American Journal of Theoretical and Applied Statistics
2017; 6(4): 182-190
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20170604.13
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Bayesian Joint Modelling of Longitudinal and Survival Data of HIV/AIDS Patients: A Case Study at Bale Robe General Hospital, Ethiopia

Ahmed Hasan Dessiso[1, *], Ayele Taye Goshu[2]

[1]Department of Statistics, College of Natural and Computational Science, Madda Walabu University, Bale Robe, Ethiopia

[2]School of Mathematical and Statistical Sciences, Hawassa University, Hawassa, Ethiopia

Email address:

ah30994@gmail.com (A. H. Dessiso), ayele_taye@yahoo.com (A. T. Goshu)

[*]Corresponding author

To cite this article:

Ahmed Hasan Dessiso, Ayele Taye Goshu. Bayesian Joint Modelling of Longitudinal and Survival Data of HIV/AIDS Patients: A Case Study at Bale Robe General Hospital, Ethiopia. *American Journal of Theoretical and Applied Statistics*. Vol. 6, No. 4, 2017, pp. 182-190.
doi: 10.11648/j.ajtas.20170604.13

Received: February 14, 2017; **Accepted**: February 25, 2017; **Published**: June 23, 2017

Abstract: Joint analysis of longitudinal and survival data has received increasing attention in the recent years, especially for AIDS. This study explores application of Bayesian joint modeling of HIV/AIDS data obtained from Bale Robe General Hospital, Ethiopia. The objective is to develop separate and joint statistical models in the Bayesian framework for longitudinal measurements and time to death event data of HIV/AIDS patients. A linear mixed effects model (LMEM), assuming homogenous and heterogeneous CD4 variances, is used for modeling the CD4 counts and a Weibull survival model is used for describing the time to death event. Then, both processes are linked using unobserved random effects through the use of a shared parameter model. The analysis of both the separate and the joint models reveal that the assumption of heterogeneous (patient-specific) CD4 variances brings improvement in the model fit. The Bayesian joint model is found to best fit to the data, and provided more precise estimates of parameters. The shared frailty is significant showing the association between the linear mixed effect (LME) and survival models.

Keywords: ART, Bayesian, CD4 Count, HIV/AIDS, Joint Model, Longitudinal Model, Survival Model

1. Introduction

The term joint modeling refers to the statistical analysis of the longitudinal and survival data while taking account of any association between the repeated measurement and time to event outcomes. The development of joint model has greatly expanded the scope of models to accommodate many data complexities, yet relatively little attention has been paid to these approaches properties and performance.

The approach that this study used to build a joint model is simultaneously modeling the longitudinal CD4 measurements and the time to death by linking them using unobserved random effects through the use of a *shared parameter* model. In the proposed model, to characterize the longitudinal CD4 measurements a linear mixed effects model (LMEM) that incorporates patient specific CD4 variability is used for the longitudinal sub-model while a Weibull model is used to describe the time-to-death data of survival sub-model. Then, the two sub-models are linked through shared parameters or shared variables [1] with different forms, since these random effects characterize the subject specific longitudinal process. Alternatively, the two models are governed by the same underlying latent process (shared variables).

In this study, we employ the joint modeling approach developed by [2]. We applied the Bayesian joint and separate modeling of the patterns of CD4 changes and time to death event to mainly characterize the relationship between the two data. The central research questions are: What are the factors for determining the longitudinal evolution of CD4 cell count of HIV/AIDS patient under ART follow up? What are the risk factors for the death of HIV/AIDS patient under ART? How strong is the association between the disease progression and the time to death of the HIV/AIDS patients?

The objective of the study was to jointly analyze and build joint model for CD4 progression and time to death of HIV/AIDS patients simultaneously linked with unobserved random effects through the use of shared parameters based on data from Hospital records. The results of this study will be very useful in the development of an effective HIV care and antiretroviral therapy (ART) patient monitoring system.

2. Data and Methodology

Data was obtained solely by reviewing medical records of 400 representative sample of HIV patients diagnosed at the Bale Robe General Hospital, Ethiopia between January, 2008 and March, 2015. The target population of our study was all adult (age ≥ 18 years) HIV/AIDS patients who had at least three CD4 measurement after the first report of HIV diagnosis regardless of clinical stage were eligible for the study. All patients who were below 18 years and those patients who started ART before January 2008 or after March, 2015 were not included in the study. So our study population consists 1214 patients who fulfilled the inclusion criteria. The ethical clearance was obtained from the Hospital.

Response Variables

Two outcome variables were considered in this study. The first response variable is longitudinal CD4 count per mm^3 of blood. It is measured repeatedly for each HIV/AIDS patient under ART. The other response variable is the survival outcome variable. It is time to death event of the patient under ART follow up.

Explanatory variables

Predictors considered for the longitudinal response were observation time, sex of patient, Age, Weight and number of opportunistic infections (OIs); and those for survival response were Age, Weight, Functional status, Tobacco and condom use.

Data collected from epidemiological studies such as clinical trials or observational cohort studies often include information for an event time of interest (e.g. survival times) and repeated measurements of one or more longitudinal processes that might be associated with patient prognosis. Let the sample in a longitudinal study be records of m subjects, and let the i^{th} subject have n_i measurements over time. For $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n_i$, let the notation y_{ij} denote the longitudinal measurement of the i^{th} subject at time t_{ij}.

Let T_i denote the observed event time for the i^{th} subject, which is taken as the minimum of the true event time T_i^* and the censoring time C_i, i.e. $Ti = min(T_i^*, C_i)$. We also define $\delta_i = I(T_i^* \leq C_i)$ which takes value 0 for a right-censored event time and value 1 for an actual event time. Therefore, the observed time to event data consist of the pairs $(T_i, \delta_i), i = 1,2,\ldots,m$.

The baseline covariates predictive of the longitudinal and survival processes are denoted by X_1 and X_2 respectively, which may or may not be the same and let U_i be a vector of person-specific latent variables. Separate models for the longitudinal and survival data, and the joint model are subsequently defined here below. The Bayesian joint model is then derived.

2.1. Longitudinal Model

Longitudinal studies typically involve following one or more cohorts of subjects or experimental units repeatedly over two or more time points. One of the major objectives of statistical analysis is to address variations in the data. For longitudinal data, there are two sources of variations: within-subject variation; the variation in the measurements within each subject, and between-subject variation; the variation in the data between different subjects [3]. Modeling within-subject variation allows studying changes over time, while modeling between-subject variation allows understanding differences between subjects [4].

Linear Mixed Effects Model

Three classes of models are commonly used for analysis of longitudinal data; mixed effects model (or random effects model), marginal models (generalized estimating equations (GEE) models) and transition models [3]. Linear Mixed effects models (LMEM) are widely used in which random effects are introduced to incorporate the between subjects variation and within subject correlation in the data. In marginal models, the mean structure and the correlation (covariance) structure are modeled separately without distribution assumptions for the data while in the transitional models, the within subject correlation is modeled via Markov structures [5]. In linear mixed effects model, the sequence of the longitudinal measurements y_{ij} at times t_{ij} for $i = 1,2,\ldots,m$ and $j = 1,\ldots,n_i$ is modeled as:

$$y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \varepsilon_{ij} \quad (1)$$

Where $\mu_i(t) = X_{1i}^T(t)\beta_1$ is the mean response, $W_{1i}(t) = Z_{1i}^T(t)U_i$ incorporates subject-specific random effects and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is a sequence of mutually independent measurement errors. $W_{1i}(t)$ can be viewed as the true individual level CD4 trajectories after they have been adjusted for the overall mean trajectory and other fixed effects. The vectors $X_{1i}^T(t)$ and β_1 represent possibly time varying explanatory variables and their corresponding regression coefficients, respectively. U_i are vectors of random effects corresponding to the explanatory variables $Z_{1i}^T(t)$ (which may be a subset of $X_{1i}^T(t)$ and are typically modeled as identically and independently distributed $N(0, D)$.

2.2. Survival Model

Survival Analysis typically focuses on time to event data. In the most general sense, it consists of techniques for positive valued random variables such as time to death, time to on set (or relapse) of a disease, length of stay in a hospital, duration of a strike etc. In order to define a survival time random variable, we need an unambiguous time origin or a time scale (e.g. real time (days, weeks, months, years) and definition of the event of interest. Survival time random variables are always non-negative, i.e. if we denote the survival time by, T then $T \geq 0$ can either be discrete (taking a finite set of values) or continuous (defined on $(0, \infty)$). We

need statistical methods that use data on all subjects, whether their survival times are observed or we only observe time until censoring. There are several equivalent ways to characterize the probability distribution of a survival random variable. Non-parametric, semi parametric and parametric models are available to model survival data. Parametric models are used in this study [2].

Weibull Distribution

Parametric models are models requiring the specification of a probability distribution for the survival times, i.e., parametric models assume that the survival data follow some probability distribution. The most commonly used parametric model is the Weibull model. In a Weibull model, the survival time for the i^{th} subject is assumed to follow a Weibull distribution:

$$T_i \sim Weibull(\rho, \mu_i(t)), log(\mu_i(t)) = X_{2i}^T(t)\beta_2 \text{ and } \rho > 0.$$

The vectors $X_{2i}^T(t)$ and β_2 represent (possibly time-dependent) explanatory variables and their corresponding regression coefficients. They may or may not have elements in common with $X_{1i}^T(t)$ and β_1 in the longitudinal model [2]. The event intensity (or hazard) at time t is given as

$$\lambda_i(t) = \rho t^{\rho-1}\mu_i(t) = \rho t^{\rho-1} exp\{X_{2i}^T(t)\beta_2\} \quad (2)$$

which is monotone in t (decreasing if $\rho < 1$, increasing if $\rho > 1$) and reduces to the exponential (constant in t) hazard if $\rho = 1$.

2.3. The Joint Model Structure

The structure of the joint modeling requires a model for the longitudinal response and a model for the event time data. These two responses should be modeled simultaneously, therefore, a structure for considering the association between them is required. This study has used the joint modeling approach developed by [2] who investigated the approach proposed by [6] from a Bayesian perspective and relying on Markov Chain Monte Carlo (MCMC) algorithms.

The association between the longitudinal outcomes and event times can arise in two ways. One way is through common explanatory variables and the other is through stochastic dependence between (W_{1i}, W_{2i}). [6] proposed to jointly model the two processes via a latent zero-mean bivariate Gaussian process on (U_{1i}, U_{2i}), which is independent across different subjects. The joint model consists of two linked sub models, which they refer to as the measurement model for the longitudinal process and the intensity model for the survival process. We can apply this joint modeling strategy to connect the classical models for longitudinal data and survival data with each other. When association between the two processes exists, we should obtain less biased and more efficient inferences by using this joint model.

2.3.1. The Longitudinal Sub Model Specification

The main goal, in this study, is to jointly model the longitudinal CD4 measurements and time to death, with a special attention to the effect of CD4 variability on the risk of death. In most joint models studied in the past decade, longitudinal data are delineated by a conventional linear mixed model assuming homogeneous within subject variance. However, such a homogeneity assumption automatically precludes the assessment of the research question "'whether individuals with different levels of CD4 variability have different susceptibility to die". In the proposed model, the CD4 trajectory is described by the LMEM that incorporates subject-specific variance [7]. Thus, the longitudinal sub model that incorporates subject specific variances is given as:

$$y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \varepsilon_{ij} \quad (3)$$

Where $\varepsilon_{ij} \sim N(0, v_i)$, $log(v_i) \sim N(\mu_v, \sigma_v^2)$. This model incorporates subject-specific variances, i.e., the random errors, ε_{ij} may not have homogeneous variance. Thus, here v_i represents the (true) within-subject variability, which follows a lognormal distribution with mean μ_v and variance σ_v^2.

2.3.2. The Survival Sub Model Specification

After specifying the longitudinal sub model, the next aim is to associate the true and unobserved value of the longitudinal outcome at time t with the survival outcome via a latent zero mean (multivariate) Gaussian process on the random effects U_i, which is independent across different subjects [6].

As shown before, both of the separate and joint models assume the longitudinal sub model has the form similar to the usual LMEM, while the survival model in the joint model includes a latent association function $W_{2i}(t)$. Thus, the survival sub-model is specified in the form as:

$$\lambda_i(t) = \rho t^{\rho-1}\mu_i(t) = \rho t^{\rho-1} exp\{X_{2i}^T(t)\beta_2 + W_{2i}^T(t)\} \quad (4)$$

The form of the association function $W_{2i}(t)$, is similar to $W_{1i}(t)$, including subject specific covariate effects and an intercept (often called a frailty). When $W_{2i}(t) = 0$, the association induced is only via shared baseline covariates. Specifically, the joint model links the LMEM that incorporates subject specific variance and model by taking:

$$W_{1i}(t) = U_{1i} + U_{2i} * t, \quad (5)$$

And

$$W_{2i}(t) = \gamma_1 U_{1i} + \gamma_2 U_{2i} + \gamma_3(U_{1i} + U_{2i}(t)) + U_{3i} \quad (6)$$

The longitudinal model (5) is of the usual [8] form, with each patient receiving random intercept and linear slope terms. The parameters, γ_1, γ_2 and γ_3 in the survival model (6) measure the association between the two sub models induced by the random intercepts and linear slope respectively. As mentioned before, the bi-variate latent variables $U^T = (U_{1i}, U_{2i})$ have a mean zero bivariate Gaussian distribution $N(0, D)$ and the subject specific variances v_i have a log-normal distribution $\log v_i \sim (\mu_v, \sigma_v^2)$ while the U_{3i} are independent frailty terms, modeled as iid $N(0, \sigma_3^2)$, independent of (U_{1i}, U_{2i}). Regarding the association function, $W_{2i}(t)$, a variety of several latent

processes are considered. Finally, the precise nature of the two sub models i.e., the exact form of $W_{1i}(t)$ and $W_{2i}(t)$ and their latent association are selected using Deviance Information Criteria (DIC).

2.4. Bayesian Joint Model Parameter Estimation

A main challenge in inference for joint models is the computational complexity, when the dimension of the random effects is not small. In our joint modeling, the longitudinal measurement and time to event process are shared some components of their multivariate Gaussian process. A simultaneous method of inference based on the joint likelihood of longitudinal measurements and times to event may be favored, but the computational problems can be extensive. A Bayesian approach can reduce the complexity of these problems. We assume that, conditional on components of a multivariate latent Gaussian process, each characteristic and event time are independent. In the proposed joint model, a Bayesian approach using Markov Chain Monte Carlo (MCMC) is implemented. One of the most important advantages of using a Bayesian approach to joint modeling may be the alleviation of the computational burdens.

The standard maximum likelihood method involves integrating out latent variables from the log likelihood function, which is difficult when dealing with high-dimensional variables [9]. As a result, the proposed joint models are estimated under a Bayesian framework using Markov chain Monte Carlo (MCMC) methods with Gibbs sampling using Win BUGS software. Various authors, including [10], [11], [12], [13], [14] and [15] have also studied Bayesian joint models. Joint models may contain many unknown parameters, which may lead to potential problems in inference.

The other important advantage of Bayesian methods is that they can incorporate additional information from similar studies or from experts guess to the model in the forms of prior distributions. Thus, Bayesian methods can be very useful for inference of joint models. For Bayesian joint models, the model parameters are assumed to follow some prior distributions, and inference is then based on the posterior distribution given the observed data. Making use of the usual joint modeling assumption that the subject specific latent variable induce all of the association between longitudinal process Y_i and survival outcome T_i, so that Y_i and T_i are conditionally independent given random effects U.

2.4.1. Joint Model Likelihood

Given the random effects, the longitudinal process is assumed to be independent of the event times. So that the full joint distribution of the longitudinal continuous response and time to event can be specified in the form of:

$$f(Y_i, T, \delta|\theta_1, \theta_2) = \int f(Y_i|\theta_1, U_i) f(T, \delta|Y_i, \theta_2, U_i) f(U_i) dU_i$$

With the corresponding likelihood function being

$$L(Y, T, \delta|\theta_1, \theta_2) = \prod_{i=1}^{n} \int f(Y|\theta_1, U_i) f(T, \delta|Y, \theta_2, U_i)^{\delta_i} \times (1 - F(T, \delta|Y, \theta_2, U_i))^{1-\delta_i} f(U_i) dU_i$$

where $U_i = \{U_{1i}, U_{2i}\}$ represents the shared underlying process, $\theta_1 = \{\beta_1, D, \mu_v, \sigma_v^2\}$ are the population parameters as given in the LMEM, $\theta_2 = \{\beta_2, \gamma, \sigma_3^2\}$ are the population parameters as given in survival models, $f(.)$ and $F(.)$ denote density and distribution functions, respectively.

2.4.2. Prior and Posterior Distributions

In a Bayesian approach, model parameters are treated as random variables and assigns probability to each, which is the major difference to the likelihood approach. The assumed distributions for the parameters are called prior distributions. Bayesian estimation and inference is based on the posterior distribution which is the conditional distribution of unobserved quantities given the observed data. The joint posterior distribution for all unknown parameters θ and random effects U is then given by

$$f(\theta, U|Y, T) = \frac{f(Y,T|\theta,U)\pi(\theta)\pi(U)}{\int f(Y,T|\theta,U)\pi(\theta)\pi(U)d(\theta)d(U)} \quad (7)$$

where $f(\theta, U|Y, T)$ is the posterior probability distribution, $f(Y, T|\theta, U)$ is the likelihood function and $\pi(\theta), \pi(U)$ is the prior probability distribution

In the Bayesian framework, inference follows from the full posterior distribution. Bayesian joint model inference is then based on samples drawn from the posterior distribution using an MCMC algorithm such as the Gibbs sampler and Metropolis Hastings. For example, the posterior means and variances of the parameters can be estimated based on these samples, and Bayesian inference can then be based on these estimated posterior means and variances. This sampling can be done using Win BUGS software. We selected very vague prior distributions in our Win BUGS analysis. That is, we chose priors and hyper parameter values in such a way that, the priors will have minimal impact relative to the data.

2.5. Diagnostics of Chain Convergence and Model Selection

For assessing convergence, we have used multiple chains. If parallel chains with varying starting values give the same solution that will increase our confidence for convergence. A simple (informal) method of assessing chain convergence is to look at the history of iterations using a time series plot. If the chains show a reasonable degree of randomness between iterations, it signifies that the Markov chain has found an area of high likelihood and is integrating over the target density [16] and hence indicating that it has converged.

Also, for model selection, we evaluate model fits by inspecting DIC [17], a hierarchical modeling generalization of the AIC (Akaike Information Criterion). The DIC approach mimics AIC by setting $DIC = \bar{D} + pD$. The first term is the posterior expectation (mean) of the deviance function and measures the goodness-of-fit. The second term

pD is the effective number of parameters and measures model complexity. Since a smaller $\overline{D}$ indicates a better fit and a smaller pD indicates a parsimonious model, small values of the sum (DIC) indicate preferred models.

3. Results

The objective of this study was to model the longitudinal measurements of CD4 counts per mm^3 of blood and the associated time to death using the Bayesian joint modelling approach. The average number of baseline CD4 count is 177.57 per mm^3 of blood with standard deviation of 104.808. The results of the analysis showed that from 400 patients included in the study, 354(88.5%) are censored while 46(11.5%) are dead.

3.1. Results of Linear Mixed Effects Models

Because of right skewness of the response variable, we have used the square root transformation for CD4 counts in our analysis. Taking advantage of the fact that the conventional LMEM (assumes homogeneous within-subject variances) described by [8]; and the LMEM that incorporates subject-specific (heterogeneous) variances, produce almost identical estimates for fixed effects [18], initially, the repeated CD4 measurements are analyzed using conventional LMEM (1). The results show that all the covariates included in the model, Observation time, Baseline Age, Baseline Weight and baseline number of opportunistic infections are statistically significant at 10% level of significance. This is based on whether or not the 90% posterior credible intervals for each estimate includes zeros.

Let y_{ij} denote the square root of j^{th} CD4 count of the i^{th} patient at time t_{ij}, $(i = 1,2,\ldots,400)$ and $(j = 1,2,\ldots,n_i \leq 11)$. Hence, the linear random effects model for square root of CD4 counts is specified as:

$$y_{ij} = \beta_{11} + \beta_{12}t_{ij} + \beta_{13}t_{ij}^2 + \beta_{14}Sex_i + \beta_{15}Age_i + \beta_{16}Weight_i + \beta_{17}Ois_i + W_{1i}(t_{ij}) + \varepsilon_{ij} \quad (8)$$

Where $W_{1i}(t_{ij}) = U_{1i} + U_{2i}(t)$. Here, $W_{1i}(t_{ij})$ includes the random effects for intercept and linear time slopes over time. Where, $U_i = (U_{1i}, U_{2i})^T \sim N_2(0, D)$. This specification allows different subjects to have different baseline CD4 counts and different time trends for CD4 counts during treatment period.

In order to examine whether the assumption of heterogeneous within-subject variance for the CD4 counts is supported, longitudinal model is fitted using Win BUGS. Table 1 below presents the posterior means and 90% credible intervals for the population parameters of the two models; for the conventional LMEM and for the model incorporating patient-specific variances. Here the results of the two models are nearly the same. In both models both the linear and quadratic time effects, Sex, baseline Age, Baseline Weight and baseline number of opportunistic infections are statistically significant at 0.1 level of significance.

Table 1. *Posterior Means and 90% Credible Intervals for the Population Parameters of the Convectional LMEM and for Model that incorporates Patient-Specific Variances.*

Parameters	Without Patient-Specific Variances		With Patient-Specific Variances	
	Posterior Mean	90% CI	Posterior Mean	90% CI
Fixed Effects	-	-	-	-
Intercept ($\hat{\beta}_{11}$)	13.92	(13.58, 14.26)	14.0	(13.66, 14.33)
time ($\hat{\beta}_{12}$)	3.125	(2.963, 3.288)	2.94	(2.787, 3.093)
time^2 ($\hat{\beta}_{13}$)	-0.2954	(-0.3241, -0.2662)	-0.2617	(-0.2888, -0.2345)
Sex ($\hat{\beta}_{14}$)	-0.9978	(-1.529, -0.4643)	-1.047	(-1.579, -0.5097)
Age ($\hat{\beta}_{15}$)	-0.5402	(-0.7912, -0.2884)	-0.5414	(-0.7841, -0.2965)
Weight ($\hat{\beta}_{16}$)	0.7043	(0.4532, 0.9539)	0.7162	(0.4622, 0.9738)
Ois ($\hat{\beta}_{17}$)	-0.6417	(-0.8841, -0.4014)	-0.6802	(-0.9226, -0.4383)
$\hat{\sigma}^2_{\varepsilon}$	9.8814	(9.3985, 10.4069)	-	-
Random Effects				
var ($\hat{U}_1$)	9.6993	(8.5543, 11.1074)	10.0402	(8.9126, 11.4181)
var ($\hat{U}_2$)	0.8190	(0.6689, 1.0241)	0.7037	(0.5750, 0.8779)
$\hat{\mu}_v$	-	-	2.039	(1.959, 2.119)
$\hat{\sigma}^2_v$	-	-	0.6549	(0.5452, 0.800)
DIC	10322.600		9939.580	

The estimated average regression coefficients of the linear and quadratic time effects are 3.125 and -0.2954 for the usual mixed effects model and, 2.94 and -0.2617 for the LMEM incorporating subject-specific variances, both of which are significantly different from zero.

In this table, the estimated subject-specific variance is $\hat{\sigma}_v^2 = 0.6549$ with 90% credible interval (0.5452, 0.800). Hence, it supports the assumption of heterogeneous variance for the repeated CD4 measurements. Also, the reduction in the DIC for the model incorporating subject-specific variances is an evident that subject-specific CD4 variances must be considered in the analysis. Hence, we use the LMEM that incorporate subject-specific variances for our joint model estimation. The estimated average regression coefficients of linear Time, baseline Age, baseline Weight and baseline Ois are 2.94, -0.5414, 0.7162 and -0.6802 respectively, which are significantly different from zero. These estimates shows that, on average the longitudinal CD4 measurement significantly increases with an increase in Time and Weight, but decrease with an increase of Age and Ois.

3.2. Results of Weibull Model

The survival data is analyzed with both Weibull and Exponential models using Win BUGS in which results are

presented in Table 2. Because none of the covariates is time varying, the regression equation for the log-relative hazard in the absence of random effects is:

$$\log(\mu_i) = \beta_{21} + \beta_{22}Age_i + \beta_{23}Weight_i + \beta_{24}Func_i + \beta_{25}Tobac_i + \beta_{26}Cond_i$$

This is the parameterizations used in Win BUGS. From Table 2 below, it is easy to observe that the parameter estimates of both the Weibull and the Exponential models differ significantly. The estimated Weibull shape parameter $\hat{\rho}$ is 2.98 with 90% CI (2.833, 3.133) which is significantly greater than one indicating that death rates increase over time.

Table 2. *Posterior Means and* 90% *Credible Intervals for Population Parameters of the Survival Model using both Weibull and Exponential Distributions.*

Parameters	Weibull Model		Exponential Model	
	Posterior Mean	**90% CI**	**Posterior Mean**	**90% CI**
Intercept ($\hat{\beta}_{21}$)	-11.44	(-12.09, -10.81)	-3.66	(-3.825, -3.496)
Age ($\hat{\beta}_{22}$)	0.107	(0.0387, 0.1746)	0.0428	(-0.0232, 0.1088)
Weight ($\hat{\beta}_{23}$)	-0.0904	(-0.157, -0.0222)	-0.03812	(-0.1045, 0.0282)
Functional Status ($\hat{\beta}_{24}$)	-0.2045	(-0.2908, -0.1169)	-0.0767	(-0.1618, 0.0079)
Tobacco addiction ($\hat{\beta}_{25}$)	0.2573	(0.1835, 0.3314)	0.1125	(0.0393, 0.1844)
Condom use ($\hat{\beta}_{26}$)	-0.7383	(-0.8698, -0.6048)	-0.3344	(-0.4664, -0.2048)
$\hat{\rho}$	2.98	(2.833, 3.133)	1.000	-
DIC	3526.500		4003.210	

Finally, the smaller DIC for the Weibull model and the significance of the shape parameter assures that it is better to use the Weibull model than the Exponential model. Thus, subsequent analysis of the survival data are based on a Weibull model. In this model, among the five covariates included in the model, all of them, Baseline Age, Baseline Weight, Functional status, Tobacco addiction and Condom use are statistically significant at 0.1 significance level.

The estimated average regression coefficients of Age, Weight, Functional status, Tobacco addiction and condom use effects are 0.107, -0.0904, -0.2045, 0.2573 and -0.7383, respectively. These estimates show that, an increase in age of the patients increases the hazard of death and an increase in weight of the patients reduces the hazard of death. Since the parameter of the covariate condom use have a negative sign implies the hazard decrease (survival improves). Which indicate condom use has a negative influence for the hazard of patients but positive influence on survival of patients. Those patients that use condom have fewer hazards but, better survival than patients that do not use condom.

3.3. Joint Model Selection

We have used LMEM that incorporate subject specific variance under longitudinal sub-model and Parametric Weibul model under survival sub-model, and then we explore several joint models with a variety of latent processes. In all cases, the results are based on three parallel MCMC sampling chains of 75,000 iterations each, following a 25,000 iteration "burn-in"' period. As mentioned above, we have chosen the precise nature of the two sub models; the longitudinal to be LMEM with subject-specific variances and the survival model to be Weibull. Hence, their association is selected via the DIC. By default, Win BUGS provides the components of DIC for the two sub-models (i.e., the terms in the log-likelihood arising from longitudinal and survival model components) to evaluate their relative contributions to the total DIC score. Table 3 below reports $\overline{D}$, pD and DIC score for 12 joint models with different random effects and different forms of the latent processes $W_{1i}(t)$ and $W_{2i}(t)$, where the LMEM that incorporates patient-specific CD4 variability is used for the longitudinal sub-model and Weibull model used for survival sub-model. The simple joint models I and II with no random effects for longitudinal sub-model is fitted first, which have a large (poor) total DIC. In Model II, we add a frailty term U_3 in $W_2(t)$, but this does not seem to improve the total DIC at all. A similar relationship exists between Models V and VI. Inclusion of frailty term U_3 leads to improved total DIC of model IV. Of the model considered this is the only instance where we found frailty to have non-negligible effect. As such, we do not consider including U_3 in subsequent models. Next, random intercepts are introduced in the longitudinal sub-model. The incorporation of random intercepts in the longitudinal sub-model improves the total DIC.

Models III to VI include random intercepts in $W_{1i}(t)$, which results in a dramatic improvement for the longitudinal sub-model and the total DIC scores. Then, different latent associations through the random intercepts and random variances are introduced. Models VII to XII have both random intercepts and slopes in the longitudinal sub-model, which results in a substantial decrement in total DIC. Because Model VIII emerges with the smallest total DIC among all other models, we select it as our final model for the HIV/AID patients data obtained from Bale Robe general hospital.

Table 3. *Joint Model Selection for a variety of candidate Joint Models when the LMEM that incorporates Patient-Specific Variances is used for the longitudinal sub-model and a Weibull Model is used for the Survival sub-model.*

Model	$W_{1i}(t)$	$W_{2i}(t))$	$\overline{D}$	pD	DIC
No Random Effects	-	-	-	-	
I	0	0	14745.700	205.766	14951.400
II	0	U_3	14740.600	210.037	14950.600
Random Intercepts	-	-	-	-	
III	U_1	0	13245.600	507.191	13752.800
IV	U_1	U_3	13241.900	511.169	13753.100

Model	$W_{1i}(t)$	$W_{2i}(t))$	$\bar{D}$	pD	DIC
V	U_1	$\gamma_1 U_1$	13243.200	507.796	13750.900
VI	U_1	$\gamma_1 U_1 + U_3$	13239.500	512.034	13751.500
Random Intercepts and Slopes	-	-	-	-	
VII	$U_1 + U_2(t)$	0	12827.100	639.225	13466.400
VIII	$U_1 + U_2(t)$	$\gamma_1 U_1$	12821.800	640.544	13462.400
IX	$U_1 + U_2(t)$	$\gamma_2 U_2$	12829.800	637.778	13467.600
X	$U_1 + U_2(t)$	$\gamma(U_1 + U_2)$	12824.300	640.487	13464.800
XI	$U_1 + U_2(t)$	$\gamma_1 U_1 + \gamma_2 U_2$	12823.000	639.544	13462.600
XII	$U_1 + U_2(t)$	$\gamma_1 U_1 + \gamma_2 U_2 + \gamma_3(W_{1i}(t))$	12824.700	638.857	13463.500

In the absence of the latent association, there is nothing to be gain by the joint analysis, unless the longitudinal measurement and the intensity sub model have parameters in common. Since γ_1 is significant in this model, the data set used for this paper support the use of joint model to relate a patient's survival time to the characteristics driving the patient's longitudinal data pattern. This is clinically reasonable, since high CD4 count represents better health status; patients with CD4 counts that are low or more rapid decline would be expected to have poorer survival. As it is evident from the output of the joint model VIII, the use of joint model is apparently justified for these data, as indicated by the significance of the γ_1 parameter (90% posterior credible interval (-0.0560, -0.0069).

3.4. Comparison of Separate and Joint Models

After selecting the final model, the results obtained under the separate (i.e., ignoring any latent association introduced by W_2) and joint models are compared. In all of the cases, the models have smaller total DIC scores when the patient-specific CD4 variability's are incorporated than those models, which do not incorporate patient-specific CD4 variability. Hence, both the separate and final joint models to be compared incorporate patient-specific CD4 variability. Both of these assume the longitudinal model has form (3), while the survival model now takes the form:

$$\log(\mu_i) = \beta_{21} + \beta_{22}Age_i + \beta_{23}Weight_i + \beta_{24}Func_i + \beta_{25}Tobac_i + \beta_{26}Cond_i + \begin{cases} 0, (separate), \\ W_{2i}(t), (joint). \end{cases}$$

The posterior estimates of the regression coefficients β_1, β_2 and their 90% confidence intervals are summarized in Table 4. Here the results in both separate and joint analysis are approximately the same for longitudinal and survival data, which are similar to the finding by [14]. In the longitudinal sub-model all covariates; linear and quadratic time, sex, Age, weight and Opportunistic infection (Ois) are statistically significant at level 0.1 while Age, weight, Functional status, Tobacco addiction and condom use are significant at this level in the survival sub-model.

The association between the longitudinal outcomes and the time-to-event outcome can be explained by parameter γ_1. We observe weak (but significant) negative association between the subject-specific random intercept of the longitudinal CD4 count and the hazard of death. In the study by [14], the posterior estimates of the association parameter in the joint analysis is insignificant, indicating that the CD4 counts is not associated with the hazard of death.

Table 4. *Comparison of Separate and Joint Models of the Longitudinal CD4 Measurements and Time-to-death of HIV/patients.*

Parameters	Separate Model		Joint Model	
	Posterior Mean	90% CI	Posterior Mean	90% CI
Longitudinal sub-model				
Fixed Effects	-	-	-	-
Intercept ($\hat{\beta}_{11}$)	13.99	(13.65, 14.33)	13.99	(13.65, 14.34)
time ($\hat{\beta}_{12}$)	2.941	(2.789, 3.096)	2.936	(2.785, 3.09)
Time^2 ($\hat{\beta}_{13}$)	-0.2617	(-0.2891, -0.2346)	-0.2605	(-0.2877, -0.2337)
Sex ($\hat{\beta}_{14}$)	-1.037	(-1.558, -0.5094)	-1.048	(-1.571, -0.52)
Age ($\hat{\beta}_{15}$)	-0.5428	(-0.7875, -0.295)	-0.5395	(-0.787, -0.2951)
Weight ($\hat{\beta}_{16}$)	0.7172	(0.4646, 0.9727)	0.7172	(0.4605, 0.9707)
Ois ($\hat{\beta}_{17}$)	0.6775	(-0.9233, -0.4347)	-0.6625	(-0.9032, -0.419)
Random Effects				
var ($\hat{U}_1$)	10.0321	(8.9127, 11.4194)	10.005	(8.8731, 11.3843)
var ($\hat{U}_2$)	0.7077	(0.5817, 0.8811)	0.7087	(0.5807, 0.8842)
$\hat{\mu}_v$	2.039	(1.958, 2.118)	2.037	(1.956, 2.116)
$\hat{\sigma}^2_v$	0.6536	(0.5449, 0.7987)	0.6575	(0.5495, 0.8013)
Survival sub-model				
Intercept ($\hat{\beta}_{21}$)	-11.44	(-12.07, -10.81)	-11.48	(-12.12, -10.84)
Age ($\hat{\beta}_{22}$)	0.1057	(0.03752, 0.1744)	0.1095	(0.04019, 0.1778)
Weight ($\hat{\beta}_{23}$)	-0.08998	(-0.1577, -0.02203)	-0.08824	(-0.1576, -0.01988)
Functional Status ($\hat{\beta}_{24}$)	-0.2054	(-0.2932, -0.1188)	-0.2029	(-0.2891, -0.1154)
Tobacco addiction ($\hat{\beta}_{25}$)	0.2568	(0.1843, 0.3288)	0.2617	(0.1879, 0.3358)
Condom use ($\hat{\beta}_{26}$)	-0.7376	(-0.8709, -0.6058)	-0.7346	(-0.8677, -0.6003)
$\hat{\rho}$	2.979	(2.832, 3.129)	2.988	(2.84, 3.141)
$\hat{\gamma}_1$			-0.0314	(-0.0560, -0.0069)
DIC	13466.400		13462.400	

When evaluating the overall performance of both the separate and joint models, the joint model performs better in terms of model goodness of fit. The effective number of parameters of the separate and joint models are 639.225 and 640.544, respectively, while the posterior means of the deviance functions are 12827.100 and 12821.800. As a result, the corresponding $DICs$ for the separate and joint models are 13466.400 and 13462.400. Therefore, the Joint model fits the data better than the separate models.

In general, the Joint model is preferred as it has a smaller total DIC than the separate models. In the study by [14], when we evaluate the overall performance of both the separate and joint models in terms of model goodness of fit, the separate model performs better.

The estimated average regression coefficients of linear Time, baseline Age, baseline Weight and baseline Ois are 2.936, -0.5395, 0.7172 and -0.6625 respectively, which are significantly different from zero. These estimates shows that, on average the longitudinal CD4 measurement significantly increases with an increase in Time and Weight, but decrease with an increase of Age and Ois.

The estimated average regression coefficients of Age, Weight, Functional status, Tobacco addiction and condom use effects are 0.1095, -0.0882, -0.2029, 0.2617 and -0.7346, respectively for the joint survival sub-model. These estimates shows that, an increase in age of the patients increases the hazard of death and an increase in weight of the patients reduces the hazard of death. Since the posterior estimates of the covariate condom use have a negative sign implies the hazard decrease (survival improves). Which indicate condom use have a negative influence for the hazard of patients but positive influence on survival of patients. Those patients that use condom have fewer hazards but, better survival than patients that do not use condom.

The estimated average regression coefficient of random effects due to the square root of CD4 measure is -0.0314, which is negative. This estimate shows that, an increase in random effect decreases hazard of death.

3.5. Assessing Gibbs Sampler Convergence

In the Bayesian method, three parallel MCMC chains are run with different initial values for 75,000 iterations each. Then, we have discarded the first 25,000 iterations as pre-convergence burn-in, thinning of 10 and retained 15,000 for the posterior inference. For checking convergence of the MCMC chains, we have used time series plot of the history of iterations of the final joint model and separate model, which shows a reasonable degree of randomness between iterations and the overlaps of the three chains indicates that the same solutions are obtained for each initial values. Therefore, the Gibbs sampler has converged to the target density.

4. Conclusions

In this study, we have used Bayesian approach to joint modeling of longitudinal CD4 measurements and survival time responses of HIV/AIDS patients under ART follow up at Bale Robe General Hospital.

In the separate analysis of the longitudinal data, the square root transformation of CD4 measurements were used to meet the normality assumption. The data were analyzed using the LMEM incorporating patient specific variability. The patient specific variability was significant which supported the assumption of heterogeneous variances. The predictors: linear and quadratic observation time, sex, baseline age, weight and baseline number of opportunistic infection were found statistically significant at 0.1 level of significance. All the covariates included in the survival sub-model: Baseline Age, baseline weight, functional status, tobacco addiction and condom use were found to be significantly associated with time to death at 0.1 level of significance.

On average, the longitudinal CD4 measurement significantly increases with an increase in Time and Weight, but decrease with an increase of Age and Ois. On the other hand, an increase in age of the patients increases the hazard of death and an increase in weight of the patients reduces the hazard of death. The parameter of the covariate condom use have a negative sign implying the hazard decrease (survival improves). This indicates condom use have a negative influence for the hazard of patients but positive influence on survival of patients. Those patients that use condom have fewer hazards but, better survival than patients that do not use condom.

Separate analysis of the longitudinal CD4 measurements proves that incorporation of patient specific CD4 variances brings an improvement in the model fit. Specifically, the assumption of heterogeneous CD4 variances among patients resulted in a reduction in the posterior mean of the deviance function of the model. The results of both the separate and joint analysis were approximately the same. However, joint model has a smaller posterior mean of the deviance function, which indicates that it fits the data better than the separate models. After the separate analyses of each data, joint models with a variety of latent processes were investigated. First, a simple joint model with no random effects in both sub-models was fitted and then several models with different random effects and various latent associations of the two sub-models were investigated. The Bayesian joint model is found to best fit to the data, and provided more precise estimates of parameters. The shared frailty is significant showing the association between the LME and survival models.

References

[1] Wu L. Mixed effects models for complex data. CRC Press; 2009 Nov 11.

[2] Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. The American Statistician. 2004 Feb 1; 58 (1): 16-24.

[3] Diggle P. Analysis of longitudinal data. Oxford University Press; 2002 Jun 20.

[4] Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. Springer Science Business Media; 2009 May 12.

[5] Viviani S. Mixed effect joint models for longitudinal responses with drop-out: estimation and sensitivity issues.

[6] Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000 Dec 1; 1 (4): 465-80.

[7] Lyles RH, Munõz A, Xu J, Taylor JM, Chmiel JS. Adjusting for measurement error to assess health effects of variability in biomarkers. Statistics in medicine. 1999 May 15; 18 (9): 1069-86.

[8] Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982 Dec 1: 963-74.

[9] Huang X, Li G, Elashoff RM, Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. Lifetime data analysis. 2011 Jan 1; 17 (1): 80-100.

[10] Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in medicine. 1996 Aug 15; 15 (15): 1663-85.

[11] Ibrahim JG, Chen MH, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. Statistica Sinica. 2004 Jul 1: 863-83.

[12] Law NJ, Taylor JM, Sandler H. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. Biostatistics. 2002 Dec 1; 3 (4): 547-63.

[13] Wang Y, Taylor JM. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. Journal of the American Statistical Association. 2001 Sep 1; 96 (455): 895-905.

[14] Buta GB, Goshu AT, Worku HM. Bayesian Joint Modelling of Disease Progression Marker and Time to Death Event of HIV/AIDS Patients under ART Follow-up. British Journal of Medicine and Medical Research. 2015 Jan 1; 5 (8): 1034.

[15] Erango MA, Goshu AT, Buta GB, Dessiso AH. Bayesian Joint Modelling of Survival of HIV/AIDS Patients Using Accelerated Failure Time Data and Longitudinal CD4 Cell Counts.

[16] Dobson AJ, Barnett A. An introduction to generalized linear models. CRC press; 2008 May 12.

[17] Spiegelhalter D, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Quality control and applied statistics. 2003; 48 (4): 431-2.

[18] Manatunga A, Schmotzer B, Lyles RH, Small C, Guo Y, Marcus M. Statistical issues related to modeling menstrual length. InProceedings of the American Statistical Association, Section on Statistics in Epidemiology 2005.