British Journal of Medicine & Medical Research
20(6): 1-12, 2017; Article no.BJMMR.32123
ISSN: 2231-0614, NLM ID: 101570965

SCIENCEDOMAIN *international*
www.sciencedomain.org

Bayesian Joint Modelling of Survival of HIV/AIDS Patients Using Accelerated Failure Time Data and Longitudinal CD4 Cell Counts

Markos Abiso Erango[1*], Ayele Taye Goshu[1], Gemeda Bedaso Buta[2] and Ahmed Hassen Dessiso[2]

[1]School of Mathematical and Statistical Sciences, Hawassa University, Ethiopia.
[2]Department of Statistics, Madda Walabu University, Ethiopia.

Authors' contributions

This work was carried out in collaboration between all authors. Authors MAE and ATG designed the study, performed the statistical analysis and wrote the manuscript. Authors GBB and AHD managed the data collection and analyses. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMMR/2017/32123
Editor(s):
(1) Gauri Mankekar, ENT Department, PD Hinduja Hospital, Mumbai, India.
Reviewers:
(1) Radosław Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Matheus Henrique Dal Molin Ribeiro, Federal Technological University of Paraná, Brazil.
(3) Zhongzhan Zhang, Beijing University of Technology, China.
Complete Peer review History: http://www.sciencedomain.org/review-history/18200

Original Research Article

Received 6th February 2017
Accepted 10th March 2017
Published 15th March 2017

ABSTRACT

Objective: This paper aims to compare various Bayesian joint models based on the accelerated failure time distributions in analyzing longitudinal observations on CD4 cell counts as growth measurements and time-to-death events of HIV/AIDS patients. Three accelerated failure time distributions, namely, Weibull, lognormal and loglogistic distributions are considered.
Methods: We consider a total of 354 random sample of HIV/AIDS patients who had been under ART follow-up at Shashemene Referral Hospital in Ethiopia from January 2006 to December 2012. Linear mixed effects model is used for the longitudinal outcomes (square root of CD4 cell counts) with normality assumption, while three parametric accelerated failure time distributions are studied for the time-to-event data. The Bayesian joint models are defined with association parameters and analyzed using Gibbs sampler algorithm. Non-informative prior distributions are assumed. The model selection criteria DIC is employed to identify the model with best fit to the data. Another data

**Corresponding author: E-mail: markos.erango73@gmail.com;*

set obtained by similar setting is also further analyzed using same models.
Results: Both data sets reveal hump-shaped hazard rate functions. The findings from all the Bayesian joint models are consistent. The association parameter in each Bayesian joint model is significant for Weibull and lognormal cases in the second data set. This implies that there is dependence between the two processes: longitudinal CD4 cell counts and the time-to-death event under Weibull and lognormal models. With investigation of the empirical hazard function and the DIC criteria, the Bayesian loglogistic and Bayesian lognormal are selected for the first and second data sets, respectively.
Conclusions: The joint models provide consistent results with higher precision as compared to their respective separate models. We recommend Bayesian joint AFT models for such data with careful consideration of shape of hazard rate functions that the data reveal.

Keywords: Accelerated failure time; Bayesian joint model; CD4 cell count; HIV/AIDS; linear mixed effects; Longitudinal; survival analysis.

1. INTRODUCTION

Globally more than 70 million people have been infected with the HIV virus and the life of about 35 million people have died of the virus since the start of the epidemic [1]. At the end of 2015, it is estimated that 36.7 million (34.0 - 39.8 million) people live with the HIV. Although the burden of the epidemic vary considerably between countries and regions, the Sub-Saharan Africa remains the most severely affected region accounting for nearly 70% of the people living with HIV worldwide. Antiretroviral therapy (ART) has a tremendous impact in improving the quality of life of the patients.

In clinical and medical studies, longitudinal and time-to-event data are considered important measures of health, and most of the time they arise together in practice [2,3]. Longitudinal observations such as CD4 cell counts are obtained from repeated measurements on the HIV/AIDS patients at a series of time points. The related data on time-to-death event describe the length of time for occurrence of the event on each individual patient during a specified study period. The event time depends on the disease markers, and so joint analysis of survival data with related repeated measures has been fruitful to capture the dependence [4,5].

In survival analysis, the proportional hazard models are often used. However, when the proportional hazard assumption is violated, the parametric accelerated failure time (AFT) models are to be used alternatively [6]. Main advantage of using AFT models is that the interpretation of risk factors on the failure time is easier, as it regresses logarithm of survival time on to covariates and random effects. Many methods exist for analyzing longitudinal and survival data separately. But separate analyses of such data may lead to inefficient results. This can be improved by joint modelling. The joint models tend to provide valid and efficient inferences [7].

The AFT models such as Weibull, lognormal, loglogistic, Gamma and Gompertz probability distributions are studied by several researchers [8,12]. One way of making choice among these distributions may be based on their hazard rate functions. Several researchers indicated survival data of HIV infected patients under some conditions might reveal hazard rate increasing at beginning and then declining later [13,17]. It is known that Weibull and Gamma have monotonic hazard rates, while lognormal and loglogistic have non-monotonic hazard rate functions.

A study by [18] disease progression and survival status of adult HIV/AIDS patients under ART follow-up at a hospital. The authors used the AFT Weibull model in Bayesian joint model settings. The authors recommended the Bayesian joint models instead of the respective separate analysis of the data. In this study, we further analyze same data using three AFT models: Weibull, lognormal, loglogistic probability distributions, and compare the joint models under Bayesian settings. Linear mixed effects model is used for the longitudinal measures of CD4 cell counts. The two models are linked by two random effects that are generated by the longitudinal process. The Bayesian modelling of joint processes is a current research area.

2. DATA DESCRIPTION

2.1 Data 1

The first data set on CD4cell counts and survival time are collected for randomly selected 354 adult HIV/AIDS patients under follow-up over the time period from January 2006 to December 2012 at Shashemene Referral Hospital. Patients who are below 16 years and those who started

ART follow-up before January 2006 or after December 2012 are excluded from the study. The CD4 cell counts per mm^3 of blood sample are measured every six month. It is measured repeatedly for each patient under ART follow-up as indicator of progression of the disease over time. The associated survival data are taken with observed time of death event and these are right censored data.

Predictor variables considered for the longitudinal response are: visit time, square of visit time, sex (0 female, 1 male), functional status (0 working, 1 ambulatory, 2 bedridden), alcohol use (-1 no, 0 no opinion, 1 yes), tobacco use (-1 no, 0 no opinion, 1 yes), number of opportunistic infections. Predictors for the survival time are: TB infection status at baseline (0 negative, 1 positive), awareness about ART (0 no, 1 yes), condom use (0 not use always, 1 use always), number of opportunistic infections, and number of living rooms at home.

2.2 Data 2

The second data set is obtained from Bale Robe General Hospital for adult HIV/AIDS patients under follow-up from January 2008 to March 2015. Patients who had at least three CD4 measurements after the first report of HIV diagnosis are eligible for the study. Both data sets are collected with similar settings. Data on repeatedly measurement of CD4 counts and associated survival data are obtained.

Predictors considered for the longitudinal response are: visit time, square of visit time, sex, age, weight and number of opportunistic infections. For survival, age, weight, functional status, tobacco, and condom use are used.

3. STATISTICAL MODEL

The joint models are defined as in [19,2]. The longitudinal and survival process are lined through stochastic dependence. Consider that we have a set of *n* patients followed over a time interval [0,T). The i^{th} patient provides a set of longitudinal measurements Y_{ij} as square root of CD4 cell counts at a follow-up time t_{ij} of visit $j = 1, 2, \cdots n_i$ with n_i number of follow-up of patient $i = 1,\ 2,\ \cdots, n$.

For the survival data, let $T_i = min(t_i, c_i)$ be the observed time for the i^{th} patient, where t_i is time-to-death event and c_i is the censoring time which is assumed independent of t_i and δ_i=1 if the event is observed, δ_i=0 otherwise.

Let the covariates of the longitudinal and survival processes be respectively denoted by $\boldsymbol{X}_{1i}$ and $\boldsymbol{X}_{2i}$. Some of these covariates may be time dependent.

3.1 Linear Mixed Effects Model for Longitudinal CD4Cell Counts

The linear mixed effects model is often applied to handle longitudinal data analysis [19,2]. Given *k* vector of predictors $\boldsymbol{X}_{1i}$, the linear mixed effects model is given as:

$$Y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \epsilon_{ij} \quad (1)$$

where $\mu_i(t_{ij})$ is mean response which is a linear function of $\boldsymbol{X_{1i}}$, $W_{1i}(t_{ij})$ is subject specific random effects having Gaussian distribution, and $\epsilon_{ij} \sim N(0, 1/\sigma_\epsilon^2)$ is a sequence of mutually independent measurement errors.

3.2 Survival Models

Accelerated failure time models are parametric models that provide alternatives to the proportional hazard model for the analysis of survival data [19,2,6,10]. The AFT models assume the direct effect of the covariates on the survival time instead of the hazard rate. Given a vector of *p* predictors $\boldsymbol{X}_{2i}$, the log-linear form of the AFT model for survival time T_i is given as:

$$log(T_i|W_{2i}) = X_{2i}^{'}\alpha + W_{2i}(t_i) + \varepsilon_i \quad (2)$$

where $\boldsymbol{\alpha}$ is a vector of unknown and fixed coefficient of the covariates, $W_{2i}(t)$ refers to subject specific random effects of the survival time having Gaussian distribution, ε_i is a sequence of mutually independent measurement errors. ε_i follows a distribution such that the time-to-event, in this case, Weibull, lognormal and loglogistic distributions. If the error has normal distribution, the time is lognormal, and the error has logistic distribution, the time is loglogistic. The Weibull distribution arises as a general linear form of the smallest extreme value distribution [10].

The three AFT models considered in this study are Weibull, lognormal and loglogistic distributions. Table 1 lists their probability densities, cumulative distributions, hazard rate functions and survival functions.

Table 1. List of accelerated failure time distributions

AFT model	Parameter	$f(t)$	$h(t)$	$S(t)$
Weibull	$\rho, \lambda = \eta^{-\rho}$	$\lambda\rho t^{\rho-1} exp(-\lambda t^{\rho})$	$\lambda\rho t^{\rho-1}$	$exp(-\lambda t^{\rho})$
Lognormal	μ, τ	$\frac{\sqrt{\tau}}{\sqrt{2\pi}\,t} \exp\left(-\frac{\tau}{2}\{\ln(t) - \ln(\mu)\}^2\right)$	$\frac{f(t)}{S(t)}$	$1 - \Phi\left(\frac{\ln(t)-\mu}{1/\sqrt{\tau}}\right)$
Loglogistic	$\rho, \lambda = \mu^{-\rho}$	$\frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^{\rho})^2}$	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$	$\frac{1}{1+\lambda t^{\rho}}$

The parametric links to the covariates and random effects are as follows. The Weibull and loglogistic take same forms.

$T_i \sim Weibull(\rho, \lambda_i(t))$ with $log(\lambda_i(t)) = -\rho\{X_{2i}^{'}\alpha + W_{2i}(t)\}$ or $log(\eta_i(t)) = X_{2i}^{'}\alpha + W_{2i}(t)$
$T_i \sim Lognormal(\mu_i(t), \tau)$ with $log(\mu_i(t)) = X_{2i}^{'}\alpha + W_{2i}(t)$,
$T_i \sim Loglogistic(\rho, \lambda_i(t))$ with $log(\lambda_i(t)) = -\rho\{X_{2i}^{'}\alpha + W_{2i}(t)\}$ or $log(\mu_i(t)) = X_{2i}^{'}\alpha + W_{2i}(t)$

Thus, for example, the Weibull event intensity or hazard rate at time t is expressed as:

$$h(t) = \lambda\rho t^{\rho-1} = \rho\, t^{\rho-1} exp(-\rho\{X_{2i}'\alpha + W_{2i}(t)\})$$

It is known that Weibull hazard rate function is monotonic: decreasing when the shape parameter $\rho < 1$, increasing when the shape parameter $\rho > 1$, and constant when $\rho = 1$. Loglogistic and lognormal distributions have hazard rate functions that are hump-shaped - that is increasing to reach a peak and then declining over time [9,10,12,20].

3.3 Joint Model

The linear mixed effects model for the longitudinal process in equation (1) and the AFT model for the time-to-event in equation (2) are linked through random effects. The joint models are defined as in [19,2,6]. The association between the two processes comes through stochastic dependence between $W_{1i}(t)$ and $W_{2i}(t)$ within a patient but assumed independent across different patients.

We assume there is a stochastic dependence between these two processes through the random effects W_{1i} and W_{2i} as follows:

$$W_{1i}(t) = U_{1i} + U_{2i}\, t \quad (3)$$

$$W_{2i}(t) = r_1\, U_{1i} + r_2 U_{2i} \quad (4)$$

The parameters r_1, r_2 measure the association between the two sub-models (1) and (2) that is expected to be induced by the longitudinal process to the time-to-event process. They represent random intercept and random slope terms in model (1). The variables U_{1i} and U_{2i} are assumed independent latent variables representing subject-specific random effects having normal distributions with mean zeros and precisions τ_{u1} and τ_{u2}. In addition to HIV/AIDS, Prostate Cancer also one of the joint modeling application area in health sciences [5].

3.4 Bayesian Joint Model

Bayesian methods are applied in many applications [20]. Using the usual joint modeling assumptions, subject specific latent variables induce random effects from longitudinal measures to survival observations. Then both sets of observations Y and T are conditionally independent given the random effects $W = \{W_1, W_2\}$ and model parameters $\theta = \{\theta_1, \theta_2\}$, and then the joint density function of the two observations is given as:

$$f(y, t, \delta|\theta_1, \theta_2, w_1, w_2) = \int f(y|\theta_1, w_1)\, f(t, \delta|\theta_2, w_2)\, f(w_2|w_1)\, f(w_1)\, dw_2\, dw_1 \quad (5)$$

3.4.1 Likelihood function

The respective likelihood function of interest is:

$$L(y, t, \delta|\theta_1, \theta_2) = \prod_{i=1}^{n} \int f(y_i|(\theta_1, w_{1i})\, f(t_i, \delta_i|(y_i, \theta_2, w_{2i})^{\delta_i} \left(1 - F(t_i, \delta_i|y_i, \boldsymbol{\theta}_2, w_{2i})^{1-\delta_i}\right) f(w_{2i}|w_{1i})\, f(w_{1i})\, dw_{2i}\, dw_{1i} \quad (6)$$

where $\theta_1 = \{\beta, \sigma_{CD4}^2, \sigma_u^2\}$ are population parameters in the linear mixed effects model, $\theta_2 = \{\alpha, \sigma_t^2, r\}$ are the population parameters in the survival model, β are regression parameters

in the mixed effects model, σ^2_{CD4} is the variance of the transformed CD4 cell count, σ^2_u are the variance of subject specific random effects, α are regression coefficients in the AFT model, σ^2_t is the variance of the transformed event time, r represent the association parameters r_1, r_2. And $f(x)$ and $F(x)$ are probability density and distribution functions, respectively.

3.4.2 Prior distribution

Non-informative joint prior distribution $\pi(\theta, w)$ of the parameters are considered: $\beta's$ and $\alpha's$ are normally distributed with mean zero and large variance 1000, i.e. small precision 0.001; association parameters r_1 , r_2 are each assumed to have normal distribution with mean zero and variance 1000, i.e. small precision 0.001; the shape parameter ρ in Weibull and loglogistic distributions follows $Gamma(2,0.5)$; all precisions follow $Gamma(2,0.5)$.

3.4.3 Posterior distribution

The joint posterior distribution $\pi(\theta, w|y, t, \delta)$ of model parameters θ and random effects W is given by:

$$\pi(\theta, w|y, t, \delta) = \frac{f(y,t|\theta,w)\,\pi(\theta,w)}{\int f(y,t|\theta,w)\,\pi(\theta,w)d\theta\,dw} \quad (7)$$

where $f(y, t|\theta, w)$ is the likelihood function, $\pi(\theta, w)$ is the joint prior probability distribution, and the $\int f(y,t|\theta,w)\,\pi(\theta,w)d\theta\,dw$ is the normalizing constant. It is a high dimensional problem that requires modern computations. Thus inference is based on the Gibbs sampler algorithm using full conditional distributions of the parameters. The Gibbs sampler algorithm is implemented in the WinBUGS software version 14.3 [22]. Inferences are made based on simulation of 60000 iterations with burn-in of 25000 and thinning of 10. Time series plots, autocorrelations and Gelman-Rubin statistics are assessed and they all confirm convergences [23].

3.5 Model Selection

In this study, we compare the three Bayesian joint models with the AFT Weibull, lognormal, loglogistic probability distributions using the deviance information criterion (DIC), Akaike's information criterion (AIC) and Bayes information criterion (BIC). The DIC is a Bayesian alternative to the other two criteria [21,22]. It measures how best the selected model can predict future observations given that it best fits to the data at hand.

DIC involves posterior mean that takes into account prior information and penalized likelihood. It is computed as:

$$\mathrm{DIC} = \mathrm{E}[\mathrm{D}(\boldsymbol{\theta}) \mid \mathrm{data}] + pD \quad (8)$$

where $\mathrm{D}(\boldsymbol{\theta}) = -2\,log(\mathrm{Likelihood}(\boldsymbol{\theta}|\mathrm{data}))$ is deviance and $\mathrm{E}[\mathrm{D}(\boldsymbol{\theta}) \mid \mathrm{data}]$ is the posterior mean of the deviance and pD is effective number of parameters. The AIC and BIC are computed as follows:

$$\mathrm{AIC} = \mathrm{E}[\mathrm{D}(\theta) \mid \mathrm{data}] + 2p \quad (9)$$

$$\mathrm{BIC} = \mathrm{E}[\mathrm{D}(\theta) \mid \mathrm{data}] + p\,log(n) \quad (10)$$

where p is the number of parameters in the model and n is the sample size. The models used in this study involve random effects and so the DIC is more relevant for the model selection.

4. RESULTS AND DISCUSSION

4.1 Case Study I

In the first data set considered 354 HIV/AIDS patients among which 58.5% are females and 41.5% are males. For the longitudinal data, the average baseline CD4 cell counts is estimated to be 156.58 per mm^3 of blood with standard deviation of 92.54. For the survival data, 94.1% are censored while 5.9% are dead. The average survival time of the patients is estimated to be 48.69 months with standard deviation of 21.27.

The proportional hazard assumption fails as the covariates such as opportunistic infection, TB status, knowledge about ART and condom use are found to be time dependent. Fig. 1 shows plots of Schoenfeld residuals against transformed time for each covariate. There is a systematic departure from a horizontal line that indicates violation of the proportional hazard assumption.

4.1.1 Empirical hazard rates

The empirical hazard rate estimates of both data sets are plotted in Fig. 2. They show non-monotonic or hump-shaped behaviors of the hazard rates, showing suitability of lognormal and loglogistic models instead of Weibull in analyzing these data sets. The maximum hazard

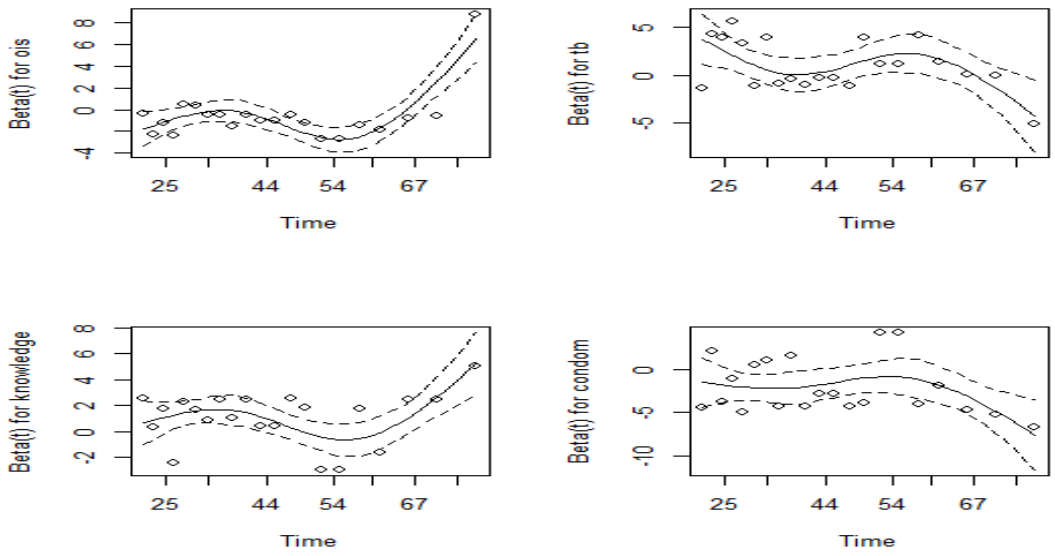

Fig. 1. Plots of scaled Schoenfeld residuals versus follow-up time

rate for Data 1 is estimated to be 0.001508 at time 56.186 months or 4.682 years. For Data 2, the maximum hazard rate is about 0.001971 at time 29.678 months or 2.473 years. The results indicate that the patients under follow-up at Bale Robe General Hospital might have higher event intensity as compared to those at Shashemene Referral Hospital.

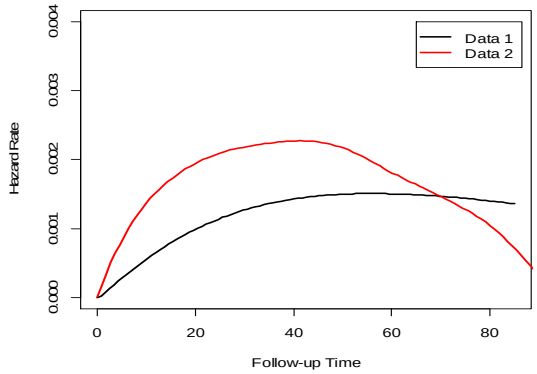

Fig. 2. Plots of empirical hazard rates estimated from data 1 and Data 2

4.1.2 Inferential analysis

The Bayesian joint AFT analysis involves the random effects $W_{1i} = U_{1i} + U_{2i} * t$ in the longitudinal and $W_{2i} = r_1 * U_{1i} + r_2 * U_{2i}$ in the time-to-event models. Prior distributions used are $\beta_j \sim Normal(0,0.001)$, $\alpha_k \sim Normal(0,0.001)$, $\rho \sim Gamma(2,0.5)$, $\tau_{LN} \sim Gamma(2,0.5)$, $r_1 \sim Normal(0,0.001)$, $r_2 \sim Normal(0,0.001)$, $\tau_{u1} \sim Gamma(2,0.5)$, $\tau_{u2} \sim Gamma(2,0.5)$, $\tau_{CD4} \sim Gamma(2,0.5)$.

The posterior means of the parameters, standard deviations, Monte Carlo errors, and 95% credible intervals are estimated.

The simulation of the posterior distribution is made using the Gibbs sampler algorithm and produced three realizations of each 60000 iterations with different initial states. A burn-in of 25000 iterations is considered and convergence diagnoses are assessed. Inferences are then made based on independent samples taken.

4.1.3 Random effects and properties of association parameters

For Data 1, the results show that association parameters r_1 and r_2 are not significant in the Weibull, lognormal and loglogistic cases. There is no significant association between the longitudinal and survival processes with the data and the joint models considered. See the results in Table 2.

4.1.3.1 Weibull case

The subject-specific random effects U_1 and U_2 are found to be significant as $\hat{\tau}_{u1} = 0.107$, 95% CI$(0.087, 0.130)$ and $\hat{\tau}_{u2} = 2.435$, 95% CI$(1.825, 3.204)$ each is having 95% credible interval that does not contain zero. The variances of the random effects are estimated to be about 9.346 and 0.411, respectively, and are

consistently the same in lognormal and loglogistic cases too.

Under the Bayesian joint Weibull model, the posterior estimates of the association parameters are $\hat{r}_1 = -0.003,\ 95\%\ \text{CI}\ (-0.014, 0.0087)$ and $\hat{r}_2 = 0.016,\ 95\%\ \text{CI}\ (-0.060, 0.0898)$. They are not significant. Their correlation is $\text{Corr}(\hat{r}_1, \hat{r}_2) = -0.444$, indicating that the association parameters r_1 and r_2 are negatively correlated.

4.1.3.2 Lognormal case

The subject-specific random effects U_1 and U_2 are again significant: $\hat{r}_{u1} = 0.107,\ 95\%\ \text{CI}(0.088, 0.12$ and $\hat{r}_{u2} = 2.433,\ 95\%\ \text{CI}(1.824, 3.204)$. Their variances are about 9.346 and 0.411, respectively. The posterior estimates of the association parameters are $\hat{r}_1 = 0.00021,\ 95\%\ \text{CI}\ (-0.00281, 0.00318$ and $\hat{r}_2 = -0.00031,\ 95\%\ \text{CI}(-0.016, 0.017)$ These are not significant. Their correlation is $Corr(\hat{r}_1, \hat{r}_2) = -0.283$ which is negative. They have weaker than that of Weibull case.

4.1.3.3 Loglogistic case

The subject-specific random effects U_1 and U_2 are again significant. The posterior estimates of the association parameters are $\hat{r}_1 = 0.00098,\ 95\%\ \text{CI}\ (-0.0095, 0.011)$ and $\hat{r}_2 = -0.00130,\ 95\%\ \text{CI}(-0.059, 0.058)$. They are negatively correlated with $Corr(\hat{r}_1, \hat{r}_2) = -0.311$. Hence the correlations are all negative with strength from lowest to highest for loglogistic, Weibull and lognormal.

Table 2. Bayesian parameter estimations of the joint models based on data 1 (Case Study I)

AFT model	Parameter	Mean	SD	MC Error	95% CI
BJ-Weibull	ρ	4.094	0.169	0.002911	(3.765, 4.439)*
	τ_{CD4}	0.106	0.004	0.000053	(0.099, 0.113)*
	τ_{u1}	0.107	0.011	0.000162	(0.087, 0.130)*
	τ_{u2}	2.435	0.354	0.006246	(1.825, 3.204)*
	r_1	-0.003	0.006	0.000103	(-0.014, 0.0087)
	r_2	0.016	0.038	0.000798	(-0.060, 0.0898)
	Corr(r_1, r_2)	-0.444			
BJ-Lognormal	τ_{CD4}	0.106	0.004	0.000048	(0.0995, 0.114)*
	τ_{u1}	0.107	0.011	0.000184	(0.088, 0.129)*
	τ_{u2}	2.433	0.355	0.008810	(1.824, 3.204)*
	τ_3	12.500	0.937	0.012220	(10.740, 14.430)*
	r_1	0.00021	0.002	0.000234	(-0.0028, 0.0032)
	r_2	0.00031	0.009	0.000189	(-0.016, 0.017)
	Corr(r_1, r_2)	-0.283			
BJ-Loglogistic	ρ	6.790	0.308	0.004860	(6.215, 7.410)*
	τ_{CD4}	0.107	0.0035	0.000060	(0.0998, 0.114)*
	τ_{u1}	0.107	0.011	0.000246	(0.088, 0.130)*
	τ_{u2}	2.412	0.350	0.012430	(1.796, 3.171)*
	r_1	0.00098	0.005	0.000095	(-0.0095, 0.011)
	r_2	-0.00130	0.029	0.000773	(-0.059, 0.058)
	Corr(r_1, r_2)	-0.311			

**significant at 5% significant level; **significant at 10% significant level*

Table 3. Model comparison among the Bayesian joint models for both case studies

Data 1						
Model	**Dbar**	**Dhat**	**pD**	**DIC**	**AIC**	**BIC**
BJ-Weibull	15044.9	14575.6	469.3	15514.2	15086.9	15168.2
BJ-Lognormal	15049.1	14577.8	471.3	15520.4	15091.1	15172.4
BJ-Loglogistic	15013.5	14541.8	471.7	15485.3	15055.5	15136.8
Data 2						
Model	**Dbar**	**Dhat**	**pD**	**DIC**	**AIC**	**BIC**
BJ-Weibull	13369.7	12891.3	478.4	13848.0	13407.7	13481.2
BJ-Lognormal	13378.8	12892.3	486.5	13865.3	13416.8	13490.3
BJ-Loglogistic	13409.9	12932.9	477.0	13886.9	13447.9	13521.4

4.1.4 Ancillary parameters in the AFT models

Weibull shape parameter is estimated $\hat{\rho} = 4.094$, $95\%\ \text{CI}\ (3.765, 4.439)$; lognormal precision is $\hat{\tau} = 12.50,\ 95\%\ \text{CI}\ (10.740, 14.430)$; loglogistic shape parameter is $\hat{\rho} = 6.79,\ 95\%\ \text{CI}\ (6.215, 7.410)$. Precision of the CD4 measure is about $\hat{\tau}_{\text{CD4}} = 0.106,\ 95\%\ \text{CI}\ (0.099, 0.113)$. These are all significant at 0.5 level.

4.1.5 Model comparison

Analysis of Data 1 for model comparison are given in Table 3. Estimates of total DIC for the three models are 15514.20 for BJ-Weibull, 15520.40 for BJ-Lognormal, and 15485.30 for BJ-Loglogistic models. The Bayesian loglogistic joint model has the smallest total DIC, hence it is considered final model for the first data set. Fortunately, the hazard rate of loglogistic distribution behaves like that of the empirical hazard rates.Sign of the coefficients of the covariates and association parameters are consistent under all the models analyzed in this study.

4.1.6 Effects of covariates under Bayesian Loglogistic model

The results from BJ-Loglogistic regression analysis show that the overall mean is significant in the longitudinal process with estimates and 0.95 credible interval {13.660 (13.060, 14.260)}. The covariates having significant effects on CD4 measure are: visit time {2.234 (2.052, 2.402)}, square of visit time {-0.129 (-0.149, -0.109)}, sex male {-0.916 (-1.644,-0.141)}, and tobacco use {1.097 (0.187, 2.021)}. The functional status, alcohol use, number of opportunistic infection are not significant. The positive coefficient of the visit time suggests positive effect of increasing CD4 measure under the ART follow-up. The parameter estimate for the square of visit time is negative, indicating a non-monotone effect of follow-up time on CD4 counts of the patients. Male patients are found to have lower disease marker than the females. Tobacco use is positively related that might indicate those severely sick patient tend to smoke.

From the survival sub-model, the intercept term is significant with estimates and 0.95 credible interval {3.996 (3.924, 4.068)}. And the

Table 4. Results of analysis of data 1 obtained from using BJ-Loglogistic model

Parameter	Mean	SD	MCE	95% CI
β_o	13.660	0.313	0.010370	(13.060, 14.260)*
β_{obt}	2.234	0.087	0.001783	(2.052, 2.402)*
β_{obt^2}	-0.129	0.010	0.000209	(-0.149, -0.109)*
$\beta_{sexMale}$	-0.916	0.375	0.012390	(-1.644, -0.141)*
β_{fun}	-0.430	0.326	0.010990	(-1.087, 0.199)
β_{alc}	-0.880	0.564	0.018260	(-1.999, 0.229)
β_{tob}	1.097	0.560	0.018880	(0.187, 2.021)**
β_{ois}	0.049	0.090	0.002914	(-0.127, 0.224)
α_0	3.996	0.036	0.001279	(3.924, 4.068)*
α_{tb}	-0.247	0.038	0.000709	(-0.322, -0.174)*
α_{artk}	0.277	0.020	0.000458	(0.238, 0.318)*
α_{cond}	-0.361	0.033	0.000962	(-0.427, -0.297)*
α_{room}	0.0047	0.012	0.000367	(-0.019, 0.029)
α_{ois}	0.0055	0.006	0.000139	(-0.008, 0.018)
ρ	6.790	0.308	0.004860	(6.215, 7.410)*
τ_{CD4}	0.107	0.0035	0.000060	(0.0998, 0.114)*
τ_{u1}	0.107	0.011	0.000246	(0.088, 0.130)*
τ_{u2}	2.412	0.350	0.012430	(1.796, 3.171)*
r_1	0.00098	0.005	0.000095	(-0.0095, 0.011)
r_2	-0.00130	0.029	0.000773	(-0.059, 0.058)
Corr(r_1, r_2)	-0.311			
Total DIC	15485.30			

**significant at 5% significant level; **significant at 10% significant level*

covariates TB infection status at baseline {-0.247 (-0.322, -0.174)}, awareness about ART {0.277 (0.238, 0.318)}, and condom use {-0.361(-0.427, -0.297)} are significant. The number of opportunistic infection and the number of living rooms are not significant. TB infection at baseline declines the survival time HIV/AIDS patient. Awareness about ART affect survival time positively related. But condom use of among the patients is negatively related to survival time - that might indicate those severely sick ones tend to condom during sex.

4.2 Case Study II

For Data 2, the average number of baseline CD4 counts is about 177.6 per mm^3 of blood sample with standard deviation of 104.8. Among the 400 sample of patients considered 88.5% of them are censored while 11.5% are dead.

4.2.1 Model comparison

Analysis of Data 2 for model comparison estimations are given in Table 3.

Estimates of total DIC for the three models are 13848.0 for BJ-Weibull, 13865.3 for BJ-Lognormal, and 13886.9 for BJ-Loglogistic models. Hence the BJ-Weibull model best fits to the data. However, the Weibull hazard rate has monotonic behaviour that does not match with the empirical hazard rate that is revealing hump-shaped. Thus the lognormal model, with next smallest DIC, is suggested for analyzing Data 2. Analysis with the BJ-Lognormal model is displayed in Tables 5 and 6.

4.2.2 Random effects and properties of association parameters

For Data 2, the results show that association parameters r_1 and r_2 are significant in some cases. There is significant association between the longitudinal to survival processes. See the results in Tables 5 and 6. Sign of the coefficients of the covariates and association parameters are consistent under all the models analyzed.

The subject-specific random effects U_1 and U_2 under BJ-Lognormal model are significant: $\hat{\tau}_{u1} = 0.103\ (0.084, 0.127)$ and $\hat{\tau}_{u2} = 1.306\ (0.918, 1.905)$. The precision parameter is estimated to be $\hat{\tau}_{CD4} = 0.101\ (0.093, 0.109)$. These parameters are consistently estimates under both Weibull and Loglogistic cases as well. Moreover, the posterior estimates of the association parameters are $\hat{r}_1 = 0.0055\ (0.00096, 0.010)$ and $\hat{r}_2 = -0.017\ (-0.040, 0.0036)$. Here r_1 is significant while r_2 is not. Their correlation is $Corr(\hat{r}_1, \hat{r}_2) = -0.484$ which is negative. The correlation is strongest compared to that of Weibull case (-0.472) and loglogistic case (-0.0031). Similarly in the Weibull case, r_1 and r_2 are not significant but it is significant when $\mathrm{r}_1 = \mathrm{r}_2$ in the model. However, they are not significant in the case of loglogistic. Strength of correlation from lowest to highest for loglogistic, and lognormal, in same order as in the first case study.

Weibull shape parameter is estimated $\hat{\rho} = 2.99\ (2.757, 3.240)$, and loglogistic shape parameter is $\hat{\rho} = 4.285\ (3.946, 4.632)$. These are all significant at 0.5 level. The shape parameters and the precision of the CD4 measure here are lower than those obtained in the first case study.

Table 5. Bayesian parameter estimations of the joint models based on data 2 (Case study II)

AFT Model	Parameter	Mean	SD	MCE	95% CI
BJ-Weibull	ρ	2.990	0.122	0.002047	(2.757, 3.240)*
	r_1	0.012	0.007	0.000125	(-0.0023, 0.026)
	r_2	-0.011	0.036	0.000647	(-0.085, 0.055)
	Corr(r_1, r_2)	-0.472			
BJ-Lognormal	τ_{LN}	6.330	0.508	0.012000	(5.407, 7.447)*
	r_1	0.0055	0.002	0.000077	(0.001, 0.010)*
	r_2	-0.017	0.011	0.000446	(-0.040, 0.0036)
	Corr(r_1, r_2)	-0.484			
BJ-Loglogistic	ρ	4.285	0.177	0.002734	(3.946, 4.632)*
	r_1	0.0022	0.021	0.000654	(-0.040, 0.042)
	r_2	-0.091	31.410	0.527000	(-62.17, 60.86)
	Corr(r_1, r_2)	-0.0031			

significant at 5% significant level

4.2.3 Effects of covariates under BJ-Lognormal model

Under the BJ-lognormal analysis of Data2, all the covariates included for the longitudinal part: visit time, square of visit time, sex, age, weight, and number of opportunistic infections are significant at 0.05 level. The results are given Table 6. The intercept term is significant. Visit time has positive effect on the AIDS progression. The square of visit time has negative effect. Male patients are found to have lower disease marker Effect of age and number of opportunistic infections on the CD4 measure are negative, while that of weight is positive.

For survival part, the intercept term, age, functional status and tobacco use are significant. Weight and condom use are not significant. Age has negative effect on survival time as it does on the CD4 cells counts. Better functional status is related to increase in survival time. Tobacco use is positively related.

Moreover, the hump-shaped hazard rates behavior of the second data is 29.7 months. This is similar to the finding by [11] - estimating the hazard rate reaching a peak of 27.6 months for HIV/AIDS patients in the San Francisco Cohort. The two models selected in this comparison study are AFT lognormal and loglogistic which are in line with the finding with [12]. Other researchers [13,16] found that in resource-poor countries access to antiretroviral therapy has improved during the last years and mortality rates among treated patients have declined significantly. However, compared to patients in high-income countries, patients in resource-poor countries are at higher risk of death in the early months of treatment.

The effects of covariates identified in this study are fairly consistent with the previous findings. For example [17] in study at South Africa found that risk of death is associated with low baseline CD4 cells count and WHO stage IV with majorly attributed causes of death to be tuberculosis, acute bacterial infections, and failure to immune reconstitution.

The three Bayesian models based on the AFT Weibull, lognormal, loglogistic distributions are developed and studied. The findings about the association parameters in the joint models are consistent with results found by [18] who studied disease progression and survival time of HIV/AIDS patients in Ethiopia. The Bayesian joint models are recommended in this study to analyze such data.

Table 6. Results of analysis of data 2 obtained from Bale Robe General Hospital using BJ-Lognormal model as final model

Parameter	Mean	SD	MCE	95% CI
β_o	13.930	0.262	0.006944	(13.42, 14.46)*
β_{obt}	3.098	0.125	0.002212	(2.853, 3.346)*
β_{obt^2}	-0.284	0.023	0.000629	(-0.328, -0.241)*
$\beta_{sexMale}$	-1.027	0.401	0.010030	(-1.835, -0.260)*
β_{age}	-0.534	0.190	0.004928	(-0.908, -0.165)*
β_{wt}	0.716	0.198	0.005424	(0.335, 1.103)*
β_{ois}	-0.615	0.191	0.006001	(-0.995, -0.243)*
α_0	1.273	0.014	0.000601	(1.245, 1.300)*
α_{age}	-0.011	0.0055	0.000114	(-0.022, -0.00049)*
α_{wt}	0.009	0.0052	0.000083	(-0.0015, 0.019)
α_{fun}	0.019	0.0071	0.000281	(0.0045, 0.032)*
α_{tobac}	0.092	0.011	0.000328	(0.071, 0.112)*
α_{cond}	-0.031	0.0062	0.000133	(-0.043, -0.019)
τ_{LN}	6.330	0.508	0.01238	(5.407, 7.447)*
τ_{CD4}	0.101	0.004	8.568E-5	(0.093, 0.109)*
τ_{u1}	0.103	0.011	2.593E-4	(0.084, 0.127)*
τ_{u2}	1.306	0.248	0.009248	(0.918, 1.905)*
r_1	0.0055	0.0023	7.717E-5	(0.00096, 0.010)*
r_2	-0.017	0.011	4.463E-4	(-0.040, 0.0036)
Corr(r_1, r_2)	-0.484			
Total DIC	13865.30			

significant at 5% significant level

5. CONCLUSIONS

The objective of this study is to develop study three Bayesian joint AFT models involving Weibull, lognormal, log logistic distributions. Two data sets are considered for which we found insignificant dependence for the first data and significant dependence between the longitudinal and time-to-event processes for the second study. The best fitting models to the data sets are different: log logistic for the first data and lognormal for the second one. The shape parameters and the precision of the CD4 measure are all lower in the second data than those obtained in the first data. It is important to consider hazard rate functions while modelling such failure time data.

For the first data, significant effects on CD4 measure are found for visit time, square of visit time, sex, and tobacco use. And the covariates TB infection status at baseline, awareness about ART, and condom use are significantly related to survival time. For the second data, the covariates visit time, square of visit time, sex, age, weight, and number of opportunistic infections are found to significantly related to the CD4 counts while age, functional status and tobacco use are significant in survival part.

Various covariates and the shared random effects would be considered to better understand the joint longitudinal and time-to-event processes. We recommend the Bayesian joint models to jointly analyze the longitudinal and survival data of the HIV/AIDS patients so as to predict future disease progression and survival of the patients. The model can be used for monitoring the patients for effective ART services.

CONSENT

It is not applicable.

ETHICAL APPROVAL

The ethical clearance for this study has been obtained from Shashemene Referral and Bale Robe General Hospitals, and Hawassa University.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank the Shashemene Referral and Bale Robe General Hospitals for providing us the data sets used in this study. Our gratitude goes to Professor Arnoldo Frigessi, Department of Biostatistics, University of Oslo, for his valuable comments on the draft document.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. WHO. Global health observatory data; 2016.
   Available:www.who.int/gho/hiv/en
   (November 2016)
2. Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000;4: 465-480.
3. Wu L, Liu W, Grace Y, Hung Y. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. Journal of Probability and Statistics. 2012; 1:1-17.
4. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica. 2004;14(3): 809-834.
5. Yu M, Law NJ, Taylor JM, Sandler HM. Joint longitudinal-survival-cure models and their application to prostate cancer. Statistica Sinica. 2004;14(3):835-862.
6. Tseng Y, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. Biometrika. 2005; 92(3):587-603.
7. Hatfield A, Hodges S, Carlin P. Combining longitudinal and survival information in Bayesian joint models: When are treatment estimates improved? Biostatistics. 2012;1-31.
8. Vallinayagam V, Prathap S, Venkatesan P. Parametric regression models in the analysis of breast cancer survival data. International Journal of Science and Technology. 2014;3(3).
9. Hayat EA, Suner A, Uyar B, Dursun Ö, Orman MN, Kitapçioğlu G. Comparison of five survival models: Breast cancer registry data from Ege University Cancer Research Center. Turkiye Klinikleri J Med Sci. 2010; 30(5).
10. Klein JP, Moeschberger ML. Survival analysis techniques for censored and truncated data. Second Edition. Statistics

for Biology and Health, Springer LLC; 2003.

11. Byers RH Jr, Morgan WM, Darrow WW, Doll L, Jaffe HW, Rutherford G, Hessol N, O'Malley PM. Estimating AIDS infection rates in the San Francisco cohort. AIDS. 1988;2(3):207-10.
12. Bennett S. Loglogistic regression models for survival data. Journal of the Royal Statistical Society – Series C (Applied Statistics). 1983;32:165–171.
13. Andinet Worku, Miguel San. Determinants of survival in adult HIV patients on antiretroviral therapy in Oromia, Ethiopia. Open Access article. Global Health Action. 2010;3:5398.
14. Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boulle A, Miotti P, Wood R, Laurent C, Sprinz E, Seyler C, Bangsberg DR, Balestre E, Sterne JA, May M, Egger M. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: Comparison between low-income and high-income countries. Lancet. 2006; 367(9513):817-824.
15. Laurent C, NgomGueye NF, Ndour CT, Gueye PM, Diouf M, Diakhate N, Toure Kane NC, Laniece I, Ndir A, Vergne L, Ndoye I, Mboup S, Sow PS, Delaporte E. Long-term benefits of highly active antiretroviral therapy in Senegalese HIV-1-infected adults. J Acquir Immune Defic Syndr. 2005;38(1):14-17.
16. Severe P, Leger P, Charles M, Noel F, Bonhomme G, Bois G, George E, Kenel-Pierre S, Wright PF, Gulick R, Johnson WD Jr., Pape JW, Fitzgerald DW. Antiretroviral therapy in a thousand patients with AIDS in Haiti. N Engl J Med. 2005;353(22):2325-2334.
17. Lawn SD, Myer L, Orrell C, Bekker LG, Wood R. Early mortality among adults accessing a community-based antiretroviral service in South Africa: Implications for programme design. AIDS. 2005;19(18):2141-2148.
18. Gemeda B. Buta, Ayele T. Goshu, Hailemichael M. Worku. Bayesian joint modelling of disease progression marker and time to death event of HIV/AIDS patients under ART follow-up. British Journal of Medicine & Medical Research. 2014;5(8):1034-1043.
19. Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. The American Statistician. 2004;58(1):16-24.
20. Farewell VT, Prentice RL. A study of distribution shape in life testing. Technometerics. 1979;19:69-75.
21. Robert CP. The Bayesian choice: From decision-theoretic foundations to computational implementation. Second Edition. Springer LLC; 2007.
22. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS version 1.4 User Manual; 2003.
Available:http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf
23. Spiegelhalter DJ, Best NG, Carlin BP. Bayesian deviance, the effective number of parameters and comparison of arbitrarily complex models. MRC Biostatistics Unit, Cambridge University; 1998.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/18200