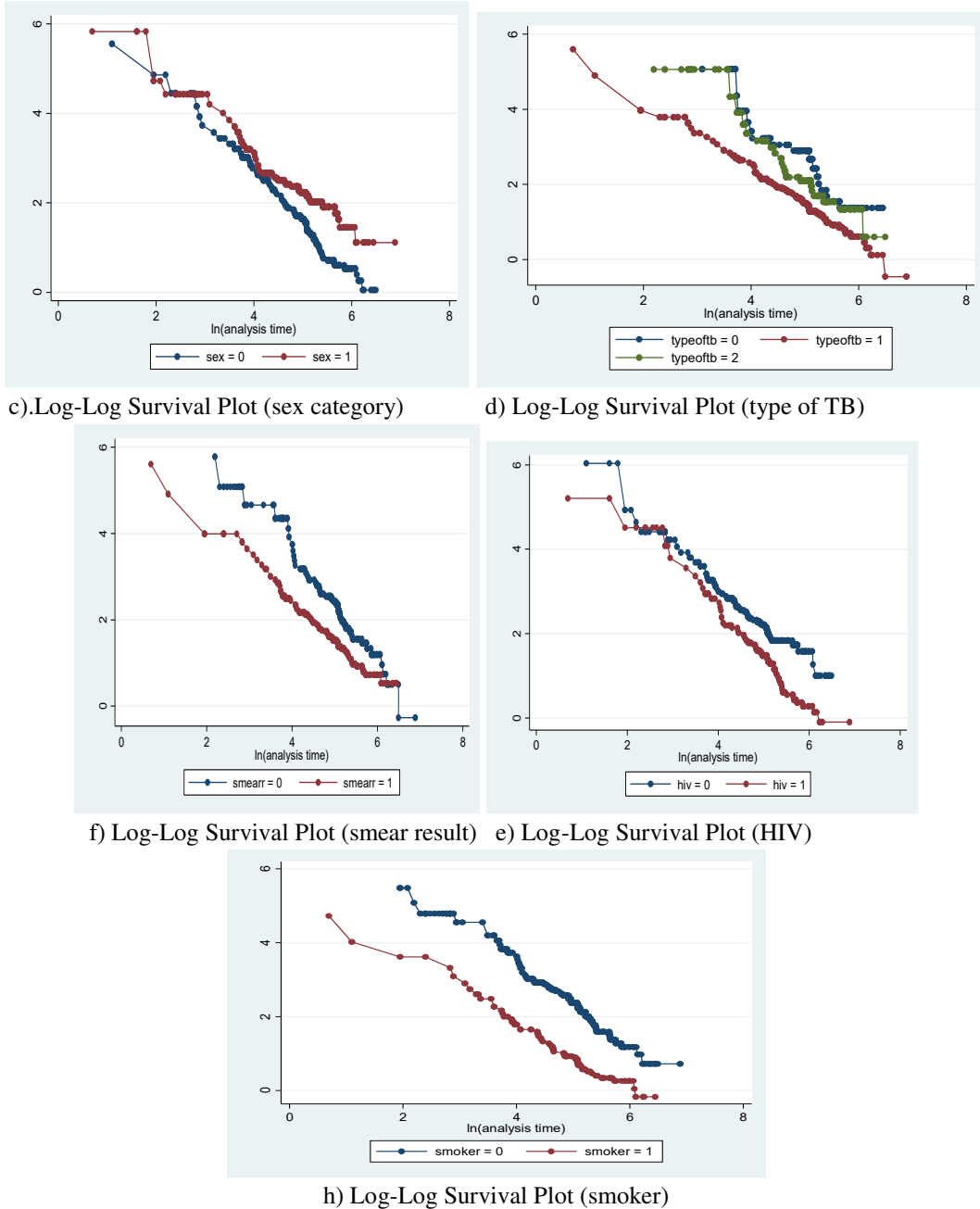


Fig 4. Shows that $\log(-\log(\text{survival}))$ plot of variable.

Figure 4 shows $\log(-\log(\text{survival}))$ plot of variables include age, sex, weight, smear result, patient category, Type of TB, HIV and smoker. For sex the plotted lines are not parallel however for age, weight, smear result, patient category, Type of TB, HIV and use Smoking the plotted lines are parallel. Although using graphs to assess the validity of the assumption is subjective, it can be a helpful tool. From the figure 3 shows that the plotted lines of the covariates age, weight, smear result, patient category, Type of TB, HIV and Smoking use are parallel this implies that assumptions of proportional hazard is satisfied over those variables. Similarly, the plotted lines of the covariates sex is not parallel, this indicates that those variables are not satisfy proportionality assumption of the Cox PH model.

The Stratified Cox Regression Model

The stratified Cox regression model is a modification of the Cox regression model by the stratification of a covariate that does not *International Journal of Basic and Applied Sciences*

satisfy the proportional hazards assumption. Covariates that are assumed to satisfy the proportional hazards assumption are included in the model, whereas the predictor being stratified is not included. As we attempted to show in the previous sections, the proportional hazards assumption is not satisfied for the covariate in the data of TB patients. So using stratified Cox regression model in which the stratification is done by using the sex category, as it is fixed by design. The only change in the model is the addition of strata statement together with the time dependent covariate in the R program and exclusion of that particular variable (category of sex patients) from the model. Stratification entails fitting separate baseline hazard functions across strata (in this study sex). A baseline hazard function represents the hazard rate over time for an individual with all modeled covariates set to zero. There are Interaction and No-interaction models defined in the stratified Cox regression model. The stratified Cox regression with No-interaction model, the

coefficients on the included covariates are common across sex categories in the No-interaction model so that the relative effect of each predictor is the same across sex. But, in the Interaction model the coefficients on the included covariates are different across (strata) sex categories. As we have attempted to show in

the previous section, the proportional hazards assumption is not satisfied for the covariate sex. Hence, the Stratified Cox regression model is used to obtain the estimated coefficients of the remaining covariates after stratification by sex.

Table:8. Result of Stratified Cox regression model use sex as strata using stata application.

Covariate	Category	Haz. Ratio	Std. Err.	P> z	Lower 95% CI	Upper CI 95%
Age	Age0®					
	Age1	1.463922	.3837623	0.146	.8757466	2.447134
	Age2	2.275414	.6461667	0.004	1.304182	3.969928
Weight	Weight0®					
	Weight1	.5123316	.1267972	0.007	.3154183	.8321764
Smear result	Smear result0®					
	Smear result 1	2.310718	.5003413	0.000	1.511596	3.532305
Patient category	Patient cat 0®					
	Patient cat 1	.4496753	.108959	0.001	.2796712	.72302
Type of TB	Type of TB0®					
	Type of TB1	2.912258	.8765826	0.000	1.614434	5.253387
	Type of TB 2	2.32563	.8239565	0.017	1.161355	4.657106
HIV	HIV 0®					
	HIV1	1.850373	.3804214	0.003	1.236685	2.768594
Smoker	Smoker 0®					
	Smoker 1	2.868362	.5903375	0.000	1.916232	4.293582

Log likelihood full = -465.31954 Prob > chi2 = 0.0000 Log likelihood Null Model= -522.31881

Parametric Regression Modeling for the Survival of TB Patient

For the data on TB patients the parametric models were fitted. Table shows the result of AIC to compare best fitted parametric

regression model .Hence, Weibull regression model has the least AIC value which shows that the data of TB patients fits for the Weibull regression model.

Table 9: The AIC values for different parametric regression models

Model type	AIC value
Exponential	1575.501
Weibull	1569.933
Log-logistic	1585.199
Lognormal	1588.677

The result of the weibull regression model in Table 9: are interpreted in terms of hazard ratios (HR), sometimes called risk ratios. The coefficient of the categorical covariates is interpreted as the logarithm of the ratio of the hazard of death to the baseline (reference group) hazard. That is, they are interpreted by comparing the reference group with others. Similarly, the coefficient for a continuous explanatory variable indicates the estimated change in the logarithm of the hazard ratio for a unit increase in the value of the respective covariate when the remaining covariates in the model are controlled. The parameter estimates of coefficients for the covariates in the final Weibull regression model along with the associated standard error, significance level and hazard ratio. The risk factors those were statistically significant included in the final Weibull regression model for the prediction of survival probability of TB patients. Survival time of TB patients was significantly related with sex categories, Age, weight, HIV status, smoker, smear result, Type of TB and Patient category. The estimated hazard ratio of death for female patients with TB diseases as compared to those male is 0.5625852 implying that the risk of death for those female patients is 0.5625852 times less than relative to those male

patients (reference category) controlling other covariates in the model.

The estimated hazard ratio of death for age grater than45 patients with TB diseases as compared to those age less than 24 is 2.199777 implying that the risk of death for those age grater than45 patients is 2.199777 times higher relative to those age less than 24 patients , In addition, the estimated hazard ratio of death for age between 24-44 patients with TB diseases as compared to those age less than 24 is 1.429397 implying that the risk of death for those age 24-44 patients is 1.429397 times higher relative to those age less than 24 patients (reference category) , Pulmonary negative (category III) compared to pulmonary positive is = 2.316292. This result indicates that patients pulmonary negative (category III) have high is 2.316292 times higher risk of death than that of pulmonary positive patients controlling for other variables in the model. The estimated hazard ratio of death for weight above 35 patients with TB diseases as compared to those weight less than 35 is 0.5262517 implying that the risk of death for those weight above 35 patients is 0.5262517 times less than relative to those weight

less than 35 patients (reference category) controlling other covariates in the model. The estimated hazard ratio of death for smoker patients with TB diseases as compared to those non-smoker is 2.758698 implying that the risk of death for those smoker patients is 2.758698 times higher relative to those non-smoker patients (reference category) controlling other covariates in the model. The estimated hazard ratio of death for HIV positive patients with TB diseases as compared to those HIV Negative is 1.86284 implying that the risk of death for those HIV positive patients is 1.86284 times higher relative to those HIV Negative patients (reference category) controlling other covariates in the mode.

The estimated hazard ratio of death for extra pulmonary patients with TB diseases as compared to those pulmonary positive is 3.004501 implying that the risk of death for those extra pulmonary patients is 3.004501 times higher relative to those pulmonary positive patients (reference category) controlling

other covariates in the model. In addition, the estimated hazard ratio of pulmonary negative (category III) compared to pulmonary positive is = 2.316292. This result indicates that patients pulmonary negative (category III) have high is 2.316292 times higher risk of death than that of pulmonary positive patients controlling for other variables in the model. The estimated hazard ratio of death for new case patients with TB diseases as compared to those non-new case is 0.4757446 implying that the risk of death for those new case patients is 0.4757446 times less relative to those non new case patients (reference category) controlling other covariates in the model.

The estimated hazard ratio of death for smear positive patients with TB diseases as compared to those smear negative is 2.35358 implying that the risk of death for those smear positive patients is 2.35358 times higher relative to those smear negative patients (reference category) controlling other covariates in the model.

Table 4.10: using stata Results of the Final Weibull Model

Covariate	Category	Haz. Ratio	Std. Err.	P-value	Lower CI 95%	Upper CI 95%
Sex	Sex 0®					
	Sex1	.5625852	.1203247	0.007	.369943	.8555428
Age	Age 0®					
	Age1	1.429397	.3688336	0.166	.8620128	2.370239
	Age2	2.199777	.6119613	0.005	1.275207	3.794691
Weight	Weight 0®					
	Weight1	.5262517	.1285547	0.009	.326031	.849431
HIV	HIV 0®					
	HIV1	1.86284	.3790174	0.002	1.250226	2.775638
Smoker	Smoker 0®					
	Smoker1	2.758698	.5632704	0.000	1.848868	4.116255
Smear result	Smear result 0®					
	Smear result 1	2.35358	.5049053	0.000	1.545694	3.583722
Type of TB	Type of TB 0®					
	Type of TB1	3.004501	.9008125	0.000	1.669415	5.4073
	Type of TB2	2.316292	.8157906	0.017	1.161456	4.619381
Patient category	Patient cat 0®					
	Patient cat 1	.4757446	.1138758	0.002	.2975963	.7605367

Models Comparison

In survival analysis, comparisons between a numbers of possible models can also be made based on the Akaike Information criterion (AIC). The guiding principle is that the smaller value of

AIC is better the fit of the model to the data. The values of AIC can be compared across different models. The values of AIC for the Cox PH model, stratified regression model using sex as strata and parametric.

Table 11: Results of stata for Model Information Criterion

Factors	Cox Standardized variability	AIC	stratified Standardized variability	AIC	Weibull AFT Standardized variability	AIC
Age	0.48(1.497309)	1192.271	0.51(1.453587)	1038.647	0.48(1.495183)	718.1206
Sex	1.08(.4382926)	1184.457			0.11(.4386178)	710.4417
Weight	0.56(.3117734)	1176.345	0.66(.3367609)	1024.717	0.05(.3002886)	700.5525
HIV	0.535(2.365339)	1182.633	0.54(2.354511)	1027.775	0.53(2.3881)	708.1594
Type of TB	1.00(1.165854)	1200.517	0.90(1.190252)	1044.904	0.97(1.171868)	726.4019
Patient category	0.795(.3617287)	1184.393	0.09(.37798)	1030.886	0.08(.3616083)	710.2038
Smear result	0.59(1.942726)	1190.52	0.58(1.973689)	1034.758	0.59(1.902022)	717.0894
Smoking	0.55(3.720126)	1160.928	0.56(3.66019)	1007.163	0.55(3.70084)	686.899

Table.11: provides the model information criteria values for Cox PH model, Stratified Cox regression model use sex as strata and Weibull AFT model. The Weibull AFT models with Smaller AIC values are a better fitting and appropriate model. This shows that using the Weibull AFT model gives more suitable results for survival data in the presence of non-proportional hazards. Weibull AFT model showed that the variables; sex, age, weight, HIV, smoker, smear result, types of TB and patient category are statistically significant and selected as significant factors for risk of death of patients with tuberculosis diseases at 0.05 level of significance.

Table 12: Results of the Model Fit

Model	Model fitting Criterion		Likelihood ratio test		
	-2 Log Likelihood	AIC	Chi-Square	DF	Sig.
Null model	-360.93181	725.8636			
Full model	-299.88445	619.7689	122.09	10	0.0000

We use R^2 as a measure of overall goodness of weibull AFT model. As it is defined in chapter three it is given as

$$R^2 = 1 - \exp\left(\frac{2}{N}(LL_0 - LL_\beta)\right)$$

Where N=601 is the total number of observation in the model.

$LL_0 = -360.93181$ is the Log likelihood for model zero

$LL_\beta = -299.88445$ is the Log likelihood for the fitted model with p covariates.

$$R^2 = 1 - \exp\left(\frac{2}{601}(LL_0 - LL_\beta)\right)$$

$$R^2 = 1 - \exp\left(\frac{2}{601}(-360.93181 - (-299.88445))\right)$$

$$R^2 = 1 - \exp\left(\frac{2}{601}(-61.04655)\right)$$

$$R^2 = 1 - \exp(-0.2031499)$$

$$R^2 = 1 - 0.81615589$$

$$R^2 = 0.18384411$$

A perfectly adequate model has low R^2 due to high percent of censored data (Cox, 1972). Thus, the value of R^2 statistic is 0.18384411 which is less than 50% shows that weibull AFT model is good fit for tuberculosis data.

A perfectly adequate model has low R^2 due to high percent of censored data (Cox, 1972). Thus, the value of R^2 statistic is 0.18384411 which is less than 50% shows that weibull AFT model is good fit for tuberculosis data.

Discussion

This study tries to estimate and compare the survival probability with a given time of TB patients and to determine major predictive factors on the survival of TB patients. The fitted weibull AFT model analysis was found that the survival of TB patient is significantly related with Age, weight, sex, Types of TB, Patient category, smoking, HIV status and Smear result of TB patients. The fitted weibull AFT model using complete case analysis found eight variables that jointly serve as predictive effect on the survival of TB patients.

Results obtained from current study were 601 TB patients registered during study time and among registered patients 17.6% died and 82.4 % were censored. According to the findings before in Addis Ababa selected six health center by (Tolosie, et

Overall Goodness of Fitted model

The researcher also needs to check the goodness of fit of the weibull AFT model based on the Tuberculosis data from Adama hospital. The model fit statistics for the final model by comparing the value -2LogL for the final proportional hazards model with that of the null model (a model without covariates) we found that the final model reduces this value by a large amount. Wald and score test to assess the adequacy of the model by testing the hypothesis that all the parameters of the final model are zero.

al., 2014); in hawassa city by (Enquoselassie, 2015); in Dangila Woreda in Northwest Ethiopia by (Abayneh Birlie, et al., 2015); in Dassei in northern part of Ethiopia by (Yeshiwork Beyene, et al, 2016). In Hawassa City and at Yirgalem Town Health Centers (Ashanfi, et al., 2014) were percentage of the deaths 12.7%,14.4%,7.4%,9.3% and 58.3% (died months of the start of treatment while the entire patients within 2 months of the start of treatment while the survival in the sixth month warranted early diagnosis) Respectively.

A median survival time of TB patients is 664 in adama hospital. A study conducted by (Tolosie, et al., 2014: Fantaw, et al., 2018: Abayneh Birlie, et al., 2015) 168, 400.5, 60 Showed that median survival time to death among TB patient respectively. In over all Kaplan Meier survival curves show that most of the event/deaths occurred in the earlier months of TB treatment initiation. Similar studies were done by (Ashenafi, et al., 2014) in hawassa and in dangla by (Abayneh Birlie, et al., 2015) the deaths of TB patients occur in the first two months of the treatment period.

Age of patients was one of a prognostic factor that significantly predicts the survival time of TB patient. The hazard rate of the TB patients having age more than 45 years was higher risk for death than patients having age less than 24 years. Our results is in concordance with studies done by (Tolosie, and sharma, 2014: Abayneh Birlie, et al., 2015: Enquoselassie, 2015 and NTBP) this indicate that there is closed relation between age and the survival of TB patients. We conclude that higher age is more hazard than lower age patients. Our study contradict with (Ashenafi, et al., 2014) there is no relation between age and the survival of TB patient.

In the current study sex of TB patient was one of the predictor of survival time of TB patients. The risk of death for those female patients is 0.5625852 times less than relative to those male patients. Our result is analogues with (Enquoselassie, 2015) the hazard rate of the male patients was 0.765 times hazard rate of female patients. Our study was contradict with study by (Tolosie, and sharma, 2014) there is no relation between the survival time of TB patient and sex. Patient category was one of the predictor of survival of TB patients. The current study revealed that the hazard rate of patients who treated for more than one times (non

new case) was higher than those patients treated for the first time (new case). The present result concurs with earlier results by (Tolosie, and sharma, 2014; Enquoselassie, 2012; Moosazadeh, et al., 2014) showing that retreated TB patients have higher hazard than new patients. Contradict early study (Ashenafi, et al., 2014).

HIV status of patients was a prognostic factor that significantly predicts the survival time of TB patients. From the total TB patients a death proportion lower for HIV negative patients (17.4%) than for HIV positive patients (18.6%) relatively seen from the categories of HIV status of the patients. a similar study were done by (Enquoselassie, 2015) in hawassa, Out of the total patients died, 113 (18%) were HIV positive and 4 (0.6 %) were HIV negative in findings of (Yeshiwork Beyene, et al., 2016) in Dasse the survival probability HIV negative test result of TB patients increased by 51.3%. Types of TB were one of the laboratories finding predictor variables used in our study. From results of our study it was one of the indicator covariates of survival time of TB patients, implying that the risk of death for those extra pulmonary patients was 3.004501 times higher relative to those pulmonary positive patients. And also a pulmonary negative TB patient (category III) was 2.316292 times higher risk of death than that of pulmonary positive patients. Our results have similar with a study done by (Fotso, et al., 2018) and (Yeshiwork Beyene, et al., 2016). But our results were contradict with a study done by (Enquoselassie, 2015; Ashenafi, et al., 2014; Tolosie, and sharma, 2014) that means types of TB were not have a significant effect on the survival time of TB patients.

In our study Initial weight of TB was one of the predictor of survival time of TB patient's. A TB patient of initial weight was less than 35 less survival time than above 35 initial weights. A similar study was conducted by (Fantaw, et al., 2018) initial weight of TB patient significantly associated with mortality of TB patients. This study contradicts with (Kabtamu Tolosie, et al., 2014) identified that body weight at initiation of treatment was not the risk factor for death in tuberculosis patients in multivariable Cox proportional regression model.

In our study smoking is one of the predictor of survival time of TB patient's. The risk of death for those smoker patients is 2.758698 times higher relative to those non-smoker patients. Similar to our finding was done by (Molalign, and Wencheko, 2015) Smoking is increase the risk of tuberculosis 2-4 times higher than that of non-smokers. Fantaw, et al., (2018) similar study with current study.

In our study smear result is one of the predictor of survival time of TB patient. The risk of death for those smear positive patients was 2.35358 times higher relative to those smear negative patients. Similar study was conducted by (Abhijit, et al., 2009). Our study was Contradict with (Enquoselassie, 2015; Ashenafi, et al., 2014; Tolosie, and sharma, 2014).

The remaining variables family size, MDR-R, Anti -TB drug, work place of TB patients are statistically not significant in our study. But, different studies reported that family size were a significant risk factor for death in TB patients (Fotso, et al., 2018). MDR-R

was a significant risk factor for death in TB patients (Fantaw, et al., 2018). Anti -TB drug were a significant risk factor for death in TB patients (Fotso, et al., 2018). Work place was a significant risk factor for death in TB patients.

In our study we use Weibull AFT model to identify the determinant factor those facilitate death for TB patients, to compare survival time of patients in categories and to estimate median survival time of TB patient during in the last five years in Adama hospital. The data follows Weibull AFT model by smallest AIC value. Our study was more or less similar in model use in Dasse Referral Hospital by (Yeshiwork Beyene, et al., 2016), in Addis Ababa health centres by (Tolosie, and sharma, 2014) and (Enquoselassie, 2015) in hawassa, Moreover another finding in survival data uses both Cox and Weibull distribution from those (Derbachew Asfaw, et al., 2015).

Conclusion

The current study in a case of Adama hospital indicated median survival time of TB patient was 664 days, with probability 0.27 (27%) TB patient died before end of study and The weibull AFT model results analysis indicated that eight covariates age, sex, type of TB, patient category, HIV status, initial weight, smear result and smoking were significantly associated with death among TB patients. Family size, work place, MDR-TB and Ant-TB drug were not significantly associated with factors that are affecting the survival of patients with tuberculosis.

From the Kaplan-Meier survival estimates there was a significant difference in survival between categories of sex, initial weight at initiation of treatment, TB patient category and smoker. However, there were no differences among survival curves of HIV status, smear result, type of TB and age. The log-rank test also showed that, there were significant difference in survival experience between the various categories of age, sex, initial weight, HIV status, and patient category, type of TB, smoker, and smear result.

Recommendations

We suggested the following recommendations:

- Most deaths occurred in the early months of TB treatment. Old age, TB/HIV co-infection and a baseline body weight of <35 kg increased the mortality during TB treatment. Therefore, a special follow up of TB patients during the intensive phase, of older patients and of TB/HIV co-infected cases, as well as nutritionally supplementing for underweight patients may be important to consider as interventions to reduce deaths during TB treatment.
- Strengthen to follow-up of patients with TB treatment from the day of anti-TB treatment initiation to completion days.
- Paying close attention to individuals with disease especially HIV co-infected and non-new TB case can prevent and decrease the death rate among TB patients.
- The country should strengthen the TB case detection and treatment programs at community level to achieve its targets during the Sustainable Development Program.

References

Abayneh Birlie, A., Tesfaw, G., Dejene, T. and Woldemichael, K., 2015. Time to death and associated factors among

- tuberculosis patients in Dangila Woreda, Northwest Ethiopia. *PLoS One*, 10(12), p.e0144244.
- Abdel, M.A.A. and Zakria, S.H.E.A., 2015. Modeling survival data by using Cox Regression Model. *American Journal of Theoretical and Applied Statistics*, 4(6), pp.504-512.
- Abhijit Mukherjee, R.S. and Saha, I., 2009. Comparing outcomes in new pulmonary sputum positive and sputum negative cases under RNTCP in rural India.
- Akaike, H., 1974. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike* (pp. 215-222). Springer, New York, NY.
- Anteneh, Y. and Kebede, D., 2017. Prevalence of Tuberculosis and its Associated Risk Factors Among Patients Visiting Bishoftu Hospital, Oromiya Region, South Eastern Ethiopia (Doctoral dissertation, Haramaya University).
- Ashenafi, S., Weldegerima, G. and Romha, G., 2014. Survival analysis and associated risk factors of tuberculosis in-hospital patients' death in Hawassa city and at Yirgalem town health centers. *World Journal of Medical Sciences*, 11(3), pp.382-388.
- Ata, N. and Sözer, M.T., 2007. Cox regression models with non proportional hazards applied to lung cancer survival data. *Haceteppe Journal of Mathematics and Statistics*, 36(2), pp.157-167
- Ata, N. and Sözer, M.T., 2007. Cox regression models with non proportional hazards applied to lung cancer survival data. *Haceteppe Journal of Mathematics and Statistics*, 36(2), pp.157-167.
- Barlow, W.E. and Prentice, R.L., 1988. Residuals for relative risk regression. *Biometrika*, 75(1), pp.65-74.
- Burnham, K.P. and Anderson, D.R., 2002. Model selection and.
- Collett, D., 1994. Model checking in the proportional hazards model. In *Modeling Survival Data in Medical Research* (pp. 149-198). Springer US.
- Cox, D.R. and Oakes, D., 1984. Analysis of Survival Data.
- Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), pp.187-202.
- Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), pp.187-202.
- DIZAJI, M.K., Kazemnejad, A., Tabarsi, P. and Zayeri, F., 2018. Risk Factors Associated with Survival of Pulmonary Tuberculosis. *Iranian journal of public health*, 47(7), p.980.
- Enquoselassie, F., 2015. Parametric Survival Modeling of Risk Factor of Tuberculosis Patients under DOTS Program at Hawassa Town, Ethiopia.
- Fantaw, D., Feyissa, M., Hamid, S. and Shibeshi, W., 2018. Assessment of the Survival Status and Risk Factors for the Mortality among Multidrug Resistant Tuberculosis Patients at Adama and Bishoftu General Hospitals, Oromia, Ethiopia: A Retrospective Cohort Study. *Adv Pharmacoeconomics Drug Saf*, 7(220), pp.2167-1052.
- Fisseha, H.Z., 2018. Trends and Treatment Outcome of Tuberculosis Patients in Assella Hospital from July 2011-June 2017, Oromia, Ethiopia, Seven Year Retrospective Study. *J Tuberc Ther*, 3(112), p.2.
- Fisseha, H.Z., 2018. Trends and Treatment Outcome of Tuberculosis Patients in Assella Hospital from July 2011-June 2017, Oromia, Ethiopia, Seven Year Retrospective Study. *J Tuberc Ther*, 3(112), p.2.
- Fotso, C.B., Vasaikar, S.D. and Apalata, T., 2018. Outcomes and Risk Factors Associated with Drug Resistant Tuberculosis in Rural Eastern Cape, South Africa. *AMERICAN JOURNAL OF INFECTIOUS DISEASES*, 14(2), pp.69-76
- Gelaw, S.M., 2016. Socioeconomic factors associated with knowledge on tuberculosis among adults in Ethiopia. *Tuberculosis research and treatment*, 2016.
- Gill, R.D., 1984. Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association*, 79(386), pp.441-447.
- Hosmer, Jr, D.W. and Lemeshow, S., 1999. Applied survival analysis: regression modeling of time to event data (1999). *Eur Orthodontic Soc*, pp.561-2.
- Hosmer, Jr, D.W., Lemeshow, S. and May, S., 2008. Applied survival analysis: regression modeling of time-to-event data (Vol. 618). Wiley-Inter science.
- Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), pp.457-481.
- Klein, J.P and Moeschberger., (1997).: *Survival Analysis: Techniques for Censored and Truncated Data*.
- Kleinbaum, D.G., *Survival analysis: a self-learning text*. 1996. View Article.
- Lawless, J.F., 2003. *Statistical models and methods for lifetime data*. Hoboken, New Jersey: A John Wiley and Sons.
- Lenjisa, J., Tolosa, B.B., Woldu, M.A., Negassa, D.E. and Wakjira, G.B., 2015. Assessment of tuberculosis retreatment case rate and its treatment outcomes at Adama Hospital Medical College, East Showa, Ethiopia. *Journal of Steroids & Hormonal Science*, 6(1).
- Leo Alexander, S.Thobias and K. Siva Naga Raju (2015). Frailty models to analyze sputum conversion of Tuberculosis Patients. *Journal homepage: <http://www.journalijar.com>*
- Molalign, S. and Wencheko, E., 2015. Risk factors of mortality in patients with multi-drug resistant TB. *Ethiop J Health Dev*, 29.
- Moosazadeh, M., Nasehi, M., Bahrapour, A., Khanjani, N., Sharafi, S. and Ahmadi, S., 2014. Forecasting tuberculosis incidence in Iran using box-jenkins models. *Iranian Red Crescent Medical Journal*, 16(5).
- Nawumbeni, N.D., Luguterah, A. and Adampah, T., 2014. Performance of Cox proportional hazard and accelerated failure time models in the analysis of HIV/TB Co-infection survival data. *Research on Humanities and Social Sciences*, 4(21), pp.94-102.
- Rao, C.R., Wu, Y., Konishi, S. and Mukerjee, R., 2001. On model selection. *Lecture Notes-Monograph Series*, pp.1-64.
- Seid, M.A., Ayalew, M.B., Mucbe, E.A., Gebreyohannes, E.A. and Abegaz, T.M., 2018. Drug-susceptible tuberculosis treatment success and associated factors in Ethiopia from 2005 to 2017: a systematic review and meta-analysis. *BMJ open*, 8(9), p.e022111.
- Tauseef Ahamad (2018) Treatment Outcome of Tuberculosis Patients Under Directly Observed Treatment Short Course and its Determinants in Shangla, Khyber Pakhtunkhwa, Pakistan: A Retrospective Study: January 29, 2019, IP: 213.55.76.173
- Teni, D.A., Enquoselassie, F. and Atsmegiorgis, C., Survival Analysis of Diabetes Mellitus Patients Using Parametric, Non-Parametric and Semi-Parametric Approaches: Addis Ababa, Ethiopia.
- Tolosie, K. and Sharma, M.K., 2014. Application of Cox

proportional hazards model in case of tuberculosis patients in selected Addis Ababa health centres, Ethiopia. Tuberculosis research and treatment, 2014.

World Health Organization, 2014. Guidance for national tuberculosis programmes on the management of tuberculosis in children (No. WHO/HTM/TB/2014.03). World Health Organization.

World Health Organization, 2014. Guidance for national tuberculosis programmes on the management of tuberculosis in children (No. WHO/HTM/TB/2014.03). World Health Organization.

World Health Organization, 2017. Global tuberculosis report 2017.

World Health Organization, 2017. Report of the guideline development group meeting on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis: a review of available World health organization. Global tuberculosis report 2018.

Yeshiwork, B., Geresu, B. and Mulu, A., 2016. Mortality among tuberculosis patients under DOTS programme: a historical cohort study. BMC public health, 16(1), p.883.