Section: Medicine

Original Article

ISSN (0): 2347-3398; ISSN (P): 2277-7253

Determinants of CD4 Cell Count among Antiretroviral Therapy (ART) Attendant HIV Positive Adults Using Longitudinal Data Analysis in University of Gondar Referral Hospital

Koyachew Yenaneh Tadege[1], Tigist Jegnaw[2], Kasim Mohammed Yesuf[2]

[1]Central Statistical Agency Gondar Branch, Gondar, Ethiopia, [2]Department of Statistics, College of Natural & Computational Sciences, University of Gondar, P.o.Box. 196, Gondar, Ethiopia.

Abstract

Background: CD4 count is the best predictor for disease status and immediate risk of death and thus should be used to identify those who have advanced HIV disease. The aim of this study to investigating the basic factors of CD4cell count and assessing the progression of CD4cell count among HIV positive adults those they were attending university of Gondar Referral ART clinic, since December 2012 up to December 2017. **Subjects and Methods:** Since, the outcome variable was measured repeatedly through time and measurements within the patient were correlated. Such type of data requires a special types of modeling strategy, there for longitudinal data analysis plays a major role on such type of data analysis. Linear mixed model with unstructured correlation matrix were used to model the data in this study by including fixed and random factors on the model. **Results:** From the total of 216 Study subjects were followed retrospectively 61.08% and 38.92 were females and males respectively. The maximum number of observation per subject was 10 and the minimum and maximum CD4cell counts were 65 and 1440 respectively. **Conclusion:** Based on the findings of the study religion (orthodox, and Muslim), weight, baseline CD4cell count, time, TB screen positive, hemoglobin level, regimen type (d4t-3TC-NVP and AZT-3TC-NVP)and the interaction effect of time with baseline CD4count was significant predictors of CD4count. We would like to recommend for the patients must be, punctual, initiated to ART with high base line CD4count and hemoglobin level.

Keywords: ART; CD4cell count; Risk Factors; Longitudinal Data Analysis.

Corresponding Author: Tigist Jegnaw Lakew, University of Gondar, P. O. Box. 196, Gondar, Ethiopia.

Received: December 2019
Accepted: December 2019

Introduction

Globally, 34.0 million people were living with HIV at the end of 2011. An estimated 0.8% of adults aged 15-49 years across the world are living with HIV, although the burden of the epidemic varies among countries and regions. Sub-Saharan Africa continuously most severely affected, with nearly 4.9% living with HIV and accounting for 69% of the people living with HIV worldwide. Although the regional distribution of HIV infection is nearly 25 times higher in sub-Saharan Africa than in Asia, almost 5 million people are living with HIV in South, South-East and East Asia combined. Next to sub-Saharan Africa, the region's most heavily affected are the Caribbean and Eastern Europe and Central Asia, where 1.0% of adults were living with HIV in 2011.[1]

Globally, the annual number of individuals newly infected with HIV continues to decrease, although this varies strongly between regions. In general HIV/AIDS is one of the major public health problems in Sub-Saharan Africa, and Ethiopia, as one of these countries has been affected by the epidemic with a prevalence of 1.5 % and 1.1% in 2015 urban are highly affected than rural areas while females are twice affected than male population with HIV.[2]

The World Health Organization (WHO) states clinical failure amongst adults and adolescents as new or recurrent clinical conditions indicating severe immunodeficiency (WHO clinical stage 4 conditions) after 6 months of effective treatment. Immunological failure as CD4 count falls to the baseline (or below); or persistent CD4 levels below 100 cells/mm3. Virologic failure as plasma viral load above 1000 copies/ml based on two consecutive viral load measurements after 3 months, with adherence support.[3]

Access to antiretroviral therapy (ART) in Africa progressed dramatically over the past decade, beginning with a few thousand people and reaching five million people by midterm 2010. This advance was because of the low cost of drugs, increased resources, distributing HIV testing, and activism. Other problems exist to limit the number of people taking ART and the ability of health techniques to effectively manage patients, including inadequate number of physicians and allied health staff and limited laboratory capacity.[4]

The two African country Malawi and South Africa apply viral load for long-term monitoring. In South Africa, discontinuation of routine CD4 cell count is now advisable

after 1 year for patients stable on ART, although CD4 cell counts are done when needed for supporting decisions regarding the stopping of prophylaxis for some AIDS-associated opportunistic infections, one of the most important obstacle to the scaling up of viral load testing remains cost: for low and middle income countries test costs ranges from around US$10 to more than US$50.[5]

In general CD4 cell count is measured repeatedly through time within 6 months interval and the observations are related to each other therefore we must be use suitable model for such type of data in order to capture the correlation between measurements but classical regressions are not capture the dependency of this observation. Since CD4 cell count is longitudinal outcome variable it must modeled by using suitable longitudinal model in order to keep variance covariance structure of the patient and to estimate the parameters significantly. This study assess the effects of demographic, socio-economic, Laboratory and epidemiologic variables on CD4 Cell count among Antiretroviral Therapy (ART) follow up HIV infected adults by using liner mixed model of longitudinal data analysis.

Subjects and Methods

The data for this study would be obtained from a retrospective cohort study based on ART data base from the review of patient charts which contains socio-demography, laboratory and clinical information of all HIV patients from Gondar Teaching Referral Hospital among Antiretroviral Therapy (ART) follow up adults whose age greater than 14 years old who initiated on ART from December 1, 2012 to December30, 2017 GC. The total number of samples included in the study was 216 patients.

Variables Include in the study

Longitudinal outcome variable in this study are CD4 Cell count measured within 6 months period interval among Antiretroviral Therapy (ART) attendant HIV infected Adults. The explanatory variables in this study are the following that are assumed to be factors that cause of CD4 cell count progression after initiation of ART with time. Sex, maritalstatus, education, occupation, religion, age, WHO clinical stage, TBscreen, OI, Functional status, baseline, CD4count, Weight, Adherence, regimen type and hemoglobin were assumed to be the predictor variables.

Modeling Longitudinal Data

Suppose that a random sample is composed of m subjects with n predesigned time points. For subject i (i = 1, 2, …, m), the observed number of time points is usually denoted that does not necessarily equal to n due to missing observations. The response mea¬surement for subject i at time point j (j = 1, 2, …,) is written as . The repeated measurements of the response variable Y for subject i can be expressed in terms of an × 1 column vector, denoted by and given by

$$Y_i = \begin{pmatrix} Yi1 \\ Yi2 \\ \vdots \\ Yimi \end{pmatrix}, \text{ i=1, 2……..m}$$

Given m subjects, there are m such vectors in a longitudinal dataset. Repeated measurements of the response variable are generally specified as a function of the time factor and some other theoretically relevant covariates. In the analysis of cross-sectional data, various regression models are generally performed by assuming con¬ditional independence of observations in the presence of specified model parameters.[6]

n_i is the numbers of repeated measures form subject i and we can assume that they can differ from subject to subject. Also, it is assumed that the time at which measurements are taken might be not common. Additionally, Xi is an associated X p covariate matrix where p is the number of covariates and Xi can be illustrated by

$$X_i = \begin{bmatrix} x_{i11} & x_{i12\ldots..} & x_{i1p} \\ x_{i21} & x_{i22\ldots..} & x_{i2p} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_{ini1} & x_{ini2\ldots..} & x_{inip} \end{bmatrix} \text{i=1,2…….. N}$$

Where X_{itp} corresponds to the p^{th} covariate for subject i at time t. Some of the elements of Xi cannot be changing across time.

Exploratory Data Analysis

Most longitudinal analyses address the relationship of a response with explanatory variables, often including time.7 Individual profiles, Mean structure, Variance function, Correlation structure are looked in longitudinal data analysis. So as to understand the possible relationships among means over time, for balanced data, graphical inspection can be used by connecting the average values computed at each time point separately.

Variance Covariance Structures

Modeling covariance structure refers to representing Var(Y) as a function of a relatively small number of parameters. Measures on the same subject at different times almost always are correlated, with measures taken close together in time being more highly correlated than measures take far apart in time. That is, repeated measures are correlated. For an analysis to be valid, the covariance among repeated measures must be modeled properly.8 The most commonly used covariance structures are compound symmetry (CS), Toeplitz (TOEP), unstructured (UN), and autoregressive (1) (AR (1)).

Linear Mixed Effects Model (LMM)

It contains fixed effects and random effects where random effects are subject-specific and are used to model between-subject variation and the correlation induced by this variation.The linear mixed models approach to repeated measurements views theanalysis as a univariate regression analysis of responses with correlated errors.One major advantage of this methodology is that it accommodates the complexities of typical longitudinal data sets. The mixed model approach permits specification of models determined by subject matter considerations rather than by limitations of the statistical methodology. It also allows explicit modeling and analysis of variation between subjects and within subjects.A commonly used mixed effects linear

model for continuous response variables was proposed by.9 In this model individuals are not assumed to be measured on the same number of time-points; individuals with incomplete data across time are included in the analysis, time is treated as a continuous variable, and both time-invariant and time-varying covariates can be included in the model.The linear mixed model (LMM) is very flexible and capable of fitting a large variety of datasets.

In general a linear mixed effect model is any model which satisfies:

$\mathbf{Y_i = X_i\beta + Z_i b_i + \mathcal{E}_i}$

where Y_i is the n_i-dimensional response vector for subject i 1=<i=<M, M is the number of subjects X_i and Z_i are $(n_i \boldsymbol{x}\ \boldsymbol{P})$ and $(n_i xq)$ dimensional vector of known covariates, β is p-dimensional vector containing the fixed effect, b_i the q-dimensional vector containing the random effects, and $\mathcal{E}_i$ is n_i-dimensional vectors of residual components. Finally, D is a general (q x q) covariance matrix with (i,j) element $d_{ij} = d_{ji}$ and $\sum_i$ is $(n_i x n_i)$ covariance matrix which depends on i only through its dimension n_i i.e the set of unknown parameters in $\sum_i$ will not depend on i. Linear mixed effect model is the most widely used methods for analyzing longitudinal data. As mentioned earlier, this model could handle the complications of incomplete measurements in a very natural way. In this study, a linear mixed model was apply based on the assumption that the vector of repeated measurements in the original scale on each patient follows a linear regression model where some of the regression parameters are the same for all patients (i.e., population-specific), while others are different across patients (i.e, patient-specific).[10]

Where: ε_i~ N(0,Σi), bi~N(0,D), b1, …, bN,$\mathcal{E}_1$, …, $\mathcal{E}_N$ are independent, D and Σi are variance components, β stands for the fixed effects, bi stands for the random effects, Xi and Zi are design matrices and the marginal mean and variance of Yi is given by

E (Yi) = Xiβ

V (Yi) = V (Z_i bi +ε_i) = Z_iDZ'$_i$ + Σi = Vi

Assumptions of LMM

There are two basic distributional assumptions for the general linear mixed effects model.

ε_i ~ N(0, δ^2I)

The between-subject errors are independent and identically normally distributed, with mean zero and variance δ^2I and they are independent of the random effects. This assumption can be relaxed by allowing to model non constant variances or special within –subject correlation structures.

b ~ N(0,D)

The random effects are normally distributed, with mean zero and covariance matrix D (Not depending on the between the subject) and are independent for different groups.

Parameter Estimation Methods

Maximum likelihood (ML) and restricted maximum likelihood (REML) methods are used to estimate D and Σ.

The general linear mixed model is given by

Yi= Xi β + Z_i b_i+ ε_i

Where bi ~ N(0, D)and ε_i ~ N(0, Σi), bi and ε_i are independent and thus the marginal model is given by

Yi ~ N($X_i\beta$, Z_iDZ_i + Σ_i)

Then marginal likelihood function is given by

l_{ML}= $\Pi_{i=1}^{N}\{(2\pi)^{-n/2}|Vi(\alpha)|^{-1/2}\exp(-\frac{1}{2}$ (Yi - Xiβ)'Vi^{-1} (α) (Yi - Xiβ))}

Where α is the vector of all variance components in D and Σi; Θ = (α'; β')' is the vector of all parameters in marginal model and β is vector of fixed effects. Then log likelihood function for subject i is

$li = \frac{-n_i}{2}\log(2\pi) - \frac{1}{2}log|Vi| - \frac{1}{2}$ (Yi - Xiβ)'Vi^{-1} (Yi - Xiβ)

$\frac{dli}{d\beta}$ = -XiV^{-1}Xiβ + Xi'$Vi^{-1}y_i$

If α were known, then the MLE of β on combining all the information from all the N subjects equals.

$\hat{\beta}(\alpha) = (\Sigma_{1=}^{N} X_i' V^{-1} X_i)^{-1} \Sigma_{i=1}^{N} X_i' V^{-1} y_i$

In most cases α is not known and needs to be estimated as say$\hat{\alpha}$, then V_i^{-1}should subsequently be replaced byV_i $(\hat{\alpha})^{-1}$. The two frequently used methods to estimate α are maximum likelihood and restricted maximum likelihood.

Maximum Likelihood Estimation

In maximum likelihood estimation $\hat{\alpha}$ is obtained by maximizing the profile likelihood L_{ML} $(\alpha, \hat{\beta}(\alpha))$ with respect to α after β is replaced. The resulting estimate of $\hat{\beta}(\hat{\alpha}_{ML})$ for β will be denoted by$\hat{\beta}$. The estimates$\hat{\alpha}$ and $\hat{\beta}_{ML}$can also be obtained from maximizing $L_{ML}(\theta)$ with respect to θ that is, with respect to both α and β simultaneously.

Restricted Maximum Likelihood Estimation

The REML estimation method applies ML estimation techniques to the likelihood function. The only difference is the REML estimation method is associated with a set of "error contrasts" rather than associated with the original observations**.**

Both ML and REML are based on the likelihood principle, which has the properties of consistency, asymptotic normality, and efficiency but REML corrects for the downward bias in the ML parameters in ***D*** and Σ. REML handles strong correlations among the responses more effectively than ML and the differences in estimating between them increases as the number of fixed effects in the model increases.

Consider a sample of N observations Y_1 ……$Y_N$from N(μ , σ^2) and for known μ MLE of σ^2 equals

$$\hat{\sigma}^2 = \sum_i \frac{(Y_i - \mu)^2}{N}$$

$\hat{\sigma}^2$is un biased for σ^2 . When μ is not known, MLE of

$\sigma^2 = \hat{\sigma}^2 = \sum_i \frac{(Y_i - \bar{Y})^2}{N}$ Note that $\hat{\sigma}^2$ is unbiased for σ^2: Ė($\hat{\sigma}^2$) =$\frac{N-1}{N}\sigma^2$

This indicating that MLE is known biased dawnward, due

to the estimation of μ

The biased expression tells us how to derive unbiased estimate

$S^2 = \sum_i \frac{(Y_i - \bar{Y})^2}{N-1}$

Apparently, having to estimate μ introduce biase in MLE of σ^2

Let Y =$(Y_1 \ldots\ldots\ldots\ldots Y_N)' \sim N(\mu \ldots\ldots\ldots\ldots \mu)', \sigma^2 I_N)$.

We transform Y such that μ vanishes from the likelihood:

U = $(Y_1 - Y_2 Y_2 - Y_3 \ldots\ldots\ldots Y_{N-2} - Y_{N-1} Y_{N-1} - Y_N)'$= A'Y and A'Y~N(0 ,$\sigma^2$A'A)

, Then MLE ofσ^2, based on U, equals:

$S^2 = \sum_i \frac{(Y_I - \bar{Y})^2}{N-1}$ A defines a set of N-1 linearly independent error contrasts and S^2 is called the REML of estimate of σ^2 and S^2 is independent of A.

Consider a sample of N observation Y_1……. Y_N from a linear regression model

Y= (Y_1……. Y_N)'~ N(Xβ , σ^2I) ,

MLE of σ^2:$\hat{\sigma}^2$ = $\frac{(Y - X\hat{\beta})(Y - X\beta)}{N}$

, note that $\hat{\sigma}^2$ is unbiased for σ^2.

E ($\hat{\sigma}^2$) = $\frac{N-p}{N}\sigma^2$

The biase expression tells us how to derive un unbiased estimate:

MSE = $\frac{(Y - X\beta)'(Y - X\beta)}{N-P}$

MSE can also be obtained from transforming the data orthogonal to X:

U= A′Y ~N(0,σ^2A'A). The MLE ofσ^2, based on U, equals the mean squared error, (MSE).The MSE is again called the REML estimate ofσ^2.

We first combine all models.

$Y_i \sim N(X\beta, V_i)$ in to one model $Y \sim N(X\beta, V)$,In which Y = (Y_1 ……Y_N)'

X= (X_1 … … … X_N)'

V (α) = $\begin{pmatrix} V_1 & \ldots. & 0 \\ \ldots. & \ldots. & \ldots \\ 0 & 0 & V_N \end{pmatrix}$

Again the data are transformed orthogonal to XU= A′Y ~N(0 , A' V(α) A). The MLE α based on U is called REML and is denoted by $\sigma^2{}_{REML}$. The resulting estimate $\hat{\beta}$ ($\hat{\alpha}$) for β denoted by β_{REML}. $\hat{\alpha}_{REML}$and$\hat{\beta}_{REML}$can also be obtained from maximizing

$L_{REML}\theta = [\sum_{i=}^{N} X' W_i(\alpha) X_i]^{-1/2} L_{ML}(\theta)$

With respect toθ which means with respect to α and simultaneously β

L_{REML} (α,$\tilde{\beta}(\alpha)$) is the likelihood of the error contrast U , and is often called REML of likelihood function. Note that $L_{REML}(\theta)$ is not the likelihood for our original data Y.

Inference for the Marginal Model

The primary interest in drawing inference on the parameters in model in order to generalize results obtained from a specific sample to the general population from which the sample was taken.

Inference for the Fixed Effects

The vector β of fixed effect is estimated by:

$\hat{\beta}(\alpha) = (\Sigma_{1=}^{N} X_i' V^{-1} X_i)^{-1} \Sigma_{i=1}^{N} X_i' V^{-1} y_i$

In which the unknown vectors of α of variance components is replaced by its ML and REML estimate. Under the marginal model and conditionally on α,$\hat{\beta}(\alpha)$ follows a multivariate normal distribution with mean vector β and variance -covariance matrix:

Var($\hat{\beta}$) = $(\sum_{i=}^{N} X' W_i X_i)^{-1} (\sum_{i=}^{N} X' W_i var(Y_i) W_i X_i) (\sum_{i=}^{N} X' W_i X_i)^{-1}$ = $(\sum_{i=}^{N} X' W_i X_i)^{-1}$

Where $W_i = V_i^{-1}(\alpha)$ the covariance matrix Var ($\hat{\beta}$) estimated by replacing α by ML and REML.

Approximate Wald Test

For each parameter β_j in β, j = 1,……,p , an approximate wald test (also termed Z-test),as well as an associated confidence interval is obtained from approximating the distribution of $\frac{(\hat{\beta}_j - \beta_j)}{\widehat{s.e}(\hat{\beta}_j)}$ by astandared univariate distiribution. In general for any known matrix L test for the hypothesis

H_0:Lβ =0

H_A : Lβ not equal to 0 and follows the distribution of

$G = (\hat{\beta}_j - \beta_j)' L' (L(\sum_{i=1}^{N} X_i' V_i^{-1}(\hat{\alpha}) X_i)^{-1} L')^{-1} L(\hat{\beta}_j - \beta_j)$

Asymptotically follows chi-square distribution with rank (L) degree of freedom.

Approximate t-Test and F-Test

The wald test statistics are based on estimated standared errors which under estimate the true variability in $\hat{\beta}$ because they do not take into account the varibility introduced by α .this dawnward biase is often resolved by using approximate t-Test and F-Testfor testing hypothesis about β.

For each parameter $\beta_j$in β , j = 1,……,p , an aproximate t-test and associated confedece interval canobtained by approximating the distribution of of $\frac{(\hat{\beta}_j - \beta_j)}{\widehat{s.e}(\hat{\beta}_j)}$ by an aproximate t- distribution.

F=$(\hat{\beta}_j - \beta_j)' L' (L(\sum_{i=1}^{N} X_i' V_i^{-1}(\hat{\alpha}) X_i)^{-1} L')^{-1} L(\hat{\beta}_j - \beta_j)$/rank (L)

The numerator degree of fredom equals rank (L). The denominator degree of fredom needs to be estmated from the data.

Likelihood RatioTest: Aclassical statistical test for the comparission of nested models with different mean stractur ,but equal variance stracturis the likelihood ratio test.

-2lnλ_N = -2ln $\left[\frac{L_{ML}(\hat{\theta}_{ML}, 0)}{L_{ML}(\hat{\theta}_{ML})}\right]$

Where$\hat{\theta}_{ML}$,0 $and\hat{\theta}_{ML}$ the maximum likelihood estimate obtained from maximazing L_{ML} over θ_β ,0 and θ_β respectively.

Inference for the Variance Componenet

Estimating the parameteres in the mariginal linear mixed effect model (the fixed effect β and the variance compnent in D and all $\sum_i$) it is often use full estimates for the random effect b_i as well , since they refelect how much the subject sesific profiles divate from the overall average proffile.

Such estimate can then be interpreted asresidual which may be helpful for detecting special profiles (i.e., outling

individuals) or group of individuals evolving differently in time.[10]

Random Coefficient Models

It is often important in a study to determine the relationship between the response and time. This is often done by including the measurement time as a covariate in the model, with a corresponding slope, say β_t. It is plausible and likely that the slope will vary with subject, so it might be useful to model a separate intercept and slope for each subject in the study. This is done by fitting the subject variable as the intercept and the subject*time interaction as the slope for each patient. These two terms could reasonably be assumed to arise at random from a distribution and, thus, would be specified as random effects. This gives rise to what is called a random coefficients model.

The Random Intercept Model

As the name indicates, the Random Intercept Model (RIM) is characterized by including an individual intercept as only random effect. Therefore, the matrix Z_i reduces to a vector of ones and the vector b_i reduces to a scalar. The model equation of the RIM as follows;

$$y_i = X_i\beta + 1_{ni}b_i + \square_i$$

$b_i \sim N(0, d^2)$, $\square_i \sim N(0, \sum_i)$

By matrix notation the formula:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_{mi} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12}....x_{1(p-1)} \\ \vdots & \ddots & \vdots \\ 1 & x_{ni,1} & x_{ni(p-2)} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} b \\ b \\ . \\ . \\ b \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_{ni} \end{pmatrix}$$

The Random Intercept and slope Model

A random intercept and slope model can be appropriate if the individual trends to differ in slope and the model equation which includes an individual slop could be as follows.[11]

Model Comparison Technique

The objective of model comparison is selecting the best model that fits the data. Methods that uses for model comparison are Akaike's Information Criterion (AIC),Bayesian information criterion (BIC and Likelihood ratio Test.Akaike's Information Criterion (AIC): often it is of interest to compare models that are not nested. One common method is using the Akaike's Information Criterion (AIC), which is also based on the maximized log-likelihood, but it includes a penalty for complexity of the covariance model assumed.

$$AIC = -2^{\ell_{model}} + 2c$$

Where ℓ_{model} is the maximized or fitted (REML) log-likelihood using the assumed model and 2c is the number of parameters included in this model, among all the interested covariance models, the one with the smallest AIC is preferred.

The basic idea behind the AIC is to strike a balance between the fit to the data and the number of parameters involved in the covariance model (if the competing models assume the same model for the mean trend). Likelihood ratio Test is best applicable to compare nested models.It is constructed by comparing the maximized log likelihoods for these full & reduced models respectively and its test statistics is defined as:

$$_{LR} = -2\ln\lambda N = -2\ln\left(\frac{L_{ML}(\hat{\alpha}_{ML,o})}{L_{ML}(\hat{\alpha}_{ML})}\right)$$

Where;$\hat{\alpha}_{ML,}$ oand $\hat{\alpha}_{ML}$ are respective maximum likelihood estimates which maximize the likelihood functions of the reduced and full model. The asymptotic null distribution of the LR test statistic is a chi-square distribution with degrees of freedom equal to the difference between the numbers of parameters in the two models.

Model Checking Technique

In our model selection we have accepted the model with the best likelihood value in relation to the number of parameters but we still do not know if the model chosen is a good model or even if the normality assumption we have made is realistic. To check this we look at two types of plots for our data, normal plots and residual plots to see if the residuals and random effects seem to follow a normal distribution, if the residuals seem to have a constant variance and to look for outliers. In linear mixed effects model, it is assumed that the random effects are normally distributed and uncorrelated with the error term. Residual plots can be used visually to check normality of these effects and to identify any outlying effect categories. The assumption of normality for the within-group error was assessed with formal tests and the normal probability plot of the residuals by covariates. Similarly, normality of the random effects is assessed using Normal Plot of each random effect. Normal plot of estimated random effects helps for checking marginal normality and to identify outlier.[12]

Results

The aim this chapter is discus about the general characteristics of the response variable and identifying the risk factors. In this study a total of 216 samples were included in order to answer the objective of research. Among this 132 (61.08%) were female and 84 (38.92%) were males . the sample was drawn from 2617HIV positive adults by using systematic random sampling those they were attending from December1, 2012 G.C up to December 30, 2017 G.C in Gondar university teaching Referral hospital. As we observe from [Table 1] the mean CD4cell count is increase from baseline up to throughout the study period with the minimum and maximum CD4cell count 65 and 1440 respectively. This indicates the progression of CD4cell count is increase with time after starting the treatment and the number of observation were decrease from 216 to 68 during the follow up time for the last month. The minimum number of observations per subject was two and maximum of 11.

Table 1: Means of continuous covariates UOGTRH, 2012-2017

Variable	Mean	Std Dev
Hgb	14.65	1.94
Weight	57.35	10.42
BaseCD4	293.88	167.91

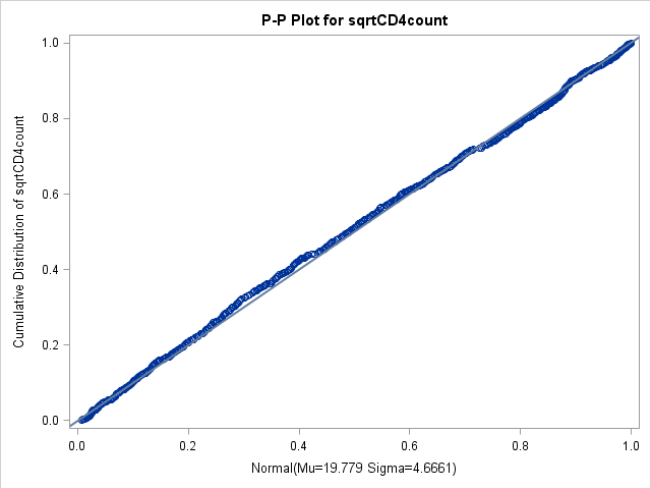

Figure 1: Normal P-P plot of square root CD4cell count UOGTRH, 2012-2017

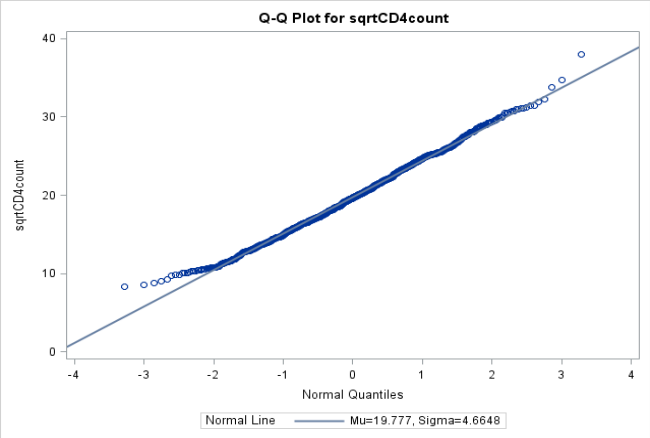

Figure 2: Normal Q-Q plot of square root CD4cell count UOGTRH, 2012-2017

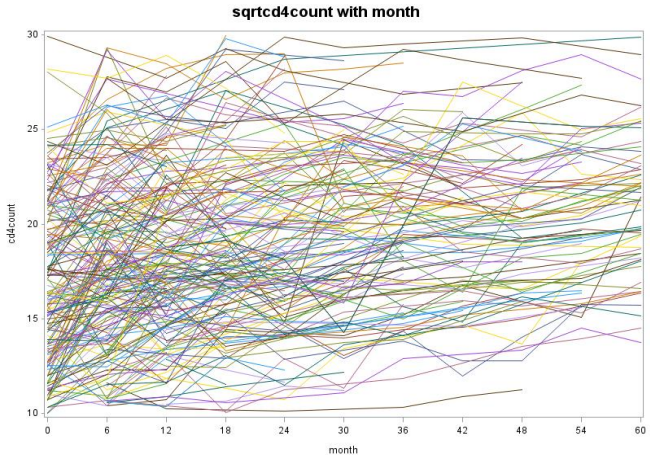

Figure 3: Individual profile of square root CD4cell count UOGTRH, 2012-2017

As we observed from [Table 1] the mean weight of the patient were 57.35 with Std Dev 10.42 and the mean baseline CD4cell count were 293.88 with Std Dev 167.91.

As we observed from [Figure 1,2] the square root of CD4cell count is show more linear trend as compared to normal CD4cell count and log CD4cell count.

The individual profiles of patients in [Figure 3], suggests there was high within and between patients variability. Increasing evolutions of CD4cell count over time were suggested from most of the individual profiles of subjects. From the above [Figure 3] the variation was high at the beginning than the last.

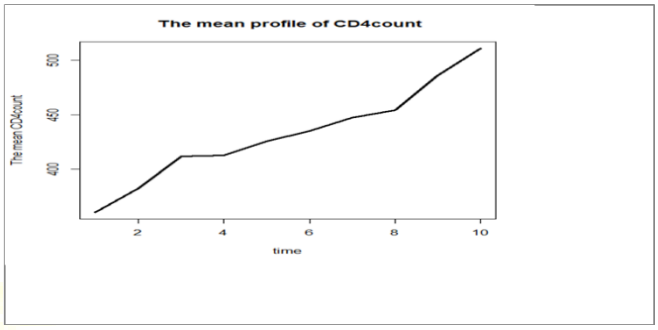

Figure 4: Mean profile of square root CD4cell count UOGTRH, 2012-2017

From the [Figure 4] we observed mean plot of CD4cell count with time follows linear structure and the relationship between CD4cell count and duration of ART is approximately linear which means the CD4cell count increases throughout the follow-up time.

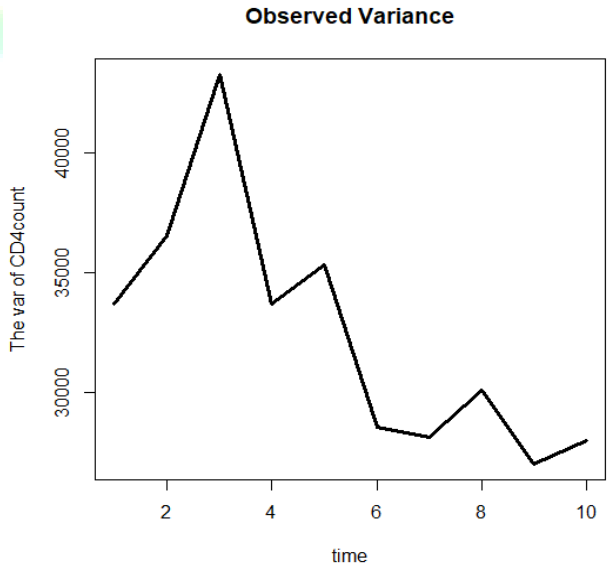

Figure 5: Variability of CD4cell count data with time UOGTRH, 2012-2017

As we observed from the [Figure 5] the observed variance was not constant, which means it needs special type of none constant variance covariance stricture to assessed the progression of CD4cell count.

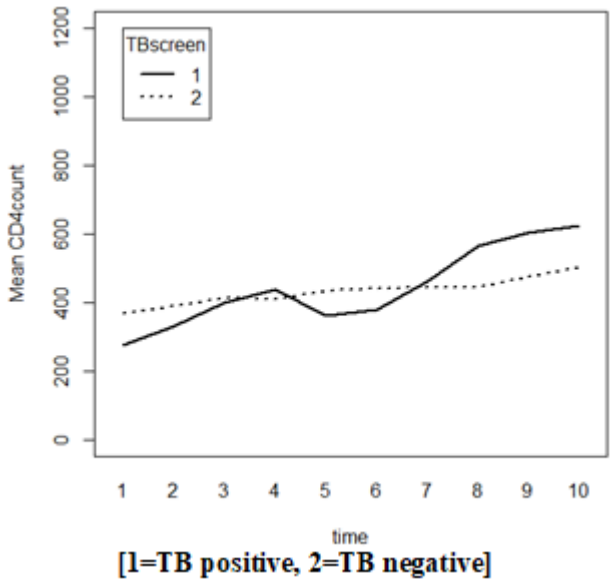

Figure 6: Mean profiles analysis of CD4cell count by TBscreen UOGTRH, 2012-2017

As we observe from the [Figure 6] patients with TBscreen negative HIV positive patients had high mean CD4cell count at the initial time than those TBscreen positive HIV patients until time 3 but after time seven the mean CD4cell count of TBscreen positive patients were increase; i.e: at the initial time there were the mean CD4cell count difference among TBscreen groups.

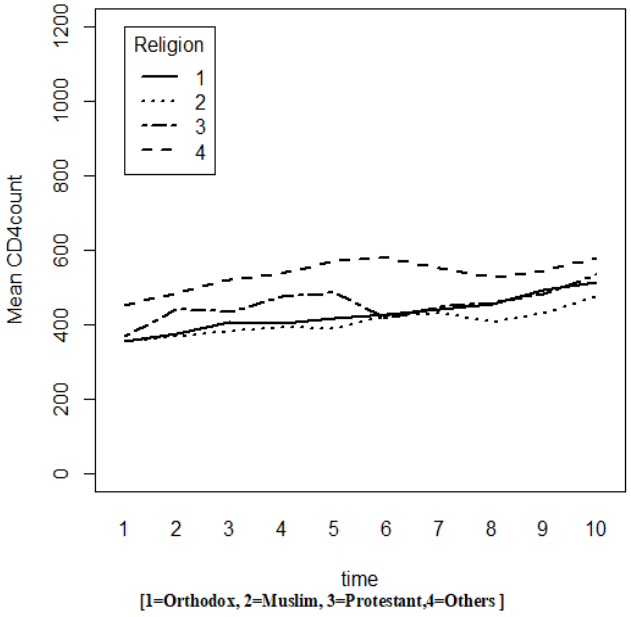

Figure 7: Mean profile plot of CD4cell count by religion UOGTRH, 2012-2017

The mean CD4cell count was variable among the four religions i.e. orthodox and Muslim had almost similar at the beginning up to time seven. Orthodox and Muslim patients had lowest mean CD4cell count throughout the study time than the other religious types like protestant and catholic [Figure 7].

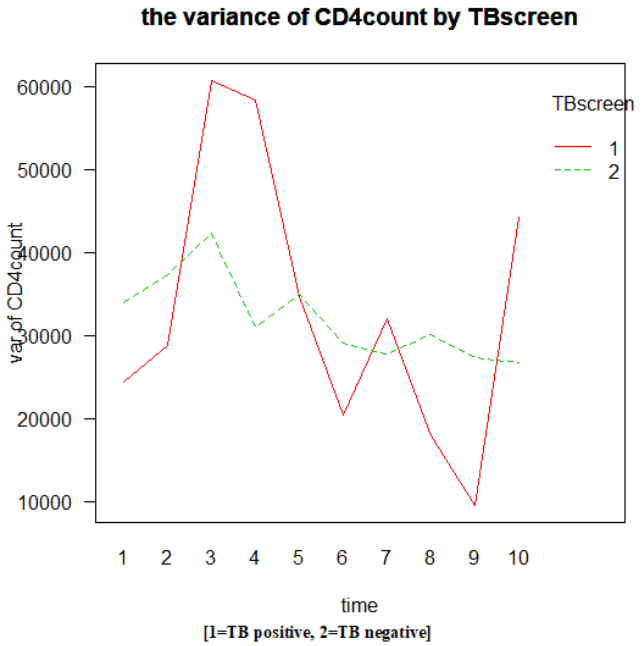

Figure 8: Variance profile of CD4cell count by TBscreen UOGTRH, 2012-2017

As we observed from [Figure 8] those patients TBscreen positive was higher variability of CD4cell count as compared to TBscreen negative HIV positive patients.

After selecting potentially variables that are included in the reduced linear mixed model and analysis sqrtCD4cell count again, the selected variance-covariance stricture were TOEP based on AIC and BIC and this indicates that the TOEP correlation structure is the best structure that shows the progression of CD4cell count data.

From the univariate model variables such as gender; Religion, weight, Base CD4, time, TBscreen, Hgb OI and Regimen type were the selected variable based on p-value less than 0.25 from the full linear mixed model. The model that has the smaller AIC and BIC is the best model that fits the data and used to further analysis for the sqrtCD4cell count data.

Table 2: Covariance structure for reduced linear mixed model UOGTRH, 2012-2017

Covariance structure	AIC	BIC	-2 Res Log Likelihood
UN	5832.5	6018.1	5722.5
AR (1)	5958.5	5965.3	5954.5
Toep	5908.3	5942.1	5888.3

From [Table 2] the best covariance structure were TOEP even if UN covariance structure had smallest AIC and -2 Res Log Likelihood but if the time point was increase the number of variance-covariance parameter to be estimated were increase.

Based on the study factors such as religion, weight, baseline CD4cell count, time, TBscreen, Hemoglobin, regimen type and the interaction of time with baseline CD4cell count were statistical significant predictors of the progression of CD4cell count. As we observe from [Table 3] bellow the

interaction of time with baseCD4cell count indicates patients that had higher base line CD4cell count has taken less time to the progression of their CD4cell count than patients that had lower baseline CD4cell count.

Table 3: Reduced linear mixed model UOGTRH, 2012-2017

Effect	$\widehat{\beta}$	S.e ($\widehat{\beta}$)	p-value
β_0	6.86	1.89	0.0003
Gender Female	-0.59	0.29	0.1172
Religion Orthodox Muslim Protestant	 -2.93 -2.8 -1.67	 1.14 1.17 1.26	 0.0106 0.0176 0.1865
Weight	0.039	0.011	0.0002
BaseCD4	0.020	0.0015	<.0001
Time	0.91	0.18	<.0001
TBscreen Positive	-0.37	0.14	0.0104
Hgb	0.35	0.089	<.0001
OI No	0.24	0.18	0.1955
Regimen d4t-3TC-EFV d4t-3TC-NVP AZT-3TC-NVP AZT-3TC-EFV TDF-3TC-EFV	 -0.64 2.39 1.56 1.51 0.97	 0.84 0.88 0.73 0.84 0.70	 0.4602 0.0162 0.0482 0.0898 0.1830
BaseCD4*time	-0.00087	0.00016	<.0001
Hgb*time	-0.020	0.013	0.1189

Random Coefficient Model

It is often important in a study to determine the relationship between the response and time. This is often done by including the measurement time as a covariate in the model, with corresponding slope, say β_t.

Table 4: Covariance structure for random intercept only model UOGTRH, 2012-2017

Covariance structure	AIC	BIC	-2 Res Log Likelihood
UN	6063.1	6069.8	6059.1
AR (1)	6065.1	6075.2	6056.5
Toep	6063.1	6069.8	6059.1

As we observed from the table UN and Toep have similar variance -covariance structure but the best one is TOEP than AR(1) and UN based on smaller AIC and BIC, since UN variance covariance structure were not advisable as the time point increase [Table 4].

Table 5: Covariance structure for random intercept and slope model UOGTRH,2012-2017

Covariance structure	AIC	BIC	-2 Res Log Likelihood
UN	5988.2	6001.7	5980.2
AR (1)	6306.5	6317.6	6302.5
Toep	6302.5	6313.2	6301.5

From [Table 5] the best Covariance structure were Toep even if UN had smallest variance covariance structure based on smaller AIC, BIC and -2 Res Log Likelihood value from random intercept and slope model.

Table 6: Covariance Parameter Estimates random intercept only model UOGTRH, 2012-2017

Covariance Parameter Estimates					
CovParm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr> Z**
Variance	Id	5.5589	0.6059	9.17	<.0001
Residual		2.8853	0.1188	24.28	<.0001

Interclass Correlation (ICC)

$$\text{ICC}= \frac{\delta_b}{\delta_b+\delta_W} = \frac{5.5589}{5.5589+2.8853} = 0.66$$

The ICC assesses the correlation of observations within the same subject. The intra class correlation coefficient ρ was found to be 0.6617indicating that 66 % of the Variability within patients is explained by the inclusion of the random effect. Hence, excluding the random effect affects the conclusions of the study [Table 6].

Model Selection

In this study, both Akaike information criterion (AIC) and Bayesian information criterion (BIC) was used as a tool for selecting the better model that fits sqrtCD4cell count data. In this study there were three types of models those are without intercept, with intercept and slope and reduced linear mixed model.

Table 7: Compare variance-covariance structure of the three models UOGTRH, 2012-2017

Model	Variance – covariance structure	AIC	BIC	-2LogLike hood
RLMM	TOEP AR(1) UN	5908.3 5958.5 5832.5	5942.1 5965.3 6018.1	5888.3 5954.5 5722.5
Without intercept	TOEP AR(1) UN	6302.5 6306.5 6063.1	6317.6 6313.2 6069.8	6301.5 6302.5 6059.1
With intercept and slope	TOEP AR(1) UN	6063.1 6065.1 5988.2	6069.8 6075.2 6001.7	6059.1 6056.5 5980.5

As we observe [Table 7] among the three models based on AIC and BIC the best model those used to investigating the progression and predictors variables of CD4cell count was reduced linear mixed model with TOEP covariance structure, which indicates reduced linear mixed model with linear time effect was considered for this study.

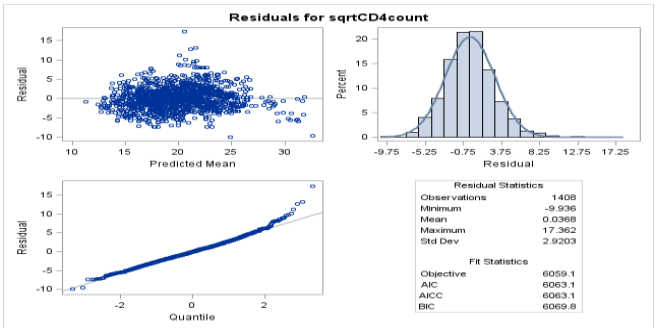

Figure 9: Residuals plot for square root CD4cell count UOGTRH, 2012-2017

As we observe from the above [Figure 9] the graph of the standardized marginal residual vs predicted mean shows most of the observation were exist around zero this indicates there is a linearity of the outcome with explanatory variables and test the homogeneity of with in patient variance. The histogram and q-q plot shows error terms are normally distributed.

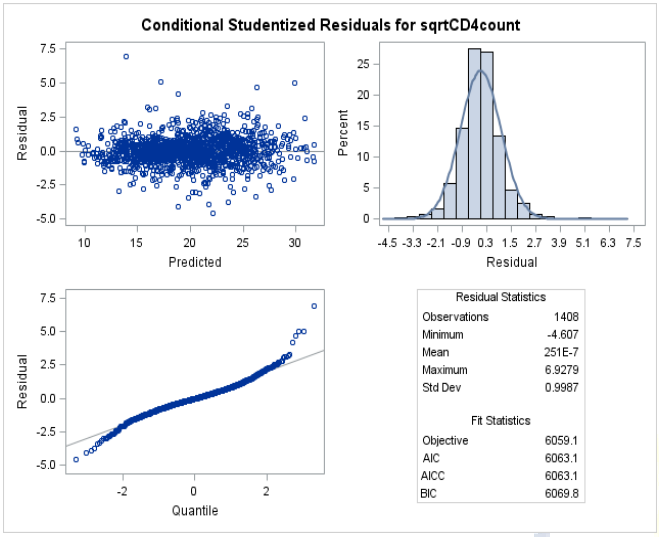

Figure 10: Studentized residual for sqrtCD4cell count UOGTRH, 2012-217

As we observe from the Studentized conditional residual vs predicted mean plot some of the observation were far from the majority of the observation this indicates there is an outlier observation and the presence of none constant variance [Figure 10].

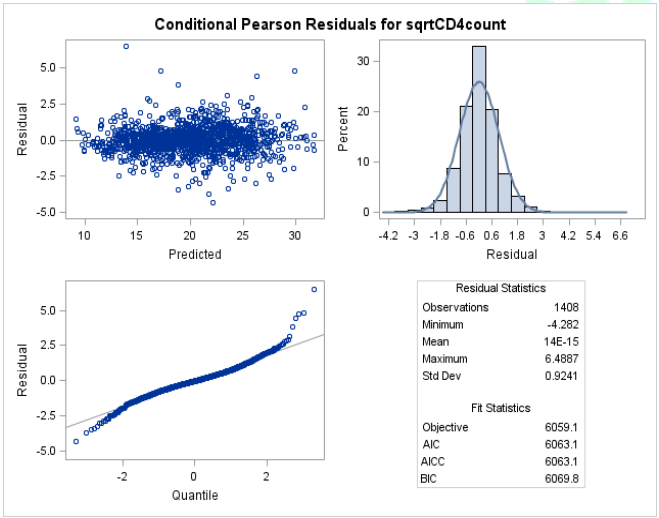

Figure 11: Pearson residual forsqrtCD4 cell count UOGTRH, 2012-2017

As we observe from [Figure 11] the plot of conditional Pearson residual and the predicted mean shows the normality assumption of the random effects using histogram and the q-q plot shows the normality of the random error term.

Discussion

Based on the study findings religion (orthodox and Muslim), regimen type(d4t-3TC-NVP and AZT-3TC-NVP), baseline CD4cell count, time, TBscreen, hemoglobin, weight and the interaction effect of time with baseline CD4cell count were predictors of CD4cell count. Based on the study findings Religion (orthodox and Muslim) were 2.93 times and 2.8 times less the progression of CD4count respectively than others like catholic. Patients whose Basline CD4cell count increase by one unit was it increases the progression of CD4cell count by 0.02 times and time were a significant effect of on the progression of CD4cell count. This results supports the study conducted by previous researchers.[13-15] This demonstrates patients they have higher baseline CD4cell count are improve their CD4cell count with in short period of time. The interaction of time with Baseline CD4cell count has a significant effect on the progression of CD4cell count but the parameter was negative, which indicates patients initiated for ART with minimum CD4cell count takes more time for the progression of CD4cell count than those who had higher CD4cell cell count. This study supported by other researchers which demonstrates the interaction of CD4cell count with time was one of the predictors of progression of CD4cell count.[13] In this study hemoglobin level was one of the predictors of CD4cell count progression, which indicates one, unite increasing of hemoglobin level increases the progression of CD4cell count by 0.35 times. This study supports the study of Mihiretu who suggests hemoglobin levels were one of the cases of the progression of CD4cell count in university of Gondar teaching Referal hospital.[16] Based on the study findings TB screen positive were significant effect on the progression of CD4cell count and one unit increased TB screen positive, it reduces the progression of CD4cell count by 0.37 times. This supports the study of Belay, which demonstrates TBscreen positive were one of the predictors of the progression CD4cell count. In his study regimen type d4t-3TC-NVP and AZT-3TC-NVP were the factors on the progression of CD4cell count.[17] Patients treated by d4t-3TC-EFV and AZT-3TC-NVP drug types were 2.39 times and 1.56 times higher than those patients treated by other types of drugs such as TDF+3TC+NVP. This study supports Kebadu who found that regimen type were one of the predictors of progression of CD4 cell count.[13]

Conclusion

As we observed from the mean of the response variable the number of observation were reduced from 216 to 68 and the number of female and male patients included in this study were 132 and 84 respectively. Investigated basic factors on the progression of CD4cell count were religion, weight, Basline CD4cell count, duration of ART, TBscreen, hemoglobin level, regimen type and the interaction effect of time with BaseCD4cell count. Religion (Orthodox and

Muslim) has a significant effect on the progression of CD4cll count, so in order to address this problem religious father must be providing advices or create awareness for their follower. Patients with low weight were associated with small number of CD4cell count, there for in order to address this problem patients must continuously check their weight and start the treatment. HIV positive patients must be taken anti-TB treatment at any time during the treatment, since TB screen positive was one of the predictor variables on the progression of CD4cell count. Patient's hemoglobin label was positively correlated with the progression of CD4cell count, it is recommended for all patients they must be start the treatment early.

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
AIC	Akaike Information Criteria
ART	Anti-Retroviral Therapy
AR	Autoregressive
CD4	Cluster Of Differentiation
HAART	Highly Active Anti-Retroviral Therapy
HIV	Human Immune Deficiency Virus
UNAIDS	Joint United Nations Program on HIV/AIDS
WHO	World Health Organization
EDHS	Ethiopian demographic and health survey
d4t	Staudinger
NVP	Neverapine
3TC	Lamivudine
AZT	Azidothymidine/ Zidovudine
TDF	Tenofovir
LDA	Longitudinal data analysis
LMM	Linear mixed model
UN	Unstructured
OI	Opportunistic infection
Toep	Toeplitz
UOGTRH	University of Gondar Teaching Referal hospital
RIM	Random intercept model
BIC	Bayesian information Criteria

Ethics, consent and permissions

The ethical clearance for the survey was approved by Ethical Review Board of College of Natural and Computational Science.

Acknowledgements

The authors of this article would like to thank University of Gondar Teaching Hospital for making available the data used in this research. We also thank Mr. Gashu Ybriedawd, he was edit the manuscript properly like grammar of the sentence and coherence of the paragraph.

Source of Finance

No fund was obtained from any source to conduct the study as well as in processing of the manuscript for publication.

Availability of data and materials

Data can be found from the corresponding author based on request.

Authorship Contributions

Idea/Concept: Koyachew Yenaneh , Kasim Mohammed Yesuf,; Design: Tigst Jegnaw, Kasim Mohammed Yesuf,; Control/Supervision: Statistics Department; Data Collection and/or Processing: University of Gondar Teaching Hospital; Analysis and/or Interpretation: Koyachew Yenaneh, Kasim Mohammed Yesuf; Lierature Review: Tigst Jegnaw, Kasim Mohammed Yesuf,; Writing The Article: Koyachew Yenaneh, Kasim Mohammed Yesuf, Critical Review: Koyachew Yenaneh, Kasim Mohammed Yesuf, Tigst Jegnaw; References and Fundings: Koyachew Yenaneh, Tigst Jegnaw; Materials: Kasim Mohammed Yesuf, Tigst Jegnaw

References

1. UNAID. Report on the global AIDS epidemic (UNAID, 2012).
2. National guidelines for comprehensive HIV prevention, care and treatment (WHO, 2015).
3. Sang R. and F. Miruka (2016). "Factors associated with virologic failure amongst adults on antiretroviral therapy in Nyanza Region, Kenya." IOSR J Dent Med Sci15(7): 108-121.
4. Jonathan Mermin, d., et al. (2011)."Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy ".
5. Nathan Ford, G. M., Anton Pozniak, Helen Bygrave, Andrew Hill, Trevor Peter, Mary-Ann Davies, Beatriz Grinsztejn, Alexandra Calmy, and P. P. N Kumarasamy, Pierre deBeaudrap, Marco Vitoria, Meg Doherty, Wendy Stevens, George K Siberry (2014). "The future role of CD4 cell count for monitoring antiretroviral therapy."
6. Xian Liu, 2016. Methods and Applications of Longitudinal Data Analysis.
7. Diggle, P. J. H., Kung-Yee Liang, Scott L. Zeger (2002). "Analysis of Longitudinal.
8. Antonello Maruotti (2013).Covariance pattern models, Lecturer in Medical Statistics, S3RI and School of Mathematics University of Southampton.
9. Ware, N. M. L. J. H. (1982). Random-Effects Models for Longitudinal Data.
10. Verbeke, Geert & Molenberghs, Geert(2000). Linear Mixed Models for Longitudinal Data.New York, Springer-Verlag. V458
11. Rudolf,2016. Multiple Imputations in Generalized Linear Mixed Models: Ludwig Maximilians University Munich.
12. Departamento de Estat´ıstica Universidade de S˜ao Paulo in collaboration with Juv˜encio Santos Nobre, UFC(2011). Residual Analysis for linear mixed models
13. Kebadu Tadesse, 2016. Modeling CD4 + Cell Counts of HIV-Positive Patients Following Antiretroviral Therapy (ART): A Case of Yekatit 12 Hospital, Addis AbebaFrom university
14. Gezie, L. (2016)Predictors of CD4 count over time among HIV patients initiated ART in FelegeHiwot Referral Hospital.
15. Ayalu A. Reda ,SibhatuBiadgilign , AmareDeribew , BetemariamGebre and KebedeDeribe (2013). Predictors of Change in CD4 Lymphocyte Count and Weight among HIV Infected Patients on Anti-Retroviral Treatment in Ethiopia: A Retrospective Longitudinal Study.
16. Mihiretu M. Kebede,Desalegn T. ZegeyeandBerihun M. Zeleke (2014).Predictors of CD4 Count Changes after Initiation of Antiretroviral Treatment in University of Gondar Hospital, Gondar in Ethiopia.
17. Belay Desyebelew (June, 2017). Application of Longitudinal Count Data Models to Progression of CD4 Count: A Case of Debre Markos Referral Hospital from Addis AbabaUniversity.

How to cite this article: Tadege KY, Jegnaw T, Yesuf KM. Determinants of CD4 Cell Count among Antiretroviral Therapy (ART) Attendant HIV Positive Adults Using Longitudinal Data Analysis in University of Gondar Referral Hospital. Asian J. Med. Res. 2019;8(4):ME18-ME28. DOI: dx.doi.org/10.21276/ajmr.2019.8.4.ME4

Source of Support: Nil, **Conflict of Interest:** None declared.