SciencePG
Science Publishing Group

Application of Longitudinal Measured CD4+ Count on HIV-Positive Patients Following Active Antiretroviral Therapy: A Case of Debre Berhan Referral Hospital

Shewayiref Geremew[1, 4, *], Dejen Tesfaw[2], Tibebu Getiye[3, 4]

[1]Department of Statistics, Debre Tabor University, Debre Tabor, Ethiopia

[2]Department of Statistics, Addis Ababa University, Addis Ababa, Ethiopia

[3]Department of Statistics, Ethiopian Civil Service University, Addis Ababa, Ethiopia

[4]Department of Statistics, Debre Berhan University, Debre Berhan, Ethiopia

Email address:
shewayirefg@gmail.com (S. Geremew), dejentm@gmail.com (D. Tesfaw), getiyetibebu@gmail.com (T. Getiye)
[*]Corresponding author

Abstract: The measurement of the CD4+ count is the predictor of evolution to AIDS, in ART. Studying the way of the CD4+ count over time provides an insight to the disease evolution. The main objective of this study was to apply statistical analysis on longitudinally measured CD4+ Cell counts of HIV-positive patients under ART. The study population consists of 647 HIV+ patients who were 16 years old or older and who were under ART follow up from 2012 to 2017 in Debre Berhan Referral Hospital, Debre Berhan, Ethiopia. The data were from the patients' chart. All patients who have initiated to ART and measured their CD4+ cell counts at least two times, including the baseline and those who started the first line ART regimen class was included in the study population. Data were explored using basic descriptive statistics and individual and mean profile plots. The methods of LMM and GLMM were used. The mean profile of CD4+ count revealed that there is an improvement in the duration of treatment in a linear pattern. From the GLMM covariates duration of treatment, sex, BMI, baseline CD4, regimen class, duration by age, duration by baseline CD4 and duration by regimen class significantly determines the change in CD4+ count overtime at 5% level of significance. There is the duration of treatment effect on the current CD4+ count. The study result suggests that HIV+ patients attending in ART improve their CD4+ count.

Keywords: ART, CD4 Count, GLMM, HIV/AIDS, Longitudinal Data Analysis, LMM

1. Introduction

For decades, scientists and humanitarian aid groups are working on pharmaceutical medicine for HIV/AIDS, but still not possible. The possible way, which is better than none- is treating patients through a clinical treatment called Highly Active Anti-Retroviral Therapy (HAART). Since HAART prolongs the lifetime of HIV/AIDS patients [1]. In clinical/medical studies, it is very common to follow a cohort of subjects evolving over a period of time to identify the relationship between one or more independent covariates with the outcome of interest and the risk of developing a disease.

After the outcome of interest is identified, it is better to study its' evolution over time and the relationship between independent covariates and also how the outcome of interest related to the risk of the disease.

There are challenges in identifying the longitudinal features of most infectious diseases, but it is better to identify the biomarker of HIV/AIDS, since there is a measurement of the CD4+ count of a disease level and whether it is high risk of infection or lower risk of infection. Due to the presence of correlated nature of observations in the repeated measurements, linear mixed model take into account these and can identify the variability of subjects within and

between, this study is considered longitudinal analysis.

2. Methods

2.1. Data and Variables

The data used in this study is obtained from Debre Berhan Referral Hospital (DBRH), ART clinic. This study used the data of HIV/AIDS patients who were undergoing Antiretroviral Therapy in the ART clinic of DBRH, Debre Berhan, Ethiopia, during the period July 1, 2012, to July 30, 2015, and were followed up through the ART routine register records up to January 31, 2017; taken from patients chart. The study population was included HIV-positive adults whose age 16 years old and above-initiated ART treatment in the hospital. All patients who have initiated to ART and measured their CD4+ cell counts at least two times, including the baseline and those who started first-line ART regimen class was included in the study population. Patients, whose age below 16 years old and/or those who started ART before July 2012 or after July 2015 were excluded and the CD4+ cell counts per mm³ of blood were taken approximately in every 6 months regardless of their visit to ART clinic. Ethical clearance was obtained from Debre Berhan University.

Response Variable: The response variable used for this study was CD4+ T-cells (CD4+ count) for each individual measured approximately in every six months interval. For the purpose of normalizing the data Square root transformation of the CD4 cell counts of observations were used for the response variable.

Independent Variables: The independent variables are baseline age, baseline CD4+ count, observation time (in months), sex of the patient, marital status of a patient at baseline, WHO clinical stage, regimen class of the patient, body mass index (BMI) at baseline and functional status of the patient.

Table 1. Variables used in the Statistical Analysis of Linear Mixed Model.

No.	Variables	Description	Values' codes (if any)
1	Sex	Sex of the patient	0=Male and 1=Female
2	Age	Age of the patient (years)	
3	Marst	Marital status of a patient	0=Married, 1=Single, 2=Others*
4	BaseCD4	Baseline CD4$^+$ count	
5	CD4	CD4$^+$ count	
6	BMI	Body Mass Index (kg/m^2)	
7	Time	Time CD4$^+$ count recorded (months)	
8	WHOSt	WHO clinical stage	1=Stage I, 2=Stage II, 3=Stage III, 4=Stage IV
9	Regcl	Regimen class of a patient	0=AZT-3TC-EFV, 1=AZT-3TC-NVP, 2=TDF-3TC-NVP, 3=d4t-3TC-NVP, 4=TDF-3TC-EFV
10	FuncSt	Funcltional status of a patient	0=Working, 1=Ambulatory, 2=Bedridden

*Others: Widowed/Divorced

2.2. Linear Mixed Effects Model

The basic important feature of a longitudinal data (i.e., individuals measurements are taken repeatedly through time) model is its ability to study changes over time within subjects and variations over time among groups. This study deals with a longitudinal data in which the CD4+ cell count of patients was measured during the ART, taken at six different time points in months. The six measurements on the same patient are, therefore not independent but correlated and grouped within the patient. The response variable CD4+ cell count in the data is continuous (when the number of CD4+ counts is large, it approximated by continuous) and set of measurements on one patient are correlated. Linear mixed effects model used to model the change in CD4+ cell count over time. Mixed models take into account both the within and between sources of variation. Mixed models are flexible enough to account for the natural heterogeneity in the population and can handle any degree of missing and drop-out in the data.

The general form of the linear mixed model is given by:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad i=1, 2, 3, \ldots, n \quad (1)$$

$$b_i \sim N(0, D)$$

$$\varepsilon_i \sim N(0, \delta^2 I_{ni})$$

Where; the random effects (b_i) and error terms (ε_i) are independent of each other. Y_i is the n_i-dimensional response vector for subject i, $1 \le i \le N$, N is the number of subjects, X_i and Z_i are (n_i x p) and (n_i x q) dimensional matrices of known covariates, respectively; β is a p-dimensional vector containing the fixed effects, b_i is a q-dimensional vector containing the random effects, and ε_i is an n_i-dimensional vector of error components, the part of y that is not explained by the model. In addition, D is a general (q x q) variance-covariance matrix. The fixed effect parameter for each predictor in the model represents the average change in CD4+ cell count for a unit increase in the predictor [2].

For model comparison, techniques like Akaike's information criterion (AIC), Bayesian information criterion (BIC) and Likelihood-ratio test were used. Model estimation techniques maximum likelihood (ML) and restricted maximum likelihood (REML) were used to estimate the covariance parameters. On the next section, statistical analysis for the longitudinally measured ART data was presented.

3. Statistical Data Analysis

In order to answer the objectives of this study, the longitudinal measurements of CD4 cell count were taken repeatedly from each subject. The total number of patients

included in this study was 647. Among the total number of patients 423 (65.38%) were females and 224 (34.62%) were males. CD4+ cell count score were measured for six repeatedly time points (at baseline: before they are under ART, Time 1, Time 2, Time 3, Time 4 and Time 5) (in months), with values ranging from 5 to 2082 (mean =377.07, standard deviation (SD) =275.47), where low count values correspond to a patient is at risk, while high scores correspond to a better health condition of the patient. At baseline the patients CD4+ count with values ranging from 11 to 1764 (mean =289.22, standard deviation (SD) =

232.95).

The loess smooth curve (Figure 1) was suggested that the average profile of square root transformed CD4+ count has a linear relationship over time. It indicates that square root transformed CD4+ counts show a slight increasing pattern, but the rate of increase is low after time point three. And also it indicates that the linear time effects in the model. We might want to consider including a random intercept for each individual patient and a random slope for time in our linear mixed model (LMM) specification.
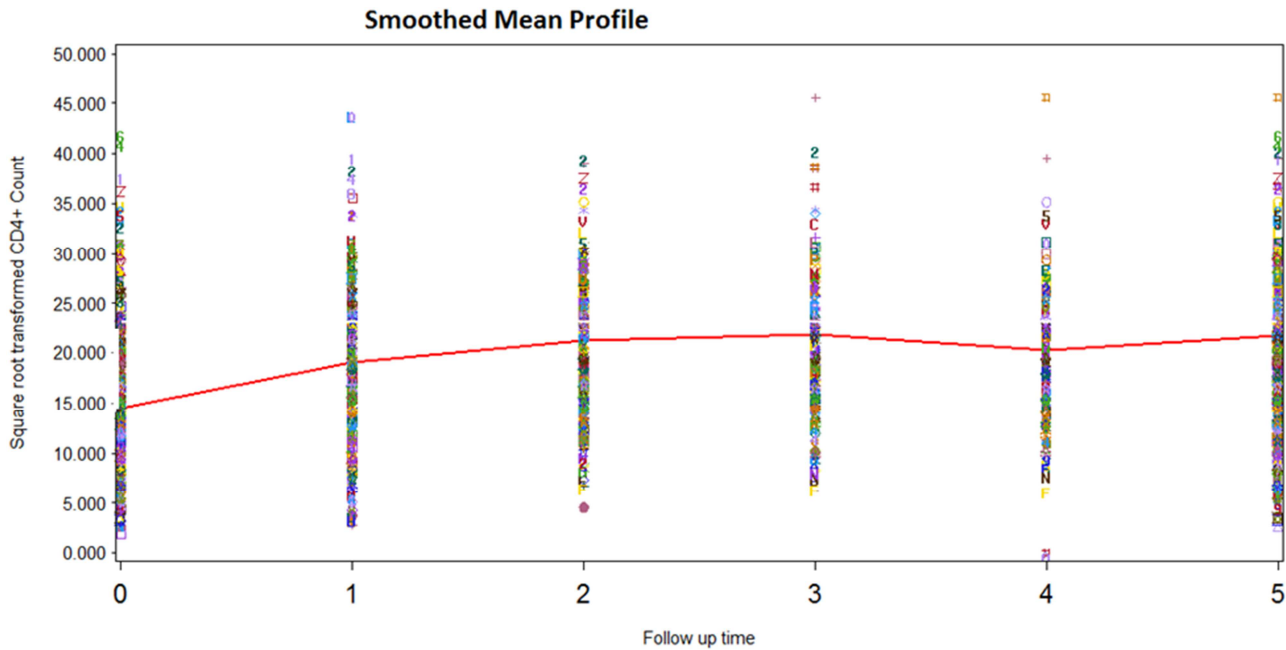

Figure 1. Mean Profile.

The Linear Mixed Effects Model Results

The descriptive statistics of variables of the data under study are shown across time points (*not presented here*). The Shapiro-Wilk test of normality presents the distribution of the CD4+ count of the data is suitable for square root transformed of CD4+ count, as well as the exploratory data analysis with graphical presentation were discussed. The unstructured (UN) covariance structure was selected based on AIC and BIC model selection criterion.

After selecting the most appropriate covariance structure and significant variables by automatic variable selection, we would do in model building is simplifying the mean structure of the model. The primary step, we have taken in the model building is to evaluate the interactions. The one recommended approach is to eliminate the interactions one at a time, starting with the least significant interaction. Based on this, we have started eliminating from least significant interaction effect. We have used the model fit statistics AIC and the estimation technique of maximum likelihood (ML) estimation method. However; after the final model is chosen; refited the model using restricted maximum likelihood (REML). REML estimators are more superior. The second

approach is to compute a likelihood ratio test to compare two models; the full model with all of the interactions and the reduced model with just a subset of terms of the full model. The difference between the -2log likelihoods for the full and reduced models is the value of the test statistic. The likelihood ratio test comparing the full and reduced models is only valid under ML estimation. Using the Unstructured (UN) covariance structure, the full model was fitted with all of the main effects, the time by main effect interactions, which were selected during the univariate analysis and using the ML-estimation method. The model reduction procedure could be based on Likelihood Ratio Tests (LRT) and AIC with corresponding p-values of independent variables.

To interpret the linear mixed model parameter estimates, since our response is square root CD4+ count we have to square for the unit change in the factor. When a unit change in time (in months) since patients initiated ART, we have that the square of the coefficient for time unit increase in CD4+ count.

The reduced fixed effects LMM for square root of CD4+ count is given by:

$$\text{Sqrt}(CD4_{ij}^+) = \beta_0 + \beta_1 t_{ij} + \beta_2 Sex_i + \beta_3 Age_i + \beta_4 BMI_i + \beta_5 FuncSt0_{1i} + \beta_6 FuncSt1_{2i} + \beta_7 MarSt0_{1i} + \beta_8 WHOSt1_{1i} + \beta_9 WHOSt1_{1i} +$$

$$\beta_{10}WHOSt22_i+\beta_{11}WHOSt33_i+\beta_{12}baseCD4_i+\beta_{13}Regcl0_{ij}+\beta_{14}Regcl1_{ij}+\beta_{15}Regcl2_{ij}+\beta_{16}Regcl3_{ij}+(\beta_{17}Age_i+\beta_{18}WHOSt1_{1i}+$$

$$\beta19WHOSt22i+\beta20WHOSt33i)+\beta21baseCD4i+\beta22Regcl00i+\beta23Regcl11i+\beta24Regcl22i+\beta25Regcl33i)\ tij \qquad (2)$$

Where; tij, FuncSt0ti, FuncSt12i, MarSt01i, MarSt12i, WHOSt11i, WHOSt22i, WHOSt33i, Regcl01i, Regcl12i, Regcl23i, Regcl34i are time; functional status: working, ambulatory; marital status: married, single; WHO stages: I, II, III and the first line ART regimen classes: AZT-3TC-EFV, AZT-3TC-NVP, TDF-3TC-NVP and d4t-3TC-NVP; respectively.

***Table 2.** Parameter Estimate for the Reduced Fixed effect Linear Mixed Model.*

Effect	Estimate	Std. error	95% CL	t-value	Pr>|t|
Intercept	608823	1.6997	(3.5509, 10.2137)	4.05	<.0001
Time	1.3857	0.4087	(0.5846, 2.1868)	3.39	0.0007*
Sex (Female)	0.5179	0.4224	(0.3100, 0.8458)	1.23	0.0204*
Sex (Male)	_	_	_	_	_
Age	0.0012	0.0233	(-0.0445, 0.0469)	0.05	0.1602
BMI	0.1232	0.1039	(0.0804, 0.3268)	1.19	0.0063*
Functional status (Working)	-0.2357	0.6493	(-1.5083, 1.0369)	-0.36	0.1167
Functional status (Ambulatory)	-0.4455	1.3152	(-3.0233, 2.1323)	-0.34	0.1349
Functional status (Bedridden)	_	_	_	_	_
Marital Status (Married)	0.4522	0.6146	(-0.7524, 1.6568)	0.74	0.1620
Marital Status (Single)	0.2945	0.6568	(-0.9928, 1.5818)	0.45	0.1539
Marital Status (Others)	_	_	_	_	_
WHO Stage I	1.0232	1.0166	(0.0693, 3.0157)	1.01	0.0144*
WHO Stage II	0.9909	1.0552	(-1.0773, 3.0591)	0.94	0.0579
WHO Stage III	0.4383	1.0231	(-1.5670, 2.4436)	0.43	0.0685
WHO Stage IV	_	_	_	_	_
BaseCD4	0.0241	0.0009	(0.0224, 0.0259)	27.26	<.0001*
Regcl (AZT-3TC-EFV)	0.8435	0.7080	(-0.5442, 2.2312)	1.19	0.2338
Regcl (AZT-3TC-NVP)	0.5709	0.7213	(-0.8428, 1.9846)	0.79	0.4289
Regcl (TDF-3TC-NVP)	-0.0946	0.7301	(-1.5256, 1.3364)	-0.13	0.8969
Regcl (d4t-3TC-NVP)	1.7801	1.6535	(-1.4608, 5.0210)	1.08	0.2820
Regcl (TDF-3TC-EFV)	_	_	_	_	_
Time*Age	-0.0167	0.0055	(-0.0275, -0.0060)	-8.05	0.0023*
Time*WHOStageI	-0.3271	0.2497	(-0.8165,- 0.0623)	-1.31	0.0104*
Time*WHOStageII	-0.2124	0.2594	(0.0208, 0.2960)	-0.82	0.0432*
Time*WHOStageIII	-0.3131	0.2521	(-0.8072, 0.1810)	-1.24	0.0545
Time*BaseCD4	-0.0011	0.0002	(-0.0015, -0.0007)	-5.48	<.0001*
Time* Regcl (AZT-3TC-EFV)	0.0336	0.1660	(-0.2917, 0.3590)	0.20	0.8394
Time* Regcl (AZT-3TC-NVP)	-0.0408	0.1647	(-0.3636, 0.2820)	-0.25	0.8045
Time* Regcl (TDF-3TC-NVP)	0.2880	0.1686	(-0.0425, 0.6185)	1.71	0.0877
Time* Regcl (d4t-3TC-NVP)	0.2957	0.3793	(0.1477, 1.0437)	1.57	0.0165*

*Significance at 5% level of significance

In longitudinal data analysis, which random effect shall be included to the model in order to account between individual variability is a critical and basic issue. The fixed effect model, we have built before is considered with all intercept and linear time effect as random part to identify the individual level variability at baseline and through time progress. In this case, we compared the models by removing each random effect one by one using AIC followed by likelihood ratio test to choose the best random effects that enable to account the between individual variability or to fit the ART data well. Therefore; to fit the random effects model we have to use all the variables that are selected in the uni-variate and fixed effects model. The selected variables in the uni-variate and fixed effect variable selection are time, sex, age, BMI, functional status, marital status, WHO-clinical stage, baseline CD4, regimen class and the interaction terms of age, WHO-clinical stage, baseline CD4 and regimen class with time.

Random effect models different from marginal or fixed effect model; since this includes parameters that are specific to the individual subject. Such parameter estimates interpreted as the residuals which may be helpful for detecting special profiles or groups of individuals evolving differently in time point. Since our interest is in the prediction of subject-specific evolution the estimates for random effects are needed.

Marginal Testing for the Need of Random Effects Model: To select the most appropriate random effect model several hierarchical or subject-specific models for studying the longitudinal evolution is illustrated and compared: (i) No random effects, (ii) Random intercept effects, (iii) Random time effects and (iv) Random intercept and time effects.

By using the mixture-chi-square inference for the variance components the results of (Table 3) is obtained.

Table 3. *Comparison of variance component models with different random effects using mixture chi-square test for the ART data taken at DBRH from 2012-2017.*

Parameter	Model 1 Estimate	Model 2 Estimate	Model 3 Estimate	Model 4 Estimate (s.e)
Var (intercept)	–	8.8136	–	1.31 (0.0041)
Var (time)	–	–	1.0507	0.1716 (0.06416)
Var (residual)	10.5449	10.5444	11.1566	10.9566 (0.4027)
-2REML LL	10845.10	10845.10	11299	10817.8

$X^2_{0:1}$=286.97, p-value=<.0001; $X^2_{0:2}$=353.25, p-value-<.0001; $X^2_{1:2}$=239.44, p-value=<.0001; REML LL (Restricted Maximum Likelihood); Model 1=No random effects, Model 2= Random intercept effects, Model 3= Random time effects, Model 4= Random intercept and time effects.

In this case, the need for random intercept was assessed with a test based on a mixture of chi-squared distributions. The test brings p-value <.0001 indicating that the random intercepts are necessary to be included in the model. And also, check for the importance of the random slope resulted in a p-value <.0001 indicating that the covariance structure should not be simplified by deleting the random slopes from the model. The model with random intercept and the random slope was selected as the final model. So, including the random intercept and slope in the model fits the data well relative to the marginal or population average model. For further illustration, the variance estimate for the intercept tells us how much the intercepts vary across subjects and the variance estimate for time represents how much the slopes for time vary across subjects or variability between subjects. The covariance estimate between the intercept and time shows how the change in the intercepts affects the slopes of time. It indicates whether the CD4+ cell count progression over time is affected by the individual subject's CD4+ cell count**.**

The variance of the random intercepts b_0 was estimated as 1.31, which is small as compared to the within-subject error variability estimated as 10.9566. This implied that the between-subject variability at baseline is smaller. The mean structure of the model remains the same across all models. Here; we use the REML estimation method because the REML test statistic performed slightly better than the ML test statistic. It suggested that REML likelihood could be further increased when we add both random intercepts and random slope. From these when we correct the boundary problem, we have got that a good and simplified covariance structure. The test is only correct when the null hypothesis is not a boundary value; the need of random effects (intercept and time) is more advantageous.

From the random effect analysis, we have got the G/D matrix. Both G and D are the representation of a matrix, which consists of the variances and covariances of the random effects. G or/and D was/were used interchangeably; in some books it was D and in others it was G, especially in statistical software: SAS use G and R use D. In our cases we had used as "G/D". From the G/D matrix the value in column 1 and row 1 represents the variance of the intercepts. The value in column 2 and row 2 represents the variance of the slopes for time. The value in column 2 and row 1 or column 1 row 2 represents the covariance of the intercepts and the slopes of time.

$$D = \begin{bmatrix} 1.3100 & 1.8621 \\ 1.8621 & 0.1716 \end{bmatrix}$$

The information from the G/D matrix above showed that the intercepts and the slopes for the time were positively correlated. The residual covariance estimate represents the error that remains after the fixed effects and random effects were accounted for. This can be represented by the R matrix, which has an independent covariance structure. The estimated V correlation matrix shows the correlations among the measurements for each subject. V matrix can be calculated by the formula (V=ZGZ'+R), where; Z matrix has time values, the correlations estimated from the V matrix are based on the variances and covariances of the random effects along with the time values of the measurements [2].

The parameter estimates for the random effects represent deviations from the fixed effects. Therefore; subject 1 deviates the magnitude of random intercept from the population intercept and the magnitude of the random slope from the population slope for time.

After selecting the appropriate random effects, we had assessed the significance of the fixed effects. From the reduced final GLMM model the linear effect of time is significant. And also sex, BMI, baseline CD4+ cell count and regimen class are the significant main effect terms on the square root CD4+count. The interaction effect of age, baseline CD4+ cell count and regimen class with time are among the significant interaction terms.

From covariance parameter estimates (Table 5), there were two estimated variance components; these were the random effects variances and the residual variance. We have random effects variance bi (i.e., Var (b_{1i})=d_{11}=1.3100, V ar (b_{2i}) = d_{22}=0.1716, cov (b_{1i}, b_{2i}) = d_{12} =d_{21} = 1.8621, and residual variance: V ar (*epsilon*$_{it}$) = *sigma*^2_R= 10.9566). The results showed that the variances and covariances of the random effects are significantly different from 0. The variances of the intercepts and linear effects of time were significantly different from 0. This indicates that the CD4+ cell count values at baseline vary across subjects and the change of CD4+ cell counts over time vary within subjects. The total variability between individuals was estimated as d_{11} + d_{12} + d_{22} = 1.310 + 1.8621 + 0.1716 = 3.3437**,** whereas the total variability within individual was 10.9566. However; the total variation in Square root CD4+count was estimated to be 3.3437+10.9566= 14.3003. The proportion of total variability that is attributed to within-person variation was given by 10.9566/14.3003 was 76.62% while the proportion of total

variability attributed to between individual variations in their general level of square root CD4+count was 3.3437/14.3003 (23.38%). Therefore; more than three quarters of the variation was explained by the residuals. All fixed effects parameters in GLMM have the subject specific interpretation, unlike the marginal model. Thus, given the random effects (b_{1i}); the intercept $(beta_0{}^2 = 6.9556^2 = 48.38)$ in GLMM is an estimate of the "i^{th}" male subject average CD4+count provided that he is bedridden, widowed/divorced, WHO-clinical stage-IV and regimen class (TDF-3TC-EFV) categories. Similarly, Time $(beta_1{}^2 = 1.4912^2 = 2.22)$, implies the mean CD4+count increases 2.22 times per month for the "i^{th}" male individual when the remaining variables kept constant and it is significantly different from zero (p-value <.0001) at 5% significance level. And also, the coefficient for

sex $(beta_2 = 0.4417)$ verifies that the mean CD4+count for "i^{th}" female individual was 0.20 $(beta_2{}^2 = 0.4417^2 = 0.20)$ times higher than male individual with the same random effects (b_i) at baseline and their difference was highly significant (p-value<.0115) at 5%. Other parameters are interpreted in the same way. But, the interaction of time by regimen class (AZT-3TC-EFV) and (AZT-3TC-NVP) and time by WHO stage-II were not significant at 5% significance level. The main effect terms except for time, sex (female), BMI, baseline CD4+ count and Regimen class (AZT-3TC-EFV) all others were not significant at 5% significance level, indicates do not showed significance differentials among groups.

The final model for generalized linear mixed model (GLMM) is given below:

$$Sqrt\ (CD4^+{}_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 Sex_i + \beta_3 Age_i + \beta_4 BMI_i + \beta_5 WHOSt1_i + \beta_6 WHOSt2_i + \beta_7 baseCD4_i + \beta_8 Regcl0_{ij} +$$

$$\beta_9 Regcl2_{ij} + \beta_{10} Regcl3_{ij} + (\beta_{11} Age_i + \beta_{12} WHOSt_{1i} + \beta_{13} WHOSt_{2i} + \beta_{14} baseCD_{4i} + \beta_{15} Regcl_{0i} + \beta_{16} Regcl_{2i} + \beta_{17} Regcl_{3i})\ t_{ij} + b_{0i} + b1_i t_{ij} \quad (3)$$

Where; b_{0i} is the random intercept and b_{1i} is the random slope for the linear effect of time.

Table 4. *Final model for GLMM for the ART data taken at DBRH from 2012-2017.*

Effect	Estimate	Std.error	95% CI	DF	t-value	Pr>|t|
Intercept	6.9556	1.3390	(4.3312, 9.5800)	512	6.13	<.0001
Time	1.4912	0.2991	(0.9050, 2.0774)	515	4.99	<.0001*
Sex (Female)	0.4417	0.2806	(0.1083, 0.8917)	482	1.57	0.0115*
Sex (Male)	_	_	_	_	_	_
Age	0.0018	0.0156	(-0.0288, 0.0324)	482	0.11	0.3099
BMI	0.1383	0.0693	(0.0025, 0.2741)	482	1.99	0.0064*
FuncSt (Working)	-0.3105	0.4296	(-1.1525, 0.5315)	482	-0.72	0.1708
FuncSt (Ambulatory)	-0.3771	0.8954	(-2.1321, 1.3779)	482	-0.42	0.1737
FuncSt (Bedridden)	_	_	_	_	_	_
MarSt (Married)	0.4109	0.4120	(-0.3966, 1.2184)	482	1.00	0.1187
MarSt (Single)	0.2967	0.4403	(-0.5663, 1.1597)	482	0.67	0.1605
MarSt (Others)	_	_	_	_	_	_
WHOStageI	1.0600	0.6707	(-0.2546, 2.3746)	482	1.58	0.1143
WHOStageII	1.0959	0.6963	(-0.2683, 2.4606)	482	1.57	0.1159
WHOStageIII	0.4903	0.6750	(-0.8327, 1.8133)	482	0.73	0.4678
WHOStageIV	_	_	_	_	_	_
BaseCD4	0.0243	0.0006	(0.0231, 0.0255)	482	40.55	<.0001*
Regcl (AZT-3TC-EFV)	0.7463	0.0691	(0.6109, 0.8817)	482	1.59	0.0112*
Regcl (AZT-3TC-NVP)	0.5416	0.4750	(0.3894, 1.4726)	482	1.14	0.0254*
Regcl (TDF-3TC-NVP)	-0.1367	0.4808	(-1.0791, 0.8057)	482	-0.28	0.0776
Regcl (d4t-3TC-NVP)	1.6486	1.0919	(-0.4915, 3.7887)	482	1.51	0.1314
Regcl (TDF-3TC-EFV)	_	_	_	_	_	_
Time*Age	-0.0119	0.0043	(-0.0203,-0.0035)	482	-2.77	0.0058*
Time*WHOStageI	0.2900	0.0663	(0.1601, 0.4199)	482	-1.09	0.0276*
Time*WHOStageII	-0.2032	0.2767	(-0.7455, 0.3391)	482	-0.73	0.4629
Time*WHOStageIII	-0.3008	0.0685	(-0.4351, -0.1665)	482	-1.12	0.0262*
Time*BaseCD4	-0.0010	0.0002	(-0.0014,- 0.0006)	482	-4.71	<.0001*
Time*Regcl (AZT-3TC-EFV)	0.0059	0.1546	(-0.2971, 0.3089)	482	0.04	0.1698
Time*Regcl (AZT-3TC-NVP)	0.0118	0.1722	(-0.3257, 0.3493)	482	0.07	0.2456
Time*Regcl (TDF-3TC-NVP)	0.3145	0.1763	(0.0310, 0.6600)	482	1.78	0.0451*
Time*Regcl (d4t-3TC-NVP)	0.4120	0.3652	(0.0138, 0.7652)	482	1.13	0.0259*

Table 5. *Covariance Parameter estimates for the ART data taken at DBRH from 2012-2017.*

Cov Parm	Subject	Estimate	Standard error	z-Value	P-value
UN (1, 1)	ID	1.3100	0.6034	2.1710	0.0041
UN (2, 1)	ID	1.8621	0.1297	14.3600	<=.0001
UN (2, 2)	ID	0.1716	0.0642	2.6700	0.0037
Residual		10.9566	0.4027	27.2100	<.0001

Remark: * Significance at 5% level of significance, functional status (FuncSt), marital status (MarSt), and regimen class (Regcl).

Heterogeneity in the R matrix was evident in the model. The residual value of 10.9566 corresponds to the variance estimate in the R matrix. The results of the Covariance Parameter Estimates table showed that the variance of the intercepts was a significant, which indicates there was significant variation of the intercepts between subjects at baseline in CD4+count. The variance of the linear effect of time was also significant, which showed that the variances of the residuals in the R matrix for the time were significant. So, variances appear to be different from each other across time.

Generally; the main objective of this study was to estimate the time progress of the change in CD4+ cell count depletion for individual subjects. When we had seen the individual profile plots the observed CD4+ count levels were highly variable over time. One of the reasons might be due to the large residual variability in the error component. Therefore; estimating individual profiles without taking it into account the error associated with residual variability in CD4+ count determinations may lead to unreliable results.

4. Discussions

This study was considered on HIV-positive patients attending ART and the data obtained from Debre Berhan Referral Hospital ART clinic between July 2012 and January 2017. The data were included, 647 HIV-positive patients. Qualitative and quantitative characteristics were used. Linear mixed model and generalized linear mixed models were used for analysis. Linear mixed model for the marginal evolution; the generalized linear mixed model for the subject-specific variation analysis purpose were considered.

In this analysis of the longitudinal data, first, the CD4+ cell count measurements are checked for normality using the Shapiro-Wilk test of normality and Q-Q plots. The plots indicate that there is a deviation from normality and needs some transformation. A square root transformation of the CD4+ cell counts was selected for the normality of the mean response.

The ART data under study were analyzed using different plots (exploratory data analysis) followed by model-based outputs. From the profile plots, we observed the existence of variability in CD4 count within and between individuals. The exploratory analysis result for the mean structure also suggested that on average, CD4+ cell count increases in a linear pattern over time. This supports the results of [3-4], who identified that after patients initiated to ART their CD4+ cell count increases due to the treatment. This means that as the CD4+ cell count increases the progression of a disease decreases since the immune system of a patient develops disease resistance. In addition to this, the mean CD4+ cell count of a patient for females is higher than males up to time 54 months and also it is significant over time. Generally, the exploratory data analysis of the mean structure supported the findings of [3-4], who put as the progression of CD4+ cell count increase at a high rate after patients initiated to ART.

The covariance structure selected for this study is the unstructured (UN) based on the minimum (AIC, BIC, and AICC). The mean response of the longitudinal square root CD4+ cell count is determined and to be linear in time. Then, the data are analyzed using both the usual LMM (marginal models) and the LMM incorporating patient-specific CD4+ cell count variability (subject-specific models)(GLMM). The estimated patient-specific variability is significant which supports the assumption of heterogeneous variances. The LMM that incorporates patient-specific variability (GLMM) have smaller AIC than the model assuming homogeneous variability (LMM). Next, For the purpose of selecting the best random effect that enable to account the variability between individuals in GLMM; with no random effect, random intercept only, random slope only and both random intercept and random slope models were compared using the mixture-chi-square test and found that the random effect term contains both intercept and slope term is selected, and these random effects were included in the model.

From the final model of GLMM, predictors such as duration of treatment, sex, BMI, baseline CD4, regimen class main effect terms and duration by age, duration by baseline CD4, duration by regimen class interaction effects (p-value=<.0001, 0.0446, 0.0370, <.0001, 0.0414, 0.0319, 0.0007, 0.04134; respectively) are among the significant predictors of CD4+ cell count progression at 5% significance level. It supports by [3] baseline CD4, age and time were significant determinants of CD4+ cell count progression but contradicts functional status, which was significant but not in our case. The significance of sex also supported by [4] and a study conducted Tamale Teaching Hospital of Ghana [5], but not supported by [6].

The importance of early treatment was evident from this study. The baseline CD4+ cell count was shown to be significantly determining the patient's disease progression following initiation of ART. A higher baseline CD4+ cell count results in a better recovery of patients on ART. This supports the findings of [6-8].

BMI was shown to significantly determine a patient's current CD4+ cell count, therefore; a higher baseline BMI predicts higher gains in CD4+ cell counts. And also the study did not show any functional status and marital status differentials. This was contradicted by [6]. However; ART regimen class was shown to significantly determine a patient's current CD4+ cell count. It was supported by [5] results.

Duration of treatment was among the significant determinant factors of the current CD4+ cell count, for patients on ART. It means that when the duration of treatment was increased patients on ART show improvement of their CD4+ cell counts; indicates that a better health condition. This result was supported by [3-4].

In the analysis of longitudinal data, the information criteria techniques (AIC and BIC) and the likelihood ratio tests were used for model comparison. The model estimation techniques REML and ML methods were used.

In general, GLMM, the within-subject variation was seen

as the deviation between individual observations. Each subject had an individual subject-specific intercept and slope. Within-subject variations were seen in the magnitude of variation in the deviation between the observations and the individual trajectory. The between-subject variation was represented by the variation among the intercepts, variation (b_{0i}) and the variation among subjects in the slopes i.e., variation (b_{1i}). The resulting estimated ($beta_{1i}$) the fixed-effect parameter for each predictor in this model, represents the average change in CD4+ cell count for a unit increase in that predictor.

5. Conclusion

The main objective of this study was to conduct statistical analysis on longitudinally measured *CD*4+ Cell counts of HIV-Positive patients treated under ART clinic. The result suggest that factors such as duration of treatment, sex, BMI, baseline CD4, regimen class main effect terms and duration by age, duration by baseline CD4, duration by regimen class interaction effects significantly determine the patient's disease progression following initiation of ART.

Based on the study results we concluded that patients CD4+ cell count was increased at different levels after put on ART at a certain initial CD4+ cell count. The determinants of CD4+ cell counts as well as the effect of the factors studied on patients CD4+ cell count was shown in this study. In this study, a generalized linear mixed model (GLMM) for the longitudinally measured CD4+ cell count fluctuates on HIV/AIDS patients under ART follow up were demonstrated. Analysis of the longitudinal CD4+ measurements including the subject-specific variability improves significantly the fit of the model. The longitudinally measured CD4+ cell counts show that variability through time evolution, and concluded that HIV/AIDS patients attending in ART improve the CD4+ cell count of patients.

List of Abbreviations

AIDS: Aquared Immuno Deficiency Syndrome
ART: Antiretroviral Therapy
BMI: Body Mass Index
DBRH: DebreBerhan Referral Hospital
GLMM: Generalized Linear Mixed Model
HAART: Highly Active Anti-Retroviral Therapy
HIV: Human Immuno Deficiency Virus
LMM: Linear Mixed Model
Sqrt: Square Root

Acknowledgements

We would like to greatly acknowledge Debre Berhan Referral Hospital ART Clinic for allowing us to use the ART data.

Competing Interests

The authors declare that no competing interests exist.

References

[1] Ugandan Antiretroviral Treatment and Care Guidelines for Adults, Adolescents, and Children. 2nd Edition, January, 2008.

[2] Verbeke, G. and Molenberghs, G., (2008), Linear Mixed Models for Longitudinal Data. Springer Series in Statistics. New-York: Springer.

[3] Lemma *et al*., (2016), Predictors of CD4 count over time among HIV patients initiated ART in Felege Hiwot Referral Hospital, northwest Ethiopia: multilevel analysis.

[4] Tekle G, Kassahun W, Gurmessa A., (2016), Statistical Analysis of CD4+ Cell Counts progression of HIV- posi tive Patients enrolled in Antiretroviral Therapy at Hossana District Queen Elleni Mohamad Memorial Hospital, South Ethiopia. Biom Biostat Int J 3 (1): 00057.

[5] Yakubu, Toyibu, Vincent K Dedu, and Patrick Owiredu Bampoh., (2016), Factors Affecting CD4 Count Response in HIV Patients within 12 Months of Treatment: A Case Study of Tamale Teaching Hospital. American Journal of Medical and Biological Research 4, no. 4, pp. 78-83.

[6] Kebede MM, Zegeye DT, Zeleke BM, (2014), Predictors of CD4 Count Changes after Initiation of Antiretroviral Treatment in University of Gondar Hospital, Gondar, Ethiopia. Clin Res HIV/AIDS 1 (2): 1008.

[7] Lemma *et al*., (2016), Predictors of CD4 count over time among HIV patients initiated ART in Felege Hiwot Referral Hospital, northwest Ethiopia: multilevel analysis.

[8] Viviane, D., Fink, L. V., Benita, Y., Robert, S. H., Harrigan, R. and Julio, S. G. M, (2009), Association between HIV-1 RNA Level and CD4+ Cell Count among untreated HIV-infected individuals, American Journal of Public Health Vol., (99) 51.

[9] Laird NM, Ware JH., (1982), Random-effects models for longitudinal data. Biometrics. 38, PP., 963-974.

[10] G, Kassahun W, Gurmessa A, (2016), Statistical Analysis of CD4+ Cell Counts progression of HIV-1-positive Patients enrolled in Antiretroviral Therapy at Hossana District Queen Elleni Mohamad Memorial Hospital, South Ethiopia. Biom Biostat Int J 3 (1): 00057.

[11] Debre Berhan Hospital annual report, 2016. [Accessed 16 February 2017].

[12] DE Gruttola and X. M. TU, (1994), Modeling progression of CD4-lymphocyte count and its relationship to survival time, Biometrics, vol. 50, no. 4, pp. 1003-1014.

[13] Diggle PJ, Heagerty P, Liang K-Y., (2002), Analysis of Longitudinal Data. 2nd ed. Oxford, United Kingdom: Oxford Uni versity Press.

[14] Hoover DR, Graham NMG, Chen B, Taylor JMG, Phair J, Zhou SYJ, Munoz A., (1992), Effect of CD4+ cell count mea surement variability on staging HIV-1 infection. Journal of Acquired Immune Deficiency Syndrome. Vol. 5, pp. 794-802.

[15] Langford SE, Ananworanich J and Cooper DA., (2007), Predictors of disease progression in HIV infection: a review. AIDS Res Ther., pp. 4:11.

[16] Macro International Inc., (2008), Ethiopia Atlas of Key Demographic and Health Indicators, 2005. Calverton: Macro In ternational, 2008, pp. 24.

[17] Mellors JW., Munoz A., Giorgi JV. (1997), Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann. Intern. Med, 126 (12):946-954.

[18] Nelson K and Williams CM., (2007) Infectious Disease Epidemiology. Theory and Practice. 2nd *ed.*

[19] Rizopoulos D, Verbeke G, Molenberhs G., (2008), Shared parameter models under random effects misspecification. Biome trika. 95, pp. 63-74.

[20] World Health Organization (WHO), (2012), World AIDS Day 2012: Closing in on global HIV targets. *In*.