

Joint Modeling of Longitudinal CD4 Count and Weight Measurements of HIV/Tuberculosis Co-infected Patients at Jimma University Specialized Hospital

Aboma Temesgen¹ · Teshome Kebede¹

Received: 2 May 2016 / Revised: 21 June 2016 / Accepted: 9 August 2016 /

Published online: 20 August 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract As HIV/TB co-infected patients are started to be visited, it is common to measure weight and CD4 repeatedly overtime to determine the health status of patients. Most of the time linear mixed modeling of weight and CD4 count cannot handle the association between the outcomes whereas the joint modeling of multivariate linear mixed model does. Thus, this study was an attempt to model jointly the longitudinal CD4 and weight measurements of HIV/TB co-infected patients. This retrospective study consists of 254 HIV/TB co-infected patients who were 18 years old and above, and on ART followup from 1st February 2009 to 1st July 2014 at Jimma University Specialized Hospital. Firstly, weight and square root of CD4 count were analyzed separately. Based on the separate model, the joint models were built to know the correlation between mean change of weight and CD4 count overtime. Finally, appropriate model was selected based on AIC and BIC values. The fit statistics showed that the joint model fitted the data better than the separate model. From the joint model sex, educational level and functional status were the factors contributing to the prediction of HIV/TB co-infected patients weight at baseline. Beside the linear time effect has a positive effect on the mean change of weight whereas the quadratic time change has negative effect. The baseline CD4 count was differ by patient status and functional status. Further, the linear time effect has a positive sign and found to be statistically significant at 5 % level of significance on the mean change of the square root of CD4 count. Nevertheless, the quadratic time effect has a significant negative effect. The

✉ Aboma Temesgen
aboma.temesgen@gmail.com

✉ Teshome Kebede
tashe.zgreat@gmail.com

¹ Department of Statistics, Haramaya University, Dire Dawa, Ethiopia

finding of the current study revealed that there was a moderate positive association between the mean change of weight and square root of CD4 count overtime.

Keywords Co-infection · Longitudinal analysis · Linear mixed model · Multivariate linear mixed model

1 Background of the Study

Tuberculosis and HIV have been closely linked since the emergence of AIDS and TB is the most common infectious disease affecting HIV-sero positive individuals and causing to their death [1,2]. Globally, the number of TB patients who had been diagnosed with HIV status reached 2.1 million in 2010 which is equivalent to 34 % of notified cases of TB. Of the 8.8 million incident cases worldwide an estimated 1.1 million (13 %) were found to be co-infected with HIV [3].

Today, HIV and TB treatments are common in many societies and the use of drugs has altered the joint dynamics of both diseases. About one third of 39.5 million HIV infected people worldwide were co-infected with TB [4] and up to 50 % of individuals living with HIV are expected to develop TB. Many TB carriers who were infected with HIV are 30–50 times more likely to develop active TB than those without HIV. For individuals infected with HIV the presence of other infections including TB tends to increase the rate of HIV replication. This acceleration may result in higher levels of infection and rapid HIV progression to the AIDS stage [5].

Over the past 20 years, HIV has fuelled TB notification rates which have increased 3–5 fold in many African countries. By 2007, the continent accounted for 79 % of the global burden of HIV associated TB [6]. Worst affected were those countries in the east and south of the continent where HIV prevalence rates are highest. In South Africa and Swaziland approximately 1 % of the population develops TB annually. Notification rates in some poor communities in South Africa have even increased to over 2 % per year rates that are almost unprecedented in the era of short course multi-drug chemotherapy [7,8].

WHO ranked Ethiopia as seventh among the 22 high burden countries with TB where the estimated annual incidence and prevalence, respectively, were 379 and 643 cases per 100,000 populations [9]. The prevalence of HIV among TB patients was estimated as high as 41 % [10,11]. In Ethiopia routine data obtained from 44 locations in the year 2005/2006 showed that about 41 % of TB patients were HIV+. In addition, another routine data set collected in 2006/2007 estimated that some 31 % of TB patients were HIV+ [12]. TB was the cause of 76,000 deaths in Ethiopia out of which 30 % were among HIV+ patients [6]. Beside the rate of mortality caused by the co-infection, the extent of the negative impact on the quality of livelihood resulting from mental disorders was studied [13].

In many medical cases more than one clinical outcome are measured longitudinally at the same time on the same subject where these measured clinical outcomes are correlated. For example, SBP and DBP for hypertensive patients and CD4 count, beta-2-macroglobulin and weight for HIV infected patients were measured longitudinally at same time. Since they are highly related changes in either often affect changes

in the other. In such cases the univariate longitudinal analysis does not take into account correlation between observations on different response variables at each time points. Beside this knowing how the evolution of one is related to the evolution of the other, as well as how the association changes or evolves overtime is difficult from univariate longitudinal analysis. Joint modeling of longitudinal data in other way round accounts two types of correlations which are known to be serial correlation and cross correlation. The serial correlation accounts correlation between observations at different time points within a subject and the cross correlation accounts correlation between different responses at each time point. If different types of outcomes are measured at each time point, the correlation structure is more complicated and hence, more difficult for drawing inference [14, 15].

Therefore, the primary objective of present study was to model jointly weight which act as a marker for TB patients and CD4 count which act as a biomarker for HIV infected individuals. Furthermore, the study deals with linear mixed modeling of weight and CD4 count measurements to know the factors that affect the mean change of each outcome variable overtime.

2 Materials and Methods

2.1 Data Source

The HIV/TB co-infection data used for this study were obtained from Jimma University Specialized Hospital HIV and TB Outpatient Clinics, South West of Ethiopia. The study population consists of all HIV/TB co-infected patients who were 18 years old and above, and who were on ART treatment any time between 1st February 2009 and 1st July 2014. Of the total 850 co-infected patients from both clinics at the hospital, 254 patients who have at least one CD4 count and weight at the same time were considered for the study. All the patients' epidemiological, laboratory and clinical information were obtained from the patients chart of ART followup retrospectively from patients card.

2.2 Ethical Consideration

Ethical clearance was obtained from Department of Statistics of Jimma University. Personal information was kept confidentially without disclosing to others during data collection from patient cards.

2.3 Variables of the Study

The two outcome variables considered for this study were the longitudinal CD4 count measurement and weight of co-infected patients which is measured in kg. The CD4 cell counts per mm^3 of blood, which is considered as a biomarker was measured approximately within 6 months interval whereas the weight of the patients were mea-

Table 1 List of covariates used in the separate and joint modeling

Variable name	Values of the variable	Type
Age	Years (baseline)	Continuous
Marital status	Single, married, separated, windowed, divorced	Categorical
Residence	Rural, urban	Categorical
Educational level	Not educated, primary, secondary, tertiary	Categorical
Working time	Partimer, working full time, unemployed	Categorical
Use of alcohol	Use, do not use	Categorical
Smoking	Smoker, nonsmoker	Categorical
Use of soft drug	Use, do not use	Categorical
Type of TB	Pulmonary TB, extra pulmonary TB	Categorical
Clinical stage	Stage-I, stage-II, stage-III, stage-IV	Categorical
Functional status	Working, ambulatory, bed ridden	Categorical
Religion	Muslim, orthodox, protestant	Categorical
Sex	Female, male	Categorical
Status of patients	Active, missed, died, transferred	Categorical

sured at each patients visit. The 14 independent covariates considered for the separate and joint modeling are listed in Table 1 with respective of their categories.

2.4 Model Specification

Longitudinal responses may arise in two common situations. One is when the measurements taken on the same subject at different times and the other is when the measurements taken on related subjects. In both cases, the responses are likely to be correlated [14]. For longitudinal data, two sources of variations were considered which were the within-subject which arises during the measurements within each subject, and between subject variation which arises during the measurement between different subjects. Modeling within subject variations help us to study changes overtime while modeling between subject variation help us to understand differences between subjects.

2.4.1 Univariate Linear Mixed Effects Model

Before going to the joint modeling of longitudinally measured weight and CD4 count of HIV/TB co-infected patients, the linear mixed model were employed for each outcome to identify an appropriate covariates that predicts the mean change of weight and CD4 count overtime. Therefore, the linear mixed-effects model handling the two of variation is

$$\mathbf{y}_j = \mathbf{X}'_j \boldsymbol{\beta}_j + \mathbf{Z}'_j \mathbf{b}_j + \varepsilon_j \quad (1)$$

where \mathbf{y}_j is the $n_j \times 1$ vector of observed response values, $\boldsymbol{\beta}_j$ is the $p \times 1$ vector of fixed-effects parameters, \mathbf{X}_j is the $n_j \times p$ observed design matrix corresponding

to the fixed-effects, \mathbf{b}_j is the $q \times 1$ vector of random-effects parameters, \mathbf{Z}_j is the $n_j \times q$ observed design matrix corresponding to the random-effects, and ε_j is the $n_j \times 1$ vector of residuals for the j th response. The corresponding assumptions for model (1) are $\mathbf{b}_j \sim N_{q_j}(\mathbf{0}, \Psi_j)$ and $\varepsilon_j \sim N_{n_j}(\mathbf{0}, \Sigma_j)$; where Ψ_j and Σ_j are the variance-covariance matrix for \mathbf{b}_j and ε_j for each outcome variable respectively.

2.4.2 Methods of Parameter Estimation

Suppose that a random sample of n_j for j th response observations is obtained from a univariate normal linear mixed effects model as defined in equation (1). Then, the likelihood of the model parameters given the vector of n_j observations is defined as

$$L = L(\beta_j, \gamma_j; \mathbf{y}_j) = \frac{\exp \left\{ -\frac{1}{2}(\mathbf{y}_j - \mathbf{X}'_j\beta_j)' \mathbf{V}^{-1}(\mathbf{y}_j - \mathbf{X}'_j\beta_j) \right\}}{(2\pi)^{\frac{n}{2}} |\mathbf{V}|^{\frac{1}{2}}} \tag{2}$$

where β_j is a vector of fixed-effect parameters and γ_j is a vector containing the variance parameters for the j th response. Hence, the log-likelihood of the model parameters is defined as

$$\begin{aligned} l &= l(\beta_j, \gamma_j; \mathbf{y}_j) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2}(\mathbf{y}_j - \mathbf{X}'_j\beta_j)' \mathbf{V}^{-1}(\mathbf{y}_j - \mathbf{X}'_j\beta_j) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \left[\log |\mathbf{V}| + (\mathbf{y}_j - \mathbf{X}'_j\beta_j)' \mathbf{V}^{-1}(\mathbf{y}_j - \mathbf{X}'_j\beta_j) \right] \end{aligned} \tag{3}$$

The required parameters of the model were obtained by maximizing the log-likelihood function with respect to β_j and γ_j . However, the maximum likelihood approach may produce variance parameters that are biased downwards since they are based on the assumption that the fixed-effects parameters are known [16]. To handle this problem the present study used restricted maximum likelihood (REML) method, which treat fixed-effects as parameters rather than constant.

2.4.3 Multivariate Linear Mixed Model

The mixed model can be easily extended to include multiple response variables by further stacking the data and defining a specific variance-covariance structure for the random effects. In our case, consider modeling weight and CD4 count measurements of co-infected patients jointly overtime by incorporating random intercepts and slopes in order to model the correlations overtime between the two responses. As we defined earlier, the primary aim of this study was to model jointly the longitudinally measured CD4 count and weight of HIV/TB co-infected patients which was modeled by multivariate linear mixed model.

Let y_{ijk} represent the i th observation from the j th subject for the k th response variable, where $i = 1, 2, \dots, n_{jk}; j = 1, 2, \dots, s; k = 1, 2$. Also define $N_k = \sum_{i=1}^s n_{jk}$ and $N = N_1 + N_2$. The vector $\mathbf{y}_{jk} = [y_{1jk}, y_{2jk}, \dots, y_{n_{jk}jk}]'$ represents the n_{jk} observations of the k th response variable from the j th subject and the vector

$\mathbf{y}_k = [\mathbf{y}_{1k}, \mathbf{y}_{2k}, \dots, \mathbf{y}_{sk}]'$ represents the N_k observations for the k th response variable across all response variables and subjects. Finally, the vector $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k]'$ represents the N observations across all response variables and subjects. In the context of modeling our response variables, the linear mixed-effects models for each response variable for subject j taken at time t can be specified as

$$\begin{aligned} y_{j1}(t) &= \mu_1(t) + a_{j1} + b_{j1}(t) + \varepsilon_{j1}(t) \\ y_{j2}(t) &= \mu_2(t) + a_{j2} + b_{j2}(t) + \varepsilon_{j2}(t) \end{aligned} \tag{4}$$

where $\mu_k(t)$ refers to the average evolution (of the k th response over time) and is a function of the fixed effects. The subject specific random intercepts a_{jk} and slopes $b_{jk}(t)$ describe how the subject specific profiles deviate from the average profile for the k th response. The two response trajectories are joined together by assuming a joint distribution for the vector of random-effects, \mathbf{b}_j , such as

$$\mathbf{b}_j = \begin{bmatrix} a_{j1} \\ b_{j1} \\ a_{j2} \\ b_{j2} \end{bmatrix} \sim N(\mathbf{0}, \Psi)$$

where the variance-covariance matrix for the random effects, Ψ , has the following structure:

$$\Psi = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 b_1} & \sigma_{a_1 a_2} & \sigma_{a_1 b_2} \\ \sigma_{b_1 a_1} & \sigma_{b_1}^2 & \sigma_{b_1 a_2} & \sigma_{b_1 b_2} \\ \sigma_{a_2 a_1} & \sigma_{a_2 b_1} & \sigma_{a_2}^2 & \sigma_{a_2 b_2} \\ \sigma_{b_2 a_1} & \sigma_{b_2 b_1} & \sigma_{b_2 a_2} & \sigma_{b_2}^2 \end{bmatrix} \tag{5}$$

The error components for each response, which are independent of the random effects can be taken to be correlated or uncorrelated ($\sigma_{12} = 0$), such that the error components are defined as

$$\begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}\right)$$

2.4.4 Special Cases for the Random Effect Variance Matrix

We obtain special cases for the variance-covariance matrix of the random effects by making specific assumptions for the variance-covariance matrix Ψ . The first assumption is when the two outcome variable could be taken to be completely independent at any point in time, we impose their variance-covariance matrix which has the following special form given by

$$\Psi = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 b_1} & 0 & 0 \\ \sigma_{b_1 a_1} & \sigma_{b_1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{a_2}^2 & \sigma_{a_2 b_2} \\ 0 & 0 & \sigma_{b_2 a_2} & \sigma_{b_2}^2 \end{bmatrix}$$

Within a response variable, the random intercept and slope induce within-subject correlations in the repeated measures overtime, while assuming independence between subjects. Moreover, this model assumes that the two responses are completely independent. The results for the model would be identical to fitting two separate random-effect models.

The second assumption is when the two response variables could be taken to be completely dependent. In this case, the two responses essentially “share” the same set of random effect parameters (intercept and slope). When two parameters are completely dependent, the correlation between them is equal to one. This occurs when the covariance between the parameters is equal to the square root of the product of their respective variances. Most notation, however, define the model with a 2×1 vector of random effects, such as

$$\mathbf{b} = \begin{bmatrix} a_j \\ b_j \end{bmatrix} \sim N(\mathbf{0}, \Psi), \text{ with } \Psi = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ba} & \sigma_b^2 \end{bmatrix}$$

Clearly, the aforementioned structure imposes strong assumptions on the relationship between the two response variables. It is very unlikely that the two responses would exhibit complete dependence in the association between the random slopes and between the random intercepts. One advantage of this model, when the assumption is tenable, is that it drastically reduces the number of random effects that must be estimated when the number of response variables is large. For models with a large number of response variables, estimation would likely be impossible if the shared-parameters (or alternative approach) were not used.

2.4.5 Association of the Evolutions (AOE)

One important question that may be addressed with a joint mixed-effects model is how the evolution of one response is associated with the evolution of another response (“association of the evolutions”). By definition, the correlation between the evolutions for the two random slopes is given by

$$r_e = \frac{Cov(b_1, b_2)}{\sqrt{Var(b_1)}\sqrt{Var(b_2)}} = \frac{\sigma_{b_1 b_2}}{\sqrt{\sigma_{b_1}^2}\sqrt{\sigma_{b_2}^2}} \quad (6)$$

It may be noted that the above expression is produced using those values from the Ψ matrix defined in (5).

2.4.6 Evolution of the Association (EOA)

A similar idea that may be investigated using a joint mixed effects model is how the association between the responses evolves overtime (“evolution of the association”). Assuming uncorrelated errors, the marginal correlation between the two responses as a function of time is given by

$$\begin{aligned}
 r_m(t) &= \frac{\text{Cov}(y_{j1}(t), y_{j2}(t))}{\sqrt{\text{Var}(y_{j1}(t))}\sqrt{\text{Var}(y_{j2}(t))}} \\
 &= \frac{\sigma_{a_1a_2} + t\sigma_{a_1b_2} + t\sigma_{a_2b_1} + t^2\sigma_{b_1b_2}}{\sqrt{\sigma_{a_1}^2 + 2t^2\sigma_{a_1b_1} + 2t^2\sigma_{b_1}^2 + \sigma_1^2}\sqrt{\sigma_{a_2}^2 + 2t^2\sigma_{a_2b_2} + 2t^2\sigma_{b_2}^2 + \sigma_2^2}}
 \end{aligned} \tag{7}$$

Assuming correlated errors, the marginal correlation between the two responses as a function of time is given by

$$\begin{aligned}
 r_m(t) &= \frac{\text{Cov}(y_{j1}(t), y_{j2}(t))}{\sqrt{\text{Var}(y_{j1}(t))}\sqrt{\text{Var}(y_{j2}(t))}} \\
 &= \frac{\sigma_{a_1a_2} + t\sigma_{a_1b_2} + t\sigma_{a_2b_1} + t^2\sigma_{b_1b_2} + \sigma_{12}}{\sqrt{\sigma_{a_1}^2 + 2t^2\sigma_{a_1b_1} + 2t^2\sigma_{b_1}^2 + \sigma_1^2}\sqrt{\sigma_{a_2}^2 + 2t^2\sigma_{a_2b_2} + 2t^2\sigma_{b_2}^2 + \sigma_2^2}}
 \end{aligned} \tag{8}$$

The delta method could be used to obtain 95 % confidence bounds for $r_m(t)$ at any particular point in time. Two observations can be made from the uncorrelated errors by noticing $t = 0$ the marginal correlation reduces to

$$r_m(t) = \frac{\sigma_{a_1a_2}}{\sqrt{\sigma_{a_1}^2 + \sigma_1^2}\sqrt{\sigma_{a_2}^2 + \sigma_2^2}} \tag{9}$$

which is essentially the correlation between the two random intercepts. In fact, when the error components are small, the closer the marginal correlation at $t = 0$ approximates the correlation between the random intercepts. Also, as t increases $r_m(t)$ converges to r_e for the case with uncorrelated errors, and to

$$r_m(t) = \frac{\sigma_{a_1a_2} + \sigma_{12}}{\sqrt{\sigma_{a_1}^2 + \sigma_1^2}\sqrt{\sigma_{a_2}^2 + \sigma_2^2}} \tag{10}$$

for the case of correlated errors, which indicates that the absolute value of the marginal correlation at $t = 0$ cannot be higher than the correlation between the random intercepts. It may also be noted that as t increases the marginal correlation converges to the correlation between the random slopes, while the variance-covariance parameters of the random effects determine the shape of the marginal correlation function [17].

2.4.7 Joint Model Estimation Techniques

In the particular context of random-effects models, so-called adaptive quadrature rules can be used [18], where the numerical integration is centered on the estimates of the random effects, and the number of quadrature points is then selected in terms of the desired accuracy. To illustrate the main ideas, we consider Gaussian and adaptive Gaussian quadrature, designed for the approximation of integrals of the form

$$\int f(z)\phi(z)dz \tag{11}$$

for a known function $f(z)$ and for $\phi(z)$ the density of the multivariate standard normal distribution. Therefore, first standardize the random effects such that they get the identity covariance matrix. Then, the likelihood contribution for subject i equals

$$f_i(\mathbf{y}_i|\beta, \gamma, \Psi) = \prod_{j=1}^{n_i} \int f_{ij}(y_{ij}|\mathbf{b}_i, \beta, \gamma) f(\mathbf{b}_i|\Psi) d\mathbf{b}_i \quad (12)$$

where \mathbf{b}_i is $q \times 1$ dimensional vector of unknown random effects and $\mathbf{b}_i \sim N(\mathbf{0}, \Psi)$, β is a vector of fixed-effects parameters and γ is a vector containing the variance parameters. $f(z)$ and for $\phi(z)$ denotes the density of the multivariate standard normal distribution.

2.4.8 Model Selection Criteria

Several model selection procedures exists but none of which were the best. To have an appropriate model for the univariate LMM and multivariate LMM most commonly known model selection criterions; Akaike Information Criterion (AIC) [19] and the Bayesian Information Criterion (BIC) [14] were considered for this study.

$$AIC = -2 \log L + 2p \quad (13)$$

$$BIC = -2 \log L + npar \log(N) \quad (14)$$

where $-2 \log L$ is twice the negative log-likelihood value for the model, p is the number of estimated parameters, $npar$ denotes the total number of parameters in the model and N is the total number of observations used to fit the model. Smaller values of AIC and BIC reflect an overall better fit.

3 Results and Discussion

3.1 Descriptive Statistics

Some demographic information and baseline characteristics of all patients disaggregated by patient status were presented in Table 2. Regarding the sex composition of patients, out of total of 254 co-infected patients 139 (54.72 %) of them were males and 18 (56.25 %) death were also occurred in male group in comparison with female group. More than half 146 (57.48 %) of the co-infected patients belongs to orthodox religious group whereas 17 (6.69 %) belongs to protestant religious group. Of the total deaths occurred in these categories small number 2 (6.25 %) of deaths were occurred in protestant religious group. When we look at the educational level category of the co-infected patients, 108 (42.52 %) attended primary education while only 17 (6.69 %) attended tertiary education. Table 2 also depicts that 8 (3.15 %), 23 (9.06 %), 123 (48.43 %) and 100 (39.37 %) co-infected patients' were at baseline clinical stage-I, stage-II, stage-III and stage-IV respectively. Totally, 40 (93.75 %) deaths were occurred in both clinical stage-III and stage-IV at baseline time in comparison with remaining two baseline clinical stages. About 13 (40.63 %) and 10 (31.25 %) death were occurred in ambulatory and

Table 2 Demographic and clinical characteristics of the study patients

Variable	Active	%	Died	%	Missed	%	Transferred	%	Total	%
Sex										
Female	61	48.41	14	43.75	24	42.86	16	40.00	115	45.28
Male	65	51.59	18	56.25	32	57.14	24	60.00	139	54.72
Religion										
Muslim	43	34.13	15	46.88	17	30.36	16	40.00	91	35.83
Orthodox	75	59.52	15	46.88	37	66.07	19	47.50	146	57.48
Protestant	8	6.35	2	6.25	2	3.57	5	12.50	17	6.69
Educational level										
Not educated	26	20.63	8	25.00	13	23.21	12	30.00	59	23.23
Primary	56	44.44	14	43.75	20	35.71	18	45.00	108	42.52
Secondary	37	29.37	8	25.00	19	33.93	6	15.00	70	27.56
Tertiary	7	5.56	2	6.25	4	7.14	4	10.00	17	6.69
Residence										
Rural	21	16.67	4	12.50	7	12.50	6	15.00	38	14.96
Urban	105	83.33	28	87.50	49	87.50	34	85.00	216	85.04
Marital status										
Divorced	6	4.76	2	6.25	7	12.50	6	15.00	21	8.27
Married	65	51.59	12	37.50	18	32.14	16	40.00	111	43.70
Separated	8	6.35	9	28.13	10	17.86	1	2.50	28	11.02
Single	33	26.19	8	25.00	19	33.93	13	32.50	73	28.74
Widowed	14	11.11	1	3.13	2	3.57	4	10.00	21	8.27
Clinical stage										
Stage-I	4	3.17	0	0	1	1.79	3	7.14	8	3.15
Stage-II	16	12.70	2	6.25	2	3.57	3	7.14	23	9.06
Stage-III	61	48.41	14	43.75	27	48.21	21	50.00	123	48.43
Stage-IV	45	35.71	16	50.00	26	46.43	13	30.95	100	39.37
Functional status										
Ambulatory	53	42.06	13	40.63	37	66.07	23	57.50	126	49.61
Bedridden	5	3.97	9	28.13	7	12.50	5	12.50	26	10.24
Working	68	53.97	10	31.25	12	21.43	12	30.00	102	40.16
Alcohol										
No	91	72.22	15	46.88	28	50.00	22	55.00	156	61.42
Yes	35	27.78	17	53.13	28	50.00	18	45.00	98	38.58
Smoke										
No	108	85.71	17	53.13	38	67.86	29	72.50	192	75.59
Yes	18	14.29	15	46.88	18	32.14	11	27.50	62	24.41
Type of TB										
Extra pulmonary	60	47.62	17	53.13	25	44.64	20	50.00	122	48.03
Pulmonary	66	52.38	15	46.88	31	55.36	20	50.00	132	51.97

Table 3 Summary measures of covaraites at each time points with respective sample sizes

Measurement time	Sample points	Square root of CD4		Weight	
		Mean	SD	Mean	SD
0	254	11.965	5.430	48.407	10.608
6	156	16.102	5.309	52.994	10.818
12	134	18.094	4.852	54.795	9.6920
18	105	19.419	5.004	55.771	10.066
24	65	19.386	5.148	56.854	11.300
30	43	19.842	5.499	57.465	11.023
36	23	18.884	5.295	56.609	8.6380
42	9	16.777	3.653	58.556	12.136
48	3	14.824	1.181	64.333	13.614

working categories of patients functional status respectively. To this end, 37 (66.07 %) and 12 (21.43 %) were missed to followup from these two groups respectively in comparison with bedridden group at baseline during the time of co-infection periods.

Mean with the corresponding standard deviation at each time points with respective sample sizes for both outcome variables was shown in Table 3. The average number of square root of CD4 count was 11.965 at baseline. There was a general increment in the mean value up to 30 months starting from baseline and starts declining. However, when we look at the standard deviations up to 36 months there was slight variation among the measurement times where smaller variation in square root of CD4 measurement was observed at 48 month. Also it can be observed from Table 3, smaller mean weight of co-infected patients were observed at the baseline of co-infection period and larger mean weight was observed at 48 month with larger weight variation among the co-infected patients.

3.2 Exploring Individual Profile and the Mean Structure

To underpin the model building and visualize the pattern of weight and CD4 measurements of the patients overtime, the overall individual plots were considered.

The graph (Fig. 1) demonstrates the variability (within and between patients) in square root of CD4 count and weight measurements of HIV/TB co-infected patients. The red line loess smoothing technique suggests that the mean structure growth of both variables was in a linear and quadratic pattern overtime. Nevertheless, the rate of increment was high in square root of CD4 count than weight.

3.3 The Separate Longitudinal Analysis

The separate longitudinal analysis were started from the fixed-effect modeling to select an appropriate covariates that predicts weight and square root of CD4 count where this linear model only considers the between source of variation among co-infected patients. First, we have used stepwise automatic variable selection method to select appropriate fixed-effects from Table 1. After arriving at an optimal subset, the selected

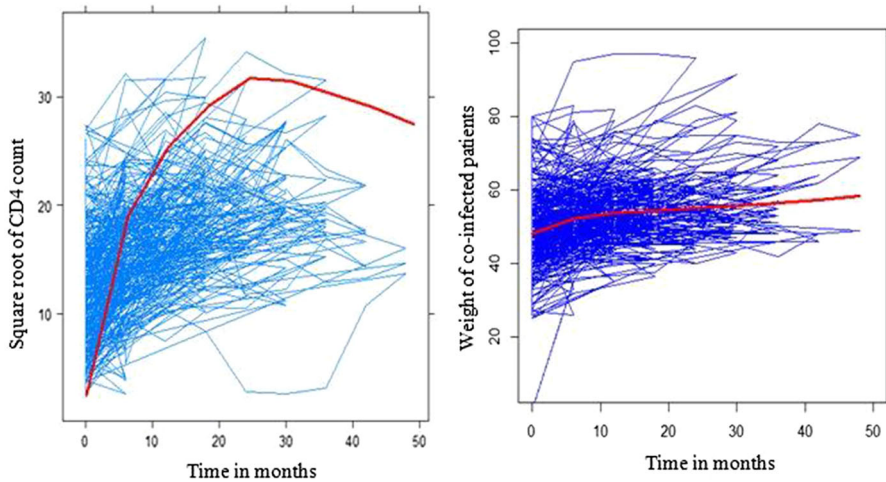


Fig. 1 Individual profile plot and the evolution of mean structure overtime for both covariates

Table 4 Selection of random effects to be included in the linear mixed-effects model

No.	Random effects	Square root of CD4		Weight	
		AIC	BIC	AIC	BIC
1	Only intercept	5195.162	5274.305	4533.621	4594.209
2	Intercept and linear time slope	5137.415	5225.869	4523.083	4592.991
3	Only linear time slope	5630.983	5710.125	4724.709	4785.297
4	Intercept and quadratic time slope	5175.464	5263.917	4536.473	4606.382
5	Only quadratic time slope	5711.422	5790.564	4794.546	4855.133
6	Linear and quadratic time slope	5583.500	5671.953	4670.797	4740.706
7	Intercept linear and quadratic time slopes	5093.714	5196.134	4486.242	4570.133

fixed-effects model was fitted with different random effects for both responses to have an appropriate linear mixed-effects models. The summary of the linear mixed-effects models which were modeled by considering different random effects were shown in Table 4.

From Table 4, for both responses only considering the quadratic time effect to the model worsen the LMM. However, considering the linear time effects to the quadratic time effect improves the LMM with only quadratic time effects since its inclusion lowers the AIC and BIC values. Finally, when we look at the improvement of the model with inclusion of random intercept to that of random linear and quadratic time effects of the LMM, there was an improvement of the model and this model have lower AIC and BIC values than the remaining six LMMs. Hence, we consider the LMM with random intercept linear and quadratic time effects as an appropriate model for the separate models.

Table 5 Separate models for each covariates

Fixed effects	Weight		Square root of CD4		
	Estimate	95 % CI	Fixed effects	Estimate	95 % CI
Intercept	35.999	[30.94, 41.057] [†]	Intercept	11.844	[10.744, 12.943]
Sex			Alcohol		
Female (ref)			Non-user (ref)		
Male	6.330	[4.040, 8.620] [†]	Users	-0.622	[-1.764, 0.520]
Age	0.244	[0.111, 0.376] [†]	Time	0.560	[0.479, 0.642] [†]
Time	0.656	[0.530, 0.783] [†]	Time ²	-0.010	[-0.011, -0.008] [†]
Time ²	-0.011	[-0.01, -0.007] [†]			
Functional status					
Ambulatory (ref)			Bedridden	-3.347	[-5.447, -1.248] [†]
Bedridden	1.543	[-2.496, 5.582]	Working	2.236	[0.982, 3.490]
Working	3.334	[0.870, 5.799]			
Patient status					
Active (ref)			Died	-0.741	[-2.579, 1.098]
Died	-3.604	[-7.256, 0.047]	Missed	-0.448	[-1.907, 1.012]
Missed	-3.530	[-6.362, -0.697] [†]	Transferred	1.461	[-0.170, 3.093]
Transferred	-1.170	[-4.344, 2.004]			
Educational level					
Not educated (ref)					
Primary	-0.653	[-3.539, 2.233]			
Secondary	4.671	[1.555, 7.788] [†]			
Tertiary	3.906	[-1.042, 8.853]			
Time × functional status					
Time × bedridden	-0.233	[-0.480, 0.013]	Time × bedridden	0.042	[-0.134, 0.217]
Time × working	-0.138	[-0.226, -0.051] [†]	Time × working	-0.046	[-0.105, 0.013]
Random-effects	Estimate	95 % CI	Random-effects	Estimate	95 % CIs
$\hat{\sigma}_{b_0}$	8.901	[8.046, 9.847]	$\hat{\sigma}_{b_0}$	4.206	[3.715, 4.761]
$\hat{\sigma}_{b_1}$	0.633	[0.525, 0.765]	$\hat{\sigma}_{b_1}$	0.364	[0.290, 0.457]
$\hat{\sigma}_{b_2}$	0.015	[0.012, 0.020]	$\hat{\sigma}_{b_2}$	0.007	[0.005, 0.010]
$\hat{\sigma}_{b_{10}b_{11}}$	-0.350	[-0.525, -0.147]	$\hat{\sigma}_{b_{10}b_{11}}$	-0.301	[-0.464, -0.118]
$\hat{\sigma}_{b_{10}b_{12}}$	0.292	[-0.009, 0.545]	$\hat{\sigma}_{b_{10}b_{12}}$	0.334	[0.090, 0.540]
$\hat{\sigma}_{b_{10}b_{22}}$	-0.996	[-1.00, -0.784]	$\hat{\sigma}_{b_{10}b_{22}}$	-0.983	[-0.995, -0.944]
$\hat{\sigma}_{\varepsilon}$	3.174	[2.931, 3.436]	$\hat{\sigma}_{\varepsilon}$	2.682	[2.491, 2.887]

[†] Indicates the significance of covariates at 5 % level of significance

As we can observe from the fitted LMM for weight of co-infected patients (Table 5), there was significance difference weight at baseline by sex, patients status and educational level since the estimated coefficient for male, missed the follow up and secondary

Table 6 Joint model selection criterion

Type of model	Random effect considered	AIC	BIC
Joint	Intercept only	9717.900	9739.100
Joint	Intercept and linear slope	9543.100	9676.000
Joint	Intercept linear and quadratic slopes	9673.900	9719.900
Univariate (separate)	Intercept and linear slope	13,989.00	13,999.70

education categories were significant ($p < 0.05$). But there was no significance difference weight at baseline by the functional status category of co-infected patients. When we look at the mean change in weight overtime, the linear time have positive significant effects whereas the quadratic time have negative significant effects. And also, the mean change overtime in weight of co-infected patients were differ by the functional status of the patients since the interaction of linear time with working functional status have significant effects on weight at 5% level of significance. When we look at the fitted LMM for square root of CD4 count, there was a significance difference in square root of CD4 measurement at baseline by the functional status of the patients since the estimated coefficient of bedridden category was significant at 5% level of significance. Likewise the case of weight, the linear time have positive effect on mean change of square root of CD4 count overtime whereas the quadratic time change have negative effect at 5% level of significance.

3.4 Joint Analysis of Square Root of CD4 Count and Weight

To have an appropriate joint model that represent the longitudinally measured CD4 count and weight of the co-infected patients, different candidate joint model with different random effect for the joint modeling were considered. The AIC and BIC were used as a guideline in selecting covariates for the model. A smaller AIC and BIC values were generally indicates a better model. The results of the fitted models information criterion were displayed in Table 6.

Table 6 elucidates that the joint model with random effects was considered as the worst model compared to the joint model with inclusion of random linear and quadratic slopes since it have larger AIC and BIC values. Moreover, inclusion of random linear slope to the random intercept for the joint modeling improves the joint model with random intercept only. However, when we include quadratic slope time effects for the joint modeling to that of with random intercept and linear slope, the joint model becomes worst since it has larger AIC and BIC. Therefore, we consider the joint model with random intercept and linear slope as an appropriate joint model since it have smaller AIC and BIC values than the remaining models. Table 6 also shows that the value of fit statistics (AIC and BIC) for all joint models were less than that of the separate model. This indicates that joint model was fitting better the data as compared to the univariate model. Therefore, our joint analysis of these two outcomes was justified and the separate analysis results were likely to be biased. The result

Table 7 Joint analysis of longitudinal square root of CD4 cell count and weight

Fixed effects	Weight		Square root of CD4		
	Estimate	95 % CI	Fixed effects	Estimate	95 % CI
Intercept	57.644	[52.331, 62.957]	Intercept	14.968	[13.208, 16.728]
Transferred (ref)					
Active	1.503	[−1.659, 4.665]	Active	−1.134	[−2.747, 0.479]
Died	−1.852	[−6.128, 2.424]	Died	−2.083	[−4.226, 0.059]
Missed	−2.122	[−5.746, 1.502]	Missed	−2.012	[−3.843, −0.180]
Sex			Alcohol		
Male (ref)			User(ref)		
Female	−7.197	[−9.452, −4.941] [†]	Non-users	0.530	[−0.602, 1.662]
Educational level					
Tertiary (ref)					
Not educated	−4.230	[−9.204, 0.744]			
Primary	−5.406	[−10.108, −0.704] [†]			
Secondary	−0.260	[−5.127, 4.608]			
Time	0.423	[0.319, 0.528] [†]	Time	0.431	[0.357, 0.506]
Time ²	−0.007	[−0.010, −0.005] [†]	Time ²	−0.008	[−0.010, −0.006] [†]
Functional status					
Working (ref)					
Ambulatory	−4.265	[−6.778, −1.752] [†]	Ambulatory	−2.329	[−3.591, −1.068] [†]
Bedridden	−1.838	[−6.070, 2.395]	Bedridden	−5.606	[−7.764, −3.448] [†]
Time × functional status					
Working × time (ref)					
Ambulatory × time	0.198	[0.086, 0.311] [†]	Ambulatory × time	0.110	[0.042, 0.178] [†]
Bedridden × time	−0.046	[−0.338, 0.245]	Bedridden × time	0.112	[−0.073, 0.297]

[†] Indicates the significance of covariates at 5 % level of significance

of selected appropriate joint model for the longitudinally measured CD4 count and weight of HIV/TB co-infected patients was shown in Table 7.

As can be seen from the fitted joint model (Table 7), covariates like sex, educational level and functional status were the factors significantly contributing to the prediction of HIV/TB co-infected patients weight at baseline. Table 7 also suggests that the linear time effect has a positive effect on the mean change of weight whereas the quadratic time change has negative effect overtime at 5 % level of significance. To this end, weight change of the patients were differ by functional status since the estimated coefficient for ambulatory functional status group with linear time interaction have significant effect.

From the fitted joint model, the baseline CD4 cell count was differ by patient status and functional status since the estimated coefficients for missed patient status category in comparison with transferred category, ambulatory and bedridden functional status category in comparison with working functional status category have significant effects at 5% level of significance. With regard to the mean change in square root of CD4 measurements overtime, the linear time effect has a positive sign and found to be statistically significant. Nevertheless, the quadratic time effect has a significant negative effect on the mean change of CD4 count overtime. Further, the mean change in CD4 cell count of the patients differ by functional status category since the estimated coefficient of ambulatory functional status with time interaction in comparison with working functional status category with time interaction have a significant effects on the mean change of CD4 count overtime at 5% level of significance.

The estimated variance-covariance matrix for the final joint model was

$$\hat{\Psi} = \begin{bmatrix} \hat{\sigma}_{a_1}^2 & \hat{\sigma}_{a_1b_1} & \hat{\sigma}_{a_1a_2} & \hat{\sigma}_{a_1b_2} \\ \hat{\sigma}_{b_1a_1} & \hat{\sigma}_{b_1}^2 & \hat{\sigma}_{b_1a_2} & \hat{\sigma}_{b_1b_2} \\ \hat{\sigma}_{a_2a_1} & \hat{\sigma}_{a_2b_1} & \hat{\sigma}_{a_2}^2 & \hat{\sigma}_{a_2b_2} \\ \hat{\sigma}_{b_2a_1} & \hat{\sigma}_{b_2b_1} & \hat{\sigma}_{b_2a_2} & \hat{\sigma}_{b_2}^2 \end{bmatrix} = \begin{bmatrix} 78.9581 & 0.1228 & -224.94 & -0.5296 \\ 0.1228 & 0.0219 & -1.3535 & 0.02124 \\ -224.94 & -1.3535 & 907.20 & -0.1613 \\ -0.5296 & 0.02124 & -0.1613 & 0.0777 \end{bmatrix}$$

and the resulting correlation matrix was

$$\mathbf{R} = \begin{bmatrix} \hat{\rho}_{a_1a_1} & \hat{\rho}_{a_1b_1} & \hat{\rho}_{a_1a_2} & \hat{\rho}_{a_1b_2} \\ \hat{\rho}_{b_1a_1} & \hat{\rho}_{b_1b_1} & \hat{\rho}_{b_1a_2} & \hat{\rho}_{b_1b_2} \\ \hat{\rho}_{a_2a_1} & \hat{\rho}_{a_2b_1} & \hat{\rho}_{a_2a_2} & \hat{\rho}_{a_2b_2} \\ \hat{\rho}_{b_2a_1} & \hat{\rho}_{b_2b_1} & \hat{\rho}_{b_2a_2} & \hat{\rho}_{b_2b_2} \end{bmatrix} = \begin{bmatrix} 1 & 0.093 & -0.841 & -0.214 \\ 0.093 & 1 & -0.303 & 0.514 \\ -0.841 & -0.303 & 1 & -0.019 \\ -0.214 & 0.514 & -0.019 & 1 \end{bmatrix}$$

From the estimated variance-covariance matrix ($\hat{\Psi}$) of the estimated joint model for the random effects, it can be seen that variability was lower for square root of CD4 count ($\hat{\sigma}_{a_1}^2 = 78.95$) than weight ($\hat{\sigma}_{a_2}^2 = 907.20$). Beside, the correlation matrix shows that (\mathbf{R}) the baseline subject specific baseline CD4 measurement and weight were negatively correlated ($\hat{\rho}_{a_1a_2} = -0.841$). On the other hand, there was a moderate positive association between subject specific change in square root of CD4 and weight of the patient overtime ($\hat{\rho}_{b_1b_2} = 0.541$). There was a negative cross correlation between the baseline weight and time slope of square root of CD4 measurement ($\hat{\rho}_{b_1a_2} = -0.303$) overtime. Similarly, there was a negative correlation between subject specific baseline weight and the patient specific change in square root of CD4 overtime ($\hat{\rho}_{b_2a_1} = -0.214$) overtime. It can thus be concluded that there was a moderate positive association between the mean change of weight and square root of CD4 counts overtime.

4 Conclusion

In this study univariate and bivariate linear mixed methods were considered for fitting two continuous response variables measured longitudinally. Based on separate analysis, the evolution of weight was significantly differ with respect to sex (female), baseline age, time, quadratic time, missed patient status category, secondary education category and the interaction of linear time with working functional status. Like wise, the baseline CD4 measurement was differ by patient status and functional status since the estimated coefficients for missed patient status category in comparison with transferred category, ambulatory and bedridden functional status category in comparison with working functional status category have significant effect. The results of this study also demonstrates that the joint modeling of longitudinally CD4 count and weight measurements fits the data better than those obtained from the separate model. To sum up, the joint model suggests that there was a moderate positive association between the mean change of weight and square root of CD4 count overtime.

Acknowledgments We would like to express our profound gratitude to the management of Jimma University Specialized Hospital for allowing us to have access to the pertinent medical registers from which we extracted the data used in this study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. AIDS Control and Prevention (2000) Project of family health internal. The francoisxavier bagnoud center for public health and human rights of the harvard school of public health, UNAIDS. The status and trends of the global hiv/aids pandemic
2. Raviglione MC, Snider DE, Kochi A (2005) Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic. *JAMA* 293:220–226
3. World Health Organization (2012) Global tuberculosis report
4. World Health Organization (2006) Stop TB partnership annual report
5. Sharma SK, Mohan A, Kadiravan T (2005) HIV/TB co-infection: epidemiology, diagnosis and management. *Indian J Medl Res* 121:550–567
6. World Health Organization (2009) HIV/TB facts
7. Lawn SD, Bekker LG, Middelkoop K, Myer L, Wood R (2006) Impact of HIV infection on the epidemiology of TB in a peri-urban community in South Africa: the need for age-specific interventions. *J Clin Infect Dis* 42(7):1040–1047
8. Middelkoop K, Bekker LG, Myer L, Dawson R, Wood R (2008) Rates of TB transmission to children and adolescents in a community with a high prevalence of HIV infection among adults. *J Clin Infect Dis* 47(3):349–355
9. World Health Organization (2008) Global tuberculosis control: surveillance, planning, financing. Geneva
10. Demissie M, Lindtjorn B, Tegbaru B (2000) HIV infection in tuberculosis patients in Addis Ababa. *Ethiop J Health Dev* 14:277–282
11. Yassin MA, Takele L, Gebresenbet S, Girma E, Lera M, Lendebo E, Cuevas LE (2004) HIV and tuberculosis co-infection in the southern region of Ethiopia: a prospective epidemiological study. *Scand J Infect Dis* 36:670–673
12. Federal Ministry of Health of Ethiopia (2008) Tuberculosis, leprosy and TB/HIV prevention and control program manual, 4th edn. Addis Ababa

13. Deribew A, Markos T, Yohannes H, Nebiyu N, Shallo D, Ajeme W, Tefera B, Ludwig A, Robert C (2009) Tuberculosis and HIV co-infection: its impact on quality of life. *Health Qual Life Outcomes* 7:105
14. Laird N, Ware J (1982) Random-effects models for longitudinal data. *Biometrika* 38:963–974
15. Olkin I, Tate R (1961) Multivariate correlation models with mixed discrete and continuous variables. *Ann Math Stat* 32:448–465
16. Brown H, Prescott R (1999) *Applied mixed models in medicine*. John Wiley and Sons Ltd, New York
17. Fieuws S, Verbeke G, Molenberghs G (2007) Random effects model for multivariate repeated measures. *Stat Methods Med Res* 16(5):387–397
18. Pinheiro J, Bates D (2000) *Mixed-effects models in S and S-PLUS*. Springer, New York
19. Sakamoto Y, Ishiguro M, Kitagawa G (1986) *Akaike information criterion statistics*. D. Reidel Publishing Company, Dordrecht