ORIGINAL PAPER

A survival analysis of prognostic determinant factors of time-to-death of HIV/TB co-infected patients under HAART followed-up in a public hospital in Ethiopia

Delelegn Eshete Gebreyes

Ethiopian Environment and Forest Research Institute, Ethiopia

Abstract

Introduction: Human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) and its' damage are prevailing at a shocking level in the world, and tuberculosis (TB) also adds to this damage, which make things "Mumps on the Goiter". In this case, highly active antiretroviral treatment (HAART) plays a great role in reducing the damage, and it is a lifetime treatment to reduce HIV-related mortality and morbidity, and prolong patients' survival time.

Material and methods: A retrospective survival study was conducted among 407 HIV-positive TB co-infected patients under HAART to observe the effects of HAART treatment and other covariates for the improvement of patient's life expectancy. Appropriate survival model was selected using AIC, BIC, and log-likelihood values.

Results: Out of the total 407 patients, 120 (29.48%) experienced the event of interest. A majority (n = 74, 61.67%) of those who died of HIV/TB co-infection were males, 108 (90%) had pulmonary TB, and 12 (10%) patients suffered from extra-pulmonary TB. For the log-normal AFT model, marital status, WHO clinical stages, functional status, antiretroviral treatment (ART) regimen, religion, sqrt CD4+, and baseline CD4+ were among significant predictors at a 5% level of significance for the change in patient's lifetime.

Conclusions: From this study, AFT models presented a better fit compared with Cox regression model. Among AFT models, the log-normal AFT model was selected, and hence, the study showed that prognostic factors, such as WHO clinical stages, functional status, sqrt CD4+ counts, ART regimen, marital status, baseline CD4+ counts, and their interactions with time, were among the significant predictors for the selected model at 5% significance level.

HIV AIDS Rev 2023; 22, 2: 110-130
DOI: https://doi.org/10.5114/hivar.2022.121399

Key words: Ethiopia, retrospective study, survival model, HIV/TB co-infection, AFT models.

Introduction

Human immunodeficiency virus (HIV) and tuberculosis (TB) are the most prevalent communicable diseases in the world, especially in sub-Saharan African countries including Ethiopia, mostly because of unsafe sex and contact with airborne droplets, so-called droplet nuclei for tuberculosis (TB). Distribution rate of both HIV and TB diseases and their damages has reached a shocking level. HIV is the most important risk factor for developing active TB, and TB is the leading cause of HIV-related deaths in the world, especially in African countries, which made things "Mumps

Address for correspondence: Delelegn Eshete Gebreyes, Ethiopian Environment and Forest Research Institute, Ethiopia, e-mail: delelegneshete@gmail.com

Article history:
Received: 03.08.2022
Revised: 22.09.2022
Accepted: 22.09.2022
Published: 08.12.2022

on the Goiter". Therefore, the relation between TB/HIV co-infection is bi-directional and synergistic. Comparing people with HIV and without HIV, HIV-positives are 15-22 times more likely to develop TB [1]. The expansion of the two epidemics has become a burning issue globally, and is accountable for economic, social, and health crises in many developing countries. Out of the worldwide infectious diseases, HIV is the first and TB is the second leading cause of death [2]. TB remains a major cause of death/ illness of people living with HIV/acquired immunodeficiency syndrome (AIDS) [3].

Out of 10 million TB incidence cases estimated in 2019 globally, people living with HIV accounted for 0.82 million (8.2%) of all new TB cases, and of 1.2 million TB-related deaths, 0.984 (8.2%) million deaths resulted from TB disease among HIV-positive individuals [1].

In the African continent, out of an estimated 1.5 million deaths from TB, 214,000 (14.30%) were HIV-positive TB patients, as stated in the WHO TB online 2021 report. Moreover, among sub-Saharan African countries and in Ethiopia alone, TB was the cause of 26,000 deaths, out of which 4,000 (15.4%) were HIV-positive patients. Ethiopia has the 10th highest TB burden those refers the 22 high burden countries, which accounts for 87% of global TB cases in the world [4, 5].

As shown in a research report from Ethiopia, from June, 2015 to June, 2017, out of the total 1,830,880 HIV-positive patients who received HIV care, 1,685,303 (92.05%) were screened for TB, and 14,152 (0.84%) were found with active form TB. As compared with the Amhara region, TB prevalence rate among HIV patients was 0.6%, and the highest TB prevalence among all HIV-positive patients was found in Somali (14.5%), followed by Gambella (9.6%), and Afar (8.5%). In addition, the lowest TB prevalence was seen in Amhara (0.6%) and Oromia (0.7%) [6].

For more than 4 decades, world scientists are working intensely to find a pharmaceutical cure for HIV; however, till now, no efficient medicine has been found. Therefore, the possible way to treat HIV-positive patients is highly active antiretroviral therapy (HAART) that can improve the survival of HIV/AIDS patients by increasing CD4+ cell counts. It was proven that HAART increases the survival of HIV/TB co-infected patients and reduces their mortality [7]. Studies of Alemu *et al.* [8] and Tiruneh *et al.* [9] confirmed that if HIV/AIDS patients take ART, it enables them to reduce the risk of developing active TB disease. WHO also recommends that all TB patients with HIV should be started on antiretroviral therapy (ART), irrespective of their CD4+ counts [10], and this survival time of patients was often studied by using non-parametric and parametric survival models, including Kaplan-Meier, Cox PH, and AFT models.

Many researchers identify the survival time of co-infected patients taking HAART with survival models. In case of survival analysis, time-to-event (survival time) endpoints can be applied, showing one-time event of interest. From the literature, there was a gap in following the required steps in analyzing survival data. Therefore, the primary objective of this study was to evaluate the survival time of HIV/TB co-infected patients using survival estimation models with required steps, and the association of TB and HIV-related mortality as well as to identify the determinants of mortality among HIV/TB co-infected patients in Debre Berhan Referral Hospital (DBRH) in Ethiopia.

Material and methods

Study population

Study population included all HIV/TB co-infected patients under HAART, who were diagnosed with TB disease in the Referral Hospital. Therefore, all HIV/TB co-infected patients in DBRH, with ages above 18 years old during the startup of their HAART since 2005, and those individuals treated for TB during HAART initiation or afterwards, up to December, 2016 were included in the current study.

Ethical consideration permission to undertake this study was obtained from Ethical Review Board of the Debre Berhan University College of Natural and Computational Science during attending the authors' MSc degree at this university. An official letter of permission and co-operation was written by Statistics Department to the DBRH, and the ethical committee of the hospital allowed for patients data to be analyzed. Any personal information regarding study subjects were replaced by a number and patients' evidence was kept confidential after data collection from clinical charts. Training on the objectives of the study was provided to the data collector to gather data accurately and honestly.

Study design

This was a retrospective cohort study conducted among 407 HIV-infected TB patients (using the survival data sample size determination rule), treated in the hospital between January 1, 2005 and December 30, 2016. Study design was employed as a part of advanced clinical monitoring of ART with a time-to-event/death study design to obtain a reasonable estimate of prognostic covariates of HIV/TB co-infected patients included in the study. All patients who started HAART on or after January 1, 2005 were included. Data collection from ART database and chart review were done retrospectively for all ART-experienced and TB patients. All data collection was performed after obtaining ethical clearance from regulatory organizations. In case of time-to-event data, survival data analysis was applied to compare survival patterns of HIV/TB co-infected patients due to long effect of HAART drugs. Adjusted effects of different prognostic factors on time-to-event/death were generated using Cox PH and/or AFT models.

Source of data

Using a standardized data collection format, data of 407 HIV/TB co-infected patients who were under HAART were obtained from the DBRH and analyzed in this study. Study population consisted of all HIV/TB co-infected patients,

who were started on ART at any time between January 1, 2005 and December 30, 2016. A retrospective data collection method was applied to collect all patients' epidemiological, laboratory, and clinical information obtained from ART intake forms, lab requests, follow-up forms, anti-TB record forms, ART database, and ART follow-up charts.

Variables used in the study

Response variable

The outcome variable considered in this study was time-to-event/death (survival time of patients) measured in months from the date of starting ART to the date of death. **Time-to-event/death** was defined as the time-to-event/death from HIV/TB co-infection and the status of an event (event or censored). *Status of an event* was the possible value of an event with two categories: either event/death or censored. *Time-to-event (event of interest)* was defined as the number of days, months, and years from the date of enrollment to ART clinic until the events of interest (event/death) occurred. *Censoring* was a condition, in which the value of a measurement or event or observation was only partially known due to the occurrence of "dropped out", transferred out to other health care centers, lost to follow-up, and "end of study". This was computed as the time difference between the time of ART initiation and the event of death occurred. The event "Death" other than HIV/TB co-infection disease was considered as censoring.

Independent variables

To meet the objectives of the study, independent variables were assessed at baseline, and included age, baseline CD4+ count, sex, residence, religion, CD4+ count, employment status, time (in months), alcohol consumption, marital status, WHO clinical stage, ART regimen, smoking status, education, type of TB, body mass index (BMI), and functional status of patients.

Method of time-to-event/ death data analysis

Survival analysis of time-to-event data was used to analyze data, in which the interest was the time to an event. It was necessary to correctly choose an appropriate time origin to determine time-to-event, which had to be easily identified for all patients. In a medical context, a single time-to-event was usually the time to recurrence of a health condition, time of response to treatment, or time to death from a certain cause that can be measured in years, months, weeks, or days. In this study, death as an event was considered.

Survival data often differ from other quantitative records, because of incomplete observations. Incomplete data resulted from only knowing an event that has not occurred in a given time period, and not knowing if or when the event would occur afterwards. These kind of observations are referred to as censored observations. Censoring is the most important characteristic that distinguishes the analysis of survival times from other areas in statistics. Subjects are said to be censored if they incomplete information about their survival time due to loss to follow-up, withdrawing from the study, or if the study ends before all subjects experience an event of interest. Mostly, there are three types of censorings: right censoring, left censoring, and interval censoring [11].

Right censoring: It happens when the true observed event has occurred to the right of our censoring time. Right censoring is the most common type of censoring, which occurs when an observation is terminated before a person experience event of interest. In right censoring, only the lower bound of the time of interest is known.

Left censoring: This appears when the event of censoring has occurred prior to the start of a study. In left censoring, only the upper bound for the time of interest is known. It is the least common type of censoring, which happens when an event is known to happen before the start of a study, but the exact time is unknown.

Interval censoring: Interval censoring is considered when the event of censoring is only known to occur within a certain interval of time. This can happen if a patient survives through the experiment and is still alive when at the end of a study. It is a slightly less common type of censoring, where an individual is known to have an event between two points in time, but the exact time is unknown.

Descriptive analysis of survival data mostly uses non-parametric methods to compare survival functions and hazard rate of two or more groups. In this case, Kaplan-Meier estimator (product-limit-estimator) proposed byKaplan and Meier [11] is commonly used to estimate the survival function and hazard rate of an individual among groups. We have discussed the survival function and hazard rate as well as their estimation method the so called Kaplan-Meier estimation as follows.

Survival function (S(t))

Survival function (*S(t)*) is a decreasing step function having jumps on the occurrence of an event, and gives the probability that an individual would survive beyond any specified time *t*; it indicates the proportion of individuals, for which the event of interest has not yet happened by time *t*. Assuming *T* is a random variable representing survival time, which has some cumulative distribution function, *F(t)*. *S(t)* is defined as the probability that an individual survives longer than *t*, where S (T = t) = P (an individual survives longer than t) = P (T > t) = 1 – P (an individual fails before t) = 1 – F(t) =1 – p (T < t) = $1 - \int_{t=0}^{t} f(u)\,du$ or $1 - \sum_{t=0}^{t} f(t)$ and takes values in between [0, 1], which equals to 1 at t = 0 and 0 at t = ∞. Generally, the survival function is most useful for comparing the survival progress of two or more strata.

Hazard rate (h(t))

Hazard function *h(t)* is the conditional failure rate (event rate) of an individual. Also, it represents the limit of the probability that an individual fails in a very short interval

(t, t + Δt),given that the individual has survived to the beginning of the time interval *t*:

$$h(T=t)=\lim_{\Delta t\to 0}\frac{1}{\Delta t}p$$ [an individual fails in the time interval (t, t + Δt) given that the individual has survived to time t]

$$=\lim_{\Delta t\to 0}\frac{1}{\Delta t}p\,[t\le T<t+\Delta t/T\ge t]=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{p[t\le T<t+\Delta t\cap T\ge t]}{p(T\ge t)}$$

$$=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{p[t\le T<t+\Delta t]}{s(t)}=\frac{f(t)}{s(t)}=-\frac{s'(t)}{s(t)}=-\frac{d}{dt}[\log(s(t))] \quad (1)$$

This hazard rate (*h(t)*) is a rate of event occurrence per unit of time, which equals to density of events at *t* (*f(t)*), divided by the probability of surviving to that duration *t* (*s(t)*) without experiencing the event, and is also expressed as force of mortality, instantaneous failure rate, age-specific failure rate, and mortality rate at conditional time interval. Generally, hazard function gives a more useful description of the risk of failure (hazard) at any time point.

Kaplan-Meier estimator

Kaplan-Meier (K-M) is a right continuous step function with jumps at the observed event times, and one of the most common and efficient means of non-parameter estimation methods of survival function of a given survival data. K-M is the commonest initial step in the analysis of un-grouped censored survival data's, and its' plot was used to show the survival pattern of categorical variables included in the present study. However, rank or Wilcoxon test could be applied to examine whether observed differences in survival pattern over time among groups were significant or not. Uncensored and censored information can be included in the plot of K-M by considering any point in time as a series of time interval defined by the observed survival and censoring times. If there is no censoring time, K-M is simply the sample proportion of observations with event times greater than *t*, for which each of these time intervals contain one death time,and this death time is assumed to occur at the start of the time interval. $t_1 < t_2 < \ldots < t_{1p}$ represent the observed failure times in a sample of size n, and r_i be the number of individuals at risk prior to time t_p. When the given survival data contains the censoring time information, then the K-M estimator of survival function by [12] formula is written as:

$$P(T\ge t)=S(t)^{\wedge}=\prod_{ti\le t}p_i=\prod_{ti\le t}\left[1-\frac{di}{ri}\right] \quad (2)$$

where d_i is the number of events (deaths) at the time t_i, and r_i is the number of subjects at risk at the time t_p.

In addition to K-M estimation, survival data can be estimated or modeled by semi-parametric and parametric survival regression models. In this regard, Cox PH and AFT models are the most commonly applied, which will be discussed in the next subsection respectively.

Cox proportional hazard (PH) model

Cox PH regression model is a statistical technique that was first proposed by Cox [13]. It enables to explore the relationship between the survival time and one or more explanatory variables in censored event-time (survival) data. It consists of two parts, the baseline hazard function, denoted by $\lambda_0(t)$, and the effect of parameters on the hazard function, denoted by $\exp(\beta^T X_{ip})$ that describe how the risk of event per time unit changes over time at baseline levels of covariates, and the hazard varies in response to explanatory covariates respectively. The Cox PH model is the most popular statistical model in censored survival data analysis, which means that the hazard function of one individual is proportional to the hazard function of the other person, and hence, the hazard ratio is constant overtime.

The popular Cox PH model defined as the hazard rate $\lambda_i(t)$ for the i^{th} subjects is:

$\lambda_i(t)=\lambda_0(t)\times\exp(\beta^T X_{ip})=\lambda_0(t)\times\exp(\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_p X_{ip})$ (3),

where t_i is the survival time, and $i = 1, 2, \ldots, n$, $\lambda_0(t)$ is the unspecified baseline hazard function, which is a non-negative function of time; $\lambda_i(t)$ is the hazard function, *P* shows the number of covariates included in the model, X_{ip} is a nxp matrix of observed covariates that represent the effect of the covariate on the outcome variable, so-called probability of survival of subjects, and β is a column vector of coefficients (p × 1 column) estimated by partial maximum likelihood that measures the covariates effect. A best fitted type of Cox PH model could be selected based on AIC, BIC, and log-likelihood, for which the minimum AIC and BIC values of the model is the better model. The coefficients of covariates in a Cox proportional model is associated with the survivorship of subjects under study, meaning a negative coefficient indicates a negative effect on survival, and a positive coefficient denotes a positive effect on survivorship.

As indicated in the above Equation (3), the Cox PH model has two parts, including the baseline hazard function, denoted by $\lambda_0(t)$, which is the hazard for the respective individual when all independent variable values equals zero, and the effect of parameters, describing how the hazard varies in response to predictor variables.

Equation (3) is a semi-parametric model since the baseline hazard is non-parametric and the relative risk function ($\exp^{(\beta)}$) is parametric. The covariates influence the hazard directly through a log-linear combination of covariates. The non-parametric element λ_0 of the Cox model makes the model flexible, since no specific distribution is assumed for the baseline group. Another advantage of the Cox model is the easy interpretation of regression parameters as log-relative risks. The β_1 would, for example, be the effect of x_{i1}, when the other covariates were corrected for. β_1 may be interpreted in terms of a relative risk, e.g., when the covariate x_{i1} is increased/decreased by one unit, the hazard function would increase/decrease by the amount of exp (β_1), respectively.

$$RR=\frac{\lambda_{0(t)}\times\exp(\beta 1\,(X_{i1}+1)+\ldots+\beta_p X_{ip})}{\lambda_{0(t)}\times\exp(\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_p X_p)}=\exp(\beta 1) \quad (4)$$

If $\beta_1 > 0$, the risk of dying increases as x_{i1} increases, if $\beta_1 < 0$, the risk of dying decreases as x_{i1} increases, and if $\beta_1 = 0$, there is no covariate effect for the change of hazards.

Model diagnostics for Cox PH model

The main assumption of Cox proportional hazards model is proportional (constant) hazards, meaning individual hazard functions ratio is constant over time. To verify Cox PH proportionality assumption, residuals are often used, and these survival data residuals are somewhat differed from other types of models residuals mainly due to censoring, and explained in two ways by graphical and non-graphical methods, as follow.

Graphical method-log cumulative hazard

In order to identify the absence or presence of non-proportional hazards in the model, it is common to use the log cumulative hazard (log (H(t)) plot. This procedure can be implemented by plotting *log(–log(s(t)))* versus *log(t)* for all groups in every categorical covariate, where *t* and *s*(*t*) are survival time and survival function at time *t*, respectively. If the plot does not yield parallel curves separated by β over the log (time), then the PH assumption is failed, and hence the Cox PH estimation is not an appropriate model for a given data set.

The difficulty of this approach would arise if a covariate has too many categories or infinite categories as in continuous covariate. This problem can be solved by dividing the subjects into a few categories.

Other methods, such as **Cox-Snell (generalized)** residuals can be used to assess the model fit for a given data, which provides the negative of natural log of the survival probability for each observation. **Schoenfeld** residuals is applied to test whether the slope of scaled residuals with respect to time is zero or not. If the slope of scaled residuals is not zero, then the proportional hazard assumption was violated. **Deviance** residual is used to check the symmetrical distribution of data about zero (detecting possible outliers in the data), and is the most common method used to verify the proportionality assumption of the Cox PH model.

Non-graphical method verifying the presence/absence of time-dependent covariates in the model

Time-dependent covariates are created by forming interactions between predictor variables and survival time, including them in the model, and finally observed the presence or absence of statistically significance covariates. If the interaction between time-dependent covariates and time is significant at significance level, then proportionality assumption is not observed. This shows that the values of time-dependent covariates changed via time, and there is no existence of Cox PH model proportionality assumption between covariates and time; therefore, the hazard ratio is not constant. The next model enables to assess the Cox PH model assumption by making time interaction for X_j for other covariates, which can be written as follows:

$$\lambda\,(t, \mathrm{x}\,(t)) = \lambda_0\,(t)\,exp^{(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j + \ldots + \beta_p x_p + \delta_j x_j \times t)} \quad (5),$$

where $(x_1; x_2, \ldots, x_p, x_j)$ are the values of the vector of explanatory variables for a particular individual, and $\delta_j x_j \times t$ represent the result of time and time-dependent covariates, with their values changed over time. The null hypothesis to verify proportionality is $\delta = 0$, tested by a test statistic either a Wald test or a likelihood ratio test.

A covariate is time-dependent, which mean that the covariate value change over time for an individual, and if the interaction between time and time-dependent covariate is statistically significant, it indicate that the Cox PH proportionality assumption is not fulfilled. However, the constant hazard assumption of Cox PH model is not always fulfilled by the given data, and a parametric PH model cannot be used as an alternative model to deal with the problem of non-proportionality hazards rather than considering another model, e.g., AFT model that does not assume a constant hazard.

Accelerated failure time model

The accelerated failure time (AFT) model is another and the most commonly and alternatively used parameter regression model to analyze the time to failure of an event (survival time) in the survival data analysis and modeling, when the proportional hazard assumptions does not hold and hence, it uses to understand the relationship between survival time and other covariates of studied subjects with maximum likelihood parametric estimation method, which is one of the interests of survival data analysis. The AFT model takes the logarithm of survival time as the response variable, and includes an error term that is assumed to follow a particular distribution, such as exponential, Weibull, standard gamma, or log distribution, for which the best-fitted distribution is chosen with AIC, BIC, or log-likelihood values, with the context of smaller values in AIC, BIC, and a larger value in log-likelihood is the best-fitted model.

Let T_i be a random variable denoting the failure time of an event for the i^{th} subject, and let $x_{i1}, x_{i2}, \ldots, x_{ip}$ be the values of *p* covariates, the AFT model is as follows:

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \sigma\varepsilon_i = X^T{}_i\beta + \sigma\varepsilon_i \quad (6),$$

where $logT_i$ is the log-transformed survival time of patients, $\beta_0, \ldots, \beta_p$ are $(p \times 1)$ the regression coefficient of interest, σ is the unknown scale parameter to be estimated, X_i is a $(p \times 1)$ vector of covariates with the first component equal to 1, and ε_i is the unobserved random disturbance (errors) term used to model the deviation of values of $\log_e(T_i)$ from the linear part of model with known mean and common variance σ_ε^2, usually assumed to be independent and identically distributed with some density function $f(\varepsilon)$ that could be selected based on model selection methods. Among some distributions that are assumed for ε_i are:

Exponential AFT model

The general equation of AFT model for any error distribution is:

$$\log(T_i) = \alpha + \delta\,\varepsilon_i \quad (7)$$

However, different assumption has produced error term distribution and its' coefficient, and hence, in the case of exponential error distribution (exponential AFT model), ε_i has a standard extreme-value distribution with one parameter λ, and assuming that $\sigma = 1$. If $T_i \sim$ E (Exponential) (λ), which can be written as $f(t) = \lambda e^{-\lambda t}$, and then equation (6) is equivalent to:

$$\log(T_i) = \alpha + \delta\,\varepsilon_i \quad (8),$$

where α represents the covariate effect, and ε_i has a standard extreme value (min) distribution with density function of $f_{\varepsilon_i}(\varepsilon_i) = e^{\varepsilon_i - e^{\varepsilon_i}}$. Generally, AFT model with exponential error distribution can be written as:

$\log(T_i) = -(\beta_p{}^T X_i) + \varepsilon_i = \beta_0{}^* + \beta_1{}^* X_{i1} + \ldots + \beta_p{}^* X_{ip} + \delta\varepsilon_i$ (9),

where $\beta_p{}^* = -\beta_p$ for all p.

The change in signs makes intuitive sense, and ε_i follows an extreme value distribution, which just means that follows a unit exponential distribution. If the hazard is high, events occur quickly and survival times are shorter.

Weibull AFT model

The Weibull AFT model is made from exponential AFT model with a slight modification. We retain the assumption that ε has a standard extreme-value distribution, but we relax the assumption of exponential model that σ = 1 allow to be estimated. It is a flexible model that the hazard rate can be one of the three functions, such as monotone increasing, constant, or decreasing. When σ > 1, the hazard decreases with time; when σ is between 0.5 and 1, then the hazard is increasing at a decreasing rate; when 0 < σ < 0.5, the hazard is increasing at an increasing rate; and when σ = 0.5, the hazard function is an increasing straight line with an origin at 0. Assuming $T_i \sim W$ (Weibull) (λ,p), with Weibull density function of $f(t) = \lambda^p p t^{p-1} e^{-\lambda t^p}$, where λ > 0, p > 0, *p* is shape parameter and then, equation (6) is equivalent to:

$\log(T_i) = \alpha + \delta\varepsilon_i$ (10),

where *α* represents the covariate effect, and ε_i has a standard extreme value distribution. Generally, AFT model with Weibull error term distribution can be presented as:

$\log(T_i) = \beta^{*T} X_i + \sigma\varepsilon_i = \beta_0{}^* + \beta_1{}^* X_{i1} + \ldots + \beta_p{}^* X_{ip} + \delta\varepsilon_i$ (11),

where $\beta_p{}^* = -\beta_p/p$, and σ = 1/p.

Gamma AFT model

Generally, there are two types of gamma distribution, such as the standard gamma with 2 parameters, and the generalized gamma with 3 parameters. If $T_i \sim$ standard gamma (λ, k), with density function of

$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-(\lambda t)}}{\Gamma(k)} = \frac{\lambda^k t^{k-1} e^{-(\lambda t)}}{\Gamma(k)}$ for λ > 0, k > 0,

$\Gamma(k) = \int_0^\infty s^{k-1} e^{-s} ds$,

then, $\log(T_i)$ can be developed in terms of standard gamma distribution, as it simply adds a scale parameter in the expression of log T, so that:

$\log T_i = \alpha + \delta\varepsilon_i$ (12),

where *α* is the covariate effect, and ε_i has a standard extreme value distribution. Standard gamma, Weibull, exponential, and log-normal models are all special cases of the generalized gamma model, when the following conditions are satisfied:

- Gamma, when p = 1;
- Weibull, when k = 1;
- Exponential, when p = 1 and k = 1;
- Log-normal as a special limiting case, when $k \to \infty$.

Log-normal AFT model

The log-normal model simply assumes that if $\varepsilon_i \sim N(0, 1)$ or $T_i \sim N(0, 1)$, then logTi has a log-normal distribution if:

$\log T_i = \alpha + \delta\varepsilon_i$ (13),

where T_i has a standard normal distribution with density function of

$f(T_i = t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\left\{\frac{\log(t)-\mu}{\sqrt{2}\sigma}\right\}^2}$, for which *μ* and *σ* are the mean and standard deviation, respectively.

Ethical consideration

Permission to undertake this study was obtained from Ethical Review Board of the Debre Berhan University College of Natural and Computational Science at the time of attending the author's MSc degree at this university. This official letter of permission and cooperation was written by the Statistics Department to the Debre Berhan Referral Hospital, and the ethical committee of the hospital allowed for reviewing patients' data. Any personal information regarding study subjects was replaced by a number, and patient evidence was kept confidential during and after data collection from clinical charts. Training on the objectives of the study was given to data collector for accurate and honest data collection.

Statistical data analysis

General descriptive statistics

As shown in Table 1, among the total of 407 HIV/TB co-infected patients included in this study, 196 (48.16%) were females. During the follow-up period, 120 (29.5%) patients died due to the diseases. Majority (*n* = 74, 61.67%) of those who died of TB/HIV co-infection were males, 108 (90%) had pulmonary TB, and 12 (10%) suffered from extra-pulmonary TB. Of those who died of HIV/TB co-infection, 3, 58, 20, and 39 cases were at stage I, stage II, stage III, and stage IV WHO clinical stage of HIV disease, respectively. Out of the 407 HIV-positive TB patients under ART followed-up, 153 (37.59%) were married, 222 (54.55%) were single, and 32 (7.9%) were divorced. Regarding the baseline functional status of patients, 24.57% were able to do their dayto-day activities, such as farming, harvesting, office work, and others; 13.27% and 62.16% presented bedridden and ambulatory status, respectively.

Descriptive statistics for survival status and associated factors

The survival endpoint was measured in months and defined as the difference between the date of death or censoring and the date of HAART initiation. Among the total of 407 HIV/TB co-infected patients, 120 (29.48%) experienced the event of interest, and the remaining 287 (70.52%) were censored. Out of the total of the 407 patents, 52.26% of female and 47.74% of male patients were censored. About 25.09% of those working, 64.81% with ambulatory and 10.10% with bedridden status patients were censored. In the case of baseline WHO clinical stages, 3.83%, of stage I, 51.22% of stage II, 17.77%, of stage III, and 27.18% of stage IV patients were censored. Based on the patients' marital status,

Table 1. Frequency distribution for baseline independent variables along with the censored observations of time-to-death

Characteristics/Categories	Total (%)	Censored (%)
Sex		
Female*	48.16	52.26
Male	51.84	47.74
Marital status		
Married*	37.59	35.89
Single	54.55	56.45
Others	7.86	7.66
Residence		
Rural*	8.84	9.06
Urban	91.15	90.94
WHO clinical stage		
Stage I*	3.44	3.83
Stage II	50.37	51.22
Stage III	17.44	17.77
Stage IV	28.75	27.18
Functional status		
Working*	24.57	25.09
Ambulatory	62.16	64.81
Bedridden	13.27	10.10
Employment status		
Unemployed*	21.87	23.00
Employed	78.13	77.00
Alcohol consumption		
No*	67.08	65.85
Yes	32.92	34.15
ART regimen		
Adult first-line*	25.30	25.09
Adult second-line	42.26	39.72
Child first-line	23.10	26.13
Child second-line	4.18	4.18
Others	5.16	4.88
Smoking status		
No*	88.45	88.15
Yes	11.55	11.85
Education		
Illiterate*	19.66	19.16
Primary	42.51	40.77
Secondary	31.44	33.80
Diploma and above	6.39	6.27
Type of tuberculosis		
Pulmonary*	86.48	85.02
Extra-pulmonary	13.52	14.98
Religion		
Orthodox*	62.90	63.41
Muslim	19.66	18.81
Protestant	10.56	10.81
Catholic	6.88	6.97

**Reference group for each covariate characteristics*

35.89% of single, 56.45% of married, and 7.66% of divorced cases were censored, and the remaining patients experience an event. The summary statistics of the others independent variables is presented in Table 1.

Plots in Figure 1A indicate that the overall survival of patients was in rather a good manner, because of the graph of survival probability of all patients was almost above the mean survival, and did not quickly descend to the lower probability stage. Figures 1B-D show that female patients in child first-line ART regimen and ambulatory functional status had slightly higher survival rate than male patients after fifty months of follow-up, and hence, female patients in child first-line ART regimen and ambulatory functional status presented slightly lower hazard rate than male patients. The survival times were found to be significantly different in sex groups.

Table 2 shows the summary results of continuous covariates, mean age, sqrtCD4, and baseline CD4+ count of the co-infected patients were 33.63, 1.65, and 125.3, respectively.

Cox proportional hazard model

It is commonly known that Cox PH model is applied in medical research, and used to associate the survival time of patients with covariates; first, it should be found which types of model (model with and without covariates) is an appropriate model for a collected data, as shown in Table 3.

As indicated in Table 3, the values of AIC and BIC were decreased, and the value of –2 logL was increased for the model estimated with covariate using exact ties handling method which implies that the chosen model was an appropriate model, and the covariates had a significant effect on hazard rate of individuals.

In addition, the significance of covariates in explaining the model can be also tested by global test method.

The results in Table 4 suggest that all covariates coefficients were differ from zero, and hence, there was some correlation between covariates and hazard rate. It was concluded that the model with explanatory variables was more effective than the null model, and hence, the Cox PH model based on covariates values could fit.

As shown in Table 5, the hazard ratio for all the reference groups was 1, which was the default value of the reference categories. The fitted Cox PH model can be written as:

$$\lambda_i(t/X) = \lambda_0(t)\exp(0.3608\text{Male}_i - 0.0122\text{Age}_i - 0.0076\text{Married}_i + 0.407\text{Others}_i + 0.1096\text{Urban}_i - 0.0809\text{Stage-II}_i - 0.0016\text{Stage-III}_i + 0.3394\text{Stage-IV}_i - 0.1452\text{Ambulatory}_i + - 0.7093\text{Bedridden}_i + 0.0422\text{Employed}_i - 0.3368\text{Consume}_i + 0.0043\text{Adult second-line}_i - 0.2442\text{Child first-line}_i - 0.2442\text{Child second-line}_i + 0.2832\text{Others}_i - 0.1293\text{Smoke}_i - 0.0204\text{Primary}_i - 0.6200\text{Secondary}_i - 0.2952\text{Diploma and above}_i - 0.4177\text{Extra-pulmonary}_i + 0.1779\text{Muslim}_i + 0.2635\text{Protestant}_i + 0.1841\text{Catholic}_i + 08485\text{Sqrt-CD4}_i - 0.0031\text{BaselineCD4}_i + 0.0620\text{BMI}_i).$$

As depicted in Table 5, the estimated hazard ratio of males was 1.434 and implied that the risk of death of males was increased by 43.4% compared with females. Risk of death could be also interpreted based on coefficients of

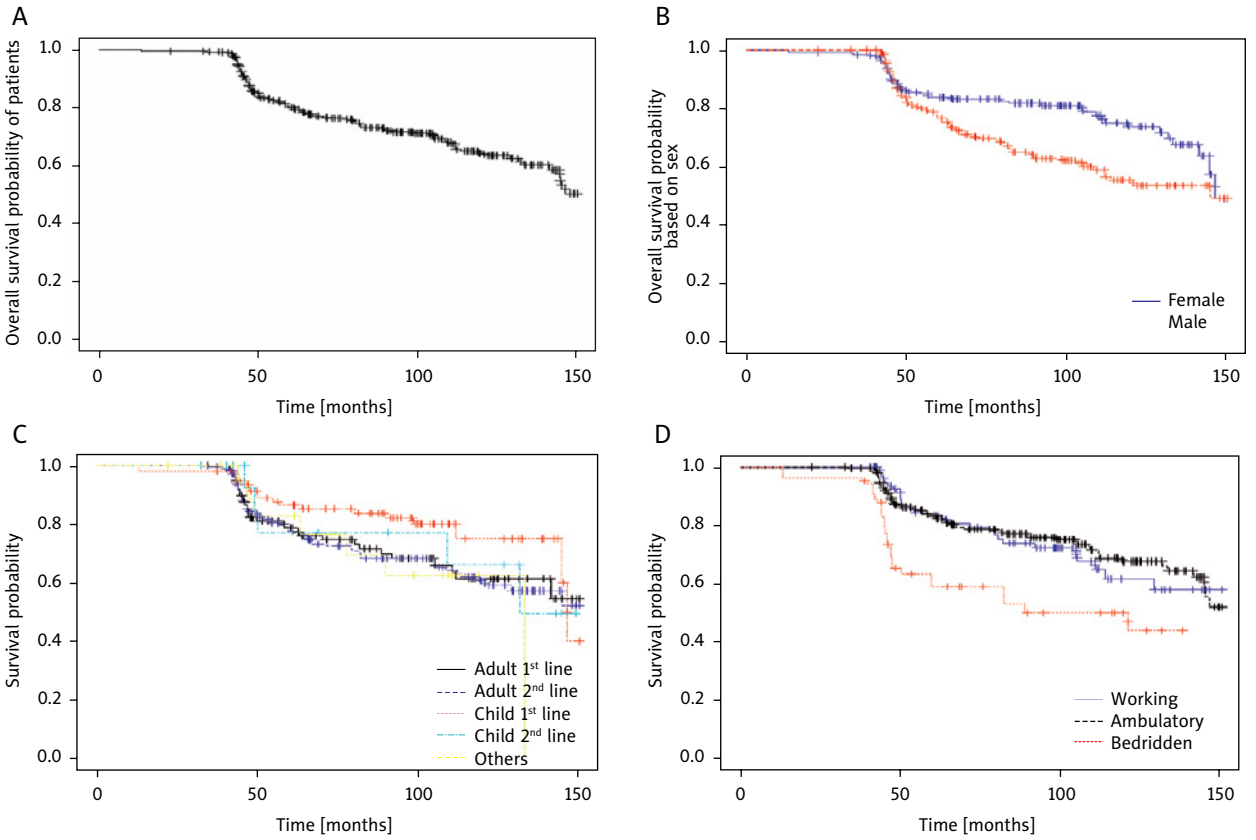

Figure 1. Kaplan-Meier survival plots of sex and antiretroviral therapy (ART) regimen of HIV-infected patients under ART

Table 2. Summary of continuous predictor variables

Covariates	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Age	18.00	27.00	32.00	33.63	39.00	66.00
Sqrt CD4	1.410	1.590	1.650	1.648	1.700	1.900
Baseline CD4+	2.00	51.00	98.00	125.3	169.0	979.0
BMI	1.20	17.04	20.20	19.47	22.56	33.28

Table 3. Model fit statistics

Criterion	With covariates	Without covariates
–2 log L	56.88	–657.17
AIC	1,261.32	1,314.35
BIC	1,311.87	1,314.35

Table 4. Testing of global null hypothesis BETA = 0

Test	χ^2	DF	Pr > χ^2
Likelihood ratio test	56.88	27	7e-04
Wald test	53.64	27	0.002
Score (log-rank) test	55.00	27	0.001

Table 5. Partial maximum likelihood estimates of Cox PH model

Covariates/Categories	Coefficients (β)	Standard error of coefficients	Hazard ratio (HR) Exp (β)	95% CI for HR	*p*-value
Sex					
Female	–	–	1.00		–
Male	0.360755	0.208484	1.434413	(0.95-2.16%)	0.04356*
Age	–0.012153	0.016759	0.987921	(0.96-1.02%)	0.46837
Marital status					
Married	–	–	1.00		–
Single	–0.007586	0.270064	0.992443	(0.58-1.68%)	0.97759
Others	0.407	0.477713	1.502561	(0.59-3.83%)	0.39403
Residence					
Rural	–	–	1.00		–
Urban	0.109635	0.370667	1.115871	(0.54-2.31%)	0.76740
WHO clinical stage					
Stage I	–	–	1.00		–
Stage II	–0.080909	0.614450	0.922278	(0.28-3.08%)	0.89524
Stage III	–0.001629	0.644368	0.998372	(0.28-3.53%)	0.99798
Stage IV	0.339431	0.623476	1.404149	(0.41-4.77%)	0.58615
Functional status					
Working	–	–	1.00		–
Ambulatory	–0.145243	0.240597	0.864812	(0.54-1.39%)	0.54606
Bedridden	0.709318	0.315480	2.032604	(1.10-3.77%)	0.02455*
Employment status					
Unemployed	–	–	1.00		–
Employed	0.042213	0.260587	1.043117	(0.63-1.74%)	0.87131
Alcohol consumption					
No	–	–	1.00		–
Yes	–0.336848	0.208357	0.714018	(0.47-1.07%)	0.10595
ART regimen					
Adult first-line	–	–	1.00		–
Adult second-line	0.004376	0.231222	1.004386	(0.64-1.58%)	0.98490
Child first-line	–0.244204	0.312166	0.783328	(0.42-1.44%)	0.43404
Child second-line	–0.112543	0.507496	0.893559	(0.33-2.42%)	0.82450
Others	0.283159	0.458018	1.327317	(0.54-3.26%)	0.53643
Smoking status					
No	–	–	1.00		–
Yes	–0.129324	0.322600	0.878689	(0.47-1.65%)	0.68851
Education level					
Illiterate	–	–	1.00		–
Primary	–0.020437	0.298255	0.979771	(0.55-1.76%)	0.94537
Secondary	–0.619954	0.324803	0.537969	(0.28-1.02%)	0.05020*
Diploma and above	–0.295195	0.438798	0.744386	(0.32-1.76%)	0.50112
Type of tuberculosis					
Pulmonary	–	–	1.00		–
Extra-pulmonary	–0.417701	0.313581	0.658559	(0.36-1.22%)	0.18285
Religion					
Orthodox	–	–	1.00		–
Muslim	0.177937	0.244482	1.194750	(0.74-1.93%)	0.46673
Protestant	0.263521	0.330892	1.301505	(0.68-2.49%)	0.42580
Catholic	0.184063	0.386996	1.202091	(0.56-2.57%)	0.63435
sqrtCD4	0.848539	1.318221	2.336232	(0.18-30.94%)	0.51977
Baseline CD4+	-0.003078	0.001143	0.996927	(0.99-1.00%)	0.00709**
BMI	0.062016	0.022966	1.063979	(1.02-1.11%)	0.00693**

**Significance at p = 0.05 level. **More significance level at 0.05 level of significance*

covariates, and hence, as the coefficient value of males was greater than zero, it suggested that the risk of dying was increased as the value of sex changed from female to male and vise-versa. The hazard ratio for a 1 year growth in the age of a patient was 0.988, and this could be interpreted as the hazard of death decreased by 1.2% for a one year growth in age. The risks of death in ambulatory and bedridden functional status were 0.864 and 2.033 times less and greater than working functional status, respectively. Based on WHO stages, the estimated HR of stage II, III, and IV were 0.922, 0.998, and 1.404, respectively. This implied that patients with stage II and stage III were 7.8% and 0.2% less at risk as compared with patients at stage I, respectively, and patients with stage IV were 1.404 times at risk than patients with stage I, or the risk of death of patients with stage IV was 40.4% more at risk as compared with patients with stage I. The estimated hazard ratio for BMI of patients was 1.063, with a 95% CI: 1.02-1.11%, which indicated that the risk of death increased by 6.3% for 1 kg/m^2 increase in BMI.

Cox PH regression model diagnosis

Graphical methods

Mostly, the graphical approaches are performed based on residuals against/versus time or to test the proportional hazard assumption of Cox PH model. The most common graphical methods are as follow.

Clog-log approach: Based on this approach, if the plot of the log (–log(s (t))) versus log (time) is parallel, then the proportional hazard assumption of Cox PH model is reasonable. As shown in Figure 2, it seems that the Cox PH assumption failed, as the clog-log (log-cumulative hazard) versus log (time) plots of all covariates were not parallel (cross-over to each other), and the Cox PH model did not adequately fit the given data.

Cox-Snell residual approach: If a plot of Cox-Snell residual is considered as a pseudoobservation time versus the cumulative hazard rate (Nelson-Aalen cumulative hazard estimator) for the goodness of fit test of Cox PH model that passes through the origin at 45º (a unit slope), then the Cox PH model adequately fit the given data set. As described in Figure 3, the plot of Nelson-Aalen cumulative hazard rate to each Cox-Snell residuals (e_i) (considered as a pseudo-observation time) that did not pass through the origin and shifted from 45 degree, which implied that the PH model did not adequately fit the data.

Schoenfeld residual approach: The scatter plot of Schoenfeld residuals, which shows the difference between each covariate observed value for an individual who failed at given risk set time minus its' expected value, are used to assess the PH assumption at every failure time. There is a separate Schoenfeld residual for each individual for each covariate. It tests the independence between each individual covariate residuals and time. Meaning, if the scatterplot of the fitted scaled Schoenfeld residuals versus time for each covariate is parallel with the horizontal reference line y = 0 and all the residuals data points are distributed around zero, then the PH model assumption holds and it could be concluded that the given Cox PH model is an appropriate model for a given data. As indicated in Figure 4, for most of the covariates, the fitted residual data points represented by blue dashed line (smooth fit of scaled Schoenfeld residual) was not parallel with the reference line y = o (red line), and residual data points represented by black doted points were not distributed around zero, and except for some covariates, it showed that the proportionality hazard assumption of Cox PH model was not satisfied as we could saw in the clog-log approach, and hence, the model did not adequately fit the data.

Deviance residual approach: The deviance residual in Cox PH model can be used in two ways, one for goodness of fit test and the other one is for proportional hazard assumption checking based on each covariate.

The goodness of fit test approach examines the influential observations in the data set as that of DF beta/s. The plot of deviance residuals distributed symmetrically about zero, and hence, this approach is used to test the effectiveness of a given model for a given data set. As shown in Figure 5, the negative values of the deviance residual were greater than its' positive values, indicating that there were more individuals that live too long as compared with their expected survival time. A very large negative or positive values showed that the presence of outliers and most of the deviance residual data points were below the horizontal line and distributed far from zero (not symmetrically distributed about zero (mean = y = 0), which indicated the weakness of Cox PH model in fitting this data set.

The second type of deviance residuals application is hazard assumption checking based on each covariate. As indicated in Figure 6, the deviance residual plots (red line) of some covariates was not parallel to the referenced horizontal line, and hence, the Cox PH proportional hazard assumption was not fulfilled in the model.

Non-graphical methods

Time-varying coefficients approach: In the case of non-graphical approach, proportional hazard assumption can be tested based on each covariate and overall model *p*-value that disproved a significant (non-significant) relationship between each individual covariate and overall Cox PH model covariates and time, which was originally proposed by Schoenfeld [14] and Harrell [15].

As shown in Table 6, the test result was not statistically significant almost for all covariates, except for alcohol consumption and education. However, the global test result showed that the overall model was statistically significant. This implies that, the coefficients of the overall model covariates changed over time and hence, the proportional hazards assumption was not fulfilled in the model.

Modeling time-dependent covariate with time interaction approach: This technique is used to include the time-dependent covariates with time interaction in the Cox PH model, and to analyze whether the interaction is statistically significant. As shown in Table 7, some covariates were statistically significant after including the time-varying covariates with time interaction, and hence, the presence of statistically significant covariates after including these interactions implies

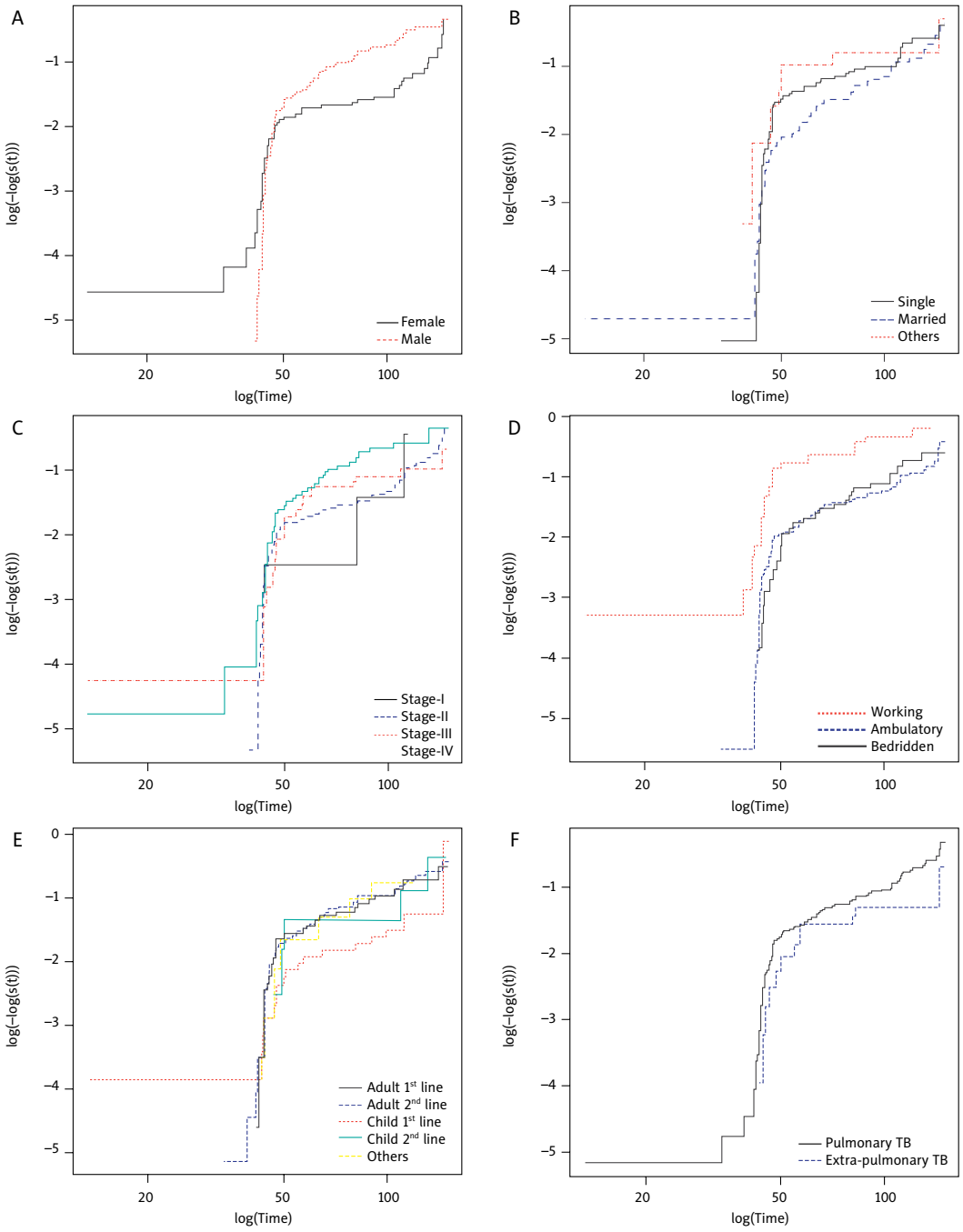

Figure 2. Clog-log versus log (time in months) plot of sex, marital status, WHO clinical stages, functional status of patients, antiretrival regimen, and type of tuberculosis

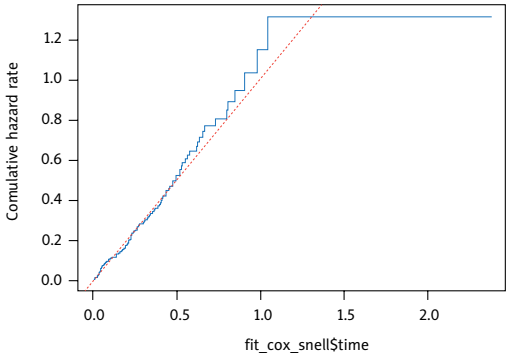

Figure 3. Plot of Cox-Snell residual for goodness of fit test of Cox PH model

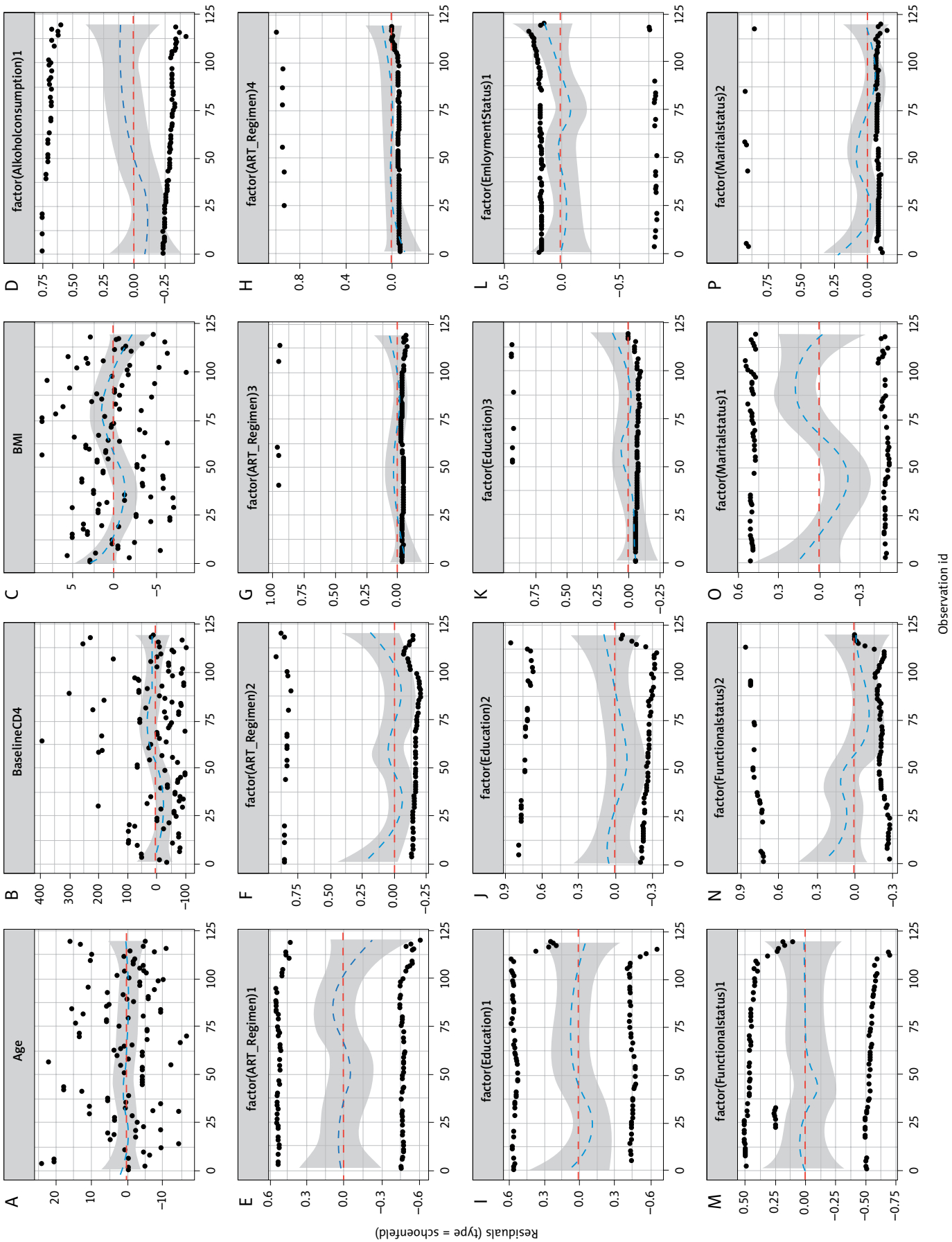

Figure 4. Scatter plot of each covariate scaled with Schoenfeld residuals against time

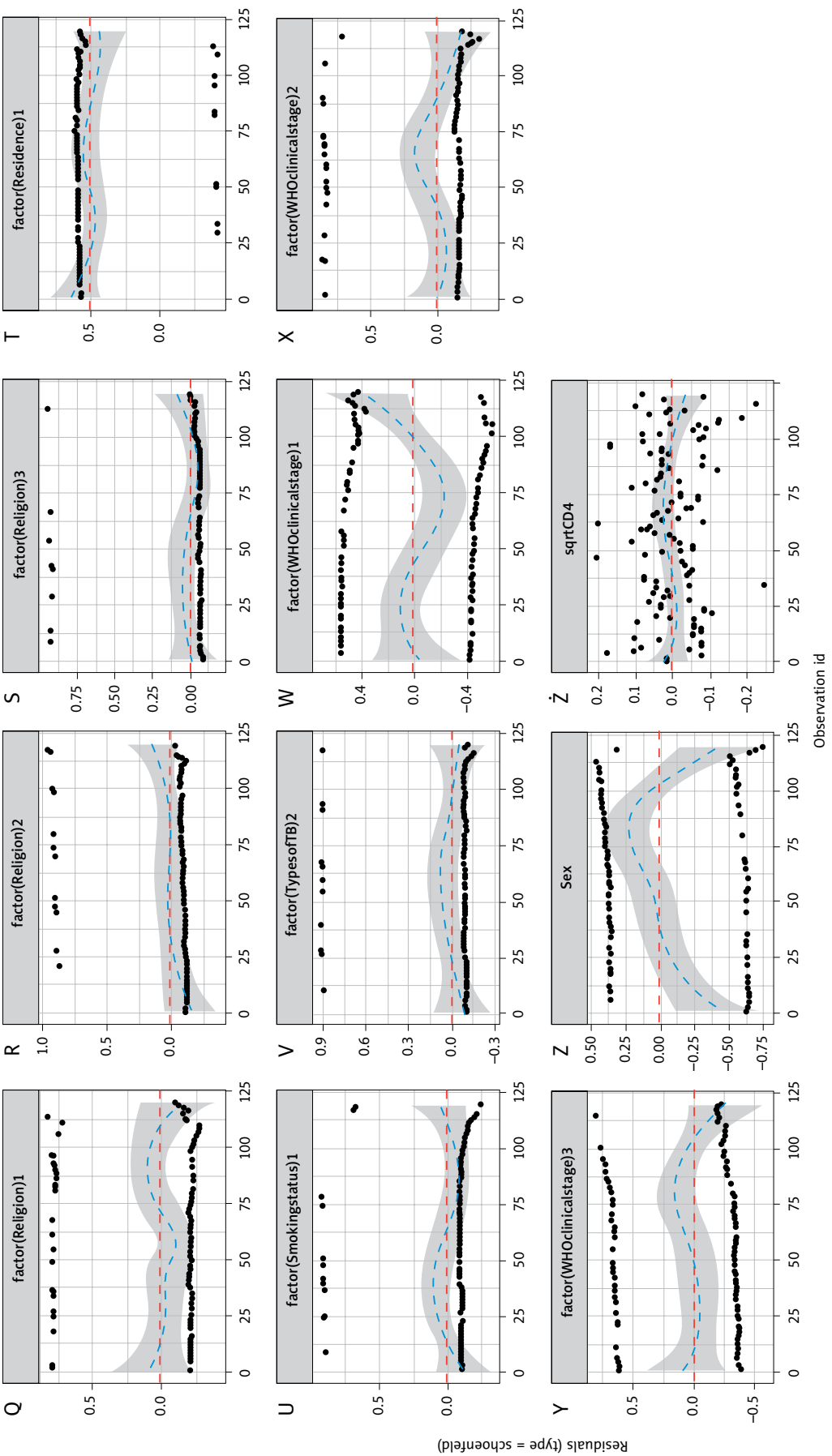

Figure 4. Cont.

that the assumption of proportional hazard in the Cox PH model was failed. Based on all Cox PH model diagnostics results, the proportional hazard assumptions of Cox PH model were not fully satisfied. Therefore, the Cox PH model was not a good model and adequately fit the co-infected patients' data; therefore, it should proceed to another best alternative model, AFT model, to fit well the given data.

Accelerated failure time model

AFT model is a parametric survival model, for which its' errors follow a certain distribution. In this case, errors may distribute as exponential, Weibull, standard Gamma, and log-normal distribution. To select the best error distribution, AIC and BIC were applied (the smaller value the better), and also maximum likelihood ratio test (the larger value the better).

As illustrated in Table 8, log-normal model seems to fit the data better, with the minimum AIC, BIC and maximum log-likelihood values of 1,534.056, 1,650.311, and -738, respectively. Log-normal model suggested as an appropriate model for the data under study. Therefore, the model was fitted using the assumed log-normal distribution survival function for survival time (T) of the co-infected patients. The parameter estimates of AFT model using log-normal distribution are given in Table 9.

Table 9 shows the results of fitted AFT model using log-normal distribution, and it was found that variables, such as marital status, WHO clinical stages, functional status, ART regimen, religion, sqrtCD$_4$, baseline CD4+, and all these covariates with time interaction and the only

Table 6. Test of proportional hazards assumption for each covariate and overall model

Covariates	χ^2 value	DF-value	*p*-value
Sex	0.54065	1	0.462
Age	0.03759	1	0.846
Marital status	2.20374	2	0.332
Residence	1.84901	1	0.174
WHO clinical stage	2.51627	3	0.472
Functional status	4.59294	2	0.101
Employment status	1.19844	1	0.274
Alcohol consumption	4.20781	1	0.040
ART regimen	4.99290	4	0.288
Smoking status	0.21576	1	0.642
Education	8.36286	3	0.039
Type of TB	0.00323	1	0.955
Religion	4.52154	3	0.210
sqrtCD4	0.13324	1	0.715
Baseline CD4+	2.10444	1	0.147
BMI	0.15716	1	0.692
GLOBAL test (overall model test)	43.84207	27	0.021

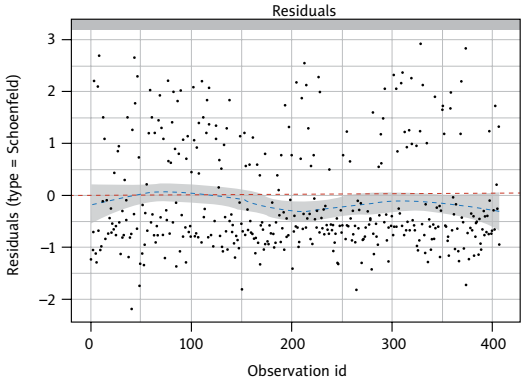

Figure 5. Mean estimation plot of deviance residual for all observations

additional covariate education with time interaction at enrollment were statistically significant. Both baseline CD4+ and sqrtCD4 had a significance effect on the survival time of the co-infected patients. In addition, the effect of ART regimen drugs were clearly shown by AFT model, which was not seen in Cox PH model; therefore, AFT model adequately fitted this data set rather than Cox PH model did, and the reduced AFT model was written as:

$$\log(\text{time}_i) = 4.30 + (3.08\text{e-}03)\,\text{Age}_i + (1.32\text{e-}01)\,\text{Maritalstatus}_{1i} + \beta_2\,(1.64\text{e-}01)\,\text{Maritalstatus}_{2i} + (-6.11\text{e-}02)\,\text{Residence} + (-1.48\text{e-}01)\,\text{WHOclinicalstage}_{1i} + (-3.83\text{e-}01)\,\text{WHOclinicalstage}_{2i} + (-3.15\text{e-}01)\,\text{WHOclinicalstage}_{3i} + (-2.32\text{e-}02)\,\text{Functionalstatus}_{1i} + (-3.26\text{e-}01)\,\text{Functionalstatus}_{2i} + (2.16\text{e-}02)\,\text{Employmentstatus}_i + (2.32\text{e-}02)\,\text{ART\_Regimen}_{1i} + (-9.64\text{e-}02)\,\text{ART\_Regimen}_{2i} + (2.24\text{e-}01)\,\text{ART\_Regimen}_{3i} + (-1.63\text{e-}02)\,\text{ART\_Regimen}_{4i} + (6.87\text{e-}02)\,\text{Smokingstatus}_i + (1.17\text{e-}02)\,\text{Religion}_{1i} + (1.88\text{e-}01)\,\text{Religion}_{2i} + (8.82\text{e-}02)\,\text{Religion}_{3i} + (-6.39\text{e-}01)\,\text{SqrtCD}_{4i} + (4.62\text{e-}04)\,\text{BaselineCD}_{4i} + (-4.06\text{e-}03)\,\text{BMI}_i + (-3.33\text{e-}05)\,\text{Time*Age}_i + (-1.73\text{e-}03)\,\text{Time*Maritalstatus}_{1i} + (-2.51\text{e-}03)\,\text{Time*Maritalstatus}_{2i} + (7.33\text{e-}04)\,\text{Time*Residence}_i + (1.79\text{e-}03)\,\text{Time}_i\text{*WHOclinicalstage}_{1i} + (5.86\text{e-}03)\,\text{Time}_i\text{*WHOclinicalstage}_{2i} + (4.68\text{e-}03)\,\text{Time}_i\text{*WHOclinicalstage}_{3i} + (1.30\text{e-}04)\,\text{Time}_i\text{*Functionalstatus}_{1i} + (4.46\text{e-}03)\,\text{Time}_i\text{*Functionalstatus}_{2i} + (-3.07\text{e-}04)\,\text{Time*Employmentstatus}_i + (-8.24\text{e-}04)\,\text{Time}_i\text{*ART\_Regimen}_{1i} + (7.35\text{e-}04)\,\text{Time}_i\text{*ART\_Regimen}_{2i} + (-4.22\text{e-}03)\,\text{Time}_i\text{*ART\_Regimen}_{3i} + (-4.80\text{e-}04)\,\text{Time}_i\text{*ART\_Regimen}_{4i} + (-1.07\text{e-}03)\,\text{Time*Smokingstatus}_i + (-2.05\text{e-}03)\,\text{Time}_i\text{*Education}_{1i} + (-1.65\text{e-}03)\,\text{Time}_i\text{*Education}_{2i} + (-2.63\text{e-}03)\,\text{Time}_i\text{*Education}_{3i} + (-4.52\text{e-}04)\,\text{Time}_i\text{*Religion}_{1i} + (-2.46\text{e-}03)\,\text{Time}_i\text{*Religion}_{2i} + (-1.79\text{e-}03)\,\text{Time}_i\text{*Religion}_{3i} + 1.05\text{e-}02\,\text{Time}_i\text{*SqrtCD}_{4i} + (-4.74\text{e-}06)\,\text{Time}_i\text{*BaselineCD}_{4i} + (-2.90\text{e-}05)\,\text{BMI}_i + (0.0913)\varepsilon_i$$

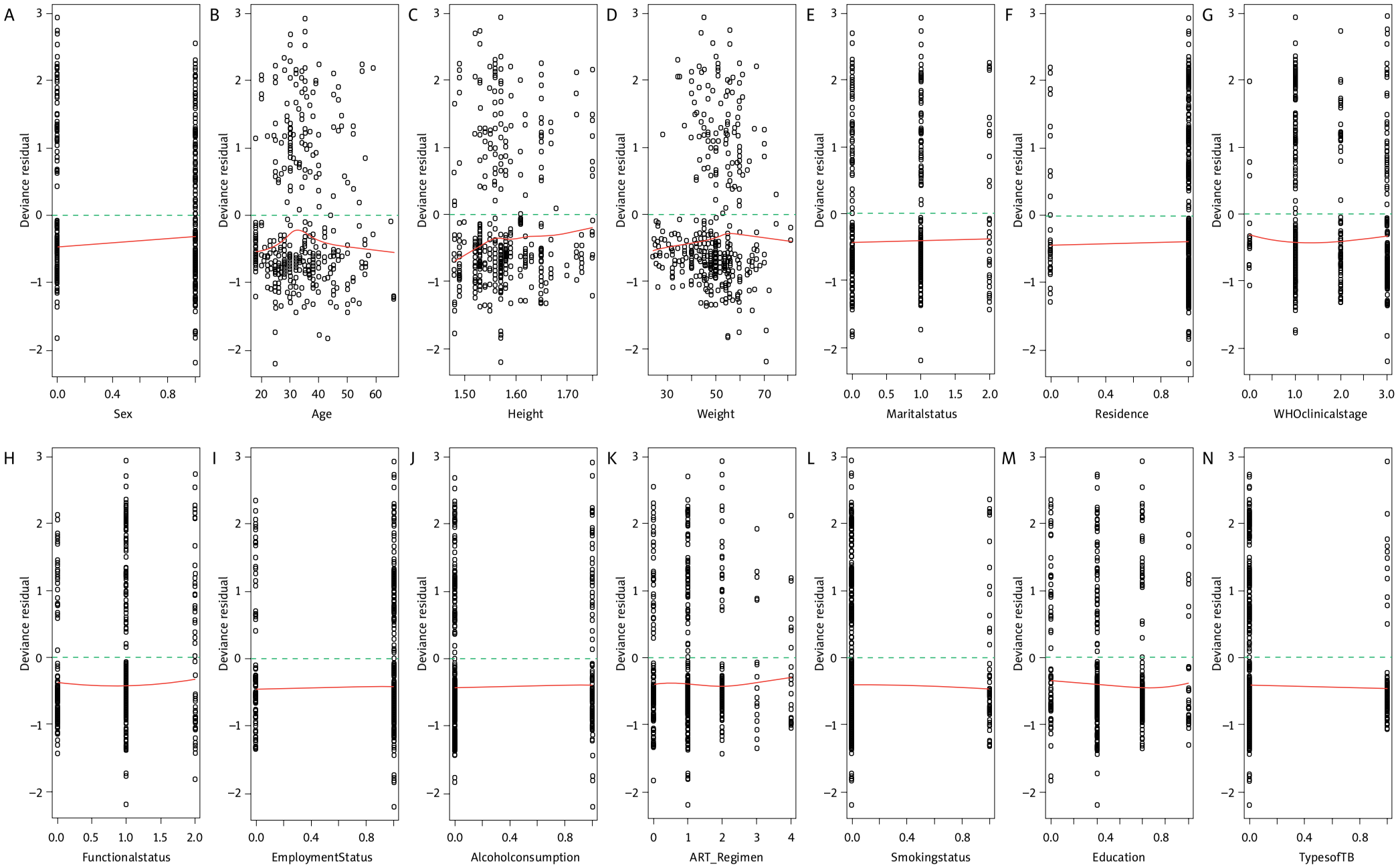

Figure 6. Scatter plot of deviance residual for each covariate at each observation (individual)

Table 7. Cox PH model with time-varying covariates with time interaction

Covariates/Categories	Coefficients (β)	Hazard ratio (HR), Exp (β)	*p*-value
Sex			
Female	–	1.00	–
Male	3.581e-01	1.431e+00	0.08526
Age	1.112e-02	1.011e+00	0.93268
Marital status			
Married	–	1.00	–
Single	1.268e-03	1.001e+00	0.99625
Others	3.935e-01	1.482e+00	0.41054
Residence			
Rural	–	1.00	–
Urban	1.245e-01	1.133e+00	0.73710
WHO clinical stage			
Stage I	–	1.00	–
Stage II	–7.427e-02	9.284e-01	0.90374
Stage III	–2.930e-03	9.971e-01	0.99637
Stage IV	3.451e-01	1.412e+00	0.57976
Functional status			
Working	–	1.00	–
Ambulatory	–1.508e-01	8.600e-01	0.53086
Bedridden	7.041e-01	2.022e+00	0.02533*
Employment status			
Unemployed	–	1.00	–
Employed	3.888e-02	1.040e+00	0.88120
Alcohol consumption			
No	–	1.00	–
Yes	–3.359e-01	7.147e-01	0.10787
ART regimen			
Adult first-line	–	1.00	–
Adult second-line	1.291e-02	1.013e+00	0.95541
Child first-line	–2.389e-01	7.875e-01	0.44396
Child second-line	–1.138e-01	8.925e-01	0.82272
Others	2.520e-01	1.287e+00	0.58211
Smoking status			
No	–	1.00	–
Yes	–1.412e-01	8.684e-01	0.66187
Education level			
Illiterate	–	1.00	–
Primary	–2.001e-02	9.802e-01	0.94648
Secondary	–6.234e-01	5.361e-01	0.05437*
Diploma and above	–2.863e-01	7.511e-01	0.51405
Type of tuberculosis			
Pulmonary	–	1.00	–
Extra-pulmonary	–4.372e-01	6.459e-01	0.16380
Religion			
Orthodox	–	1.00	–
Muslim	1.823e-01	1.200e+00	0.45552
Protestant	2.739e-01	1.315e+00	0.40779
Catholic	1.752e-01	1.191e+00	0.65036
sqrtCD4	1.422e+01	1.503e+06	0.38919
Baseline CD4+	–3.106e-03	9.969e-01	0.00673**
BMI	2.268e-01	1.255e+00	0.45148
tt (age)	–5.192e-03	9.948e-01	0.85986
tt (sqrtCD4)	–3.039e+00	4.787e-02	0.41617
tt (BMI)	–3.773e-02	9.630e-01	0.58161

**Significance at p = 0.05 level. **More significance level at 0.05 level of significance.*

Table 8. Parametric survival model selection

Parametric survival model	AIC	BIC	LogLik (model)
Weibull model	1,551.729	1,667.985	–746.9
Exponential model	1,609.213	1,721.459	–776.6
Gamma model	1,624.577	1,740.00	–783.3
Log-normal model	**1,534.056**	**1,650.311**	**–738.00**

Table 9. Estimated parameters for parametric survival model with log-normal distribution

Fixed effect/Categories	Coefficients (β)	Standard error	*p*-value
(Intercept)	4.30e+00	1.65e-01	< 2e-16*
Age	3.08e-03	3.13e-03	0.32543
Marital status			
Married	–	–	–
Single	1.32e-01	5.01e-02	0.00827*
Others	1.64e-01	8.70e-02	0.05919*
Residence			
Rural	–	–	–
Urban	–6.11e-02	8.00e-02	0.44494
WHO clinical stage			
Stage I	–	–	–
Stage II	–1.48e-01	1.36e-01	0.27912
Stage III	–3.83e-01	1.47e-01	0.00901*
Stage IV	–3.15e-01	1.40e-01	0.02462*
Functional status			
Working	–	–	–
Ambulatory	–2.32e-02	4.73e-02	0.62383
Bedridden	–3.26e-01	6.73e-02	1.3e-06**
Employment status			
Unemployed	–	–	–
Employed	2.16e-02	5.21e-02	0.67793
ART regimen			
Adult first-line	–	–	–
Adult second-line	2.32e-02	4.82e-02	0.62996
Child first-line	–9.64e-02	5.64e-02	0.08766
Child second-line	2.24e-01	9.31e-02	0.01629*
Others	–1.63e-02	8.82e-02	0.85334
Smoking status			
No	–	–	–
Yes	6.87e-02	5.40e-02	0.20265
Education			
Illiterate	–	–	–
Primary	4.78e-02	6.66e-02	0.47289
Secondary	6.38e-02	6.78e-02	0.34679
Diploma and above	1.44e-01	1.06e-01	0.17340
Type of tuberculosis			
Pulmonary	–	–	–
Extra-pulmonary	7.49e-02	5.47e-02	0.17062
Religion			
Orthodox	–	–	–
Muslim	1.17e-02	5.02e-02	0.81532
Protestant	1.88e-01	6.50e-02	0.00382*
Catholic	8.82e-02	7.73e-02	0.25372

Table 9. Cont.

Fixed effect/Categories	Coefficients (β)	Standard error	*p*-value
sqrtCD4	–6.39e-01	1.67e-01	0.00013**
Baseline CD4+	4.62e-04	1.98e-04	0.01982*
BMI	–4.06e-03	5.31e-03	0.44387
Age: Time	–3.33e-05	4.27e-05	0.43488
Marital status: Time			
Married	–	–	–
Single	–1.73e-03	6.79e-04	0.01104*
Others	–2.51e-03	1.28e-03	0.050*
Residence: Time			
Rural	–	–	–
Urban	7.33e-04	1.03e-03	0.47601
WHO clinical stage: Time			
Stage I	–	–	–
Stage II	1.79e-03	1.68e-03	0.28880
Stage III	5.86e-03	1.94e-03	0.00257*
Stage IV	4.68e-03	1.79e-03	0.009*
Functional status: Time			
Working	–	–	–
Ambulatory	1.30e-04	5.98e-04	0.82847
Bedridden	4.46e-03	1.06e-03	2.4e-05*
Employment status: Time			
Unemployed	–	–	–
Employed	–3.07e-04	7.34e-04	0.67540
ART regimen: Time			
Adult first-line	–	–	–
Adult second-line	–8.24e-04	6.64e-04	0.21464
Child first-line	7.35e-04	7.56e-04	0.33109
Child second-line	–4.22e-03	1.18e-03	0.00036**
Others	–4.80e-04	1.11e-03	0.66656
Smoking status: Time			
No	–	–	–
Yes	–1.07e-03	7.06e-04	0.13094
Education: Time			
Illiterate	–	–	–
Primary	–2.05e-03	1.08e-03	0.05712*
Secondary	–1.65e-03	1.08e-03	0.12825
Diploma and above	–2.63e-03	1.39e-03	0.05762*
Type of tuberculosis: Time			
Pulmonary	–	–	–
Extra-pulmonary	–6.98e-04	6.98e-04	–0.31746
Religion: Time			
Orthodox	–	–	–
Muslim	–4.52e-04	6.98e-04	0.51780
Protestant	–2.46e-03	8.08e-04	0.00233*
Catholic	–1.79e-03	1.10e-03	0.10346
sqrtCD4: Time	1.05e-02	1.73e-03	1.4e-09**
Baseline CD4+: Time	–4.74e-06	2.36e-06	0.04510*
BMI: Time	–2.90e-05	7.20e-05	0.68749
Log (scale)	2.39e+00	6.66e-02	< 2e-16**
Scale: 0.0913	–	–	–

**Significance at p = 0.05 level. **More significance level at 0.05 level of significance.*

Discussion

HIV and TB diseases are a major public health problem, and their damages are still at a shocking level. Therefore, a retrospective cohort study was performed among 407 patients in order to assess the effect of HAART treatment on the improvement of HIV/TB co-infected patients' lifetime using survival data analysis methods. These survival data analysis method may include Cox PH or AFT model (if Cox PH proportional assumption is failed). A study conducted in South Africa [16] revealed that the log-logistic AFT model was the most appropriate model for fitting the co-infected patients' data. But, in the present study, the lognormal AFT model was the best fit for the data. The followup period of this study also revealed that there were 29.5% deaths among HIV/TB co-infected patients, which might be comparable with a study of Sanzana [17] and higher and lower than studies of [18-23] and [1, 16, 24-26], respectively. In this study, female patients presented slightly higher survival time than male patients as indicated in the K-M plot of patients, which might show a comparable result with studies of Sanzana [17] and Kosgei *et al*. [27]. The mean and median age of the co-infected patients' were 33.63 and 32 years old, respectively, which implies that youths were the most infected with HIV persons, leading to TB co-infection, since this age represents sexually active stages with a high possibility of unprotected sex or sharing drug injection equipment that are the most common modes of HIV transmission [17, 28].

In the current study, both the Cox PH and AFT models for analyzing of the co-infected patients' data set were demonstrated. But, based on several types of statistical tests, the proportionality hazard assumption of the Cox PH model failed, and analyzing this data set using the Cox PH model would lead to misleading and erroneous scientific findings to proceed with an alternative survival data analysis method, AFT model. AFT model is another alternative survival model for survival datasets with censored observations when the Cox PH model assumption fails, with a different assignment of error term distribution or survival time. As indicated in Table 8, out of four parametric assumptions for survival time or error term distribution, log-normal AFT model distribution was selected as a wellfitted error term distribution based on AIC, BIC, and log-likelihood values. In the final log-normal AFT model applied, its' output revealed that WHO clinical stages and their interaction with time were the significant determinant factors of survival time of the co-infected patients at a 5% significant level, presenting a higher risk of developing TB and other HIV opportunistic diseases, while patients WHO clinical stages increased, as demonstrated by previous studies [29-31]. Also, it was found that the functional status and its' interaction with time had a significant effect on the survival time of patients at a given significant level, which agreed with another studies' results [16, 21, 32, 33]. The level of CD4+ counts of co-infected patients' and their interaction with time were the significant predictor variables for the survival time of patients, which showed that a higher CD4+ counts of patients would increase their survival time, according to the literature [25, 32, 34-37].

The current study also found that ART regimen had a positive statistical significance at a 5% significant level, which showed an improved level of CD4+ counts of co-infected patients, and as a result, improved the survival time of patients, as shown in the AFT model rather than in the Cox PH model. Therefore, a long time of HAART exposure could be important factor in reducing the incidence of TB, and as a result, mortality of the co-infected patients [8]. Inconsistent with a study of [38], in women's first birth interval (FBI) using the log-normal AFT model, education was found as a significant covariate for the survival time of co-infected patients in the hospital, because HAART improves CD4+ counts necessary to fight with HIV/AIDS viruses replication [37, 39], and reduces the risk of developing TB [8, 40, 41].

In addition to the abovementioned covariates, the log-normal AFT model results reveal that ART regimen, marital status, religion, and baseline CD4+ count with their interactions with time, were significant predictor variables at the 5% levels of significance. However, covariates, such as age, residence, employment status, smoking status, education, types of TB, and BMI were not statistically significant, with p-value greater than 0.05.

The main strength of the study is the inclusion of numerous paired baseline and followup predictor variables with their interaction with time, and data collected with follow-up time of 11 years, which yields relatively consistent records and reliable findings. Finally, AFT model provided a better fit, and was found that it was the more suited model for this study compared with the Cox proportional hazards model [34, 42], due to the following reasons. The first reason is that the AFT model analyses directly the time-to-event/ death rather than hazard ratios as that of the Cox proportional hazard model, which makes the interpretation of output clinically relevant and meaningful. It provides a more informative results with realistic interpretation [43]. The second reason is that AFT (log-normal) model was also found to perform better prediction, and lead to a more precise result compared with the Cox proportional hazard model for the right-hand distribution of the model [44, 45]. Additionally, a recent study conducted in Malaysia reported that AFT models were the more fitted models as compared with the Cox models [42]. In the current study, based on AIC, BIC, and log-likelihood values, the log-normal parametric model was the selected AFT model, when the proportionality assumption of Cox's regression model was not fulfilled.

Conclusions

This study was a retrospective cohort study of HIV/TB co-infected patients, and data from January 1, 2005 till 30 December, 2016 were taken from the Debre Berhan Referral Hospital, with permission given to undergo the main

objectives of the study. From the results of this study, early start and lifelong HAART treatment is necessary to improve lifetime of the co-infected patients, and hence, provide a great concern on the treatment of patients upon this lifelong drug.

AFT models showed a better fit compared with the Cox regression model for the time-to-event/death data. To model the determinant factors of time-to-event/death of HIV/TB co-infected patients, four parametric models were assumed for the error term distribution, and the log-normal model was selected as the best model based on AIC, BIC, and log-likelihood values. According to the selected AFT (log-normal) model, the current study showed that prognostic factors, such as WHO clinical stages, functional status, CD4+ counts, ART regimen, marital status, religion, baseline CD4+ counts, and their interactions with time, were among the significant predictors of log-normal AFT model at 5% significance level.

Limitation of the study

This study has the following limitation. Some important prognostic factors, including diabetes mellitus, viral load, hemoglobin, liver function tests, and hypertension that could potentially affect TB/HIV co-infection survival rate, were not considered. Despite that, our study results have policy implications, and can be used as a reference point for further studies.

Acknowledgment

The author would like to thank the Almighty GOD giving strength to us from start to the end of this work, and the Debre Berhan Referral Hospital staff members for facilitating in obtaining the research dataset.

Conflict of interests

The author declares no conflict of interest.

References

1. World Health Organization. Updated Recommendations on HIV Prevention, Infant Diagnosis, Antiretroviral Initiation and Monitoring; 2021.
2. World Health Organization. 'HIV-Associated Tuberculosis: fact-sheet'. Retrieved from: https://www.who.int/tb/areas-of-work/tb-hiv/tbhiv_factsheet.pdf?ua=1; 2018 (Accessed: 11.12.2021).
3. UNAIDS. Fact Sheet – World Tuberculosis Day 2022. Unaids. Published online 2022:1-2.
4. Federal Democratic Republic of Ethiopia, Ministry of Health. Ethiopia – National Strategic Plan Tuberculosis and Leprosy Control 2013-2020. Moh. 2017; 13 (November 2017): 92.
5. WHO. TB Burden Report 2018. Vol 63; 2018. Available at: https://apps.who.int/iris/handle/10665/274453.
6. Alene KA, Viney K, Moore HC, Wagaw M, Clements ACA. Spatial patterns of tuberculosis and HIV coinfection in Ethiopia. PLoS One 2019; 14: 1-15.
7. Onyango DO, Yuen CM, Cain KP, Ngari F, Masini EO, Borgdorff MW. Reduction of HIV-associated excess mortality by antiretroviral treatment among tuberculosis patients in Kenya. PLoS One 2017; 12: 1-13. doi: 10.1371/journal.pone.0188235.
8. Alemu A, Bitew ZW, Yesuf A, Zerihun B, Getu M. The effect of long-term haart on the incidence of tuberculosis among people living with hiv in addis ababa, ethiopia: A matched nested case–control study. Infect Drug Resist 2021; 14: 5189-5198. doi: 10.2147/IDR.S345080.
9. Tiruneh F, Deyas Y. Effect of highly active antiretroviral treatment on TB incidence among HIV infected children and their clinical profile, retrospective cohort study, South West Ethiopia. Sci Rep 2020; 10: 1-6. doi: 10.1038/s41598-020-78466-0.
10. WHO. WHO publishes new clinical and service delivery recommendations for HIV prevention, treatment and care. Published online 2021. Available at: https://www.who.int/news/item/17-03-2021-who-publishes-new-clinical-recommendations-on-hiv-prevention-infant-diagnosis-antiretroviral-therapy-initiation-and-monitoring.
11. Kaplan EL, Meier P. Nonparametric estimation from incomplete samples. J Am Stat Assoc 1958; 53: 457-481.
12. Greenwood M. The errors of sampling of the survivorship Table, vol. 33 of report on the public health and medical subjects. London: His Majesty's Stationery Office; 1926.
13. Cox DR. Regression Models and Life-Tables. Royal Statistical Society, Series B, 1972; 34: 187-220.
14. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982; 69: 239-241.
15. Harrell F. The Phglm Procedure. In: SAS Supplemental Library User's Guide, Version 5. Cary, NC: SAS Institute Inc; 1986.
16. Bengura P, Ndlovu P, Managa MA. Accelerated failure time modelling of tuberculosis predictors in HIV/AIDS patients in Albert Luthuli Municipality of South Africa. Published online 2020. Available at: https://lens.org/004-557-456-543-091.
17. Sanhueza-Sanzana C, Kerr L, Kendall C. Mortality from AIDS and tuberculosis-HIV coinfection in the Chilean AIDS cohort of 2000-2017. Cad Saude Publica 2021; 37. doi: 10.1590/0102-311X00212920.
18. Tancredi MV, Sakabe S, Waldman EA. Mortality and survival of tuberculosis coinfected patients living with AIDS in São Paulo, Brazil: a 12-year cohort study. BMC Infect Dis 2022; 22: 1-13. doi: 10.1186/s12879-022-07232-6.
19. Camara A, Sow MS, Toure A, et al. Treatment outcome, survival and their risk factors among new tuberculosis patients co-infected with HIV during the Ebola outbreak in Conakry. Epidemiol Public Health 2017; 1–8. doi: 10.1016/j.respe.2017.05.011.
20. Gezae KE. Predictors of accelerated mortality of TB/HIV co-infected patients on ART in Mekelle, Ethiopia: an 8 years retrospective follow-up study. Ann Biostat Biometric Appl 2019; 3: 2-6. doi: 10.33552/abba.2019.03.000572.
21. Deres G, Nigussie ZM, Chanie MG, Worku N. Survival time and associated factors among adults living with HIV after initiation of HAART in South Gondar, Northwest Ethiopia: a retrospective cohort. J Multidiscip Healthc 2021; 14: 1463-1474.
22. Brand ÉM, Rossetto M, Hentges B, et al. Survival and predictors of death in tuberculosis/HIV coinfection cases in Porto Alegre, Brazil: a historical cohort from 2009 to 2013. PLoS Glob Public Health 2021; 1: e0000051. doi:10.1371/journal.pgph.0000051.
23. Wondimu W, Dube L, Kabeta T. Factors affecting survival rates among adult TB/HIV co-infected patients in Mizan Tepi University Teaching Hospital, South West Ethiopia. HIV AIDS (Auckl) 2020; 12: 157-164
24. Birhan H, Derebe K, Muche S, Melese B. Statistical analysis on determinant factors associated with time to death of HIV/TB co -infected patients under HAART at Debre Tabor Referral Hospital: an application of accelerated failure time-shared frailty models. HIV AIDS (Auckl) 2021; 13: 775-787.
25. de Melo MC, Donalisio MR, Cordeiro RC. Survival of patients with AIDS and co-infection with the tuberculosis bacillus in the South

and Southeast regions of Brazil. Cien Saude Colet 2017; 22: 3781-3792.
26. Azeez A, Mutambayi R, Akinwumi O, Ndege J. Survival model analysis of tuberculosis treatment among patients with human immunodeficiency virus coinfection. Int J Mycobacteriol 2019; 8: 244-251.
27. Kosgei RJ, Callens S, Gichangi P, et al. Gender difference in mortality among pulmonary tuberculosis HIV co-infected adults aged 15-49 years in Kenya. PLoS One 2020; 15(12 December): 1-15. doi: 10.1371/journal.pone.0243977.
28. World Health Organization. HIV-Associated Tuberculosis: factsheet, 2022. Available at: https://www.who.int/news-room/fact-sheets/detail/hiv-aids.
29. Lelisho ME, Teshale BM, Tareke SA, et al. Modeling survival time to death among TB and HIV co-infected adult patients: an institution-based retrospective cohort study. J Racial Ethn Health Disparities 2022; doi: 10.1007/s40615-022-01348-w.
30. Seyoum D, Degryse JM, Kifle Y, et al. Risk factors for mortality among adult HIV/AIDS patients following antiretroviral therapy in Southwestern Ethiopia: an assessment through survival models. Int J Environ Res Public Health 2017; 14: 296. doi: 0.3390/ijerph14030296.
31. Erango M, Goshu A, Buta G, Dessiso A. Bayesian joint modelling of survival of HIV/AIDS patients using accelerated failure time data and longitudinal CD4 cell counts. Br J Med Med Res 2017; doi: 10.9734/bjmmr/2017/32123.
32. Kenesa Umeta A, Yermosa SF, Dufera AG. Bayesian parametric modeling of time to tuberculosis co-infection of HIV/AIDS patients under antiretroviral therapy treatment at Jimma University Medical Center, Ethiopia. Sci Rep 2022; 12: 16475; doi: https://doi.org/10.21203/rs.3.rs-1151811/v2.
33. Wondimu W, Dube L, Kabeta T. Time to death and its predictors among adult TB/HIV co-infected patients in Mizan Tepi University Teaching Hospital, South West Ethiopia. medRxiv 2019; doi: https://doi.org/10.1101/19004234.
34. Saikia R, Pratim M. A review on accelerated failure time models. Int J Stat Syst 2017; 12: 311-322.
35. Gebremichael SG. Determinants of survival time among HIV-infected patients receiving care at antiretroviral therapy (ART) clinic of a public hospital, Ethiopia. MOJ Public Health 2020; 9: 201-207.
36. Gebremichael SG. AIDS-duration predictors of HIV/AIDS patients on antiretroviral therapy at Debre Berhan referral hospital, north-central Ethiopia. MOJ Public Health 2020; 9: 99-105.
37. Kaplan R, Hermans S, Caldwell J, Jennings K, Bekker LG, Wood R. HIV and TB co-infection in the ART era: CD4 count distributions and TB case fatality in Cape Town. BMC Infect Dis 2018; 18: 356. doi: 10.1186/s12879-018-3256-9.
38. Faruk A. The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. J Phys Conf Ser 2018; 974; doi: 10.1088/1742-6596/974/1/012008.
39. Sarfo B, Vanderpuye NA, Addison A, Nyasulu P. HIV case management support service is associated with improved CD4 counts of patients receiving care at the antiretroviral clinic of Pantang Hospital, Ghana. AIDS Res Treat 2017; 2017: 4697473; doi: 10.1155/2017/4697473.
40. Legese H, Degefa H, Gebrewahd A, Gebremedhin H. Utilization of isoniazid prophylaxis therapy and its associated factors among HIV positive clients taking antiretroviral therapy at Fre Semaetat primary hospital, Hawzien districts, Tigrai, Northern Ethiopia. Trop Dis Travel Med Vaccines 2020; 6: 11. doi: 10.1186/s40794-020-00106-2.
41. Abdu M, Ali Y, Anteneh S, et al. Determinant factors for the occurrence of tuberculosis after initiation of antiretroviral treatment among adult patients living with HIV at Dessie Referral Hospital, South Wollo, Northeast Ethiopia, 2020. A case-control study. PLoS One 2021; 16: e0248490. doi:10.1371/journal.pone.0248490.
42. Mardhiah K, Wan-Arfah N, Naing N, Hassan M, Chan HK. Comparison of Cox proportional hazards model, Cox proportional hazards with time-varying coefficients model, and lognormal accelerated failure time model: application in time to event analysis of melioidosis patients. Asian Pac J Trop Med 2022; 15: 128-134. doi: 10.4103/1995-7645.340568.
43. Liu E, Lim K. Using the Weibull accelerated failure time regression model to predict time to health events. bioRxiv 2018; doi: https://doi.org/10.1101/362186.
44. Ogungbola OO, Akomolafe AA, Musa AZ. Accelerated failure time model with application to data on tuberculosis/HIV co-infected patients in Nigeria. Am J Epidemiol Public Health 2018; 2: 21-26.
45. Khamis A, Azmeeza Che Hamat C, Asrul Affendi Abdullah M. Modeling students performance using Cox and parametric accelerated failure time models. Sci Res J 2020; 8: 44-49.