



eISSN: 2281-7824

<https://www.pagepressjournals.org/index.php/hls/index>

Publisher's Disclaimer. E-publishing ahead of print is increasingly important for the rapid dissemination of science. The **Early Access** service lets users access peer-reviewed articles well before print / regular issue publication, significantly reducing the time it takes for critical findings to reach the research community.

These articles are searchable and citable by their DOI (Digital Object Identifier).

The **Healthcare in Low-resource Settings** is, therefore, e-publishing PDF files of an early version of manuscripts that undergone a regular peer review and have been accepted for publication, but have not been through the typesetting, pagination and proofreading processes, which may lead to differences between this version and the final one.

The final version of the manuscript will then appear on a regular issue of the journal.

E-publishing of this PDF file has been approved by the authors.

Healthc Low-resour S 2026 [Online ahead of print]

To cite this Article:

Tadesse Cheire K. **Linear mixed models for evaluating CD4⁺ count progression and its predictors: a four-year longitudinal study at Yekatit 12 Referral Hospital, Ethiopia.** *Healthc Low-resour S* doi: 10.4081/hls.2026.14227

 ©The Author(s), 2026

Licensee [PAGEPress](#), Italy

Note: The publisher is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries should be directed to the corresponding author for the article.

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Submitted: 6 August 2025

Accepted: 23 September 2026

Early access: 23 October 2026

Linear mixed models for evaluating CD4⁺ count progression and its predictors: a four-year longitudinal study at Yekatit 12 Referral Hospital, Ethiopia

Kebadu Tadesse Cherie

Department of Statistics, Debre Markos University, Debre Markos, Ethiopia

Correspondence: Kebadu Tadesse Cherie, Department of Statistics, Debre Markos University, P.O.Box, 269, Debre Markos, Ethiopia. E-mail: kebadu.tadesse@dmu.edu.et

Key words: ART, CD4⁺, longitudinal data, missing data, linear mixed model, missing at random.

Abstract

Despite significant improvement in the use of antiretroviral therapy, HIV/AIDS remains to be one of the major public health challenges in Sub-Saharan African countries including Ethiopia. The main objective of the study was to identify factors associated with progression of CD4⁺ cell counts of HIV patients on ART. A longitudinal data was extracted from SMARTCARE of patients on ART from September 2009 to January 2016 in Yekatit 12 referral Hospital. The data was analyzed using linear mixed model taking correlation between CD4⁺ counts of patients into account. We included 217 patients of whom 65% were females and only 51.61% of patients were followed until the last follow up time. The missingness in the data was handled using direct likelihood and multiple imputation approaches. Results from the two methods were found to be comparable. Thus, baseline WHO clinical stage, time on treatment and the interaction effects of sex, age, baseline CD4⁺ count and regimen class with time were found significantly associated with progression of CD4⁺ count. We concluded that HIV/AIDS patients should be put on ART at early stages despite their baseline CD4⁺ cell count and treatments should be taken continuously to improve the progression of CD4⁺ count.

Introduction

The Human Immunodeficiency Virus (HIV) causes Acquired Immune-Deficiency Syndrome (AIDS), a condition in which the immune system begins to decline, exposing infected individuals to life-threatening opportunistic infections.¹ It is an epidemic that affects every part of the globe. Thus HIV/AIDS epidemic remains to be one of the major challenges of development in both developed and developing countries of the world. It affects peoples' welfare, undermining social cohesion and security, incapacitating global workforces, and diminishing social and economic productivity.² Ethiopia is among the most affected countries with HIV/AIDS in the world and the disease is one of the top ten causes of death in the country. The epidemic affects the population as the number of persons living with the disease, the number of new infections each year, the number of deaths related to HIV and the number of persons in need of ART is high in Ethiopia especially in urban areas of the country.³

As of 2024, Ethiopia has achieved a stable national HIV prevalence of 0.9%, though significant disparities persist between urban (2.9%) and rural (0.4%) settings, and between females (1.2%) and males (0.6%).³⁻⁵ While the country has made substantial progress toward the global 95-95-95 targets, identifying the drivers of immune recovery remains essential for optimizing long-term clinical outcomes. This study evaluates the longitudinal progression of CD4⁺ counts among patients on Antiretroviral Therapy (ART) at Yekatit 12 Referral Hospital to provide evidence-based insights into these regional dynamics.

HIV is associated with immune suppression which can be measured by the levels of T cell and has been shown to reduce levels of CD4⁺ and CD3⁺ counts in addition to other hematological values.⁴ Its infection leads to severe depletion of CD4⁺ counts in the gut- associated lymphoid tissue with subsequent reduced levels of circulating CD4⁺ lymphocytes in the peripheral blood.

CD4⁺ count is one of the laboratory markers used to track the progression of HIV to AIDS and is among predictors for risk of death and AIDS at the time of initiating therapy. CD4⁺ counts are reduced precipitously in acute HIV infection, but usually rebound over several weeks as HIV-specific CD8 T cells help to lower plasma viremia.

Significant depletion of CD4⁺ counts leads to opportunistic infections and mortality in untreated patients. ART involves taking a combination of ART in combination of HIV regimen. ART initiation time was used to be determined by the count of CD4⁺. ART cannot cure HIV, but it can help the patient live a longer, healthier life and reduce their risk of HIV transmission.

ART is the mainstay of treatment for those infected with HIV.⁶ Since its introduction in 1996, mortality and morbidity rates in HIV-infected individuals in countries with widespread access to ART have plummeted.⁷ The main effect of ART is to suppress viral replication, allowing the individual's immune system to recover and protecting him/her from progressing to AIDS and/or death.

While numerous studies in resource-limited settings have investigated factors influencing CD4⁺ count progression, many rely on classical statistical approaches that may not fully account for the longitudinal nature of the data. This can lead to statistical inefficiency, where the variability of the estimates is higher than necessary and longitudinal trends are underutilized.⁸ To address this, the current study utilizes linear mixed model, which enhances statistical efficiency by accounting for within-subject correlations and individual growth trajectories.⁹

Classical statistical models are not efficient for longitudinal CD4⁺ count data as the CD4⁺ counts are not expected to be independent within a patient over time. Linear mixed model is an appropriate model to assess the progression of CD4⁺ count and to identify associated risk factors as it accounts for within and between patients correlations.¹⁰

As it is the case in most longitudinal studies, missing data were observed in our studies. Hence, we applied statistical methods that could handle the missing values in the data. Direct likelihood and multiple imputations are among the methods used to handling missing data under the plausible assumption of missing at random.

The objective of the study was to assess the progression of CD4⁺ counts of HIV-positive patients following ART for a period of four years in Yekatit 12 Referral Hospital and to identify factors that affect the change in CD4⁺ cell count.

Materials and Methods

Data

The data was obtained from the ART case unit of Yekatit 12 referral Hospital. HIV positive patients who initiated ART between September 2009 and September 2012 and followed until January 2016 were considered. Longitudinal data of 217 HIV positive patients was extracted from the SMART CARE of patients' chart. Patients with at least three times CD4⁺ count measurements and those on first line regimens were included. The chart includes epidemiological, laboratory and clinical information of all HIV patients under ART follow-up and a detailed antiretroviral therapy history.

Variables in the study

The response variable was the number of CD4⁺ count of HIV positive patients. Explanatory variables that were expected to influence the progression of CD4⁺ count after ART initiation are: sex, age at ART initiation, marital status, baseline CD4⁺ count, baseline weight in kgs, regimen type (first line ART regimens), Adherence, WHO staging at ART initiation, functional status and time in months.

Data analysis

Exploratory data analysis

In longitudinal studies like this study, data exploration is a very helpful tool in the selection of appropriate models and understanding of the nature of the data. Examining the data for clues about the likely nature of the mean structure and to see how the mean profile changes over time is essential for specifying the functional form of the mean response of the model. It is also useful to choose the appropriate fixed-effects structure of the linear mixed model. Graphical methods were used to explore the magnitude of person-to person variability in outcomes over time and the correlation structures were studied through the correlation matrix and scatter plot matrix.

Variance covariance structures

Measurements made on the same subject are likely to be more similar than measurements made on different individuals. For the analysis to be valid, the co-variances among repeated measures must be modeled properly. The following four most commonly used covariance structures Compound Symmetry (CS), Toeplitz (TOEP), Unstructured (UN), and autoregressive (1) (AR (1)); were considered in this study.

Linear mixed model for longitudinal data

Patients' CD4⁺ count was measured every six months giving rise to a cluster of observations from each patient and the within subject CD4⁺ count are subject to biologic variation and measurement error. Linear mixed model with maximum likelihood and restricted maximum likelihood method of estimation was used to identify covariates that affect the progression of CD4⁺ cell counts of HIV positive patients on ART.⁹

Random coefficient models were also applied, as they are useful for highly unbalanced data with many repeated measurements per subject. In this model, the fixed effect parameter estimates represent the expected values of the population of intercept and slopes. The random effects for intercept represent the difference between the intercept for the i^{th} patient and the overall intercept. The random effects for slope represent the difference between the slope for the i^{th} patient and the overall slope. Random coefficient models also have a random effect for the within-subject variation. It is assumed that the random effects are independently and identically distributed random variables.¹¹

Missing data that may arise due to loss to follow up, refusal, attrition, measurement errors or simply ignorance of interviews are common in longitudinal studies. In this study, the missingness were mainly due to dropout, lost to follow up, transfer out and death. We used direct likelihood and multiple imputation approaches to handle the missingness in the CD4 count data under the most plausible Missing At Random (MAR) assumption.¹²

Model comparison was done using information criteria, Likelihood Ratio Test (LRT) and mixture chi-square distribution to select the best model. All the data analysis was done using SAS version 9.2 considering variables with p values under 0.05 to be significant.¹³

Results

Exploratory data analysis of the CD4⁺ count of HIV positive patients treated in the ART case unit of Yekatit 12 Hospital was carried out to understand the data structure and determine modeling approaches relevant to the data. The total number of patients included in this study was 217, of whom 141 (65%) were females and the remaining 76 (35%) were males.

Descriptive statistics of continuous variables

The average baseline age of the HIV positive patients is 39 with a standard deviation of 10.888; the average baseline CD4⁺ count is 264.3641 with a standard deviation of 210.93 and the mean baseline weight is 56.945Kg with a standard deviation of 12.376 (Table 1).

The mean CD4⁺ count of patient's increases with an increasing rate until the third visit, then increases slowly to visit six and increases at an increasing rate to visit seven to the maximum average 466.722 with a standard deviation of 233.172 at the last time point and follows almost a linear trend as depicted in Figure 1. The number of patients per visit was not constant for CD4⁺ count indicating that the data have both intermittent and dropout missingness.

Descriptive statistics of categorical covariates by time

The number of patients coming for CD4 measurement decreased from 217 at the baseline to 97 at visit seven. The number of female patients were greater than male patients at all follow-up times, where time of follow up indicates (0 = Baseline, 1 = Six Months, 2 = 12 Months, 3 = 18 Months, 4 = 24 Months, 5 = 30 Months, 6 = 36 Months, and 7 = 42 Months). Only 51.61% of the patients were followed until visit time seven (Month 42) inducing missingness of 48.39%.

Distribution of CD4⁺ count data

Shapiro-wilk test and q-q plots were used to assess the normality of the data, fundamental assumption of linear mixed model. Count data can be well approximated by a normal distribution when the number of the count is large. From the Shapiro-wilk tests of normality, we found that

the actual CD4⁺ count were not normal at all-time points as the test showed significant deviation from normality. Hence, square root transformation was chosen to normalize the data. This transformation is a well-established approach in HIV research¹⁴ as it stabilizes the variance and effectively normalizes the distribution of CD4⁺ counts, which are often right-indexed. All statistical inferences and p-values reported in the Results are based on this transformed scale.

Exploring individual profiles

Figure 2 presents individual profiles of patients. Plotting observed profiles over time helps to identify general trends within patients and may detect change over time by providing information about the variability at given times. Connecting the repeated measurements for each patient's CD4⁺ count over time shows a non-constant pattern common to most patients. These individual profiles also provide some information on between patient's CD4⁺ count variability and indicate change among patient's over time CD4⁺ count.

There is a fluctuation in CD4⁺ count after patients began ART and the variability seemed bigger at the beginning compared to at the end.

Associations between the response variable and continuous predictors (weight, age and baseline CD4⁺ count) were assessed using scatter plots of CD4⁺ count. The plots suggested strong correlation between baseline CD4⁺ count and CD4⁺ count at all follow up times. Weak negative correlation was observed between the baseline age and CD4⁺ count of patients. The correlation between CD4⁺ count and baseline weight seems weak at all-time points.

Exploring mean profiles

From the average profile of CD4⁺ counts of patients displayed in Figure 1, we observed an almost sharp linear increase up to visit time two and then goes by following some slight increase in mean CD4⁺ count up to the last time of follow up. As depicted in Figure 3, the mean CD4⁺ count profiles of female patients seems higher than that of males at all visit times except the first visit. Generally, the mean CD4⁺ count of patients on ART seems to show an increasing trend for

all covariates with some difference on the amount of variation among categories of the covariates.

Exploring correlation structure and scatter plot of correlation matrix

The correlation matrix can be used to show the correlation between repeated measurements. In this study, we used the available data for computing the correlation matrix due to missingness. The correlation matrix shown revealed a positive correlation between any two repeated measurements except for the last three time points when the correlation decreased over time (Table 2).

Pair wise scatter plots were used for exploring the correlation between any two repeated measurements of CD4⁺ count of patients and it appears that, there was a positive and linear relationship between CD4⁺ counts at different time points (Figure 4).

Linear mixed models

Univariate marginal models

We started by fitting univariate marginal models to explore the relationship between the square root of CD4⁺ count and each of the covariates. In this study, six categorical and three continuous variables were considered. Variables significant at 25% level of significance were time, sex, age, marital status, baseline CD4⁺ count, regimen class, functional status, who stage and the interaction terms of age, marital status, baseline CD4⁺ count, regimen class, functional status and WHO clinical stage with time.

The full marginal model including variable selection

This section is devoted to selection of correlation structure and variables using REML and ML method of estimation respectively.

Selection of covariance structure

Covariance structures should be carefully selected in order to obtain valid inferences for the parameters of fixed effects in the model. If one ignores important correlations by using a model that is too simple, one risks increasing type I error rate and underestimates standard errors. On the other hand, if the model is too complex, the power and efficiency is sacrificed.¹³ The covariance structure that fits data best based on the AIC was used to estimate the fixed effect parameters. In fitting the linear mixed model, a series of correlation structures of the longitudinal data of CD4⁺ counts on HIV positive patients were considered. From the possible covariance structures for fitting the model, the one with the smallest AIC and convergence of the model in REML and ML was considered. Hence, AR (1) was selected for assessing the progression of CD4⁺ counts by linear mixed model.

Variables selection

The model development process was carried out by selecting covariates that have the potential to be included in the linear mixed models. Variable selection was done using the manual backward selection criteria. It was carried out based on maximum likelihood method of estimation and using the selected covariance structure, AR (1). Accordingly, the selected variables were: time, age, baseline CD4⁺ count, WHO clinical stage and the interaction terms of sex, age, baseline CD4⁺ count and regimen class with time.

Checking the correlation structure

After selecting variables of linear mixed models, we refitted the linear mixed effects model of Square root transformed CD4⁺ count under all possible and proposed correlation structures. So, by fitting the model using the REML estimation method, Unstructured (UN) correlation structure was chosen due to its smallest AIC (Table3).

Possible two ways interaction terms were considered in the refined model and interaction between sex, age, baseline CD4⁺ count and regimen class and time were found to be significant.

The significant interaction of sex with time indicates that females have a significant higher progression in CD4⁺ count than males over time which coincides with the mean profile plot shown in Figure 4. The positive significant interaction of age with time indicates that the rate of progression in CD4⁺ count is higher among older patients.

Baseline CD4⁺ count has negative interaction with time indicating that patients who initiated ART at minimum CD4⁺ count made better progression than those who initiated ART at higher baseline CD4⁺ counts.

The final linear mixed model is stated as follows:

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 S_i + \beta_2 A_{ij} + \beta_3 B_i + \beta_4 WHO_{0i} + \beta_5 WHO_{1i} + \beta_6 WHO_{2i} + \beta_7 R_{0i} + \beta_8 R_{1i} + \beta_9 R_{2i} + \beta_{10} R_{3i} + (\beta_{11} S_i + \beta_{12} A_{ij} + \beta_{13} B_i + \beta_{14} R_{0i} + \beta_{15} R_{1i} + \beta_{16} R_{2i} + \beta_{17} R_{3i} + b_{1i}) t_{ij} + \varepsilon_{ij} \dots (1)$$

where, A_{ij} , B_i , WH_{0i} , WH_{1i} , WH_{2i} , S_i , A_{ij} , R_{0i} , R_{1i} , R_{2i} , R_{3i} represent age, baseline CD4⁺ count, WHO clinical stage1, WHO clinical stage2, WHO clinical stage3, and first line ART regimen classes AZT-3TC-EFV, AZT-3TC-NVP, TDF-3TC-NVP and d4t-3TC-NVP respectively.

According to the result of type three tests of fixed effects, base line CD4⁺ count, WHO clinical stage and the overall interaction effects of sex, age, baseline CD4⁺ count and regimen class with time are statistically significant (Table 5).

From the random effect models, the random intercept only model was fitted lastly (Table 6). The model thus fitted was:

$$Y_{ij} = \beta_0 + b_{0i} + \varepsilon_{ij} \dots \dots \dots (2)$$

The intra class correlation coefficient ρ was found to be 0.5466 indicating that 54.66% 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 ($\beta_0 = 18.8196$ with P-value = 0.0001).

Random intercept and slope models that are convenient for highly unbalanced data with many repeated measurements per subject were fitted. The covariance estimates for the random effects associated with the intercept and the linear time effect shown was found to be significant (Table 7). This suggests that the intercepts and slopes vary across patients which are in line with the univariable random effects model. The primary objective of model comparison is to choose the simplest model that provides the best fit to the data. So, including the random intercept and slope in the model fits the data well relative to the marginal one.

Handling of missing data

Direct likelihood and Multiple Imputation (MI) were applied as missing data handling mechanisms under the plausible assumption of MAR. Multiple imputation consists of three distinct phases where the missing data were filled in 10 times to generate ten complete data sets, the ten data sets were analyzed using linear mixed models and lastly, the results from the ten complete data sets were combined for inference.

Comparing the parameter estimates of direct likelihood and Multiple Imputation, we observed comparable result. The variables baseline CD4⁺ count, WHO clinical stage 3, time and the interaction terms of baseline CD4⁺ count with time were statistically significant in both models.

Discussion

This study attempted to examine the impact of some demographic and clinical factors on the progression of CD4⁺ count in Yekatit 12 Referral Hospital. The results revealed a significant linear progression of CD4⁺ counts of patients on ART. This coincides with the results of a study conducted by Reda *et al.*¹⁵

The mean profile of patients showed that females had a higher progression of CD4⁺ counts than male patients. This pattern of female patients having better CD4⁺ count progression was also confirmed by the fitted linear mixed model.

When selecting a covariance structure for a model, only structures that make sense for the data should be considered.¹⁶ In the linear mixed model, after considering all possible covariance

structures, variable selection was done using AR (1) correlation structure due its smallest value by using the manual backward selection criteria. It was carried out based on maximum likelihood method of estimation and using the selected covariance structure and hence, the selected variables were: time, age, baseline CD4⁺ count, WHO clinical stage and the interaction terms of sex, age, baseline CD4⁺ count and regimen class with time. The linear mixed model showed that, the interaction term of sex with time was significant which indicates that females had a significantly higher progression in square root CD4⁺ count than males. This is in contrast to a study conducted by Adams and Luguterah since 2013.¹⁷

Baseline CD4⁺ count had statistically significant interaction with time and the parameter estimate for this interaction term is negative indicating that patients that initiated ART at minimum CD4⁺ counts have higher progression than those who start at higher baselines. Our study revealed that baseline CD4⁺ count was significant predictor of CD4⁺ count progression which is in line with a retrospective cross-sectional study on CD4⁺ cell count trends after commencement of ART among HIV infected patients in Tigray, Northern Ethiopia.¹⁸

Moreover, the findings demonstrate that immunological recovery in Addis Ababa is driven by a complex interplay of baseline clinical markers and demographic factors. Our analysis identified baseline CD4⁺ count and WHO clinical stage as the most potent independent predictors of recovery. These results are highly consistent with longitudinal studies conducted at the University of Gondar Hospital and Jimma University Specialized Hospital, which similarly reported that patients initiating ART at WHO stages III or IV experienced significantly slower CD4⁺ increment rates compared to those in stages I or II.^{8,19}

A notable finding in our study was the statistical significance of the interaction effects (Sex by Time and Age by Time). While some cohorts in Southern Ethiopia have identified male sex as a standalone risk factor for poor recovery,²⁰ our longitudinal model suggests that the disparity lies in the velocity of recovery over time which might be attributed to a combination of late clinical presentation and socio-behavioral barriers to early ART initiation in the Ethiopian context. Specifically, younger patients and those starting on TDF-based regimens showed a more robust slope of immunological improvement. This time-dependent variation in recovery highlights the

necessity of using Linear Mixed-Effects Models to capture the nuance of patient progress that traditional cross-sectional studies might overlook.

The significance of baseline CD4⁺ count as a predictor across all reviewed Ethiopian cohorts—including those in the Amhara and Oromia regions reinforces the Test and Treat strategy. Studies from Northwest Ethiopia (Amhara) have shown that patients starting ART with a CD4⁺ count above 200 cells/μl achieve significantly faster immunological restoration compared to those in advanced immunosuppression.²¹ Similarly, longitudinal data from Southwest Ethiopia (Oromia) indicate that early initiation is the strongest predictor of reaching an optimal CD4⁺ threshold within the first year of treatment.²² By initiating ART before the immune system is severely compromised, the likelihood of achieving an optimal immunological response within the first 14 months is significantly enhanced.

Among the interaction of first line ART regimen classes with time, only AZT-3TC-NVP with time was found to have a significant difference from the reference group TDF-3TC-EFV with time. Among WHO clinical stages, only stage2 was found to have a significant difference from the reference group stage 4.

In a retrospective longitudinal study on predictors of change in CD4 Lymphocyte count and weight among HIV infected patients on ART in Ethiopia by Reda *et al.*, since 2013,¹⁵ functional status had a significant effect on CD4 lymphocyte count. In our study functional status was not significantly associated with the progression of CD4⁺ counts over time.

In this study, the interaction term of age with time was found to have a significant positive effect implying that, the rate of progression in square root CD4⁺ count for older patients was higher than that of younger ones which was also in contrast with a study by Picat *et al.* 2013.²³

Lastly, we conducted Multiple Imputation (MI) and found that the results of the direct likelihood and MI approaches are comparable. This coincides with finding of simulation study reported.^{24,25}

Strengths and limitations

This study provides a robust longitudinal analysis of CD4 count recovery among ART patients in the Ethiopia, filling a critical data gap in the country. A key strength lies in the use of Linear Mixed-Effects Models, which explicitly accounted for the inherent correlation between repeated measures within the same individual and handled unbalanced data efficiently. Furthermore, by transforming the CD4 counts using a square root function, we ensured that the statistical assumptions of normality and homoscedasticity were met, providing more reliable p-values and confidence intervals for the clinical predictors.

Limitations

Despite its strengths, as is characteristic of clinical settings in Ethiopia, missing data and loss to follow-up were observed, with some patients having incomplete records or exiting the study before the third time. While the linear mixed-effects model used is robust to data Missing at Random (MAR), there remains a potential for "Missing Not at Random" (MNAR) bias if those who dropped out possessed significantly different immunological outcomes than those who remained and we recommend future studies. Moreover, the geographical focus on the capital city Addis Ababa may limit the generalizability of these findings to other contexts, such as the pastoralist regions or major conflict zones like Oromia, Amhara and Tigray regions where socio-economic dynamics and access to specialized ART regimens differ.

Conclusions and recommendation

CD4⁺ counts are among laboratory markers used to track the progression of HIV to AIDS. It is commonly used predictor for risk of death and AIDS²⁶ and used to decide when to start ART. The objective of this study was to assess the progression of CD4⁺ counts of patients and to identify determinant factors for the progression of CD4⁺ counts of patients on ART. This study therefore tried to assess the progression of CD4⁺ count and identify factors affecting progression of CD4⁺ count of patients using linear mixed model.

The results of this study showed that age, baseline CD4⁺ count, WHO clinical stage, time and the interaction effects of age, sex, baseline CD4⁺ count and regimen class with time are significantly associated with the progression of CD4⁺ counts over time. From the results of this study, the following recommendations are forwarded for clinicians and patients as well: i) comparing the progression of CD4⁺ counts, males were found to have the worst CD4⁺ count progression compared to females; hence, actions should be taken to improve CD4⁺ count of male patients; ii) based on the findings of our study and the recent WHO recommendation, patients should initiate ART regimens at an early WHO clinical stage; iii) ART regimens should be taken continuously to improve the progression of CD4⁺ count over time; iv) other factors that were not considered in this study might affect the progression of CD4⁺ counts. So, future studies in the area should consider possible covariates such as educational status, adherence, ART second-line regimens and other socioeconomic factors.

References

1. Callaway DS, Ribeiro RM, Nowak MA. Virus phenotype switching and disease progression in HIV 1 infection. *Proc Royal Soc London Series B* 1999;266:2523-30.
2. World Bank. Africa Regional Office. Intensifying Action Against HIV/AIDS in Africa: Responding to A Development Crisis. 2000: Praeger/Greenwood.
3. Ethiopian Public Health Institute, HIV related estimates and projections in Ethiopia for the year 2021–2022. 2022, The Ethiopian Public Health Institute Addis Ababa, Ethiopia.
4. Sack U, Tárnok A, Rothe G. Cellular diagnostics: basic principles, methods and clinical applications of flow cytometry. Karger Medical and Scientific Publishers; 2009.
5. UNAIDS, World AIDS Day Report 2025: Overcoming Disruption: Transforming the AIDS Response. 2025: Stylus Publishing, LLC.
6. Jayasuriya A, Robertson C, Allan P. Twenty-five years of HIV management. *J Royal Soc Med* 2007;100:363-6.
7. Mocroft A, Ledergerber B, Katlama C, et al. Decline in the AIDS and death rates in the EuroSIDA study: an observational study. *Lancet* 2003;362:22-9.
8. Kebede MM, Zegeye DT, Zeleke BM. Predictors of CD4 count changes after initiation of antiretroviral treatment in University of Gondar Hospital, Gondar in Ethiopia. *Clin Res HIV or AIDS* 2015;1:1-15.

9. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;963-74.
10. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. John Wiley & Sons; 2012.
11. Verbeke G. Linear mixed models for longitudinal data, in *Linear mixed models in practice: A SAS-oriented approach*. Springer; 2000:63-153.
12. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Vol. 81. John Wiley & Sons; 2004.
13. Littell RC, Milliken GA, Stroup WW, et al. *SAS for mixed models*. 2006: SAS publishing.
14. Diggle P. *Analysis of longitudinal data*. 2002: Oxford university press.
15. Reda AA, Milliken GA, Stroup WW, et al. Predictors of change in CD4 lymphocyte count and weight among HIV infected patients on anti-retroviral treatment in Ethiopia: a retrospective longitudinal study. *PLoS One* 2013;8:e58595.
16. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986 Mar;42(1):121-30. PMID: 3719049.
17. Adams M, Luguterah A. Longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on antiretroviral therapy (ART) in the Builsa District Hospital. *Eur Sci J* 2013;9(33).
18. Asfaw A, Ali D, Eticha T, et al. CD4 cell count trends after commencement of antiretroviral therapy among HIV-infected patients in Tigray, Northern Ethiopia: a retrospective cross-sectional study. *PloS One* 2015;10:e0122583.
19. Workneh N, Girma T, Woldie M. Immunologic and clinical outcomes of children on HAART: a Retrospective cohort analysis at Jimma University specialized hospital. *Ethiopian J Health Sci* 2009;19:69422.
20. Osman FT, Yizengaw MA. Virological failure and associated risk factors among HIV/AIDS pediatric patients at the ART clinic of Jimma University Medical Center, Southwest Ethiopia. *Open AIDS J* 2020;14.
21. Lawn SD, Myer L, Bekker LG, Wood R. CD4 cell count recovery among HIV-infected patients with very advanced immunodeficiency commencing antiretroviral treatment in sub-Saharan Africa. *BMC Infect Dis* 2006;6:59.

22. Asefa A, Asaye Z, Girma A, Hiko D. Predictors of Clinical and Immunological Failure Among Patients on First-Line Antiretroviral Therapy (ART) in Southwest Ethiopia. HIV AIDS (Auckl) 2019;11:377-86.
23. Picat MQ, Lewis J, Musiime V, et al. Predicting patterns of long-term CD4 reconstitution in HIV-infected children starting antiretroviral therapy in sub-Saharan Africa: a cohort-based modelling study. PLoS Med 2013;10:e1001542.
24. Verbeke G, Molenberghs G, Thijs H, et al. Sensitivity analysis for nonrandom dropout: a local influence approach. Biometrics 2001;57:7-14.
25. Molenberghs G, Kenward M. Missing data in clinical studies. John Wiley & Sons; 2007.
26. Kagan JM, Sanchez AM, Landay A, Denny TN. A brief chronicle of CD4 as a Biomarker for HIV/AIDS: a tribute to the memory of John L. Fahey. For Immunopathol Dis Therap 2015;6:55-64.

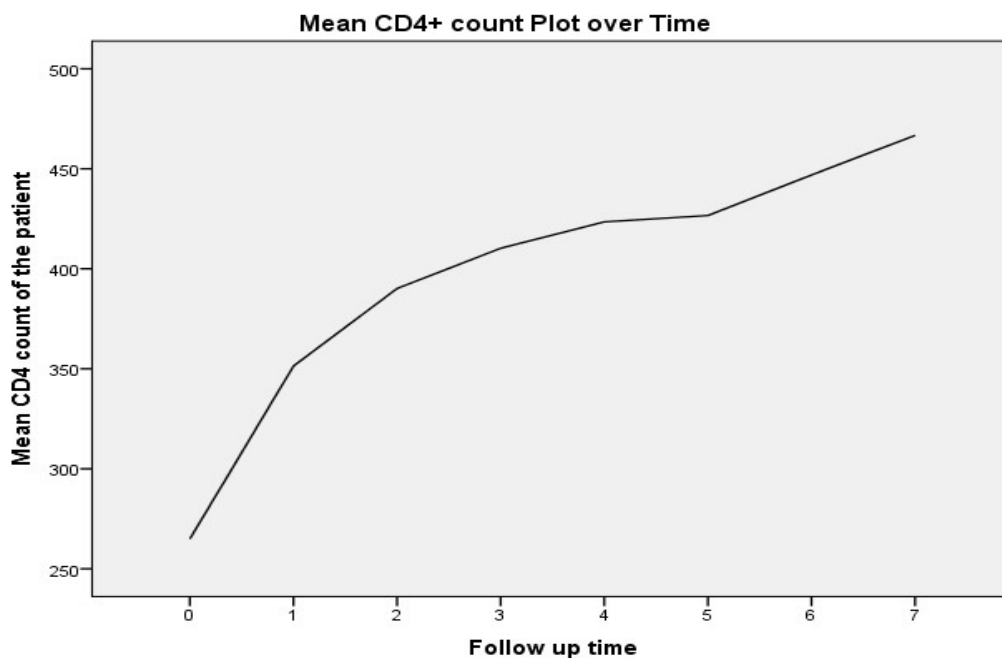


Figure 1. Mean CD4⁺ count plot by time.

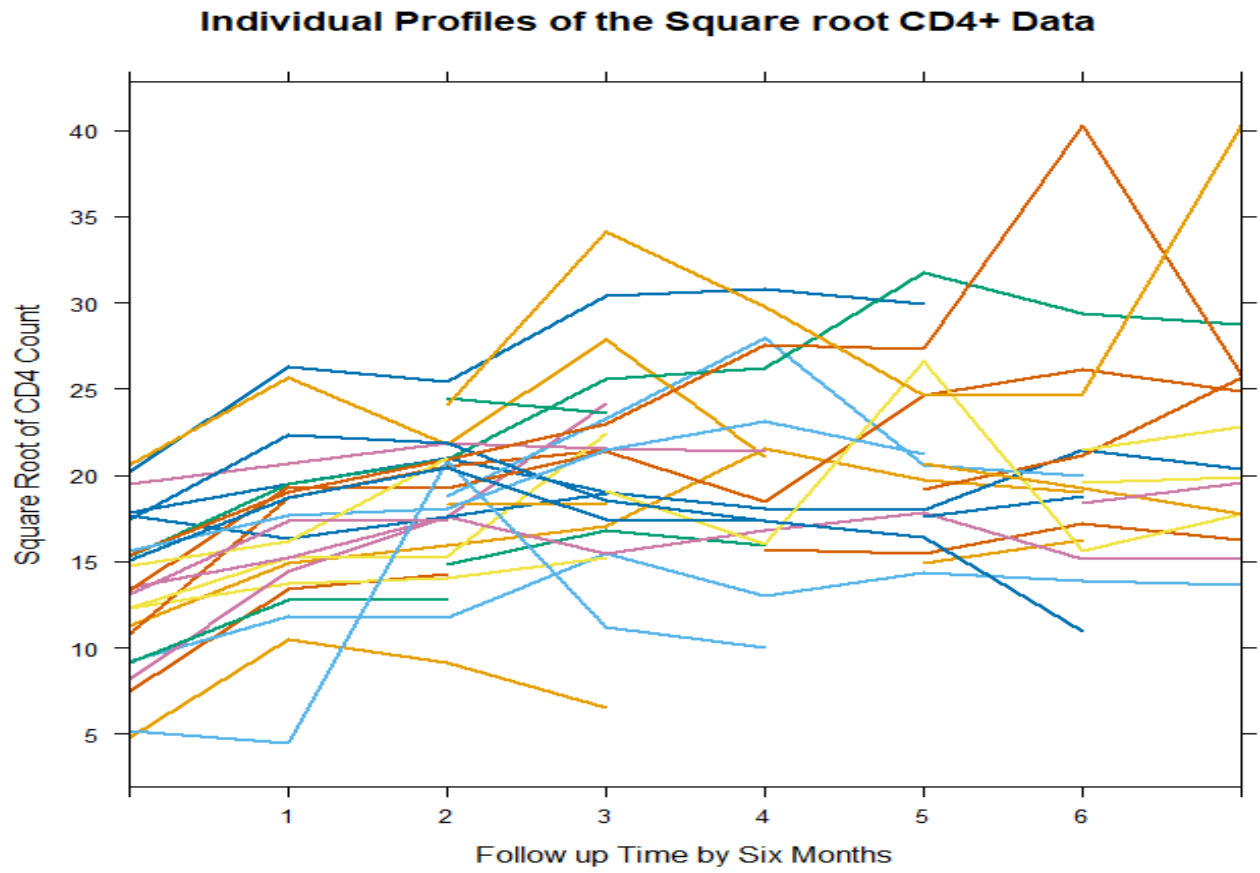


Figure 2. Individual profile plot of square root CD4⁺ count by time.

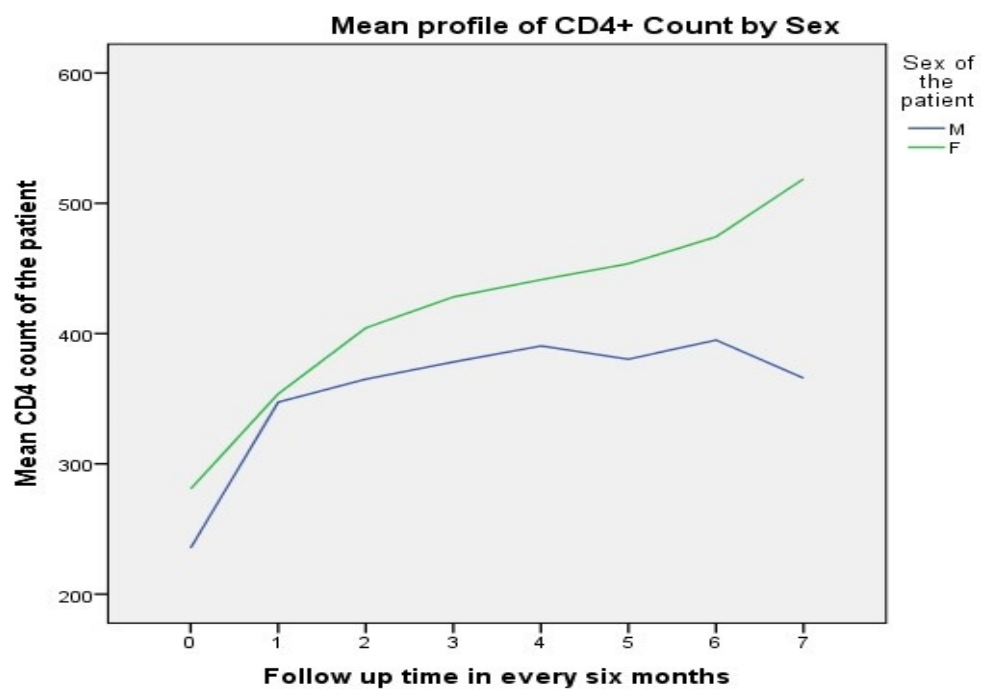


Figure 3. Mean profile of patients CD4⁺ count by sex over time.

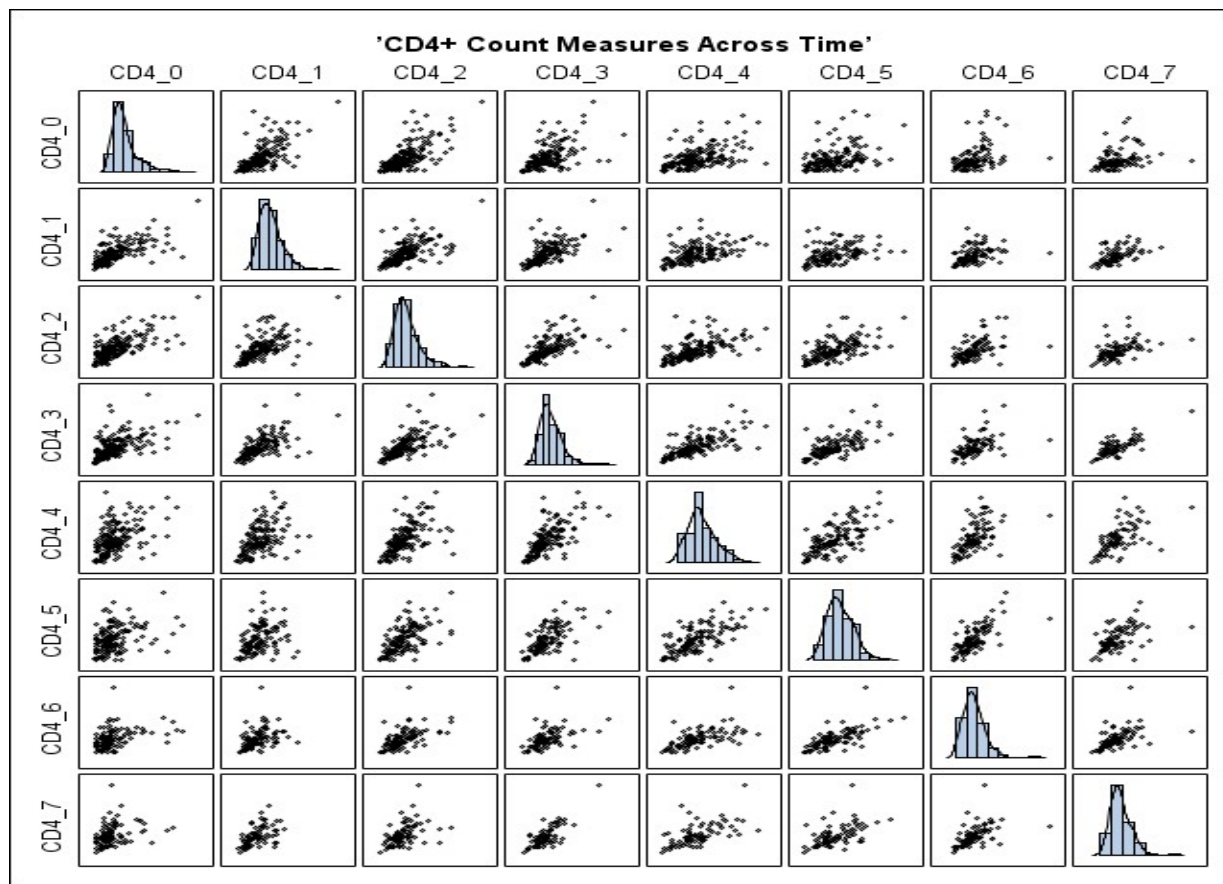


Figure 4. Scatter plot of correlation matrix.

Table 1. Descriptive statistics of continuous variables.

Time	Baseline	1	2	3	4	5	6	7
N	217	179	191	162	168	141	116	97
Mean (CD4)	264.8848	351.3519	390.1518	410.2716	423.4643	426.6525	446.9138	466.7217
Std Dev	210.9259	210.5959	231.6899	233.8812	209.3979	197.2816	209.0785	233.1719

Table 2. Correlation matrix of CD4⁺ count over time.

	Base line	Time1	Time2	Time3	Time4	Time5	Time6	Time7
Baseline	1.00	0.8383	0.7455	0.4858	0.1812	-	-	0.0576
Time1	0					0.0265	0.1189	
Time2		1.000	0.6029	0.5438	0.1251	-	-	0.2061
Time3			1.000	0.5765	0.2246	0.0536	-	-0.1577
Time4				1.000	0.1548	-	0.0143	-0.0995
Time5					1.000	0.1493	0.0252	-0.1177
Time6						1.000	-	0.1731
Time7							0.0899	1.0000

Table 3. Fit statistics for checking correlation structure.

Information Criteria	Correlation Structure	
	AR (1)	UN
-2loglikelihood	6918.3	6598.0
AIC	6924.3	6670.0
AICC	6924.3	6672.1
BIC	6934.4	6791.6

Table 4. Description of Antiretroviral Therapy (ART) Regimens Used in the Study.

Regimen Class Abbreviation	Full Generic Composition
AZT-3TC-NVP	Zidovudine + Lamivudine + Nevirapine
AZT-3TC-EFV	Zidovudine + Lamivudine + Efavirenz
TDF-3TC-NVP	Tenofovir + Lamivudine + Nevirapine
d4T-3TC-NVP	Stavudine + Lamivudine + Nevirapine
TDF-3TC-EFV	Tenofovir + Lamivudine + Efavirenz

Table 5. Parameter estimates for full linear mixed model.

Effect	Estimate	St.Err	DF	T.val	Pr> t
Intercept	8.3364	0.5502	211	15.1	<.0001
				5	
Sex	-0.6568	0.2537	211	-2.59	0.0103
AGE	-0.00810	0.009591	211	-0.85	0.3991
BASECD4	0.02666	0.000938	211	28.4	<.0001
				2	
REGIMENCLASS (AZT-3TC-EFV)	0.1626	0.2686	211	0.61	0.5456
REGIMENCLASS (AZT-3TC-NVP)	0.1890	0.2637	211	0.72	0.4744
REGIMENCLASS (TDF-3TC-NVP)	0.1326	0.3424	211	0.39	0.6990
REGIMENCLASS (d4t-3TC-NVP)	0.3678	0.5356	211	0.69	0.4930
WHOSTAGE1	0.2767	0.3154	211	0.88	0.3813
WHOSTAGE2	0.6282	0.3164	211	1.99	0.0484
WHOSTAGE3	-0.02595	0.3471	211	-0.07	0.9405
TIME	1.3320	0.2096	211	6.35	<.0001
TIME*SEX(Male)	-0.2144	0.09904	211	-2.16	0.0315
AGE*TIME	0.009795	0.004708	211	2.08	0.0387
BASECD4*TIME	-0.00248	0.000252	211	-9.83	<.0001
TIME*REGIMENCLASS (AZT-3TC-EFV)	-0.2382	0.1241	211	-1.92	0.0562

TIME*REGIMENCLASS (AZT-3TC-NVP)	-0.2697	0.1309	211	-2.06	0.0405
TIME*REGIMENCLASS (TDF-3TC-NVP)	-0.2592	0.1803	211	-1.44	0.1522
TIME*REGIMENCLASS(d4t-3TC-NVP)	0.3060	0.1994	211	1.53	0.1263

Table 6. Covariance parameter estimates for random intercept only model.

Cov Parm	Subject	Estimate	std. Error	Z. Value	Pr> Z	ICC
Intercept	ID	18.2868	2.0470	8.93	<.0001	0.5466
Residual		15.1671	0.6620	22.91	<.0001	

Table 7. Covariance parameter estimates.

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	ID	3.0523	0.6597	4.63	<.0001
TIME	ID	0.3325	0.0659	5.05	<.0001
Residual		9.2265	0.4237	21.78	<.0001

Conflict of interest: the author declare that they have no competing interests.

Availability of data and materials: the anonymized data used to produce results reported in this manuscript could be obtained from the corresponding author Mr. Keadu Tadesse Cherie (Email: kebadu.tadesse@dmu.edu.et).

Ethics approval: ethical clearance and approval for this study were obtained from the Institutional Review Board (IRB) of the College of Natural and Computational Sciences, Addis Ababa University. Following this, a formal letter of support and administrative permission were granted by the Yekatit 12 Hospital Medical Case Team. To ensure patient confidentiality, all data were extracted from the ART clinic registries by trained data clerks using anonymized abstraction forms. The research team did not have access to any personal identifiers or original patient medical cards. The study was conducted in accordance with the Declaration of Helsinki and the national guidelines for research involving secondary clinical data.

Funding: this work was supported by Addis Ababa University under grant GSR/1172/07.

Acknowledgments: the author sincerely acknowledge the valuable guidance and advice provided by the late Dr. Birhanu Teshome Ayele during the development of this manuscript. His scholarly contribution and thoughtful guidance are gratefully remembered. The authors also express their sincere gratitude to Wolkite University, Addis Ababa University, and Yekatit 12 Referral Hospital for their support and for providing the data used in this study.