International Journal of Multidisciplinary Research and Modern Education (IJMRME)
Impact Factor: 7.315, ISSN (Online): 2454 - 6119
(www.rdmodernresearch.org) Volume 6, Issue 1, 2020

MODELLING THE CD4+ COUNT PROGRESSION OF HIV+ PATIENTS USING GENERALIZED ESTIMATING EQUATIONS: A CASE OF DEBRE BERHAN REFERRAL HOSPITAL, ETHIOPIA

Abere Wondimu Kassie* & Damenech Syum G / Yohanes**
Lecturer in Statistics, Department of Statistics, College of Natural and Computational Sciences, Debre Berhan University, Debre Berhan, Ethiopia

Cite This Article: Abere Wondimu Kassie & Damenech Syum G / Yohanes, "Modelling the CD4+ Count Progression of HIV+ Patients Using Generalized Estimating Equations: A Case of Debre Berhan Referral Hospital, Ethiopia", International Journal of Multidisciplinary Research and Modern Education, Volume 6, Issue 1, Page Number 43-52, 2020.

Abstract:

CD4+ count is used to assess the clinical status of HIV-infected patients and making informed decisions about the initiation of Antiretroviral Therapy. To determine the progression of patient characteristics, the CD4+ count should be measured repeatedly per individual who used to evaluate the evolution of CD4+ count over time from baseline. Thus, the main objective of the study was modeling the progression of CD4+ counts of HIV^+ patients under ART in Debre Berhan Referral Hospital, Ethiopia. A retrospective cohort study is used; and out of a population of HIV-patients who were taking ART in the hospital during September, 2013 to February, 2019, data on 322 patients was taken to the study from which 225 (69.88%) were females and 97(30.12%) were males. Generalized Estimating Equations model is used to analyze the non-independent data measured for each patients for specified period of time. The result of this study showed that CD4+ count patients improved from average baseline, 335.7 to 408.6 after all follow-up times. The average CD4+ count of female patients was changed 0.5793 squared times than average CD4+ count of male patients (β = 0.5793 , p-value = 0.02618). HIV+ patients' age had negatively significant effect on the average progression of CD4+ count(β = -0.0321, p-value=0.0139). Its negative estimate indicated that the average CD4+ count older patients reduced by 0.0321 squared times than younger individuals. The predictors of patients' baseline CD4+count, sex, age, functional status, marital status, treatment regimen class, WHO clinical stage and follow up time found to be significant association with the progression of CD4+ count. HIV+ patients who had lowest initial CD4+ count need special attention as patients with higher baseline CD4 count tend to have higher rate of change over time than those with lower baseline CD4+ count.

Key Words: AIDS, Antiretroviral Therapy, CD4+ count, HIV+ & Generalized Estimating Equations

1. Introduction:

CD4 count is a blood test to check the amount of CD4 cells in the body. It is used in assessing the clinical status of HIV-infected patients, in making informed decisions regarding the initiation of antiretroviral therapy (ART).Measuring CD4 count is a strong predictor of progression to Human Immunodeficiency Syndrome (AIDS), as well as a means of monitoring the success of such antiretroviral therapy. The possible increase or decrease in CD4+ counts are directly related to HIV replication. Low CD4 counts are associated with a greater risk of patients living with HIV developing opportunistic infections, which may then progress to advanced disease and death (Langford et al., 2007 and Hoffman et al., 2010). The CD4+ count is also used to decide when to start antiretroviral therapy. Initiation on Highly Active Antiretroviral Therapy (HAART) at higher CD4+ counts has been demonstrated to "the risk of death, opportunistic infections and non HIV related co-morbidities" (Lichtenstein et al, 2008). In a healthy adult, a normal CD4 count can vary enormously (by population, age group, etc.) but is typically around 500 to 1500 cells per cubic milliliter of blood (mm^3). When it falls below 200, however, then the disease is technically classified as AIDS.

A number of attempts have been made to determine the nature of progression in CD4+T-cells after HIV infection. A study by (Tekele et al, 2016) found the average CD4+ count increases after patients initiated to ART program (the disease rate declines) and the progression depends on patient's baseline socio-demographic characteristics. A study on modelling patients CD4+ count using linear mixed model concluded that the mean rates of increment of CD4 counts for patients with ambulatory/bedridden and working baseline functional status were 17.4 and 30.6 cells/mm3 per year, respectively. After adjusting for other variables, for each additional baseline CD4 count, the gain in CD4 count during treatment was 0.818 cells/mm3 (p value <0.001). Patient's age and baseline functional status were also statistically significantly associated with CD4 count (Gezie et. al, 2016).

A study conducted to identify Predictors of CD4 Count Changes after Initiation of Antiretroviral Treatment conducted in University of Gondar Hospital, Gondar in Ethiopia (Kebede et al., 2015). The result of this study showed, the median CD4 count has increased from 139 cells/ml at the initiation of ART to 356 cells/ul at the most recent visit. A median CD4 count change of 208 cells/microliter was observed after 194.4 weeks on ART. And finally, Age when starting ART, educational status, marital status, WHO clinical staging, baseline hemoglobin level, baseline CD4 count, ART adherence status, functional status, and recent follow up CD4 are significant predictors of CD4 count change.

In order to study the progression of HIV infection, the CD4+T-cells should be measured repeatedly per individual known as longitudinal study. Longitudinal studies are defined as studies in which the outcome variable is repeatedly measured; that is the outcome variable is measured in the same individual on several different occasions. In longitudinal studies the observations of one individual over time are not independent of each other, and therefore it is necessary to apply special statistical techniques, which take into account the fact that the repeated observations of each individual are correlated (Twisk, 2003).

One of the approaches for extending generalized linear models to longitudinal data that leads to a class of regression models is Generalized Estimating Equations (GEE) models. The Generalized Estimating Equations approach projected (Liang and Zeger, 1986) is a class of estimating equations which take into account the correlation arising due to a longitudinal study design, to increase efficiency of standard error estimates. The GEE is quasi likelihood based and can be used for continuous as well as for discrete outcome. The GEE method is a multivariate generalization of quasi-likelihood, and this method is mainly proposed for marginal modeling with GLM. It avoids the use of multivariate distribution by assuming a functional form for marginal distribution at each time, making it useful for non-Gaussian outcomes. The generalized estimating equations deal with the correlation caused by collecting numerous samples from each individual via adjusting the standard error to compensate for the lack of independence among samples.

Statement of the Problem:

HIV infected patients have advised to start ART in order to reduce AIDS related morbidity and mortality by increasing their CD4+ count and to improve their quality of life. Researchers have carried out studies on the change of CD4+cell counts of HIV patients. But, most of these studies have look at change from the cross sectional point of view without considering how the CD4+cell counts evolve over time after patients initiated to ART and considering possible correlation among successive CD4+ counts as it were independent. This may lead to incorrect inferences concerning regression parameters due to inappropriate estimated standard errors and inefficient estimators. For instance, a study on CD4+ count trends after commencement of ART among HIV infected patients in Northern Ethiopia was conducted by Asfaw et al. (2015). The study used logistic regression and this might be used to identify factors associated with the changes in CD4+ count. Thus, this study tried to examine the progression of the CD4 count over time using Generalized Estimating Equations model which incorporates correlations come from repeated measurements among same subjects.

Objectives of the Study:

The main aim of the study was modeling the progression of CD4+ counts of HIV-positive patients under ART from September 11, 2013 up to February 8, 2019 in Debre Berhan Referral Hospital. The study had also the following specific objectives

- To model the rate of change in CD4 count and estimate average time of progression after start of ART
- To identify determinant factors associated with rate of decline or increase after use of ART
- To characterize the degree of correlation across patients in the rate of progression of CD4+ count among specified point of times.

2. Material and Methods:

Data and Variables:

The data for this study obtained from Debre Berhan Referral Hospital (DBRH), ART clinic by extracting from HIV-patient cards. This retrospective cohort study has used data of HIV/AIDS patients who initiated on Antiretroviral Therapy in the ART clinic of DBRH, Debre Berhan, Ethiopia, during the period of September 11, 2013 up to February 8, 2019. The study population was HIV-positive adults whose age 16 years old and above initiated on ART treatment in the hospital. All patients who have initiated to ART and measured their CD4+ counts at least two times, including the baseline and those who started first line ART regimen class were included in the study population. Patients, whose age below16 years old and those who started ART before and after the start of the study were excluded in the study and CD4+ counts permm^3 of blood are taking approximately in every six months.

The outcome variable for this study is CD4+ count for each individual measured in every six months interval. The CD4+ count is a continuous variable that is measured on individual patients and expressed as cells per cubic millimeter (cells/mm^3) of blood (Diggle et al, 2002).And the independent variables are baseline age of

patients, baseline CD4+ count, observation time (in months), sex of the patients, marital status of patients at baseline, WHO clinical stage, regimen class of patients, baseline body mass index (BMI), functional status and educational level of patients.

Table 2.1: Description of the study variables and codes in the analysis

Number	Variables	Value 'Code'
1	Sex of the patients	0 = Male,1 = Female
2	Age of the patients(years)	Continuous
3	Marital status	0 = Married, 1 = never married, 2 = divorce, 3 = others
4	Body mass index(kg/m2)	continuous
5	Base line CD4+ count	count
6	CD4+ count at each follow up	count
7	WHO clinical stage	1 = stage I, 2 = stage II, 3= stage III, 4 = stage IV
8	Time(in month)	continuous
9	Regimen class of patients	0 = AZT-3TC-EFV, 1 = AZT-3TC-NVP, 2 = TDF-3TC-NVP, 3 = TDF-3TC-EFV, 4 = others
10	Functional status of patients	0 = Working, 1 = Ambulatory, 2 = Bedridden
11	level of education	0 = No Education, 1 = Primary, 2 = Secondary, 3 = Tertiary

Exploratory Data Analysis:

Longitudinal data analysis contains exploratory analysis (EDA) and confirmatory analysis. EDA is the detective work to make data visualized in graphical presentation before confirmatory analysis to check evidence in the data using hypothesis. Thus, individual profiles plot were used to observe subject specific evolution over time and to decide on the random effects to be included in the model. To choose the fixed effects of the model mean profile plot is explored and to study the possible differences between groups, plotting the mean profiles for each subgroup separately is done. Finally, variance structure is plotted to explore variations of CD4+ count over time from baseline values.

Generalized Estimating Equations (GEEs):

The Generalized Estimating Equations were introduced by Liang and Zeger (1986) as a method of dealing for the correlation among responses. The GEE models are termed marginal models, and they model the regression of response on explanatory and the within-subject dependency. Marginal models specify a generalized linear model for the longitudinal responses at each occasion but also include a model for the within-subject association among the responses. In particular, the regression parameters in the marginal model have so-called population-averaged interpretations.

Let the vector of measurements on the i[th] subject be $Y_i = [Y_{i1}, \dots, Y_{ini}]'$ with corresponding vector of means $\boldsymbol{\mu_i} = [\mu_{i1}, \dots, \mu_{ini}]'$, where Y_{ij}, j=1, …,n_i, i=1, …,N represent the j^{th} measurements on the i^{th}subject and let V_i be an estimate of the covariance matrix of Y_i. The Y_{ij} (CD4 count) is a continuous response and it is of interest to relate changes in mean response over time to the covariates. Thus, the GEE requires the following specifications for a marginal model.

The mean of CD4 count is related to covariates by an identity link function,

$$\mu_{ij} = \eta_{ij} = X_{ij}^{'}\beta \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (1)$$

Where$\mu_{ij} = E(Y_{ij}/X_{ij})$ is a conditional expectation of the response (CD4 count) variable and β is a $p \times 1$ vector of regression parameter.

Because of the CD4+ count is a continuous, the variance of eachY_{ij}, given the effect of covariates does not depend on the mean response, as $var(Y_{ij}) = \phi v(\mu_{ij}) = \phi$, where the variance function $v(\mu_{ij}) = 1$ and ϕ is scale parameter that may be known or to be estimated.

Correlation among repeated measurements is a function of the means, μ_{ij}, and a set of parameters, α, which characterize the within-subject correlation and need to be estimated. The "working" covariance matrix for Y_i is given by

$$V_i = \emptyset A_i^{1/2} Ri(\alpha) A_i^{1/2} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2)$$

Where, $Ri(\alpha)$ is an$_i \times n_i$ "working" correlation matrix that is fully specified by the vector of parameters α (correlation parameters). A_i is a $n_i \times n_i$ diagonal matrix with $v(\mu_{ij})$ as the j[th] diagonal element and $\emptyset$ is the dispersion parameter.

Working Correlation Structure:

The generalized estimating equations method does not require specification of the distributional assumptions of the repeated responses, but requires specification of the covariance structure of the outcome variable. The within-subject associations or correlations among the repeated measures are taken into account by defined as "working" correlation structure and incorporating that structure into the estimation. Thus, independent working correlation, exchangeable, first autoregressive correlation and unstructured working correlation were proposed and then unstructured correlation was used after comparing AIC and BIC values of these working correlation structures.

Quasi-Likelihood in GEE Model:

The Quasi-likelihood used in GEE based on (Zeger & Liang, 1986) is a method for regression that requires the specification of relationships between mean response and covariates, mean response and variance. The GEE for β estimating is an extension of the independence estimating equation to correlated data and is given by

$$U(\beta;\alpha,\emptyset)=\sum_i^N\left(\frac{\partial\mu_i}{\partial\beta_j}\right)^T\left(A_i^{\frac{1}{2}}Ri(\alpha)A_i^{\frac{1}{2}}\emptyset\right)^{-1}\left(Y_i-\mu i(\beta)\right)=0\ldots\ldots\ldots\ldots\ldots(3)$$

Which is known as quasi score, where, $\mu_i(\beta)$ is the fitted mean, A_i =diag {var (Y_{ij})}, $Ri(\alpha)$ is the correlation matrix of Y_i and $\emptyset$ is the over-dispersion parameter.

The variance of the GEE estimator of β can be made robust to the misspecification of correlation structure (in a sense that it provides valid standard errors when correlation structure is not correct) by incorporating the empirical variance estimator, called "sandwich" covariance estimator (Diggle et al. 2002).The quasi-likelihood information criterion' which is counterpart to the AIC is used in model selection process.

3. Results:

Among the total number of patients initiated on ART within specified period those were fulfilling inclusion criteria and included in the study were 322 subjects. And thus each 322 patients followed for six occasions with six month interval, and then we have 1932 observations.

Descriptive Summary:

Table 3.1: Socio-Demographic characteristics of patients on ART in Debre Berhan referral hospital

Variables		Frequency	Percent
Gender	Male	97	30.12
	Female	225	69.88
Marital Status	Married	155	48.14
	Single	66	20.50
	Divorce	45	13.98
	Widowed and separated	56	17.39
Educational Level	No education	90	27.95
	Primary	132	40.99
	Secondary	70	21.74
	Tertiary	30	9.32

From Table 3.1, of the total 322 patients who were included in the study, 225 (69.88%) were females and the remaining 97 (30.12%) were males. Regarding the marital status composition of patients, of the total of 322 HIV-infected patients in the study 155 (48.14%) were married, 66 (20.50%) were never married, 45 (13.98%) were divorced and 56(17.39%) were others respectively. When we look at the educational level categories of the HIV-positive patients, 90(27.95%) were no educated, 132 (40.99%) were primary school educated and the remaining 31.06 % were secondary and tertiary educated.

Table 3.2: Baseline Clinical characteristics of patients on ART in Debre Berhan referral hospital

Variables		Frequency	Percent
Functional Status	Working	296	91.93
	Ambulatory	23	7.14
	Bedridden	3	0.93
WHO Clinical Stage	Stage I	141	43.79
	Stage II	72	22.36
	Stage III	98	30.43
	Stage IV	11	3.42
Regimen Class	AZT-3TC-EFV	22	6.83
	AZT-3TC-NVP	36	11.18
	TDF-3TC-NVP	27	8.39

	TDF-3TC-EFV	215	66.77
	Others	22	6.83

In Table 3.2, out of the total HIV-positive patients who have included in the study 296 (91.93%) were in 'working' functional status, whereas the less than one-tenth remaining, 8.07% were in ambulatory and bedridden status. Table 3.2 also depicts, 141 (43.79%) of the study patients were in WHO clinical stage I, 72(22.36%) were found in stage II, 98(30.43%) were in stage III and only 3.42% of patents were in stage IV. Regarding the composition of baseline ART regimen classes, 22(6.83%) of the patients were started by AZT-3TC-EFV class of ART regimen, 36(11.18%) were by AZT-3TC-NVP ART regimen type, 27(8.39%) were by TDF-3TC-NVP, 215(66.77%) were by TDF-3TC-EFV class of ART regimen and the remaining 22(6.83%) were by other different ART regimen classes.

Table 3.3: Descriptive Statistics for continuous covariates

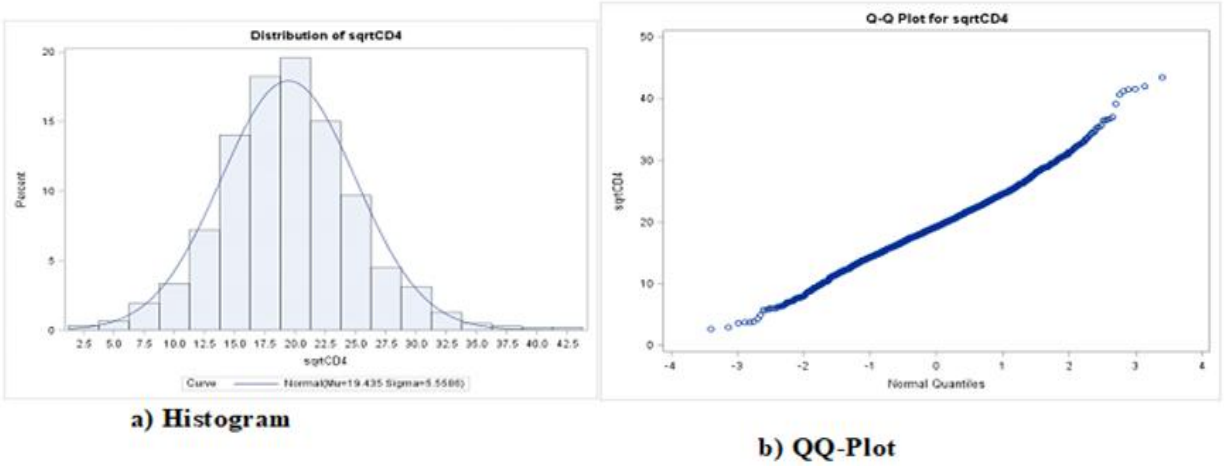
Variable	N	Min	Max	Mean	Median	Std. Dev
Age	1932	16	66	36.8	35	10.12
BMI	1932	17.26	26.92	21.34	21.17	2.78
BaseCD4	1932	7	1344	335.7	310.5	222.8
CD4	1889	7	1894	408.6	370.0	230.89

From table 3.3, the baseline age of patients ranged from 16 to 66 with mean 36.8 and standard deviation 10.12; the mean baseline body mass index of patients was 21.34 with standard deviation 2.78. The range of baseline CD4+ count was 7 to 1344 and the average CD4+ count 335.7 with standard deviation 222.8; and the average CD4+ count after all follow-up time was 408.6 (230.89).

Table 3.4: Descriptive Statistics for sqrtCD4 count at each time

Time (Month)	N	Min	Max	Mean	Median	Std. Dev.
0	322	2.646	36.661	17.239	17.621	6.215
6	322	3.742	41.497	18.974	18.775	5.749
12	322	3.464	37.027	19.578	19.079	5.343
18	314	2.828	43.520	20.046	19.685	5.028
24	308	7.616	42.011	20.304	19.557	5.141
30	301	6.000	41.641	20.597	20.542	5.081

In a similar way, mean with the corresponding standard deviation at each time points with respective sample sizes for square root CD4+ count was shown in Table 3.4. Thus, the average number of CD4 count at baseline was 17.239. There was an increasing up to twelfth months starting from baseline and then the rate of increment become constant in between twenty and twenty-one. However, look at the standard deviations, from six months up to 30 months there was similar variation among the measurement times where higher variation in square root of CD4+ count was observed at baseline time point.



a) Histogram b) QQ-Plot

Figure 3.1: Test of Normality for sqrtCD4+ count data

Figure 3.1 depicts CD4+ count was check its normality assumption using histogram (Figure 3.1a) and QQ-plot (Figure 3.1b). Since the actual CD4+ count was not satisfied normality and then it need to be transformed. Thus, logarithmic and square root transformation were carried out; but logarithmic transformation doesn't make

visible change of normality, however square root method made it approximately normally distributed as shown Figure 3.1.

Table 3.4: Test of Normality of CD4+ count

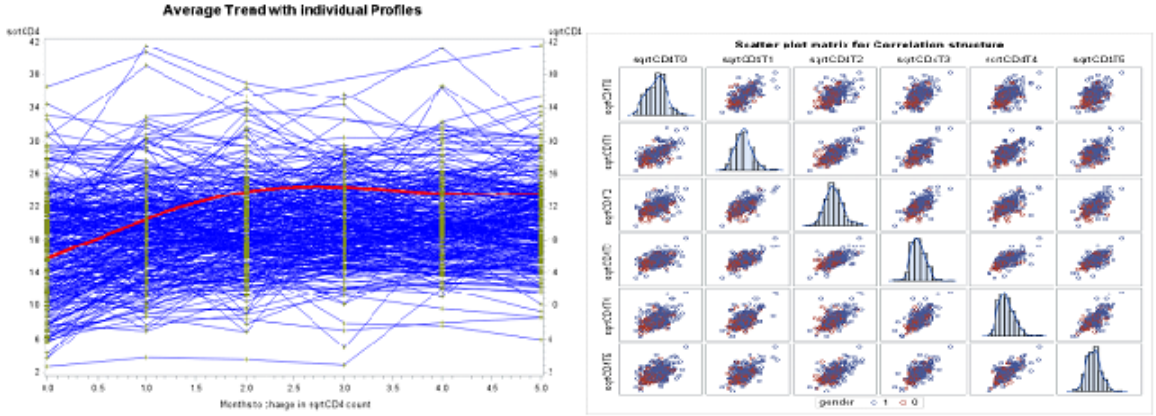
	Actual CD4+		Log transformed CD4		Sqrt CD4 transformed	
Test	Statistic	p-value	Statistic	p-value	statistic	p-value
Shapiro-Wilk	0.91258	<0.0001**	0.938595	<0.0001**	0.991197	0.432
Kolmogorov-Smirnov	0.084069	<0.0100**	0.074793	<0.0100**	0.030059	0.273

*P-value <0.05, **P-value <0.01

In Table 3.2 CD4+ count data is also normal after taking square root transformation as its insignificant p-values in Shaprio-Wilk and Kolmogorov-Smirnov tests.

Exploratory Analysis:

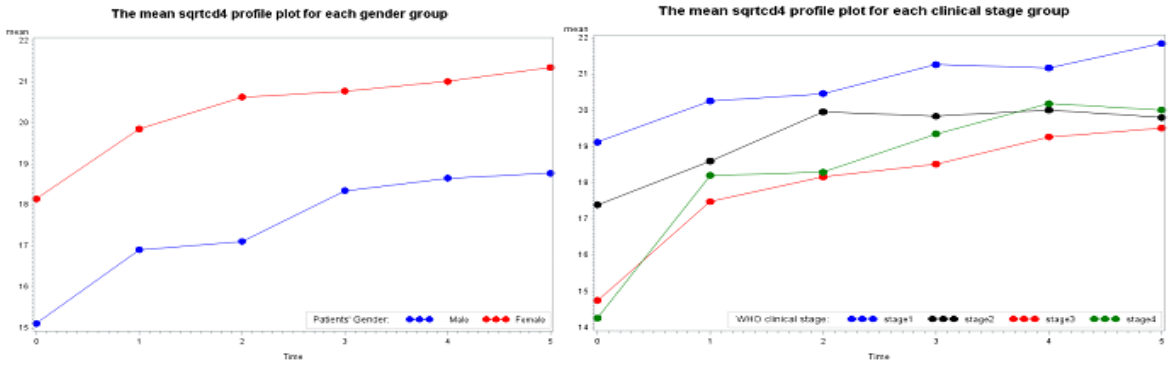
The smoothed cure in Figure 3.4 (a) suggested the average square root CD4+ count has linear relationship with time. It increases up to twelfth months but the rate of increment become lower after twelfth months' time point and then it is fairly constant. The other important plotting analyses in longitudinal data was correlation structure in response among consecutive measurement occasions and variance structure between subjects in their responses. Thus, in Figure 3.2(b) the pair wise scatter plot showed that positive and linear relationship between square root CD4+ counts across follow up time periods.



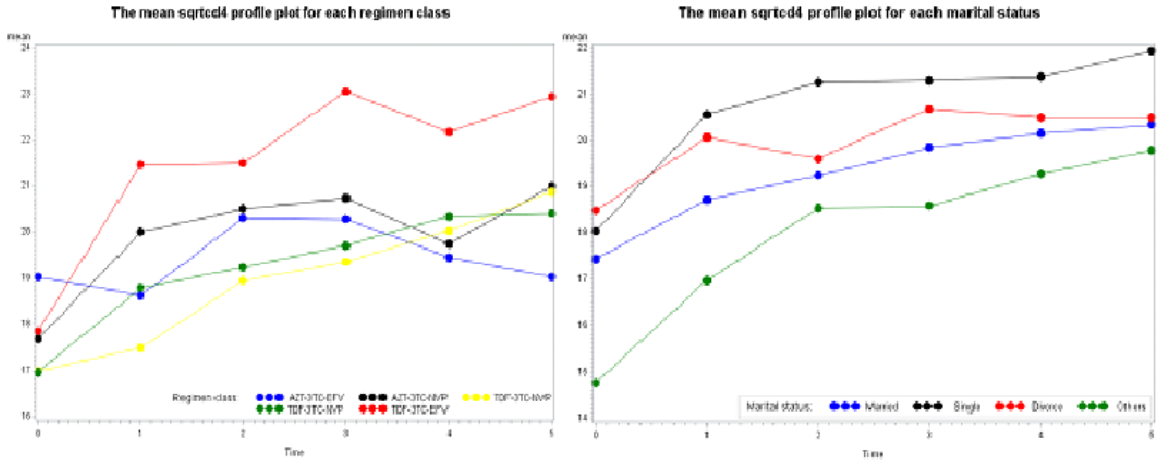
(a) Mean Profile plot sqrtCD4+ count **(b) Pairwise Scatter plot for sqrtCD4+ count**

Figure 3.2: Mean Profile and Pairwise Scatter plot for sqrtCD4+ count

In Figure 3.3, we observed the heterogeneity among patients in the rate of progression of CD4+ count between categories of gender, WHO clinical stages, ART regimens, marital status and functional status. Female patients have higher change in CD4+ count than males overtime (Figure 3.3(a)). Here the effect of gender on evolution depends on the patients' baseline CD4+ count values, which implies patients with high baseline CD4+ would have higher rate of change over time. The plot for each clinical stages in Figure 3.3(b) verified the HIV-positive patients in stage one have higher increment rate in their CD4+ count as compared to patients in other stages.



(a). Mean profile plot by Gender (b). Mean profile plot by WHO Clinical Stages



(c). Mean profile plot by Regimen Class **(d). Mean profile plot by Marital Status**

Figure 3. 3. Average profile plots for each categorical variable

Generalized Estimating Equations (GEEs):

Unstructured working correlation structure was selected based on its smallest quasi likelihood information criteria in empirical GEE parameter estimate as shown in Table 3.5 and its smallest standard error as compared to other correlation structures.

Table 3.5: Selection of correlation structure

Model Fit Criteria	Independent	Exchangeable	AR(1)	Unstructured
QIC	1969.64	1971.46	1954.16	1927.70
QI Cu	1914.00	1914.00	1914.00	1914.00

We started by fitting univariate marginal models to explore the relationship between square root of CD4+count and each of the covariates. For this analyses sex, age, marital status, baseline CD4+ count, regimen class, functional status, educational level, time and WHO clinical stage were considered as significant variables at 0.25 level of significance.

The GEE Parameter Estimates:

Table 3.6 showed the generalized estimating equations parameter estimates of CD4 count of patients. The change in average CD4+ count of female patients' was 0.5793 squared times higher than average change in CD4+ count of male patients (β=0.5793 ,p-value=0.02618) i.e square of the estimate higher because the data in the analysis used by square root transformation. Age had significant effect on average change CD4+ count of HIV-positive patients (β=-0.0321, p-value=0.0139). Its negative coefficient indicated that the average change CD4+ count for older patients reduced by 0.0321 squared times than younger patients.

The change in average CD4+ count of actively working patients' were 1.9701 squared times lower than patients with bedridden status (β = -1.9701, p-value = 0.0004). While the coefficient of ambulatory patients is -2.3820(0.7336) with its p-value 0.0012. Similarly, patients with ambulatory status were in 2.3820 squared times lower than patients with bedridden status in average CD4+ count progression (β = -2.3820, p-value = 0.0012), when the effect of other variables fixed. Patients' average CD4+ count in AZT-3TC-EFV first line regimen class were 1.9839 times changed than patients with ABC-3TC-EFV/NVP baseline regimen class(β =1.4085, p value = 0.0462). HIV+ patients who were never married was 0.7756 squared times higher in average CD4+ count progression than those who were widowed and separated patients (β = 0.7756, p-value = 0.0352). The estimate for baseline CD4+ count is (β = 0.0179 , p-value,<0.0001), which indicates when baseline CD4+ count higher a cell per mm^3, the average CD4+ count of patients increases in 0.0179 squared times overtime. The average CD4+ count of patients in clinical stage I is changed in 0.0155 squared times higher as compared to patients in clinical stage IV (β=0.0155, p-value, 0.0487). The follow –up time itself is significantly associated with the progression of patients' average CD4+count as shown in Table 3.6.

Table 3.6: Empirical GEE parameter estimates

Variables	Estimate(β)	Std. Error	Z	Pr > |Z|
Intercept	13.0345	1.7741	7.35	<.0001**
Sex(female)	0.5793	0.2618	2.21	0.0269*

Sex(male) Reference	.	.	.	.
Age	-0.0321	0.0131	-2.46	0.0139*
Functional Status(working)	-1.9701	0.5554	-3.55	0.0004*
Functional Status (ambulatory)	-2.3820	0.7336	-3.25	0.0012*
Functional Status (bedridden) Reference	.	.	.	.
BMI	0.0615	0.0443	1.39	0.1652
Regimen Class(AZT-3TC-EFV)	1.4085	0.7667	-1.84	0.0462*
Regimen Class (AZT-3TC-NVP)	0.1255	0.7089	-0.18	0.8595
Regimen Class (TDF-3TC-NVP)	0.5030	0.7050	-0.71	0.4755
Regimen Class (TDF-3TC-EFV)	0.9349	0.5951	-1.57	0.1162
Regimen Class (Others) Reference	.	.	.	.
Educational Level (no educated)	0.2724	0.4107	0.66	0.5072
Educational Level (primary)	0.5358	0.3781	1.42	0.1564
Educational Level (secondary)	0.1800	0.4098	0.44	0.6606
Educational Level (tertiary Reference	.	.	.	.
Marital Status(married)	0.0918	0.3626	0.25	0.8002
Marital Status (no married)	0.7756	0.4506	1.72	0.0352*
Marital Status (divorced)	0.4981	0.4453	1.12	0.2633
Marital Status (widowed and separated) Reference	.	.	.	.
Baseline CD4+ count	0.0179	0.0007	24.57	<.0001**
WHO Clinical Stages(stage I)	0.0155	0.4940	-0.03	0.0487*
WHO Clinical Stages (stage II)	0.4087	0.5395	-0.76	0.9750
WHO Clinical Stages (stage III)	0.3272	0.4833	-0.68	0.4984
WHO Clinical Stages (stage IV) Reference	.	.	.	.
Time(6^{th}month)	1.7278	0.2797	6.18	<.0001**
Time(12^{th}month)	2.3288	0.3239	7.19	<.0001**
Time(18^{th}month)	2.7608	0.3227	8.56	<.0001**
Time(24^{th}month)	3.0511	0.3368	9.06	<.0001**
Time(30^{th}month)	3.3614	0.3431	9.80	<.0001**
Time(baseline)Reference	.	.	.	.

*P-value <0.05, **P-value <0.0001

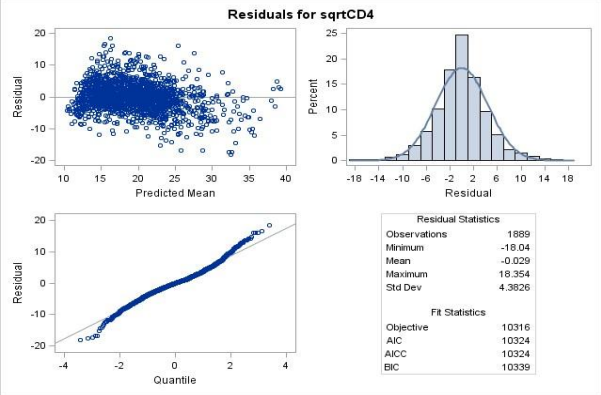
Thus, the GEE model could be stated as:

$$\sqrt{\text{CD4ij}} = \beta_0 + \beta_1 Sex_i + \beta_2 Age_i + \beta_3 FuncSt_{0i} + \beta_4 FuncSt_{1i} + \beta_5 MarSt_{1i} + \beta_6 RegCl_{0i} + \beta_7 WHOSt_{1i} + \beta_9 BaseCD4 + \beta_{11} Time_{ij}$$

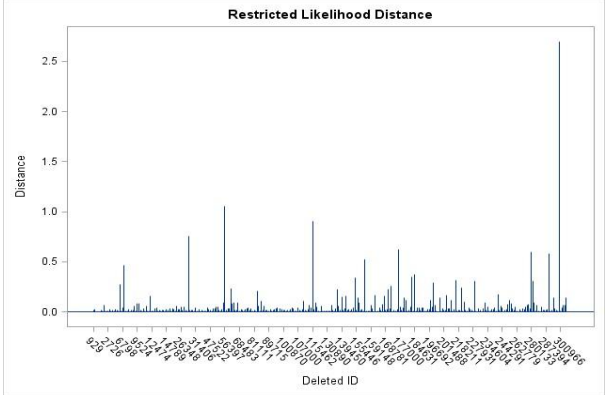
Where, $FuncSt_{0i}$, $FuncSt_{1i}$, $MarSt_{1i}$, $WHOSt_{2i}$, $RegCl_{0i}$ and Time are functional status: actively working and ambulatory; none married; WHO clinical stage I; the first line ART regimen classes: AZT-3TC-EFV and finally follow up time for each occasion.

Model Adequacy Checking:

In Figure 3.4(a), the marginal residuals versus predicted value plots have a random behavior as scattered around reference line that linearity is satisfied. Figure 3.4(b) showed an overall influence, since near by the last subject stands out as most overall influential compared to the other patients on the CD4 count progression and the CD4+ count of some middle patients also seemed to be influential observation.



(a) Residual plot for Marginal effect



(b) overall influential plot

Figure 3.4: Model Checking

4. Discussion:

The study was aimed to investigate the progression of CD4+ count among patients under ART in Debre Berhan Referral Hospital. It was also aimed in determining whether the change in CD4+ count differs among selected patient characteristics. On the basis of those objectives, exploratory analyses which provide some initial basis followed by subsequent model-based results were conducted. The normality of data had been tested using Q-Q plots, histogram and Shapiro-Wilk tests before going to detail analyses. As these tests indicated, CD4+ count needed to be transformed and then after log and square root transformations carried out, CD4+ count was found to be normal in square root transformation**.**

From exploratory analysis, the mean profile plot suggested that average CD4+ count increases up to twelfth month and then fairly shows constant increment rate over time. This supports the result of (Abrogoua et al, 2012**)** which identified that after patients initiated to ART their CD4+ count increases up to third time points from baseline. There is a positive and linear relationship between square root CD4+ counts within each individual across follow up time periods as we observed from pair wise scatter plot matrix in exploratory analysis. Whereas in descriptive summary, the average baseline CD4+ count after patients initiated to ART was 335.7 and changed to 408.6 overtime. When we look the average CD4 count at each follow-up time, there was an increasing progression up to twelfth months starting from baseline and then the rate of increment become constant over the last times.

The unstructured working correlation structure was selected after considering all possible correlation structures based on minimum of QIC and smaller standard error during model comparison process in GEE. The result of the study showed, female HIV-positive patients under ART had higher mean change CD4+ count compared to males. Here the effect of gender on progression depends on the patients' baseline CD4+ count that patients with high initial CD4+ would have higher rate of change over time. However, the result contrary with study (Gezie et al., 2016), which found that when adjusted for other variables, sex was not found to be significant predictor of CD4+ count. In addition, patients with higher baseline CD4 count tend to have higher rate of change over time than those with lower initial CD4+ count.

Baseline age was negatively associated with CD4+ count which implied the rate of progression in CD4+ count for older patients was lower than that of younger ones. The finding of the study also showed never married patients were 0.7756 squared times higher mean change in CD4+ count as compared to those with widowed and separated patients over time (β=0.7756, p-value= 0.0352). The estimate for baseline CD4+ count is (β = 0.0179, p-value, <0.0001), which indicates when baseline CD4+ count higher a cell per mm^3, the average CD4+ count of patients increases in 0.0179 squared times overtime.

The predictors of patients' baseline CD4+count, sex, age, functional status, marital status, treatment regimen class, WHO clinical stage and follow up time found to be significant association with the progression of CD4+ count. Whereas educational level doesn't had significant effect on average progression of CD4+ count. This was in contrary with the findings of the study (Kebede et al., 2015) showed that baseline CD4 count, age, patient's functional status when commencing ART, advanced WHO clinical stages, educational status, marital status were significant predictors of CD4 count change.

5. Conclusion:

The study revealed, the average CD4+ count showed better increment after HIV-positive patients initiated to ART than that of patients CD4+ count at baseline. In the study we also observed, patients with higher baseline CD4+ count tend to have higher rate of change over time than those with lower initial CD4+ count. Females had higher CD4+ count progression as compared to patients who were males, the change depends on the patients'

baseline CD4+ count that females had better initial status. The rate of progression in CD4+ count for older patients was lower than that of younger ones that baseline age was negatively associated with CD4+ count.

From the result of the study heterogeneous rate of increment in CD4+ count had been observed among patients under different characteristics such as patients' marital status, functional status, WHO clinical stages etc. In the study patients' baseline CD4+count, sex, age, functional status, marital status, treatment regimen class, WHO clinical stage and follow up time found to be significant association with the progression of CD4+ count. HIV+ who had lowest initial CD4+ count need special attention as patients with higher baseline CD4 count tend to have higher rate of change over time than those with lower baseline CD4+ count.

6. References:

1. Abrogoua, D.P., Kablan, B.J., Kamenan, B.A.T., Aulagner, G., N'Guessan, K. and Zohoré, C., 2012. Assessment of the impact of adherence and other predictors during HAART on various CD4 cell responses in resource-limited settings. Patient preference and adherence, 6, p.227.
2. Asfaw, A., Ali, D., Tadele Eticha, A.A., Alemayehu, M. and Kindeya, F., 2015. CD4 cell count trends after commencement of antiretroviral therapy among HIV-infected patients in Tigray, Northern Ethiopia: a retrospective cross-sectional study. PloS one, 10(3).
3. Hoffman, J., Van Griensven, J., Colebunders, R. and McKellar, M., 2010. Role of the CD4 count in HIV management. HIV therapy, 4(1), pp.27-39.
4. Langford, S.E., Ananworanich, J. and Cooper, D.A., 2007. Predictors of disease progression in HIV infection: a review. AIDS research and therapy, 4(1), p.11
5. Diggle, P.J. et al. (2002). Analysis of Longitudinal Data Second edi., New York: Oxford UniversityPress.
6. Kebede, Mihiretu M., Desalegn T. Zegeye, and Berihun M. Zeleke. "Predictors of CD4 Count Changes after Initiation of Antiretroviral Treatment in University of Gondar Hospital, Gondar in Ethiopia." Clin Res HIV/AIDS 1, no. 2 (2015): 1-15.
7. Gezie, L.D., (2016). Predictors of CD4 count over time among HIV patients initiated ART in Felege Hiwot Referral Hospital, northwest Ethiopia: multilevel analysis. BMC research notes, 9(1), p.377.
8. Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.
9. Lichtenstein, K.A., Armon, C., Buchacz, K., Chmiel, J.S., Moorman, A.C., Wood, K.C., Holmberg, S.D., Brooks, J.T. and HIV Outpatient Study (HOPS) Investigators, 2008. Initiation of antiretroviral therapy at CD4 cell counts≥ 350 cells/mm3 does not increase incidence or risk of peripheral neuropathy, anemia, or renal insufficiency. JAIDS Journal of Acquired Immune Deficiency Syndromes, 47(1), pp.27-35.
10. Tekle G. (2016). Statistical Analysis of CD4+ Cell Counts progression of HIV-1- positive Patients enrolled in Antiretroviral Therapy, in Ethiopia. Biom Biostat Int J 3(1): 00057. DIO:10.15406.
11. Twisk, J.W.,(2013). Applied longitudinal data analysis for epidemiology: a practical guide. Cambridge university press, New York.