British Journal of Medicine & Medical Research
5(8): 1034-1043, 2015, Article no.BJMMR.2015.114
ISSN: 2231-0614

SCIENCEDOMAIN *international*
www.sciencedomain.org

Bayesian Joint Modelling of Disease Progression Marker and Time to Death Event of HIV/AIDS Patients under ART Follow-up

Gemeda Bedaso Buta[1], Ayele Taye Goshu[*1] and Hailemichael M. Worku[2]

[1] *School of Mathematical and Statistical Sciences, Hawassa University, Ethiopia.*
[2] *Department of Methods and Statistics, Leiden University, The Netherlands.*

Article Information
DOI: 10.9734/BJMMR/2015/12907
Editor(s):
(1) Chan Shen, Department of Biostatistics, MD Anderson Cancer Center, University of Texas, USA.
Reviewers:
(1) Fasakin Kolawole, Department of Haematology, Federal Medical Centre, Ido Ekiti, Nigeria.
(2) Nlida Virginia Gmez, Medicine Department, University of Buenos Aires, Argentina.
(3) Mohammad Kamali, Rehabilitation Research Centre, Iran University of Medical Sciences, Iran.
(4) Shalini Malhotra, Microbiology, Delhi University, India.
Complete Peer review History:
http://www.sciencedomain.org/review-history.php?iid=715id=12aid=6672

Original Research Article

Received: 22 July 2014
Accepted: 03 September 2014
Published: 24 October 2014

Abstract

Objectives: To develop separate and joint statistical models in the Bayesian framework for longitudinal measurements and time to death event data of HIV/AIDS patients.
Study design: Longitudinal study.
Place and Duration of Study: The population of study includes all HIV/AIDS patients who had been under follow up of Antiretroviral Therapy (ART) from January 2006 to December 2012 at Shashemene Referral Hospital in Ethiopia.
Methodology: The posterior model was analyzed using Gibbs sampler by sampling from the distributions of the parameters given the data. Convergence of each sample was maintained.

**Corresponding author: E-mail: ayele_taye@yahoo.com*

Results: The results indicated that the joint model was not significant indicating that the CD4 count did not have significant effect on the patient's survival time. The results of both the separate and joint analyses were consistent. The separate model was better interims of goodness of fitness than the joint model, while the final joint model was found to be simpler (less complex) model than the separate models. In the longitudinal sub-model, the predictors: linear time, squared time, sex, and tobacco addiction were statistically significant at 0.05 level of significance. For the survival sub-model, knowledge of ART and condom use were significantly related with time to death.
Conclusion: The Bayesian Joint model provides results consistent with that of the separate models.

Keywords: ART; Bayesian; CD4 count; Joint model; Longitudinal model; Survival model.
2010 Mathematics Subject Classification: 53C25; 83C05; 57N16

1 Introduction

Often longitudinal and survival data are generate together with related covariates over periods of time. A typical example is a clinical trials from which CD4 lymphocyte counts as a biomarker of disease progression is measured intermittently and time to death event of a patient is recorded under Antiretroviral Therapy (ART) follow-up at a health institution. In such cases, there are often possible early dropouts or failures to occurrence of the event by the end of study period [1]. Data analysis can mainly focus on either the longitudinal data or the survival data or both. When the analysis focuses on longitudinal data, we often need to address informative dropouts since dropouts are very common in such studies. When the analysis focuses on survival data, we often need to incorporate time-dependent covariates such as CD4 counts since the times to event may be associated with the covariate trajectories. It is interesting to investigate the association between the two processes. The joint models are can handle to the association between the longitudinal and the survival data [2]. The joint models can accommodate complexities in data observed simultaneously.

In this study, we employ the joint modeling approach developed by [3]. We applied the Bayesian joint and separate modelling 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 death? 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 analyze and model disease progression as measured by biomarker and time to death of HIV/AIDS patients based on data from Hospital records. The results of such study is useful in the developing an effective strategy for ART and monitoring system.

2 Methods

The population of our study includes all HIV/AIDS patients with ages 16 years and older and under ART follow-up from January 2006 to December 2012 at Shashemene Referral Hospital, Ethiopia. Secondary data were collected from 354 HIV/AIDS patients' records at the Hospital, after determining appropriate sample size and applying simple random sampling technique. Patients with age 16 years and older were included, while those patients with age below 16 years or those who started ART before January 2006 or after December 2012 were excluded. The ethical clearance was obtained from the Hospital.

Response variables: Two response variables are considered in this study. The first response variable is longitudinal CD4 count. Number of CD4 counts per mm^3 of blood were measured approximately every 6 month. Square root transformation of the CD4 cell counts used as observation for the response variable. The second response variable is the survival outcome, which is time in months to a death event for a patient calculated by subtracting date of ART start from date of the event.

Predictor variables: predictors considered for the longitudinal response are observation time, sex of patient, tobacco addiction, functional level, alcohol addiction and number of opportunistic infections; and those for survival response are condom use, number of living room, knowledge, TB status and number of opportunistic infections.

Separate models for the longitudinal and survival data, and the joint model are subsquetly defined here below. The Bayesian joint model is then derived.

2.1 Longitudinal Model

The longitudinal data, CD4 counts, are measurements on the response variable taken from same individuals over several observation times. These set of observations on a subject tends to be intercorrelated [4]. Let us denote CD4 count as Y_{ij} of the i^{th} patient and j^{th} observation time, where $i = 1, 2, 3, , n; j = 1, 2, 3, , a$.

Two sources of variations are expected for the longitudinal data: within-subject and between-subjects variations. Analysis of within-subject variation allows studying of changes over time, while analysis of between-subjects variation allows understanding differences between subjects. We apply the linear mixed effects models. The 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. It is given as:

$$Y_{ij} = \mu_i(s_{ij}) + \mathbf{W}_{1i}(s_{ij}) + \epsilon_{ij} \tag{2.1}$$

where $\epsilon_{ij} \sim N(0, V_i)$, $\log(V_i) \sim N(\mu_v, \sigma_v^2)$. The error term ϵ_{ij} is random and has normal distribution with mean zero and subject-specific variance, V_i which is random by itself. Th variance V_i is assumed to follow log-normal distribution with mean μ_v and variance σ_V^2

2.2 Survival Model

The survival time T is random variable defined on non-negative real numbers. We apply the Weibull model as a parametric distribution model of T. That is, we assume that the survival time for the i^{th} subject follows the Weibull distribution:

$$T_i \sim Weibull(p, \mu_i(t)), \log(\mu_i(t)) = \mathbf{X}_{2i}^T(t)\boldsymbol{\beta}_2 + \mathbf{W}_{2i}(t) \tag{2.2}$$

where with parameters $p > 0$ and $\mu_i(t)$ function of time t, the vectors of predictors $\mathbf{X}_{2i}(t)$ and $\boldsymbol{\beta}_2$ are the corresponding regression coefficients. Note that the predictors in this model can have elements in common with those in the longitudinal model. The form of $\mathbf{W}_{2i}(t)$ is similar to $\mathbf{W}_{1i}(s)$, including subject-specific covariate effects and an intercept representing a frailty. The event intensity or hazard at time t is given as

$$\lambda_i(t) = pt^{p-1}\mu_i(t) = pt^{p-1}exp\left\{\mathbf{X}_{2i}^T(t)\boldsymbol{\beta}_2 + \mathbf{W}_{2i}(t)\right\} \tag{2.3}$$

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

2.3 Joint Model

The joint modeling approach given by [3] is used in the current study. It is assumed that the association between the longitudinal and survival processes arises through a stochastic dependence between $\mathbf{W}_{2i}$ and $\mathbf{W}_{1i}$. Then the joint model constituted of two linked submodels: the longitudinal process measurements model and the survival process model. The joint model links longitudinal model (1) and survival model (2) by taking

$$\mathbf{W}_{1i}(s) = \mathbf{U}_{1i} + \mathbf{U}_{2i} * s \tag{2.4}$$

and

$$\mathbf{W}_{2i}(t) = \gamma_1 \mathbf{U}_{1i} + \gamma_2 \mathbf{U}_{2i} + \gamma_3(\mathbf{U}_{1i} + \mathbf{U}_{2i} s) + \mathbf{U}_{3i} \tag{2.5}$$

where form of the association function, $\mathbf{W}_{2i}(t)$, is similar to $\mathbf{W}_{1i}(s)$, including subject specific covariate effects and an intercept. We adopt the usual joint modeling assumption that the $\mathbf{W}_{2i}(t)$ induce all of the association between the two processes. The longitudinal submodel is the linear mixed effects model that includes subject specific heterogeneous variance with each patient receiving random intercept and linear slope terms. The form in $\mathbf{W}_{1i}(s)$ is linear in time s, which is motivated while exploring the longitudinal data. The parameters, γ_1, γ_2 and γ_3 respectively measure the association between the two submodels induced by the random intercepts, slopes and longitudinal term at the event time $\mathbf{W}_{1i}(t)$. Note that the pair of latent variables $(\mathbf{U}_{1i}, \mathbf{U}_{2i})$ has a mean-zero bivariate Gaussian distribution $\mathbf{N}(0, \Psi)$, while the frailty term $\mathbf{U}_{3i})$ distributed as i.i.d. $N(0, \sigma_3^2)$ independent of $(\mathbf{U}_{1i}, \mathbf{U}_{2i})$. The subject specific variance V_i has a lognormal distribution $\log(V_i) \sim (\mu_v, \sigma_v^2)$ as defined earlier. Regarding the association function, $\mathbf{W}_{2i}(t)$, a variety of several latent processes are considered. The final model or form of $\mathbf{W}_{1i}(s)$ and $\mathbf{W}_{2i}(t)$ with their latent association are selected using Deviance Information Criteria (DIC).

2.4 The Bayesian Joint Model

The Bayesian approach can be thought for the parameter estimation of the joint model. The standard maximum likelihood method involves integrating out latent variables from the log likelihood function which is difficult when the parameters are of high-dimensional [5]. The Bayesian method can overcome such difficult as it can be computed by generating Markov chains with the Gibbs sampler. Bayesian joint models have been studied by several researchers including [6],[7].

The Bayesian model is defined as the product of likelihood function and prior distribution, and hence it can incorporate additional prior information through prior distributions. Here we assume the association between longitudinal response **Y** and survival response **T** are conditionally independent given the random effects $\mathbf{U}_i$'. So the full joint distribution can be specified as:

$$f(\mathbf{Y}, \mathbf{T}, \boldsymbol{\delta} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int f(\mathbf{Y} | \boldsymbol{\theta}_1, \mathbf{U}_i) f(\mathbf{T}, \boldsymbol{\delta} | \mathbf{Y}, \boldsymbol{\theta}_2, \mathbf{U}_i) f(\mathbf{U}_i) d\mathbf{U}_i \tag{2.6}$$

The Likelihood Function

The likelihood function for the full joint distribution of the longitudinal continuous response and time to event variable is given as.

$$L(\mathbf{Y}, \mathbf{T}, \boldsymbol{\delta} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \prod_{i=1}^{n} \int f(\mathbf{Y} | \boldsymbol{\theta}_1, \mathbf{U}_i) f(\mathbf{T}, \boldsymbol{\delta} | \mathbf{Y}, \boldsymbol{\theta}_2, \mathbf{U}_i)^{\delta_i} X (1 - F(\mathbf{T}, \boldsymbol{\delta} | \mathbf{Y}, \boldsymbol{\theta}_2, \mathbf{U}_i))^{1-\delta_i} f(\mathbf{U}_i) d\mathbf{U}_i \tag{2.7}$$

where

- $\mathbf{U}_i = \{\mathbf{U}_{1i}, \mathbf{U}_{2i}, \mathbf{U}_{3i}\}$ represents the shared underlying effects;
- $\boldsymbol{\theta}_1 = \left\{\boldsymbol{\beta}_1, \Psi, \mu_v, \sigma_v^2\right\}$ are the population parameters as defined in the linear mixed effects model;

- $\boldsymbol{\theta}_2 = \left\{\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \sigma_3^2\right\}$ are the population parameters as given in survival model;
- $f(.)$ and $F(.)$ denote density and distribution functions, respectively.

Prior Distribution

Prior specification for the parameters is important in the Bayesian approach. Thus the regression parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are assumed to be random variables and having normal distributions with mean zeros and constant variances. The shape parameter *p* in the Weibull model and the association parameters γ_1, γ_2 and γ_3 in the joint model are assumed to follow gamma distributions. Moreover, the shared effects are assumed to have normal distributions with mean zeros and constant variances.

Posterior Distribution

The Bayesian model is defined as the product of likelihood function and prior distribution including normalizing constant. The joint posterior distribution for all the unknown parameters $\boldsymbol{\theta}$ and random effects **U** is then given by:

$$f(\boldsymbol{\theta}, \mathbf{U}|\mathbf{Y}, \mathbf{T}) = \frac{f(\mathbf{Y}, \mathbf{T}|\boldsymbol{\theta}, \mathbf{U})\pi(\boldsymbol{\theta}, \mathbf{U})}{\int f(\mathbf{Y}, \mathbf{T}|\boldsymbol{\theta}, \mathbf{U})\pi(\boldsymbol{\theta}, \mathbf{U})d(\boldsymbol{\theta}, \mathbf{U})} \tag{2.8}$$

where

- $f(\boldsymbol{\theta}, \mathbf{U}|\mathbf{Y}, \mathbf{T})$ is the required posterior probability distribution;
- $f(\mathbf{Y}, \mathbf{T}|\boldsymbol{\theta}, \mathbf{U})$ is the likelihood function; and
- $\pi(\boldsymbol{\theta}, \mathbf{U})$ is the prior probability density

Then the Bayesian inference is based on samples drawn from the posterior distribution using the Gibbs sampler. Estimate of the posterior distributions including posterior means and variances of the parameters given the data are obtained based on the samples generated. Here the simulation was conducted using the WinBUGS software.

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 counts was 156.58 per mm^3 of blood with standard deviation of 92.535. The results of the analysis showed that from the 354 patients included in the study, about 5.9% of them were dead while 94.1% were censored.

3.1 Results for Analysis of Linear Mixed Effects Model

The results are displayed in Table 1. It shows that among the covariates included in the longitudinal model, observation time, squared observation time, gender, and tobacco addiction were statistically significant at 5% level of significance. However, functional level, alcohol addiction and number of opportunistic infections were insignificant. This is based on whether or not the 95% posterior credible intervals for each estimate includes zeros. The estimates, $\beta_{12} = 2.149$ and $\beta_{13} = -0.118$ indicate that the average CD4 counts of the patients may have the parabolic shape of increasing and then decreasing over time with maximal point. In the Table 1, the estimated mean subject-specific precision is $(1/\hat{\sigma}_v^2) = 2.158$ with 95% credible interval $(1.596, 2.897)$. Hence, it supports the assumption of heterogeneous variance for the repeated CD4 measurements. Use of the linear mixed effect model that incorporate subject-specific variances is justified.

Table 1: Results of LME Model that incorporates Patient-Specific Variances Analysis

Parameters	Posterior Mean	st.dev	MC error	95% CI
Fixed Effects	-	-	-	-
Intercept (β_{11})	13.670	0.3171	0.0043	(13.05, 14.29)
$Obstime$ (β_{12})	2.149	0.0834	0.0007	(1.987, 2.314)
$(Obstime)^2$(β_{13})	-0.118	0.0010	0.0000	(-0.137, -0.099)
Sex (β_{14})	-0.915	0.3862	0.0059	(-1.677, -0.160)
Functional(β_{15})	-0.422	0.3331	0.0047	(-1.073, 0.231)
Alcohol(β_{16})	0.932	0.5627	0.0092	(-2.023, 0.180)
Tobacco(β_{17})	1.119	0.5610	0.0092	(0.026, 2.212)
OIS(β_{18})	0.050	0.0890	0.0014	(-0.124, 0.225)
Random Effects	-	-	-	-
tau1	0.106	0.1061	0.0000	(0.087, 0.128)
tau2	2.381	0.3410	0.0035	(1.793, 3.123)
μ_v	2.033	0.0532	0.0005	(1.927, 2.136)
tau_v	2.158	0.3345	0.0048	(1.596, 2.897)
DIC	12359.700	-	-	-

3.2 Results of Analysis for the Weibull Model

For the Weibull model, all the five variables: number of living room, TB status, condom use, knowledge of ART and number of opportunistic infections were significant under separate model analysis and so all were selected to be included in the survival model. The model used was:

$$\log(\mu(time_i)) = \beta_{21} + \beta_{22}tb_i + \beta_{23}know_i + \beta_{24}cond_i + \beta_{25}ois_i + \beta_{26}rom_i \tag{3.1}$$

The results are displayed in Table 2. It can be seen that TB status, knowledge of ART and condom use are statistically significant at 0.05 level of significance. In Bayesian sense, the 95% posterior credible intervals for coefficients of TB status, knowledge of ART and condom use exclude 0 while that of number of living room and number of opportunistic infections include 0.

Table 2: Results of Analysis for Weibull Model

Parameters	Posterior Mean	st.dev	MC error	95% CI
Intercept(β_{21})	-16.360	0.6982	0.0276	(-17.750,-15.040)
TB status(β_{22})	1.182	0.1431	0.0020	(0.899,1.459)
Knowledge(β_{23})	-0.708	0.0704	0.0009	(-0.845,-0.570)
Condomuse(β_{24})	1.278	0.1322	0.0022	(1.020,1.540)
OIS(β_{25})	-0.023	0.0288	0.0002	(-0.079,0.030)
Living room(β_{26})	0.053	0.0473	0.0004	(-0.148,0.037)
p (Shape parameter)	3.979	0.1625	0.0064	(3.671,4.305)

3.3 Model Selection for the Joint Models

The literature on model selection for joint models is quite limited. In practice, the best longitudinal model can be selected based on the analysis of observed longitudinal data, and the best survival model can be selected based on the analysis of survival data, using standard model selection procedures for these models. Then, we specify reasonable link between the two models such as shared random effects. As mentioned above, we have chosen the precise nature of the two sub models; the longitudinal to be LME model with subject-specific variances and the survival model to be Weibull. Hence, their association is selected via the DIC (Deviance Information Criterion) and a hierarchical modeling generalization of the AIC (Akaike Information Criterion). Taking $\boldsymbol{\theta}*$ and $\mathbf{Y}*$ as the entire collections of model parameters and data, respectively, DIC is computed as:

$$DIC = \bar{D} + pD \tag{3.2}$$

where D is deviance and pD is effective number of parameters. Here, the fit of a model is summarized in the first term by the posterior expectation of the deviance, $\bar{D} = E_{\boldsymbol{\theta}*|\mathbf{Y}*}[D]$ while the complexity of the model is captured in the second term by the effective number of parameters pD as in [8]. The pD is defined

$$pD = E_{\boldsymbol{\theta}|\mathbf{Y}}[D] - D(E_{\boldsymbol{\theta}|\mathbf{Y}}[\bar{\boldsymbol{\theta}}]) = \bar{D} - D([\bar{\boldsymbol{\theta}}]) \tag{3.3}$$

Small value of $\bar{D}$ indicates goodness of fit, while small value of pD indicates a parsimonious model. Hence small values of the sum (DIC) indicates preferred models. Several joint models with different form of latent processes are explored in order to identify the joint model that fit data well. In all cases, the results are based on three parallel MCMC sampling chains of 50,000 iterations each, following a 25,000 iteration burn-in period. By default, WinBUGS provides the components of DIC for the two submodels (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; hence the DIC for the longitudinal and survival sub models are denoted as DIC_1 and DIC_2, respectively.

Table 3 displays $\bar{D}$, pD and DIC scores where the linear mixed effects model that incorporates patient-specific CD4 variability is used for the longitudinal submodel and Weibull model used for survival submodel are joined by taking different forms of the latent processes $\mathbf{W}_{i1}(s)$ and $\mathbf{W}_{2i}(t)$. The simple joint models M_1 and M_2 with no random effects for longitudinal submodel is fitted first, which have a large (poor) total DIC. Next, random intercepts are introduced in the longitudinal submodel. The incorporation of random intercepts in the longitudinal submodel improves DIC_1 and also the total DIC. Models $M3$ to $M6$ include random intercepts in $\mathbf{W}_{i1}(s)$, which results in high improvement in DIC_1 for the longitudinal submodel and the total DIC scores. Then, different latent associations through the random intercepts and random variances are introduced. Models $M7$ to $M12$ have both random intercepts and slopes in the longitudinal submodel which results in a substantial decrement in DIC_1. But, the incorporation of a frailty term, U_{3i}, in $\mathbf{W}_{i2}(t)$ increased the value of DIC_2 in general as compared to models which does not include frailty term. Hence, the inclusion of frailty term does not seem to improve the total DIC at all.

Table 3: Results for Joint Model Selection

Model	$\mathbf{W}_{1i}(s)$	$\mathbf{W}_{i2}(t)$	$\mathbf{DIC}_1$	$\mathbf{DIC}_2$	$\bar{D}_{Total}$	$\mathbf{pD}_{Total}$	$\mathbf{DIC}_{Total}$
M_1	0	0	14171.800	3413.350	17364.900	220.263	17583.200
M_2	0	U_3	14172.100	3429.960	17373.100	228.950	17602.800
M_3	U_1	0	12789.400	3413.370	15717.800	485.029	16202.800
M_4	U_1	U_3	12790.000	3429.970	15726.100	493.896	16220.000
M_5	U_1	$\gamma_1 U_1$	12790.200	3415.330	15719.400	486.141	16205.000
M_6	U_1	$\gamma_1 U_1 + U_3$	12789.700	3431.860	15726.800	494.804	16221.600
M_7	$U_1 + U_2 s$	0	12359.600	3413.350	15139.600	633.365	15773.000
M_8	$U_1 + U_2 s$	$\gamma_1 U_1$	12359.300	3415.250	15141.100	633.503	15774.600
M_9	$U_1 + U_2 s$	$\gamma_2 U_2$	12359.300	3415.290	15141.500	633.078	15774.600
M_{10}	$U_1 + U_2 s$	$\gamma(U_1 + U_2)$	12359.600	3415.310	15141.600	633.340	15774.900
M_{11}	$U_1 + U_2 s$	$\gamma_1 U_1 + \gamma_2 U_2$	12359.600	3417.250	15142.800	634.034	15776.800
M_{12}	$U_1 + U_2 s$	$\gamma_1 U_1 + \gamma_2 U_2 + \gamma_3(\mathbf{W}_{1i}(s))$	12360.900	3417.372	15142.800	635.372	15778.100

Generally, Model $M9$ emerges with the smallest effective number of parameters (less complex or more parsimonious model) among the candidate models. Model $M7$ has the smallest total DIC (fits the data well) among all other models. Since $\mathbf{W}_{2i}(t) = 0$ in model $M7$, the data set used for this paper does not support the use of joint model to relate a patients survival time to the characteristics driving the patients longitudinal data pattern. This is clinically not 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 $M9$, the use of joint model is apparently not justified for these data, as indicated by the increase in the DIC score and the insignificance of the association parameter γ_2 with 95% posterior credible interval (-0.174,0.161) that includes zero.

The posterior estimates of the regression coefficients β_1 and β_2 with their 95% confidence intervals for final joint model $M9$ are summarized in Table 4. Here the results in both separate and joint analyses are the same for longitudinal data. In the longitudinal submodel linear and quadratic time, sex and tobacco addiction are statistically significant, and knowledge of ART and condom use are significant in the survival submodel. However, the posterior estimates of the association parameter γ_2 in the joint analysis is insignificant, indicating that the CD4 counts is not associated with the hazard of death.

Table 4: Analysis of the Final Joint Model: $M9$

Parameters	Posterior Mean	st.dev	MC error	95% CI
Longitudinal Submodel	-	-	-	
Intercept (β_{11})	13.670	0.3136	0.0060	(13.050, 14.280)
$Obstime(\beta_{12})$	2.151	0.0836	0.0009	(1.987, 2.314)
$(Obstime)^2(\beta_{13})$	-0.118	0.0096	0.0001	(-0.137, -0.099)
Sex (β_{14})	-0.932	0.386	0.0080	(-1.679, -0.167)
Functional(β_{15})	-0.417	0.3423	0.0065	(-1.089, 0.250)
Alcohol(β_{16})	-0.959	0.5559	0.0120	(-2.057, 0.137)
Tobacco(β_{17})	1.148	0.5640	0.0122	(0.038, 2.270)
OIS(β_{18})	0.053	0.0873	0.0018	(-0.120, 0.222)
tau1	0.106	0.0100	0.0001	(0.087, 0.128)
tau2	2.398	0.3469	0.0048	(1.802, 3.165)
μ_v	2.033	0.0530	0.0007	(1.929, 2.136)
tau_v	2.143	0.3346	0.0065	(1.588, 2.891)
Survival submodel	-	-	-	-
Intercept(β_{21})	-4.010	0.1456	0.0019	(-4.298,-3.727)
TB status(β_{22})	0.266	0.1398	0.0008	(-0.012,0.535)
Knowledge(β_{23})	-0.257	0.0717	0.0006	(-0.399,-0.117)
Condomuse(β_{24})	0.347	0.1321	0.0014	(0.089,0.607)
OIS(β_{25})	-0.008	0.026	0.0002	(-0.060,0.042)
Living room(β_{26})	-0.016	0.0476	0.0005	(-0.112,0.075)
γ_2 (Assc. parameter)	-0.006	0.0850	0.0004	(-0.174,0.161)
p (shape parameter)	3.995	0.1633	0.0070	(3.692,4.346)

When evaluating the overall performance of both the separate and joint models in terms of model goodness of fit, the separate model performs better. However, the joint model is found to be better in terms of the effective number of parameters. The effective number of parameters of the separate $M7$ and joint model $M9$ are 633.365 and 633.078, respectively, while the posterior means of the

deviance functions are 15139.60 and 15141.50. The corresponding $DICs$ for the separate and joint models are 15773 and 15774.600, respectively. Hence, the posterior mean of the deviance function of the separate model is smaller, which results in smaller total DIC score, than that of the joint model. Therefore, the separate model fits the data better than joint model $M9$.

Regarding to the submodels, the $DICs$ of the longitudinal submodel in the separate and joint models are 12359.60 and 12359.30, respectively, which is some what lager in the separate model. The respective $DICs$ of the survival submodel in the separate and joint models are 3413.350 and 3415.290, the survival submodel has smaller DIC value. In general, the separate model is preferred as it has a smaller total DIC than the joint model. The statistical insignificance of the association parameter γ_1 is also another evidence that the separate model is better than the joint model.

Assessing Chain Convergence

In all of the joint models, three parallel MCMC sampling chains, 50000 iteration each and 25000 burn-in, with different starting values are used. One of the initial values is obtained from the separate analysis, the other is by randomly selecting from the corresponding prior distributions and the third one is set to be null for all parameters.

Time series plot of the history of iterations of the final joint model and separate model shows a reasonable degree of randomness between iterations and also the overlaps of the three chains indicates that the same solutions are obtained for each initial values. Therefore, the Gibbs sampler has been converged to the target density. Moreover, MC error can be checked. Since the values of MC errors are very low in comparison to its posterior summaries especially its standard error, thus the posterior density has converged to target density.

4 Conclusions

The objective of this study was to investigate the Bayesian joint model of the longitudinal CD4 measurements and time-to-death event of HIV/AIDS patients. The patients had been under ART follow-up at ShaShemene Referral Hospital, Ethiopia. The method includes shared random effects which induce association between the two models by incorporating subject specific variances which possesses some attractive features on modeling longitudinal response. Both separate and joint analysis were conducted.

In the separate analysis of the longitudinal data, the square root transformation CD4 measurements were used to meet the normality assumption. The data were analyzed using the linear mixed effects model incorporating patient specific variability. The patient specific variability was significant which supported the assumption of heterogeneous variances. The predictors: observation time, squared time, sex and tobacco addiction were statistically significant. Out of the covariates included in the survival submodel of the joint model, knowledge of ART and condom use were found to be significantly associated with time to death at 0.05 level of significance.

The Bayesian analyses of the joint models with a variety of latent processes were investigated. First, a simple joint model with no random effects in both submodels is fitted and then other 11 models with different random effects and various latent associations of the two submodels were investigated. The results showed that the separate models were found to be statistically significant, while the Bayesian joint model was not for the data considered in this study. This indicates that the CD4 count has no significant effect on the patient's survival time. It can be concluded that the statistical results obtained from the separate analyses are consistent with those obtained from the joint model and so the joint model is still important for predictions.

Competing Interests

The authors declare that no competing interests exist.

References

[1] Yao F. Functional principal component analysis for longitudinal and survival data. Statistica Sinica. 2007;11:965-983.

[2] 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.

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

[4] Viviani S. Mixed effect joint models for longitudinal responses with dropout: estimation and sensitivity issues. 2012.

[5] Xin H, Gang L, Robert M, Jianxin PA. General joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. Probability and Statistics Group; 2009.

[6] Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in Medicine. 1996;15:1663-1685.

[7] Huang Y, Dagne G, Wu L. Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. Statistics in Medicine. 2011;30:2930-2946.

[8] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit: with discussion. J. R. Stat. Soc. Ser. 2002;64:583-616.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=715&id=12&aid=6672