Bayesian Parametric Modeling of Time to Tuberculosis Co-infection of HIV/AIDS Patients Under Antiretroviral Therapy Treatment at Jimma University Medical Center, Ethiopia

Abdi Kenesa Umeta (✉ abdiikenesa@gmail.com)

Haramaya University

Samuel Fikadu Yermosa

Jimma University

Abdisa G. Dufera

Jimma University

Research Article

Keywords: Tuberculosis, Co-infection, INLA, Bayesian, Parametric

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1151811/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Bayesian parametric modeling of time to Tuberculosis co-infection of HIV/AIDS patients under antiretroviral therapy treatment at Jimma University Medical Center, Ethiopia

Abdi Kenesa Umeta^{1*}, Samuel Fikadu Yermosa² and Abdisa G. Dufera²

Abstract

Background: Tuberculosis is the most common opportunistic infection among HIV/AIDS patients, including those following Antiretroviral Therapy treatment. The risk of Tuberculosis infection is higher in people living with HIV/AIDS than in people who are free from HIV/AIDS. Many studies focused on prevalence and determinants of Tuberculosis in HIV/AIDS patients without taking into account the censoring aspects of the time to event data. Alternatively, this study used INLA methodology to build Bayesian parametric survival models for modeling time to Tuberculosis co-infection of HIV/AIDS patients following Antiretroviral Therapy treatment.

Methods: A data of a retrospective cohort of HIV/AIDS patients under Antiretroviral Therapy treatment follow-up from January 2016 to December 2020 until Tuberculosis was clinically diagnosed or until the end of the study was collected from Antiretroviral Therapy treatment center of Jimma University Medical Center, Ethiopia. In order to identify the risk factors which have association with Tuberculosis free survival time, we fitted Exponential, Weibull, Log-logistic, Gamma and Log-normal models to our data set..

Results: About 26.37% of the study subjects had been co-infected with tuberculosis during the study period. Among the parametric Accelerated failure time models, the Bayesian log-logistic Accelerated failure time model was found to be the best fitting model for the data.

Conclusions: Tuberculosis co-infection survival time was significantly associated with place of residence, smoking, drinking alcohol, family size, WHO clinical stages, functional status, CD4 count, BMI and hemoglobin level. The finding of this study provide timely information on the risk factors associated with TB co-infection survival time for healthy policy makers and planners.

Keywords: Tuberculosis; Co-infection; INLA; Bayesian; Parametric

Background

Tuberculosis is one of the infectious diseases that affects the lungs and other sites[1]. Tuberculosis has been the main public health problem affecting millions worldwide and it remains the top infectious killer in the world causing close

*Correspondence: abdiikenesa@gmail.com

¹Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia

Full list of author information is available at the end of the article

to about 4000 lives a day[2]. Around 10.0 million people estimated to have developed TB disease in 2019 worldwide, and there were around 1.2 million TB deaths among HIV-negative people and an additional 208,000 deaths among people living with HIV[3].

Tuberculosis is the most common opportunistic infection among HIV positive people, including those following ART treatment, and it is the major cause of HIV-related death[4]. UNAIDS report of 2018 showed that Sub-Saharan Africa is the hardest hit region of the world, as it has around 70% of all people living with HIV/AIDS and TB co-infection in the world[5].

Ethiopia is one of the highest 30 TB and HIV co-infection countries in the world, with TB incidence rate of about 164 per 100,000 TB, 112/100,00 TB among people living with HIV[3]. Research shows that the TB co-infection in Ethiopia is high in areas where HIV is highly distributed[6]. The majority of the People living with HIV were from the Amhara (30%), Oromia (26%) and Addis Ababa (18%) regions[7].

The HIV virus infects CD4 cells causing reduction of the number of immune cells which causes the body fail to control viral multiplication which increases the chance of opportunistic infection with tuberculosis being the most common opportunistic infection at HIV diagnosis[8]. The treatment outcome of HIV-positive following ART treatment has remarkably changed with a reduction of plasma viral copies and an increase of CD4 counts[9]. It had been reported that the ART treatment has reduced the incidence of TB in HIV patients by about 70 - 90% [10]. Even with the advantages of ART treatment, still HIV/AIDS patients on ART treatment develop TB with about prevalence rate of 2.5–30.1[8].

Though, Tuberculosis can affect everyone, the risk of Tuberculosis infection is higher in people living with HIV than in people who are free from HIV[11]. Studies revealed that certain HIV-infected people develop TB, while others do not. The risk of developing active TB in people living with HIV is about 21 times higher than those free from HIV which may be due to the altered immune state of HIV positive patients[18]. There are various factors that increase the chance of TB infection among HIV/AIDS patients including CD4 cell count and the number of vi-

ral loads[19, 20], household family size, cigarette smoking, baseline CD4 cell counts, WHO clinical stages, having a history of diabetics[58], and etc. However, these factors have not been studied in the context of Survival Analysis, where association between risk factors and time to TB co-infection might be of interest.

Majority of the study focused on prevalence and predictors of TB in HIV patients. In order to determine the important determinants of TB co-infection in HIV patients, most of the methodologies in the literature used logistic regression with outcome being the TB's viability through follow up time of HIV/AIDS patients taking ART treatment[12, 14, 17]. In logistic regression, our interest is to study how risk factors were associated with the presence or absence of a disease (or an event) without taking into account the effect of time[15]. These approaches help to provide odds ratios for significant variables associated with the risk of TB infection but reject the censoring aspects of time-to-event data.

The aim of this study was to model the predictors of time to TB co-infection in HIV/AIDS patients following ART treatment using Bayesian parametric survival analysis approach based on INLA methodology. There have been advances in computational and modeling techniques using Bayesian approach of survival data[23, 44]. Due to the complex likelihood functions to accommodate censoring, survival models are generally very difficult to fit. Bayesian approach to survival analysis may overcome this by using the MCMC techniques and other numerical integration methods like INLA[34]. Additionally, the Bayesian approach can help to capture prior information and incorporate it seamlessly via a rigorous, probabilistic framework[35]. Therefore, Bayesian Survival Analysis approach is preferred over the usual frequentist technique as the power of information obtained from the approach is much better as it is the combination of likelihood of the data and prior information about the distribution of the parameter[37].

Due to the fact that Bayesian inference cannot be performed analytically, MCMC is particularly useful in Bayesian inference because of the focus on posterior distributions which are often difficult to work analytically[41].

The popular software packages used with MCMC

method include BUGS, JAGS and STAN[48]. The first problem with these packages is that they are slow and some times need to wait days to get the result. The second difficulty with MCMC methods is that it requires high skill of programming language to specify the likelihood functions of different models.

The integrated nested Laplace approximation method for approximate Bayesian inference was developed by Rue, Martino, and Chopin as an alternative to the MCMC method[23]. INLA is an alternative method for Bayesian inference on latent Gaussian models when the focus is on posterior marginal distributions. It substitutes MCMC methods with accurate, deterministic approximations to posterior marginal distributions. Integrated nested Laplace approximation provides a fast and exact approach to fitting latent Gaussian models which include many statistical models, including survival models[48].

Survival models can be written into a latent Gaussian model which allows us to perform Bayesian inference using integrated nested Laplace approximations[44]. Survival analysis consists of a great body of work using latent Gaussian models and it is one of the statistical models on which INLA has been successfully applied[27, 28]. The main advantage of INLA over MCMC techniques is its simplicity of computation[28]. Using INLA results are generated within seconds and minutes even for models with a large dimensional latent field, where as MCMC algorithm would take hours or even days. The other advantage of INLA is that INLA treats latent Gaussian models in a unified way, thus allowing greater automation of the inference process.

Even though Bayesian approaches to the analysis of survival data can provide a number of benefits, they are less widely used than the classical approaches[29]. Therefore, the motivation to apply Bayesian Survival Analysis for this study stems from the above mentioned advantages of Bayesian survival analysis approach over the classical survival analysis approach.

Methods

Data source

The nature of the data set used for this study was survival data. In the data set, time until TB infection was clinically diagnosed in HIV/AIDS patients was investigated.

Inclusion and exclusion criteria

All adult HIV patients following ART treatment and who were 18 years old and above during the study period and TB free at the inception of the study with at least two follow up period were included in the study. Patients whose date of TB co-infection was unknown were excluded from the study. Also, Patients with insufficient information about one of the variables in the study were not included.

Study design, population and sample size

A data of a retrospective cohort of adult HIV/AIDS patients under ART follow-up from January 2016 to December 2020 until TB was clinically diagnosed or until the end of the study was collected from ART clinic of JUMC, Ethiopia. In this study, the source population was all adult HIV/AIDS patients who were 18 years old and above at Jimma University Medical Center. There were a total number of 3069 HIV/AIDS patients. Among the total patients 421 of the patients were included in the study based on the inclusion and exclusion criteria.

Study variables

The response variable for this study was time to TB infection in HIV/AIDS patients under ART treatment. Time is measured in months. The predictor variables which were assumed to have effect on time to TB infection in HIV/AIDS patients were Age, sex, place of residence, family size, alcohol usage status, smoking status, marital status, education level, WHO clinical stages, functional status, CD4 count, body mass index and hemoglobin level.

Survival data analysis

Survival analysis is a collection of statistical techniques for data analysis for which the outcome variable of interest is the 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. By event, we mean death, disease incidence, relapse from remission, recovery from disease or any designated experience of interest that may happen to an individual. Censoring is one of the common features that makes survival analysis unique from

another statistical analysis. Censoring is present when we have some information about a subject's event time, but we don't know the exact event time[30]. The general reasons why censoring might occur are: A subject does not experience the event before the study ends, the patient is lost to follow-up during the study period, or the patient withdraws from the study.

Survival function

Assume that the survival time, T , is a continuous random variable. The distribution of T can be described by the usual cumulative distribution function

$$F(t) = P(T \leq t) = \int_0^t f(u)du, \text{ where; } t \geq 0$$

which is the probability that a subject from the population will die (or a specific event of interest for a subject has occurred) before time t [31]. The corresponding density function of T is

$$f(t) = \frac{\partial}{\partial t} F(t)$$

In survival analysis, it is common to use the survival function

$$S(t) = P(T \geq t) = 1 - F(t), t \geq 0$$

The relationship between $f(t)$ and $S(t)$ is given as follows;

$$f(t) = \frac{\partial}{\partial t} F(t) = \frac{\partial}{\partial t} (1 - S(t)) = -\frac{\partial}{\partial t} S(t)$$

Hazard function

It is also of interest, in analyzing survival data, to assess which periods having high or low chances of the event among those still active at the certain time. A suitable method to characterize such risks is the hazard function, $h(t)$, defined by the following equation[31].

$$h(t) = \lim_{s \rightarrow 0} \frac{P(t \leq T \leq t+s | T \geq t)}{s}$$

It is the instantaneous rate of failure (experiencing the event) at the time t given that a subject is alive at the time t .

The definition of the hazard function implies that

$$h(t) = \frac{f(t)}{S(t)} = -\frac{\partial}{\partial t} \log(S(t))$$

A related quantity is the cumulative hazard function, $H(t)$, defined by

$$H(t) = \int_0^t h(u)du = -\log(S(t))$$

And thus,

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(u)du\right)$$

The Kaplan–Meier estimator of survival function

The Kaplan–Meier (KM) method is a non-parametric(distribution-free) survival analysis method which is used to make comparisons of the survival rates between two or more groups[32]. To estimate the survivor function, $S(t)$, without covariates, we can use the Kaplan Meier estimator. Let there be n individuals with observed survival times $t_1, \dots, t_n$ and r be death times amongst the individuals, where $r \leq n$, $j = 1, \dots, r$. The r ordered death times are $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Let n_j denotes the number of individual who are alive just before time $t_{(j)}$, including those who are about to die at this time, and let d_j denotes the number who die at this time. The Kaplan–Meier estimator of the survival function at any time in the k^{th} time interval from $t_{(k)}$ to $t_{(k+1)}$, $k = 1, \dots, r$ is given by [36].

$$\hat{S}_{(t)} = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right)$$

Parametric Survival Models

In a parametric survival models, survival time is assumed to follow a known distribution[33]. Parametric models play an important role in Bayesian survival analysis, since many Bayesian analyses in practice are carried out using parametric models and parametric modeling offers straightforward modeling and analysis techniques[34].

Parametric proportional hazard models

Let $h(t/x)$ be the hazard function at time t for a subject given the covariate vector $x = (x_1, \dots, x_p)^T$. The basic model proposed by Cox is as follows:

$$h(t/x) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p)$$

where $h_0(t)$ is the baseline hazard function and β_i 's are the unknown regression parameters to be estimated. In Parametric proportional hazard model, the baseline hazard function $h_0(t)$ is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data. The hazard ratio is hence given by $HR = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$.

Accelerated failure time models

Although parametric PH models are very useful to analyze survival data, there are relatively few probability distributions for the survival time that can be used with these models [39]. 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 we measure the direct effect of the explanatory variables on the survival time instead of hazard, as we do in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the covariates on the survival time.

In accelerated failure time (AFT) models, the natural logarithm of the survival time, $\log(t)$, is expressed as a linear function of the covariates, which yields therefore a linear model:

$$\log(t_i) = \mu + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \sigma \epsilon_i = X_i \beta + z_i$$

We interpret the effect of the AFT model as the change in the time scale by a factor of $\exp(x_j \beta)$. Based on whether this factor is greater or less than 1, survival time is interpreted to either accelerate or decelerate. Accelerated failure time does not imply a positive acceleration of time with the increase of a covariates but rather a deceleration, or, in other words, an increase in the expected waiting time until failure. AFT models have the

opposite sign from similar estimates in proportional hazard models, due to the fact that the PH models predict the hazard and the AFT model predicts time.

An advantage of the AFT approach is that the effect of the covariates is described in absolute terms (i.e. number of months or years) instead of in relative terms (i.e. a hazard ratio). The acceleration factor is the central measure of association obtained in AFT models and allows you to evaluate the effect of covariates on the survival time.

Bayesian Modeling Approach for survival data

The Bayesian paradigm is based on specifying a probability model for the observed data D , given a vector of unknown parameters θ , leading to the likelihood function $L(\theta/D)$. Then we assume that θ is random and has a prior distribution denoted by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution [34], which is obtained by Bayes' theorem. The posterior distribution of θ is given by:

$$\pi(\theta/D) = \frac{L(\theta/D)\pi(\theta)}{\int_{\Theta} L(\theta/D)\pi(\theta)d\theta}$$

where Θ denotes the parameter space of θ

The quantity $m(D) = \int_{\Theta} L(\theta/D)\pi(\theta)d\theta$ is the normalizing constant of $\pi(\theta/D)$, and is often called the marginal distribution of the data or the prior predictive distribution. In most models and applications, $m(D)$ does not have an analytic closed form, and therefore $\pi(\theta/D)$ does not have a closed form. The Bayesian survival analysis approach considers the parameters of the model as random variables and requires that prior distributions specified for them and data are considered as fixed [40].

Likelihood Function in Bayesian Survival Analysis

Suppose we observe n independent vectors of (T_i, δ_i) , where T_i is time to the event and δ_i is indicator variable telling us whether T_i is censored or not, i.e, $T_i = 0$ for censored observation ($\delta_i = 0$) and $T_i = 1$ for uncensored observation ($\delta_i = 1$).

The likelihood function of the set of unknown parameters

θ in the presence of right censoring is given as

$$L(\theta/D) = \prod_{i=1}^n f(t_i, \theta)^{I(\delta_i=0)} * S(t_i, \theta)^{I(\delta_i=1)} * \pi(\theta/D)$$

The Integrated Nested Laplace Approximation Methodology for Bayesian Inference

For long time, Bayesian statistical inference has relied on MCMC methods to compute the joint posterior distribution of the model parameters which is usually computationally very expensive[41]. An alternative approach and fast estimation methods to MCMC which allows user to easily perform approximate Bayesian inference using Integrated Nested Laplace Approximation was proposed by Havard Rue, Martino, and Chopin[44]. INLA computes posterior marginals for each component in model, from which posterior expectation and standard deviations can easily be found.

The Integrated Nested Laplace Approximation Procedure

In order to approximate the posterior marginals of $\pi(x_i/y)$, $\pi(\theta/y)$ and $\pi(\theta_j/y)$ the latent Gaussian models is used [44]. Latent Gaussian models are subset of all Bayesian additive models with a structured additive predictor say η_i . In these models, the observed variable y_i is assumed to belong to an exponential family, where the mean μ_i is linked to this structured additive predictor η_i through a link function $g(\cdot)$, so that $g(\mu_i) = \eta_i$. The structured additive predictor η_i accounts for effects of various covariates in an additive way:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k Z_{ki} + \varepsilon$$

Here, the $\{f^j(\cdot)\}$ are unknown functions of covariates $\mathbf{u}$, the $\{\beta_k\}$ represent the linear effect of covariates $\mathbf{z}$ and the ε_i 's are unstructured terms. A Gaussian prior is assigned to $\alpha, \{f^j(\cdot)\}, \{\beta_k\}$ and ε_i . We denote $\pi(\cdot/\cdot)$ as the conditional density of its arguments, and let $\mathbf{x}$ denote the vector of all n Gaussian variables $\eta_i, \alpha, \{f^j(\cdot)\}$ and $\{\beta_k\}$, and θ denotes the vector of hyper-parameters, which are

not necessarily Gaussian. The density $\pi(x/\theta_1)$ is Gaussian with (assumed) zero mean and precision matrix $Q(\theta_1)$ with hyperparameter θ_1 .

The distribution for the n_d observational variables $\mathbf{y} = \{y_i : i \in I\}$ is denoted by $\pi(y/x, \theta_2)$ and we assume that $\{y_i : i \in I\}$ are conditionally independent given $\mathbf{x}$ and θ_2 . For simplicity, we denote by $\theta = (\theta_1^T, \theta_2^T)^T$ with $\dim(\theta) = m$. The posterior then reads (for non singular $Q(\theta)$),

$$\pi(x, \theta/y) = \pi(\theta) \pi(x/\theta) \prod_{i \in I} \pi(y_i/x_i, \theta)$$

The imposed linear constraints (if any) are denoted by $\mathbf{A}\mathbf{x} = \mathbf{e}$ for a $k \times k$ matrix $\mathbf{A}$ of rank k . The main aim is to approximate the posterior marginals of the latent field, $\pi(x_i/y)$ and the posterior marginals of the hyperparameters, $\pi(\theta/y)$ and $\pi(\theta_j/y)$. We can write the posterior marginal of interest as

$$\begin{aligned} \pi(x_i) &= \int \pi(x_i/\theta, y) \pi(\theta/y) d\theta, \\ \pi(\theta_j) &= \int \pi(\theta/y) d\theta_{-j} \end{aligned}$$

The importance of INLA is to use the above form to construct nested approximations, as this approach makes Laplace approximations very accurate when applied to latent Gaussian models.

$$\begin{aligned} \tilde{\pi}(x_i/y) &= \int \pi(x_i/\theta, y) \tilde{\pi}(\theta/y) d\theta \\ \tilde{\pi}(\theta_j/y) &= \int \tilde{\pi}(\theta/y) d\theta_{-j} \end{aligned}$$

Here, $\tilde{\pi}(\cdot/\cdot)$ is an approximated (conditional) density of its arguments. Approximations to $\pi(x_i/y)$ are computed by $\pi(\theta/y)$ and $\pi(x_i/\theta, y)$ and using numerical integration to integrate out θ . The approximation of $\pi(\theta_j/y)$ is computed by integrating out θ_{-j} from $\tilde{\pi}(\theta/y)$. The posterior marginal $\pi(\theta)$ of the hyperparameters θ is approximated using a Laplace approximation

$$\tilde{\pi}(\theta/y) \propto \frac{\pi(x, \theta, y)}{\pi_G(x/\theta, y)} \Big|_{x=x^*(\theta)}$$

Prior Distributions in INLA

Bayesian statistical inference depends on the posterior distribution which is obtained by updating the prior beliefs by new evidence. Prior distribution can be broadly classified into non-informative, weakly informative and informative prior distributions. Non-informative prior distributions, also known as objective prior distributions, are designed to have minimal impact on the posterior distribution

so that the data alone can be the source of inference[45]. The non-informative prior distribution often produce the same results as maximum likelihood estimates. On the other hand, the informative prior distributions that aim to construct a prior distribution that reflect the current knowledge on the values of the parameters and the uncertainties that surround the knowledge about the parameters in question[46]. In INLA, it is assumed that fixed effects follow Gaussian distribution with mean zero and small number of precision matrix $\mathbf{Q}(\theta_1)$ and only the parameters in the precision matrix of the random effect need a prior which was considered as a hyper-parameter[47]. For this study, Gaussian prior distribution (non-informative) with mean zero and variance equal to 1000 (precision equal to 0.001) was used for the fixed effects and the intercept[23]. And for hyper-parameters a non-informative prior of Gamma distribution prior is a common non-informative prior to be assigned[48]. In INLA, the Latent component of the model, $\eta_i = \beta_0 + \beta_1 z_1 + \dots + \beta_p z_p$ must follow a Gaussian distribution[48]. In this study, it was assumed that fixed effect (coefficients) associated with covariates have a Normal distribution with mean 0 and variance 10^2 , i.e., $\beta_p, p = 0, \dots, p$, i.e., $\beta_p \sim N(0, 10^2)$ [34]. Then for this study, to complete the model we have assigned a non-informative Gamma prior for the hyperparameter of the model $\tau_i \sim \Gamma(a, b)$ and $\alpha \sim \Gamma(a, b)$ with $a=1$ and $b=0.001$ which is similar with prior distribution used by many researchers worked on Bayesian survival analysis[23, 34, 44, 49].

Bayesian parametric survival models

Exponential Model

The exponential model is the most fundamental parametric model in survival analysis[34]. Suppose we have independent identically distributed (i.i.d.) survival times $\mathbf{t} = (t_1, t_2, \dots, t_n)'$, each having an exponential distribution with parameter λ . Denote the censoring indicators by $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$, where $\delta_i = 0$ if T_i is right censored and $\delta_i = 1$ if T_i is a failure time. Let $f(t_i/\lambda) = \lambda \exp(-\lambda t_i)$ denote the density for t_i , $S(t_i/\lambda) = \exp(-\lambda t_i)$ denotes the survival function. We build a regression model by introducing covariates through λ , and write $\lambda_i = \varphi(x_i' \beta)$, where

x_i' is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of regression coefficients, $\varphi(\cdot)$ and is a known function and $\mathbf{D} = (n, \mathbf{t}, \mathbf{X}, \delta)$ denotes the observed data for regression model. Using these, we get the likelihood function[44].

$$\begin{aligned} L(\beta/D) &= \prod_{i=1}^n f(t_i/\lambda_i)^{\delta_i} S(t_i/\lambda_i)^{1-\delta_i} \\ &= \exp \left\{ \sum_{i=1}^n \delta_i x_i' \beta \right\} \exp \left\{ - \sum_{i=1}^n t_i \exp(x_i' \beta) \right\} \end{aligned}$$

Suppose we specify a normal prior for β with mean μ_0 and variance σ_0^2 . Then the posterior distribution of β is given by

$$\pi(\beta/D) \propto L(\beta/D) \pi(\beta/\mu_0, \sigma_0)$$

where $\pi(\beta/\mu_0, \sigma_0)$ is the normal density with mean μ_0 and variance σ_0^2 .

Weibull model

The Weibull model is perhaps the most widely used parametric survival model[34]. Suppose we have independent identically distributed survival times $\mathbf{t} = (t_1, t_2, \dots, t_n)'$, each having a Weibull distribution, denoted by $\omega(\alpha, \gamma)$. It is often more convenient to write the model in terms of the parameterization $\lambda = \log(\gamma)$, leading to $f(t_i/\alpha, \lambda) = \alpha t_i^{\alpha-1} \exp(\lambda - \exp(\lambda) t_i^\alpha)$. Let $S(t_i/\alpha, \lambda) = \exp(-\exp(\lambda) t_i^\alpha)$ denote the survival function. We can write the likelihood function of (α, λ) as

$$\begin{aligned} L(\alpha, \lambda/D) &= \prod_{i=1}^n f(t_i/\alpha, \lambda)^{\delta_i} S(t_i/\alpha, \lambda)^{(1-\delta_i)} \\ &= \alpha^d \exp \left\{ d\lambda + \sum_{i=1}^n (\delta_i(\alpha - 1)) \log(t_i) \right\} \\ &\quad \exp \left\{ - \exp(\lambda) t_i^\alpha \right\} \end{aligned}$$

Where $d = \sum_{i=1}^n \delta_i$ and δ is the indicator variable taking value 1 if t_i is failure time and 0 if t_i is right censored.

To build the Weibull regression model, we introduce covariates through λ and write $\lambda_i = x_i' \beta$. Where x_i is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of regression coefficients. Assuming a normal prior with parameters (μ_0, σ_0^2) for λ and gamma prior with parameters (α_0, κ_0) , the joint

posterior distribution of (α, λ) is given by

$$\pi(\beta, \alpha/D) \propto \alpha^{\alpha_0+d-1} \exp \left\{ \begin{aligned} & \sum_{i=1}^n (\delta_i x'_i \beta + \delta_i (\alpha - 1) \log(t_i)) \\ & - (t_i^\alpha \exp(x'_i \beta)) \\ & - \kappa_0 \alpha - \frac{1}{2} (\beta - \mu_0)' \frac{1}{\sigma_0^2} (\beta - \mu_0) \end{aligned} \right\}$$

Where $D = (n, t, \mathbf{x}, \delta)$ denote the observed data for regression model.

Log-logistic model

The log-logistic model possesses a rather supple functional form[50]. The Log-logistic distribution is among the parametric survival models where the hazard rate initially increases and then decreases. If we have independent identically distributed survival times $\mathbf{t} = (t_1, t_2, \dots, t_n)$, each having an log-logistic distribution, denoted by $T \sim LL(\alpha, \lambda)$, with density

$$f(t_i/\alpha, \lambda) = \frac{\alpha \lambda^\alpha t_i^{\alpha-1}}{(t_i^\alpha + \lambda^\alpha)^2}$$

for $\alpha > 0, \lambda > 0$ and $t \geq 0$.

And, the survival function is given by

$$S(t_i/\alpha, \lambda) = \frac{\lambda^\alpha}{(t_i^\alpha + \lambda^\alpha)}$$

for $t > 0$.

We can write the likelihood function of (α, λ) as

$$\begin{aligned} L(\alpha, \lambda|D) &= \prod_{i=1}^n f(t_i/\alpha, \lambda)^{\delta_i} S(t_i/\alpha, \lambda)^{(1-\delta_i)} \\ &= \alpha^d \lambda^{n\alpha} t_i^{(\alpha-1)d} (t_i^\alpha + \lambda^\alpha)^{-d} \end{aligned}$$

Where $d = \sum_{i=1}^n \delta_i$ and Where δ is the indicator variable taking value 1 if t_i is failure time and 0 if t_i is right censored.

To build the regression model, we introduce covariates through λ , and write $\lambda_i = x'_i \beta$. Where x'_i is $p \times 1$ vector of covariates, and β is $p \times 1$ regression coefficients. If we assume gamma prior with parameters (α_0, κ_0) for α , we will

have the following joint posterior

Log-Normal Model

Another commonly used parametric survival model is the log-normal model[34]. For this model, we assume that the logarithms of the survival times are normally distributed. If t_i has a log-normal distribution with parameters (μ, σ) , denoted by $\ln(\mu, \sigma)$, then

$$f(t_i/\mu, \sigma) = (2\pi)^{-1/2} (t_i \sigma)^{-1} \exp \left\{ \frac{-1}{2\sigma^2} (\log(t_i) - \mu)^2 \right\}$$

The survival function is given by

$$S(t_i/\mu, \sigma) = 1 - \Phi \left(\frac{\log(t_i) - \mu}{\sigma} \right)$$

We can thus write the likelihood function of (μ, σ) as

$$L(\mu, \sigma/D) = \prod_{i=1}^n f(t_i|\mu, \sigma)^{\delta_i} S(t_i|\mu, \sigma)^{1-\delta_i}$$

Then,

$$\begin{aligned} L(\mu, \sigma/D) &= (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \delta_i (\log(t_i) - \mu)^2 \right\} \\ &\times \prod_{i=1}^n t_i^{-\delta_i} \left(1 - \Phi \left(\frac{\log(t_i) - \mu}{\sigma} \right) \right)^{1-\delta_i} \end{aligned}$$

To construct the regression model, we introduce covariates through μ , and write $\mu_i = x'_i \beta$. Assuming $\beta/\tau \sim N_p(\mu_0, \tau^{-1} \varsigma_0)$, the joint posterior for β, τ is given by

$$\begin{aligned} \pi(\beta, \tau|D) &\propto \tau^{\frac{\alpha_0+d}{2}-1} \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^n \delta_i (\log(t_i) - x'_i \beta)^2 \right] \right. \\ &\quad \left. + (\beta - \mu_0)' \frac{1}{\sigma_0^2} (\beta - \mu_0) + \lambda_0 \right\} \\ &\times \prod_{i=1}^n t_i^{-\delta_i} \left(1 - \Phi \left(\tau^{1/2} (\log(t_i) - x'_i \beta) \right) \right)^{1-\delta_i} \end{aligned}$$

Gamma model

The gamma model is a generalization of the exponential model[34]. For this model, $t_i \sim \zeta(\alpha, \lambda)$ and its density function is given by:

$$f(t_i/\alpha, \lambda) = \frac{1}{\Gamma(\alpha)} t_i^{\alpha-1} \exp(\alpha\lambda - t_i \exp(\lambda))$$

The survival function is given by

$$S(t_i/\alpha, \lambda) = 1 - \frac{1}{\Gamma(\alpha)} \int_0^{t_i \exp(\lambda)} u^{\alpha-1} \exp(-u) du,$$

We can thus write the likelihood function of (α, λ) as

$$L(\alpha, \lambda | D) = \prod_{i=1}^n f(t_i | \alpha, \lambda)^{\delta_i} S(t_i | \alpha, \lambda)^{(1-\delta_i)}$$

To construct the regression model, we introduce covariates through λ , and write $\lambda_i = x_i' \beta$. Assuming $\beta \sim N(\mu_0, \sigma_0^2)$, we are lead to the joint posterior

$$\begin{aligned} \pi(\beta, \alpha | D) \propto & \frac{\alpha^{\alpha_0-1}}{(\Gamma(\alpha))^d} \exp \left\{ \sum_{i=1}^n \delta_i \left[\alpha(x_i' \beta) + \log(t_i) \right] \right. \\ & \left. - t_i \exp(x_i' \beta) \right\} \\ & \times \prod_{i=1}^n t_i^{-\delta_i} (1 - IG(\alpha, t_i(x_i' \beta)))^{1-\delta_i} \\ & \times \exp \left(-\kappa_0 \alpha - \frac{1}{2\sigma_0^2} (\beta - \mu_0)' (\beta - \mu_0) \right) \end{aligned}$$

Where, $IG = \frac{1}{\Gamma(\alpha)} \int_0^{t_i \exp(\lambda)} u^{\alpha-1} \exp(-u) du$ is the incomplete gamma function.

Model Comparison Methods

Integrated Nested Laplace Approximation computes a number of Bayesian criteria for model assessment and model selection[51]. Model selection criteria will be of help when selecting among different models. The following methods of model selection techniques was used in this study.

Marginal likelihood

The marginal likelihood of a model is the probability of the observed data under a given model[42]. The marginal likelihood approximation provided by INLA is computed

as.

$$\tilde{\pi}(y) = \int \frac{\pi(\theta, x, y)}{\tilde{\pi}_G(x/\theta, y)} \Big|_{x=x^*(\theta)} d\theta$$

Information-based criteria (DIC and WAIC)

The deviance information criterion (DIC) is a popular criterion for model choice[52]. It takes into account goodness-of-fit and a penalty term that is based on the complexity of the model via the estimated effective number of parameters. The DIC is defined as

$$DIC = D(\hat{x}, \hat{\theta}) + 2pD$$

where, $D(\cdot)$ is the deviance, $\hat{x}$ and $\hat{\theta}$ the posterior expectations of the latent effects and hyperparameters, respectively, and pD is the effective number of parameters. The effective number of parameters pD can be computed as

$$pD = E[D(\cdot)] - D(\hat{x}, \hat{\theta})$$

The Watanabe-Akaike information criterion, also known as widely applicable Bayesian information criterion, is similar to the DIC but the effective number of parameters is computed in a different way. The final formula to calculate WAIC is.

$$WAIC = -2 \sum_{i=1}^n \log p_{post(y_i)} + 2pD$$

Where, $\sum_{i=1}^n \log p_{post(y_i)}$ is the sum of predictive density for each data point and pD is the effective number of parameters.

Model Diagnosis

Diagnosis for the accuracy of INLA approximation for the models

The Kullback-Leibler divergence (kld): This value describes the difference between the normal approximation and the simplified Laplace approximation. Small values indicate that the posterior distribution is well-approximated by a normal.

Effective number of parameters(pD): The posterior summary results from INLA also contain, effective number of

parameters which is another measure of the accuracy of approximation. In particular, if the effective number of parameters is low compared to the sample size, then one expects the approximation to be good. A two types of "Goodness of fit" reported by INLA are:

Conditional predictive ordinates (CPO): Conditional predictive ordinates are a cross-validatory criterion for model assessment[53]. It is computed for each observation as

$$CPO_i = \pi(y_i/y_{-i})$$

Unusually small or large values of CPO_i indicate a surprising observation.

Predictive integral transform (PIT): The predictive integral transform (PIT) measures the probability of a new value to be lower than the actual observed value for each observation [54]. It is computed as

$$PIT_i = \pi(y_i^{new} \leq y_i/y_{-i})$$

An unusual large or small value indicates possible outliers.

Due to how $\tilde{\pi}(x_i/y_{-1}, \theta)$ are computed there may be cases where this computation fails due to inaccurate tail behavior of $\tilde{\pi}(x_i/y_i, \theta_j)$. To monitor the reliability of the CPO and PIT values computed, failure variable computed for each i (or y_i) is defined as follows.

- If $\tilde{\pi}(x_i/y_i, \theta_j)$ is monotone increasing or decreasing, then failure is set to 1 and then $\tilde{\pi}(x_i/y_i, \theta_j)$ is set to the 0-function. In this case, $\tilde{\pi}(x_i/y_i, \theta_j)$ is known to be just wrong.
- If $\tilde{\pi}(x_i/y_i, \theta_j)$ is has a (local) maximum either at $\min x_i$ or at $\max x_i$, then $\tilde{\pi}(x_i/y_i, \theta_j)$ is set to zero in that part where $\tilde{\pi}(x_i/y_i, \theta_j)$ is decreasing (starting from $\min(x_i)$) or increasing (starting from $\max x_i$). When the expected failure is 0 then the computed value of CPO and PIT seems to be reliable, and when the expected failure is 1 then the computed value of CPO and PIT is known to be completely unreliable.

Results

Summary of descriptive statistics results

A total of 421 adult HIV patients following ART treatment at Jimma University Medical Center, Ethiopia were included in the analysis. During the follow-up period, 111(26.37%) of the study subjects had experienced the event(had been co-infected with TB).

Descriptive results of demographic and clinical characteristics of patients were presented in Table 1 and Table 2. The percentage of female patients who had been co-infected with TB was 53.15% which is larger compared to male patients. About 76.58% of the HIV patients with TB cases were urban residents. Among the patients, who had TB co-infection, about 36.94% were smokers and 57.67% were alcoholics. Patients with no education accounted for 14.41% of experiencing TB, patients with primary education accounted for 22.52%, patients with secondary education accounted for 39.63%, patients with Tertiary education accounted for 18.91%, and patients with education level of Diploma and above accounted for 4.50% of experiencing TB during follow up time. Among 111 HIV patients, who had TB co-infection, 16.22% of them were in WHO clinical I, 19.82% were in WHO clinical stage II, 27.02% were in WHO clinical stage III, and 36.94% were in WHO clinical stage IV. About 28.83%, 28.83%, and 42.34% co-infection of TB were occurred in working, ambulatory and bedridden HIV patients respectively.

Kaplan-Meier estimate of survival functions

From the plot of the overall Kaplan-Meier survival curve given in the figure 1 below, it can be seen that, a large number of TB co-infection recorded at the earlier time of the initiation of ART treatment and there is a decreasing pattern of TB co-infection through the follow up period. In order to explore differences between TB co-infection free survival time between or among groups, separate KM survival function curves are constructed for categorical covariates and results are given in figure 2,3,4 and 5. In general, if the pattern of one survivor-ship function is above the other,

Table 1: Descriptive results of the demographic and clinical characteristics of patients

Covariates	Category	Patient Status		Total
		Censored	Event	
Sex	Male	128(71.1%)	52(28.9%)	180
	Female	182(75.5%)	59(24.5%)	241
Residence	Urban	159(65.2%)	85(34.8%)	244
	Rural	151(85.3%)	26(14.7%)	177
Smoking	No	256(78.5%)	70(21.5%)	326
	Yes	54(56.8%)	95(43.2)	95
Alcohol	No	204(81.3%)	47(18.7%)	251
	Yes	106(62.4%)	64(37.6%)	170
Education level	No formal education	42(72.4%)	16(27.6%)	58
	Primary education	71(74.0%)	25(26.0%)	96
	Secondary education	105(70.5%))	44(29.5%)	149
	Tertiary education	56(72.7%)	21(27.3%)	77
	Other	36(87.8%))	5(12.2%)	41
Family size	≤ 2	173(83.6%)	34(16.4%)	207
	3-4	116(69.5%)	51(31.5%)	167
	≥ 5	21(44.7%)	26(55.3%)	47

Table 2: Continued from Table 1

Covariates	Category	Patient Status		Total
		Censored	Event	
Marital status	Single	43(63.23%)	25(36.77%)	65
	Married	152(76.8%)	46(23.2%)	198
	Widowed/Divorced	115(74.2%)	40(25.8%)	155
WHO disease stage	I	98(84.5%)	18(15.5%)	116
	II	80(78.4%)	22(21.6%)	102
	III	80(72.7%)	30(27.3%)	101
	IV	52(55.9%)	41(54.1%)	93
Functional status	Working	148(82.2%)	32(17.8%)	180
	Ambulatory	90(73.8%)	32(26.2%)	122
	Bedridden	72(60.5%)	47(39.5%)	119
CD4 count	< 200	71(61.2%)	45(38.8%)	116
	200 - 349	71(64.5%)	39(35.5%)	110
	350 - 499	70(84.3%)	13(15.7%)	83
	≥ 500	98(87.5%)	14(12.5%)	112
Hemoglobin level	Anemic	18(45.0%)	22(55%)	40
	Moderate anemic	59(70.2%)	25(29.8%)	84
	Normal	233(78.5%)	64(21.5%)	297
BMI	Underweight	57(55.3%)	46(44.7%)	103
	Normal	190(82.3%)	41(17.7%)	231
	Overweight	63(72.4%)	24(17.6%)	87
Total		310(78.37%)	111(26.37%)	421(100%)

it means the group defined by the upper curve had a better survival than the group defined by the lower curve.

Comparison of Bayesian parametric survival models

Parametric AFT survival models (Exponential, Weibull, Log-logistic, Log-normal and Gamma modes) based on Bayesian paradigm considering all covariates were fitted for the data. In order to compare and select the best model among different parametric models, we used DIC, WAIC and Marginal likelihood of the models. The model with the smallest values of DIC and WAIC, and largest value marginal log-likelihood is selected as the best model. The five parametric survival models with their corresponding values of DIC, WAIC and Marginal Likelihood values are displayed in table 3 below. The Bayesian Log-logistic AFT model was found to be the best fitting model for our data set as it has the smallest values of DIC and WAIC, and has the largest values of marginal loglikelihood among the five models.

Results of Bayesian Log-logistic AFT model

The table 4 and Table 5 below shows the posterior summary results of Bayesian Log-logistic AFT model. The decision about the significance of the variables is based on the 95% Credible Interval for the posterior mean of the coefficients.

In our study, it appeared that residence, smoking status, alcohol consumption status, WHO clinical stages, functional status, family size, CD4 count, Body Mass Index and hemoglobin level of the patients were significant risk factors associated with TB co-infection free survival time of HIV patients following ART treatment. The interpretation of the estimated posterior mean of parameters of the model was done using estimated acceleration factor ($\hat{\gamma} = \exp(\beta_j)$). In order to decide the significance of the covariates in the model, the 95% credible interval was used. The factors whose credible intervals for posterior mean of parameters contained 0, or whose credible intervals for acceleration factor contained 1, implied that these factors were not sig-

nificant. The final model can be written as:

$$\begin{aligned} \log(T_i) = & 6.460 - 1.258I_{(\text{Residence}=2)} - 0.667I_{(\text{Smoking}=2)} \\ & - 0.879I_{(\text{Alcoholics}=2)} - 0.472I_{(\text{Clinstages}=3)} \\ & - 0.849I_{(\text{Clinstages}=4)} - 0.672I_{(\text{Funstat}=3)} \\ & - 0.980I_{(\text{Famsize}=3-4)} - 1.131I_{(\text{Famsize} \geq 5)} - 1.534I_{(\text{CD4}=3)} \\ & - 0.980I_{(\text{CD4}=4)} - 0.950I_{(\text{BMI}=2)} - 1.192I_{(\text{Hgb}=2)} \end{aligned}$$

Where, T represents time to TB co-infection for each subject. I is an indicator variable for categories of variables where $I_{(.=1)}$ is considered as a reference category.

In Log logistic AFT model, the positive estimated posterior β coefficients indicate a longer TB co-infection free survival time, where as the negative estimated posterior β coefficients indicate shorter TB co-infection free survival time for the patients.

The estimated acceleration factor for patients who reside in urban was $\hat{\gamma} = \exp(-1.258) = 0.2842$ with 95% CI of [0.1571, 0.5035] which does not include one. This means that, keeping all other factors constant the expected TB co-infection free survival time of patients who reside in urban area decreases by a factor 0.2842 as compared to patients residing in rural area.

The second significant predictor of TB co-infection free survival time of HIV/AIDS patient was smoking status, in which acceleration factor of smoker patients was estimated to 0.5137 with 95% CI of [0.2814, 0.9361]. Thus, keeping all other factors constant, the expected TB co-infection free survival time of smoker patients decreases by a factor of 0.5137 as compared to non-smoker patients.

If we look at the alcohol usage status of the patients, the estimated acceleration factor for posterior mean of its coefficient was estimated to be 0.4151 with 95% CI of [0.2324, 0.7408] which can be interpreted as, keeping all other factors constant, the expected TB co-infection free survival time of alcoholics patients decreases by a factor of 0.4151 as compared to non-alcoholics patients.

The other significant risk factor for TB co-infection in HIV/AIDS patients was WHO clinical stages III and IV. The estimated acceleration factor of WHO clinical Stages III and IV with 95% CI's were 0.4278[0.2009, 0.8932] and 0.3308[0.1556, 0.6900] respectively. This means that keep-

Fig.1 The overall Kaplan-Meier survival curve estimate of TB free co-infection survival time of HIV/AIDS patients.

Fig.2 Kaplan-Meier survival curves for TB free co-infection survival time of patents by residence of patients and family size of patients.

Fig.3 Kaplan-Meier survival curves for TB free co-infection free survival time of patients by smoking status and alcohol status of patients.

Fig.4 Kaplan-Meier survival curves for TB free co-infection free survival time of patients by disease stages and functional status of patients.

ing all other factors constant, the expected TB co-infection free survival time of HIV patients with baseline clinical stages III and IV decreases by a factor of 0.4278 and 0.3308 respectively as compared to baseline clinical stage I.

When we look at the functional status of the patients, HIV/AIDS patients with baseline bedridden functional status was found to have shorter TB co-infection free survival time compared to patients with working functional status at baseline. The estimated acceleration factor for posterior mean of the coefficients of bedridden functional status patients was 0.5107 with 95% CI of [0.2750, 0.9455], which can be interpreted as keeping all other factors constant, the expected TB co-infection free survival time of HIV patients with bedridden functional status decreases by a factor of 0.5107 as compared to patients with working baseline functional status.

Family size was also found to be the significant risk fac-

tors of TB co-infection in HIV/AIDS patients with the estimated acceleration factor of 0.3933 with 95% CI [0.2137, 0.7174], and 0.3227 with 95% CI of [0.1518, 0.6845] for patients with family size of 3 - 4 and patients with family size of greater than 5 respectively. This means that, keeping all other factors constant, the TB co-infection free survival time of HIV/AIDS patients with family size of 3 - 4 decreases by a factor of 0.3933 as compared to patients with family size of less than or equal to 2. Similarly, for patients with family size of greater than or equal to 5, keeping all other factors constant, the expected TB co-infection free survival time of HIV/AIDS decreases by a factor of 0.3227 as compared to patients with family size of less than or equal to 2.

The other clinical factors which was risk factors of TB co-infection free survival time in HIV/AIDS patients were baseline CD4 count less than 200 and baseline CD4 count

Table 3: Parametric survival models with their corresponding DIC, WAIC and Marginal log-likelihood

Models	pD	DIC	WAIC	Marginal loglikelihood
Exponential	32.36	664.26	676.03	-432.32
Weibull	29.97	659.73	667.72	-427.32
Log-logistic	27.99	655.65	659.64	-421.66
Log-normal	25.05	663.06	663.69	-432.07
Gamma	27.68	664.19	670.29	-430.14

pD:Effective number of parameters, DIC: Deviance Information Criteria, WAIC:Widely Applicable Information Criteria

in the range 200 - 349 cells/ mm^3 . It was estimated that the acceleration factor for patients with baseline CD4 count of less than 200 cells/ mm^3 was 0.2156 with 95% CI of [0.0948, 0.4742]. Thus, keeping all other factors constant, the TB co-infection free survival time of HIV/AIDS patients with CD4 count of less than 200 cells/ mm^3 decreases by a factor of 0.2156 as compared to patients with baseline CD4 count of greater than or equal to 500 cells/ mm^3 . Similarly, the estimated acceleration factor for patients with baseline CD4 count of 300 - 349 cells/ mm^3 was 0.3753 with 95% CI of [0.1620, 0.8453], which means, keeping all other factors constant, the TB co-infection free survival time of HIV/AIDS patients with baseline CD4 count of 200 -349 cells/ mm^3 decreases by a factor of 0.3753 as compared to patients with baseline CD4 count of greater than or equal to 500 cells/ mm^3 .

It can also be seen that patients with low body mass index and patients with severe anemic status was found to be the significant risk factors for TB co-infection in HIV patients. The estimated acceleration factor for underweight patients was 0.3667 with 95% CI of [0.2141, 0.6955], and the estimated acceleration factor of severe anemic patients was estimated to be 0.3036 with CI of [0.1408, 0.6564]. Thus, keeping all other factors constant, the TB co-infection free survival time of underweight HIV patients decreases by a factor of 0.3667 as compared to patients with normal weight, and, the TB co-infection free survival time of severe anemic HIV patients decreases by a factor of 0.3036 as compared to patients with normal anemic status.

Model Diagnosis results

The Kullback-Leibler divergence (kld) From the tables 4 and Table 5, we can see that the values of kld is zero which means the marginal posterior densities of regression coefficients were well approximated by the Normal distribution. **Effective number of parameters(pD)** In this study, the ratio of sample size (421) and effective number of parameters (28.93) was found to be 14.55, suggesting a reasonably good approximation. The ratio can be interpreted as the number of equivalent replicates corresponding to the number of observations for each expected number of effective parameters.

Discussions

The finding of this study revealed that being urban resident, smoker, alcoholics, in WHO clinical stages III and IV, in bedridden functional status, in larger family size, in lower CD4 count, underweight and anemic patients shortens TB co-infection free survival time of HIV/AIDS patients under ART treatment at Jimma University Medical Center.

Among the study participants who fulfilled the inclusion criteria, about 26.37% of subjects had been co-infected with Tuberculosis. The proportion of TB co-infection in this study cohort is smaller compared to other study settings in different parts of Ethiopia in which it was found to be 62.3% and 40.1% in a retrospective study conducted in seven ART clinics located at Addis Ababa and in North-east Ethiopia with respectively. The proportion of TB co-infection observed in this study is consistent with the findings from North-west Ethiopia(26.4%) [55], Amhara re-

Table 4: Summary results of Bayesian Loglogistic AFT model

Covariates	Categories	Mean	SD	$\hat{\gamma}$	95%CI	kld
	Intercept	6.460	0.937		[4.658, 8.333]	0
Age	Continuous	0.007	0.014	1.0070	[-0.020, 0.035]	0
Sex	Female	-0.189	0.285	0.8278	[-0.748, 0.373]	0
	Male(Ref.)					
Residence	Urban	-1.258	0.297	0.2842	[-1.851, -0.686]*	0
	Rural(Ref.)					
Smoking	Yes	-0.667	0.308	0.5137	[-1.268,-0.061]*	0
	No(Ref.)					
Alcoholics	Yes	-0.879	0.295	0.4151	[-1.459, -0.300]*	0
	No(Ref.)					
Education levels	No education(Ref.)					
	Primary education	0.033	0.427	1.0335	[-0.811, 0.866]	0
	Secondary education	0.102	0.398	1.1074	[-0.688, 0.874]	0
	Tertiary education	0.610	0.469	1.8404	[-0.308, 1.531]	0
	Diploma & above	1.055	0.670	2.8720	[-0.213, 2.417]	0
WHO clinical stages	Stage I(Ref.)					
	Stage II	-0.472	0.408	0.6238	[-1.278, 0.323]	0
	Stage III	-0.849	0.380	0.4278	[-1.605, -0.113]*	0
	Stage IV	-1.106	0.379	0.3308	[-1.860, -0.371]*	0
Functional status	Working(Ref.)					
	Ambulatory	-0.502	0.345	0.6053	[-1.181, 0.173]	0
	Bedridden	-0.672	0.315	0.5107	[-1.291, -0.056]*	0

SD:standard deviation, γ : Acceleration fator, * shows significance, CI: Credible Interval, Kld: Kullback-leibler divergence

Table 5: Summary results of Bayesian Loglogistic AFT model(Continued from Table 4)

Covariates	Categories	Mean	SD	$\hat{\gamma}$	95% CI	kld
Family size	≤ 2 (Ref.)					
	3 - 4	-0.933	0.308	0.3933	[-1.543, -0.332]*	0
	≥ 5	-1.131	0.384	0.3227	[-1.885, -0.379]*	0
CD4 count	< 200	-1.534	0.410	0.2156	[-2.356, -0.746]*	0
	200 - 349	-0.980	0.421	0.3753	[-1.820, -0.168]*	0
	350 - 499	-0.220	0.479	0.8025	[-1.157, 0.723]	0
	≥ 500 (Ref.)					
BMI	Underweight	-0.950	0.300	0.3867	[-1.541, -0.363]*	0
	Normal(Ref.)					
	Overweight	-0.427	0.350	0.6525	[-1.110, 0.264]	0
Marital Status	Married(Ref.)					
	Single	-0.543	0.389	0.5810	[-1.300, 0.224]	0
	Widowed/Divorced	-0.120	0.307	0.8869	[-0.722, 0.482]	0
Hemoglobin	Severe anemic	-1.192	0.392	0.3036	[-1.960, -0.421]*	0
	Moderate anemic	0.080	0.354	1.0832	[-0.604, 0.785]	0
	Normal(Ref.)					

SD:standard deviation, $\hat{\gamma}$:Accelaration factor,* Shows statistical significance, CI: Credible Interval, Kld: Kullback-leibler divergence

gion(27.7%) [63]. However, The proportion of TB/HIV infection reported in this study is larger than the reported prevalence of TB among HIV patients in Albert Luthuli municipality of South Africa(18.2%) [56].

The findings of this study showed that patients' residence place were significantly associated with TB co-infection free survival time of HIV/AIDS patients. Patients who reside in urban areas are more likely to be infected with TB as compared to patients residing in rural areas. This result is

consistent with the report of the retrospective study conducted by Beshir *et al.* at Adama Referral Hospital and Medical College, Oromia, Ethiopia [57]. They reported residence place as one of the significant risk factors of TB co-infection in HIV/AIDS patients. However, the finding of this study is inconsistent with those of other studies undertaken in Ethiopia [55, 58, 59].

The finding of this study have shown that smoking was an important predictor of TB co-infection free survival time of

HIV/AIDS patients following ART treatment. Smoker patients are more likely to be infected with TB compared to non-smoker patients. This result agrees with the result reported by Anye *et al.* based on four year retrospective data of 1077 HIV patients in the Bameda regional hospital of Cameroon [60]. Our result also agrees with report of studies undertaken in Ethiopia [55, 59, 61]. Their results suggested that being smoker is significantly associated with TB co-infection free survival time in HIV/AIDS patients.

Our result also revealed that being alcoholics is associated with TB-co-infection free survival time of HIV/AIDS patients. Being alcoholics leads to shorter TB co-infection free survival time of HIV/AIDS patients attending ART treatment. Our findings showed alcoholic patients were more likely to co-infected with TB compared non-alcoholic patients. This finding is consistent with the reports of Ahmed *et al.* and Abdu *et al.* [55, 61].

The result of our study suggested that baseline WHO clinical stages were one of the clinical factors associated with TB co-infection free survival time of HIV/AIDS patients. Patients with WHO clinical stages III and IV are more likely to be co-infected with TB than those with clinical stage I. This finding supports the findings of the study undertaken by Chang *et al.* [62]. Their study suggested that being in advanced WHO clinical stages is associated with higher risk of developing TB compared WHO clinical stages I and II [62]. Our finding also coincides with the study conducted in Amhara region of Ethiopia by Aweke *et al.* [63].

Similarly, in this study, patients' functional status at baseline was found to be the predictor of TB co-infection free survival time of HIV/AIDS patients. This result is consistent with the report of the study conducted by Aemro *et al.* at Debra Markos referral hospital, Northwest Ethiopia [64]. According to our study, patient with bedridden functional status at baseline had shorter TB co-infection free survival time compared to patients with working functional status at baseline.

Conclusion

Bayesian survival analysis approach with INLA method was applied to fit the parametric survival models to our data set. Among the parametric AFT survival models, Bayesian

Log-logistic AFT model was found to be the best fitting model for our data set.

The TB-free survival time of HIV/AIDS patients was significantly affected by residence place, smoking, drinking alcohol, family size, WHO clinical stages, functional status, CD4 count, body mass index and hemoglobin level of the patients. Thus, the finding of this study suggests that all stake holders are expected to work on prevention of TB by giving awareness on the risk factors of TB co-infection in HIV/AIDS patients.

Abbreviations

AIC: Akaike Information criteria; AFT: Accelerated Failure Time Model; ART: Anti-Retroviral Therapy; AIDS: Human Immuno Deficiency Syndrome; BIC: Bayesian Information Criteria; BMI: Body Mass Index; CPT: Cotrimoxazole Preventive Therapy; Cox-PH: Cox Proportional Hazard model; DIC: Deviance Information Criteria; JUMC: Jimma University Medical Center; FMOH: Federal Ministry of Health; HIV: Human Immunodeficiency Virus; INLA: Integrated Nested Laplace Approximation; IPT: Ionized Preventive Therapy; KLD: Kullback Leibler Deviance; MCMC: Markov Chain Monte Carlo; MLE: Maximum Likelihood Estimate; OI: Opportunistic Infection; SE: Standard Error; PH: Proportional Hazard; PLHIV: People Living with HIV; SSA: Sub-Saharan Africa; TB: Tuberculosis; WAIC: Widely Applicable Information Criteria; WHO: World Health Organization;

Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

The study was funded by Jimma University.

Availability of data and materials

The data set used in this study can be obtained from Jimma University Medical Center.

Ethics approval and consent to participate

Letter of ethical clearance was obtained from Department of Statistics of Jimma University and submitted to Jimma University Medical Center to get permission to undertake the research. This study was developed in accordance with established legislation and complies with the norms of good clinical practice, and informed consent was being not necessary as personal identifying information was kept separate from the research data.

Author's contributions

AKU, SFY and AGD conceived and designed the study. AKU did the analysis under supervision of AGD and SF. AKU, SF and AGD interpreted the results. The final draft of the manuscript was prepared by AKU under the supervision of AGD and SF. Each of the three authors read and approved the final manuscript.

Consent for publication

Not applicable.

Acknowledgements

We thank Jimma University for financing this study. We are also grateful to Jimma University Medical Center for providing us the data set.

Author details

¹Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia. ²Department of Statistics, College of Natural Sciences, Jimma University, Ethiopia.

References

- Gube, A. A., Debalkie, M., Seid, K., Bisete, K., Mengesha, A., Zeynu, A., ... & Gebremeskel, F. (2018). Assessment of anti-TB drug nonadherence and associated factors among TB patients attending TB clinics in Arba Minch Governmental Health Institutions, Southern Ethiopia. *Tuberculosis research and treatment*, 2018.
- Harding, Emilia. "WHO global progress report on tuberculosis elimination." *The Lancet Respiratory Medicine* 8.1 (2020): 19.
- Chakaya, Jeremiah, et al. "Global Tuberculosis Report 2020—Reflections on the Global TB burden, treatment and prevention efforts." *International Journal of Infectious Diseases* (2021).
- Kasibante, J., Rutakingirwa, M. K., Kagimu, E., Ssebambulidde, K., Ellis, J., Tugume, L., ... & Meya, D. B. (2020). Tuberculosis preventive therapy (TPT) to prevent tuberculosis co-infection among adults with HIV-associated cryptococcal meningitis: A clinician's perspective. *Journal of clinical tuberculosis and other mycobacterial diseases*, 20, 100180.
- Raviglione, M., & Maher, D. (2017). Ending infectious diseases in the era of the Sustainable Development Goals. *Porto biomedical journal*, 2(5), 140-142.
- Alene, K. A., Viney, K., Moore, H. C., Wagaw, M., & Clements, A. C. (2019). Spatial patterns of tuberculosis and HIV co-infection in Ethiopia. *PloS one*, 14(12), e0226127.
- Abubeker, F. A., Fanta, M. B., & Dalton, V. K. (2019). Unmet need for contraception among HIV-positive women attending HIV care and treatment service at Saint Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia. *International journal of reproductive medicine*, 2019.
- Gunda, D. W., Maganga, S. C., Nkandala, I., Kilonzo, S. B., Mpondo, B. C., Shao, E. R., & Kalluvya, S. E. (2018). Prevalence and risk factors of active TB among adult HIV patients receiving ART in northwestern Tanzania: a retrospective cohort study. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 2018.
- Chiacchio, T., Petruccioli, E., Vanini, V., Cuzzi, G., La Manna, M. P., Orlando, V., ... & Goletti, D. (2018). Impact of antiretroviral and tuberculosis therapies on CD4+ and CD8+ HIV/M. tuberculosis-specific T-cell in co-infected subjects. *Immunology letters*, 198, 33-43.
- Dravid, A., Natarajan, K., Medisetty, M., Gawali, R., Mahajan, U., Kulkarni, M., ... & Dravid, M. (2019). Incidence of tuberculosis among HIV infected individuals on long term antiretroviral therapy in private healthcare sector in Pune, Western India. *BMC infectious diseases*, 19(1), 1-12.
- Letang, E., Ellis, J., Naidoo, K., Casas, E. C., Sánchez, P., Hassan-Moosa, R., ... & García-Basteiro, A. L. (2020). Tuberculosis-HIV co-infection: progress and challenges after two decades of global antiretroviral treatment roll-out. *Archivos de bronconeumología*, 56(7), 446-454.
- Nugus, G. G., & Irena, M. E. (2020). Determinants of Active Tuberculosis Occurrences after ART Initiation among Adult HIV-Positive Clients in West Showa Zone Public Hospitals, Ethiopia: A Case-Control Study. *Advances in Public Health*, 2020.
- Mitku, A. A., Dessie, Z. G., Muluneh, E. K., & Workie, D. L. (2016). Prevalence and associated factors of TB/HIV co-infection among HIV Infected patients in Amhara region, Ethiopia. *African health sciences*, 16(2), 588-595.
- Belew, H., Wubie, M., Tizazu, G., Bitew, A., & Birlew, T. (2020). Predictors of tuberculosis infection among adults visiting anti-retroviral treatment center at east and west Gojjam, northwest, Ethiopia, 2017. *BMC Infectious Diseases*, 20(1), 1-10.
- Tolles, J., & Meurer, W. J. (2016). Logistic regression: relating patient characteristics to outcomes. *Jama*, 316(5), 533-534.
- Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, 64(3), 441-448.
- Mitku, A. A., Dessie, Z. G., Muluneh, E. K., & Workie, D. L. (2016). Prevalence and associated factors of TB/HIV co-infection among HIV Infected patients in Amhara region, Ethiopia. *African health sciences*, 16(2), 588-595.
- World Health Organization. (2019). Global status report on alcohol and health 2018. World Health Organization.
- Fenner, L., Atkinson, A., Boule, A., Fox, M. P., Prozesky, H., Zürcher, K., ... & International Epidemiology Database to Evaluate AIDS in Southern Africa (IeDEA-SA). (2017). HIV viral load as an independent risk factor for tuberculosis in South Africa: collaborative analysis of cohort studies. *Journal of the International AIDS Society*, 20(1), 21327.
- Azanaw, M. M., Derseh, N. M., Yetemegn, G. S., & Angaw, D. A. (2019). Incidence and Predictors of tuberculosis among HIV patients after initiation of ART treatment in Ethiopia: A systematic review and Meta-analysis.
- Temesgen, B., Kibret, G. D., Alamirew, N. M., Melkamu, M. W., Hibstie, Y. T., Petrucka, P., & Alebel, A. (2019). Incidence and predictors of tuberculosis among HIV-positive adults on antiretroviral therapy at Debre Markos referral hospital, Northwest Ethiopia: a retrospective record review. *BMC public health*, 19(1), 1-9.
- Akerkar, R., Martino, S., & Rue, H. (2010). Implementing approximate Bayesian inference for survival analysis using integrated nested Laplace approximations. *Preprint Statistics, Norwegian University of Science and Technology*, 1, 1-38.
- Rue, H., Lindgren, F., Simpson, D., Martino, S., Teixeira Krainski, E., Bakka, H., ... & Fuglstad, G. A. (2019). INLA: Full Bayesian analysis of latent Gaussian models using integrated nested Laplace approximations. *R package version*, 19(03).
- Klein, J. P., Van Houwelingen, H. C., Ibrahim, J. G., & Scheike, T. H. (Eds.). (2016). *Handbook of survival analysis*. CRC Press.
- Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. Springer Science and Business Media.
- Wang, X., Yue, Y., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.
- Tefera, T. A., Muleta, G., & Tadesse, K. (2021). Bayesian Survival Analysis of Heart Failure Patients: A Case Study in Jimma University Medical Center, Jimma, Ethiopia.
- Martino, S., & Rue, H. (2010). Case studies in Bayesian computation using INLA. In *Complex data modeling and computationally intensive statistical methods* (pp. 99-114). Springer, Milano.
- Brilleman, S. L., Elci, E. M., Novik, J. B., & Wolfe, R. (2020). Bayesian survival analysis using the rstanarm R package. *arXiv preprint arXiv:2002.09633*.
- Klein, J. P., Van Houwelingen, H. C., Ibrahim, J. G., & Scheike, T. H. (Eds.). (2016). *Handbook of survival analysis*. CRC Press.

31. Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
32. Etikan, 'I., Abubakar, S., & Alkassim, R. (2017). The Kaplan-Meier estimate in survival analysis. *Biom Biostatistics Int J*, 5(2), 00128.
33. Moore, D. F. (2016). *Applied survival analysis using R*. New York, NY: Springer.
34. Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. Springer.
35. Alterovitz, G., Liu, J., Afkhami, E., & Ramoni, M. F. (2007). Bayesian methods for proteomics. *Proteomics*, 7(16), 2843-2855.
36. Collett, D. (2015). *Modelling survival data in medical research*. CRC press.
37. Ferreira, D., Barthoulot, M., Pottecher, J., Torp, K. D., Diemunsch, P., & Meyer, N. (2020). Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review. *British journal of anaesthesia*, 125(2), 201-207.
38. Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic bulletin & review*, 25(1), 143-154.
39. Islam, M. K. (2014). Introducing Survival and Event History Analysis. *Canadian Studies in Population [ARCHIVES]*, 41(1-2), 223-224. Science and Business Media.
40. Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin and review*, 25(1), 155-177.
41. Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic bulletin and review*, 25(1), 143- 154.
42. Fong, E., & Holmes, C. C. (2020). On the marginal likelihood and cross-validation. *Biometrika*, 107(2), 489-496.
43. Wang, X., Yue, Y., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.
44. Akerkar, R., Martino, S., & Rue, H. (2010). Implementing approximate Bayesian inference for survival analysis using integrated nested Laplace approximations. *Preprint Statistics, Norwegian University of Science and Technology*, 1, 1-38.
45. Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science and Business Media. [82] Raftery, A. E., Madigan, D
46. Gelman, A. (2013). Two simple examples for understanding posterior p-values whose distributions are far from uniform. *Electronic Journal of Statistics*, 7, 2595-2602.
47. Opitz, T. (2017). Latent Gaussian modeling and INLA: A review with focus on spacetime applications. *Journal de la société française de statistique*, 158(3), 62-85.
48. Wang, X., Yue, Y., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.
49. Mlynarczyk, D., Armero, C., Gómez-Rubio, V., & Puig, P. (2021). Bayesian analysis of population health data. *Mathematics*, 9(5), 577.
50. Al-Shomrani, A. A., Shawky, A. I., Arif, O. H., & Aslam, M. (2016). Log-logistic distribution for survival data analysis using MCMC. *SpringerPlus*, 5(1), 1-16.
51. Held, L., Schrödle, B., & Rue, H. (2010). Posterior and cross-validated predictive checks: a comparison of MCMC and INLA. In *Statistical modelling and regression structures* (pp. 91-110). Physica-Verlag HD.
52. Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation* (Vol. 13). John Wiley and Sons.
53. Gomez-Rubio, V., Bivand, R. S., & Rue, H. (2017). Estimating spatial econometrics models with integrated nested Laplace approximation. *arXiv preprint arXiv:1703.01273*.
54. Marshall, E. C., & Spiegelhalter, D. J. (2003). Approximate cross-validated predictive checks in disease mapping models. *Statistics in medicine*, 22(10), 1649-1660.
55. Ahmed, A., Mekonnen, D., Shiferaw, A. M., Belayneh, F., & Yenit, M. K. (2018). Incidence and determinants of tuberculosis infection among adult patients with HIV attending HIV care in north-east Ethiopia: a retrospective cohort study. *BMJ open*, 8(2), e016961.
56. Bengura, P., & Managa, M. A. (2020). Accelerated failure time modelling of Tuberculosis predictors in HIV/AIDS patients in Albert Luthuli Municipality of South Africa.
57. Beshir, M. T., Beyene, A. H., Tlaye, K. G., & Demelew, T. M. (2019). Incidence and predictors of tuberculosis among HIV-positive children at Adama Referral Hospital and Medical College, Oromia, Ethiopia: a retrospective follow-up study. *Epidemiology and health*, 41.
58. Temesgen, B., Kibret, G. D., Alamirew, N. M., Melkamu, M. W., Hibstie, Y. T., Petrucka, P., & Alebel, A. (2019). Incidence and predictors of tuberculosis among HIV-positive adults on antiretroviral therapy at Debre Markos referral hospital, Northwest Ethiopia: a retrospective record review. *BMC public health*, 19(1), 1-9.
59. Alemu, A., Yesuf, A., Zerihun, B., Getu, M., Worku, T., & Bitew, Z. W. (2020). Incidence and determinants of tuberculosis among HIV-positive individuals in Addis Ababa, Ethiopia: A retrospective cohort study. *International Journal of Infectious Diseases*, 95, 59-66.
60. Anye, C. S., Nkfusai, C. N., Yankam, B. M., Wirsiy, F. S., Tsoka-Gwegweni, J. M., & Cumber, S. N. (2020). A Four-Year Hospital-Based Retrospective Study of the Predictors of Tuberculosis in People Living with HIV and Receiving Care at Bamenda Regional Hospital, Cameroon. *International Journal of Maternal and Child Health and AIDS*, 9(2), 167.
61. Abdu, M., Ali, Y., Anteneh, S., Yesuf, M., Birhanu, A., Mohamed, S., & Hussien, A. (2021). 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*, 16(3), e0248490.
62. Chang, C. A., Meloni, S. T., Eisen, G., Chaplin, B., Akande, P., Okonkwo, P., ... & Kanki, P. J. (2015, December). Tuberculosis incidence and risk factors among human immunodeficiency virus (HIV)-infected adults receiving antiretroviral therapy in a large HIV program in Nigeria. In *Open forum infectious diseases* (Vol. 2, No. 4, p. ofv154). Oxford University Press.
63. Aweke Abebaw Mitku, Zelalem Getahun Dessie, Essey Kebede Muluneh, & Demeke Lakew Workie. Prevalence and associated factors of TB/HIV co-infection among HIV Infected patients in Amhara region, Ethiopia. *African health sciences*, 16(2):588-595, 2016.
64. Aemro, A., Jember, A., & Anlay, D. Z. (2020). Incidence and predictors of tuberculosis occurrence among adults on antiretroviral therapy at Debre Markos referral hospital, Northwest Ethiopia: retrospective follow-up study. *BMC infectious diseases*, 20(1), 1-11.
65. Lartey, M., Asante-Quashie, A., Essel, A., Kenu, E., Ganu, V., & Neequaye, A. (2015). Causes of death in hospitalized HIV patients in the early anti-retroviral therapy era. *Ghana medical journal*, 49(1), 7-11.