

## Modeling of Multivariate Longitudinal Factors on Human Immunodeficiency Virus Infections Cell Counts: Longitudinal Study Design

### Research Article

DOI: 10.59152/ESJCR/1038

Alebachew Abebe\*

Department of Statistics, College of Computing and Informatics, Haramaya University, P.O.Box: 138, Dire Dawa, Ethiopia

Received: April 07, 2023; Accepted: June 02, 2023; Published: June 05, 2023

\*Corresponding author: Alebachew Abebe, Department of Statistics, College of Computing and Informatics, Haramaya University, P.O.Box: 138, Dire Dawa, Ethiopia

**Copyright:** © 2023 Alebachew Abebe. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

**Background:** Modeling of multivariate longitudinal data provides a unique opportunity in studying the joint evolution of multiple response variables over time. This study was modeling of multivariate longitudinal data cell counts on human immunodeficiency virus infections at Gondar districts were used two different multivariate repeated measurement with a kronecker product covariance and random coefficient mixed models.

**Methods:** The study was based on data from 566 per four visits human immunodeficiency virus infections were enrolled in the first 4 visits of the 5-year secondary data with retrospective longitudinal study design.

**Results:** The results revealed that both models of human immunodeficiency virus infections cell counts values of CD4 and CD8 increase over time while hemoglobin decreases over time. Those models results reveals that a strong positive correlation between CD4 and CD8 cells, but the correlation between CD4 and hemoglobin as well as the correlation between CD8 and hemoglobin are not statistically significant at 5% level of significance.

**Conclusion:** According to this study in the early stage of human immunodeficiency virus infections, CD4 and CD8 steadily increase over time while the values of hemoglobin decrease over time. Consequently, the study suggests that concerned bodies should focus on awareness creation to increase cell counts of CD4 and CD8 over time while hemoglobin decrease over time for HIV infections at Gondar districts, Ethiopia.

**Keywords:** HIV Infections; Cell counts; Multivariate longitudinal data; Kronecker product covariance; Random coefficients

**Abbreviations:** CD4 Count: T-Cell Test; CD8 Count: +T-Cell Test; HMG: Hemoglobin Levels; HIV: Human Immunodeficiency Virus; SAS: Statistical Analysis System; AIC: Akaike Information Critical; BIC: Bayesian Information Critical; -2LnL: Likelihood Ratio Test; CS: Compound Symmetry; AR(1): Autoregressive Order One; TOEP: Toeplitz; UN: Unstructured; ICC: Intra-Cluster Correlation; PHID: Patient Health Identification; Proc: Procedure.

### Introduction

Longitudinal data consist of repeated measurements that belong to same subjects/units. This type of data might arise in many research fields such as medical sciences [1-

4]. The repeated measurements are typically dependent and these dependencies (associations) must be taken into account to have valid statistical inferences [5].

Longitudinal data are often collected on multiple responses. In such a case, in addition to the within-subject association, two additional association structures, multivariate response association at a specific time point and cross-temporal response association, occur. The multivariate responses could either belong to the same response family or different response families. In literature, there are two common approaches to handle multivariate longitudinal data rather than jointly analyzing them. First one is to ignore the multivariate response association and construct univariate models [6] or carry out univariate cluster analyses [7,8]. The second one is to consider one of the responses as the dependent variable and the other(s) as independent one(s) together with the other independent variables and to construct a univariate model [9].

From the clinical point of view, what is required for the present model is that an essential component of the immune system is the Lymphocytes which destroy invaders. Lymphocytes are of two types B-cells and T-cells, being B-cells antibody factories producing antibodies as fast as they can and also clone themselves, whereas T-cells either direct the activity of B-cells (called CD4+T-cells) or act as suppressors (called CD8+T-cells) destroying infected cells and thus damp out the activity of the immune system. The AIDS has three stages, one including the initial infection, a second one of latency and the third one corresponding to a runaway destruction of the immune system. During the latency lapse healthy T-cells are infected although its number remains high. When the concentration of T-cells decreases and that of HIV virus cells increases, it is the AIDS stage. In practice, as an example, cytometry is a procedure useful to count the concentration of healthy, latently infected and infected T-cells in the blood stream, for instance using the dispersion of laser light that depends on the enzymes covering the cells [23]. Knowing the concentration of CD4+T-cells it is possible to wonder about modeling the evolution of the different cell populations. Among the various models trying to describe the dynamics of CD4+T-cells (see for instance [24,25]), a simple one describing the phases of latency and the destruction of the immune system is Perelson's model [26-28].

A total of 566 per four visits (2264 individuals) were enrolled in HIV infections for this study and followed up to 5-years with in 4 repetitions. In this study, we only considered modeling of multivariate longitudinal response variables for HIV infections have three responses. These

are CD4 counts (CD4), CD8 counts (CD8) and Hemoglobin levels (HMG). Our primary interest was to assess the inter-relationships among these multivariate responses. For an easy notation, throughout this study we let X, Y, and Z represents for CD4, CD8 and HMG respectively. We note that, as ubiquitous in longitudinal studies, not all multivariate responses are measured at all occasions in HIV infections data. In spite of their practicability, these approaches do not yield valid statistical inferences. In this study, we consider the accommodation of the aforementioned three association structures with a multivariate modeling approach for multiple responses belonging to the same response family.

However, the analysis of such a modeling of multivariate longitudinal data can be challenging because (i) the variances of errors are likely to be different for different subjects, (ii) the errors are likely to be correlated for the same subjects measured at different occasions, and (iii) the errors are also likely to be correlated among subjects measured at the same time. In the SAS software, 2 different approaches have been provided to analyze modeling of multivariate longitudinal data: modeling of multivariate repeated measurement models with a kronecker product covariance structure [15] and random coefficient mixed models [12].

In this study, we first present the HIV infections data that motivates our study. Then we illustrate and compare these 2 approaches in studying the joint evolution of the modeling of multivariate longitudinal data for HIV infections at an early stage of HIV infections data. The objectives of this study was to explore the joint evolution of CD4 counts, CD8 counts and hemoglobin levels of HIV infections and to compare modeling of multivariate repeated measurement models (kronecker product covariance and random effects mixed model) using HIV infections data of North Gondar, Ethiopia. In this study, model development procedures were Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Likelihood Ratio Test (-2LnL).

## Materials and Methods

### Study area and data descriptions

This study was based on data from 566 per four visits (2264 individuals) persons with HIV infections were enrolled in the first four visits of the five year multivariate longitudinal study of HIV infections using the data of North Gondar, Ethiopia. The study design was retrospective on

multivariate longitudinal data setting that go back in time to compare multivariate repeated measurement models with a kronecker product covariance structure and random mixed effects model and to explore the joint evolution of CD4, CD8 counts and hemoglobin levels of HIV infections. The secondary data were collected from different part of North Gondar districts that measured repeatedly in four visits between the years on May 30, 2013 to May 30, 2017. Each repeated measures were conducted on the average six month interval in the study period. In this study all multivariate longitudinal data response variables were time variant and independents variables were time variant and invariant.

### Eligible Criteria

Outstanding to some natural and unknown reasons few person fail to continue up to the end of the study period. This study was considered both inclusion and exclusion criteria.

### Inclusion Criteria

Throughout the data collection, we were selected person who have HIV infections that could have four repeatedly measured over time data were included in the study period. That is, the repeatedly measured data with four follow-ups (visits) person with HIV infections were included in the study.

### Exclusion Criteria

HIV infections who dropout (withdraw), intermittent missing values and loss to follow-up within the study period was excluded from the study. Hence, person who may have died from other cause, refuse and others who do not participate at the previous measurement (attrition) time of the study period was excluded from the study. Although, the time period before May 30, 2013 and after May 30, 2017 HIV infections would been excluded from the study.

### Variables Considered in the Study

HIV infections were considered as multivariate longitudinal data act as the response variables. That variable has three categories were CD4 counts, CD8 counts and hemoglobin levels. Even if, they were used as a continuous variable to maximize the amount of information available in the data set.

The independent variables such as cite of HIV infections (Chilga, Debark, Gondar and Metema), ART number,

visits (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> & 4<sup>th</sup>), value of HIV infections and time of HIV infections were included in the study. The first two variables were used for descriptive statistics and the remaing variables were applied for inferential statistics (modeling part).

### Methods of Statistical Analysis

Multivariate longitudinal data is a special case of repeatedly measured data, the observations are not independent and are characterized as having both between-subject and within-subject variation, time dependent covariates and missing data [10]. Furthermore, multivariate longitudinal data response variables have become increasingly popular, more accessible and good in missing data handling through statistical software such as SAS Virson-9.4 [13].

### Exploring Data Analysis

#### Exploring the Correlation Structure

Each covariance model has a corresponding correlation model. Even though there are so many models for covariance structure. There are 4 types of covariance models were considered and compared in this study [29].

#### Compound Symmetry (CS)

Suppose that a parameter representing the common correlation for any two time points, then the correlation here,  $-1 < \rho < 1$ .  $Cov(\sum_j, \sum_k) = \rho, \forall j \neq k$ ; where the single correlation parameter  $\rho$  is generally referred to as the intra-class correlation coefficient.

Let us assume that, with  $m$  repeated measures.

$$\mathbf{\Sigma} = \sum = \delta^2 \begin{pmatrix} 1 & \rho^2 & \dots & \rho^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^2 & \rho^2 & \dots & 1 \end{pmatrix}$$

#### Autoregressive Order One (AR(1))

Autoregressive models express the current observation as a linear function of previous observations plus a homoscedastic noise term,  $\alpha_t$ , centered at  $(E[\alpha_t] = 0)$  and assumed independent of the previous observations.

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + \alpha_t$$

$$h(K, \phi) = \phi^K, K \in \{0, 1, \dots, m-1\}$$

$$\mathbf{A} (1) = \sum = \delta^2 \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \dots & 1 \end{pmatrix}$$

$$\mathbf{Toeplitz (TOEP)} \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \dots & 1 \end{pmatrix}$$

If there were  $m$  measures, then there will be two distances: one unit distance, two units distance. And, hence two parameters were estimated.

$$TOEP = \sum = \delta^2 \begin{pmatrix} 1 & \rho_1^2 & \cdots & \rho_{m-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m-1}^2 & \rho_{m-2}^2 & \cdots & 1 \end{pmatrix}$$

### Unstructured (UN)

Unstructured covariance matrix allows  $m$  different variances, one for each time point and  $\frac{m(m-1)}{2}$  distinct off-diagonal elements representing the possibly different covariance for each pair of times, for a total of

$$m + \frac{m(m-1)}{2} = \frac{m(m+1)}{2} \text{ variances and co-variances.}$$

$$N = \sum = \begin{pmatrix} \sigma_1 & \sigma_2 & \cdots & \sigma_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m \end{pmatrix}$$

Simple (VC): The simple model assumes independent observations and homogeneous variance. This is the default.

$$C = \sum = \delta^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

UN (1): Different variance and 0 covariance.

$$N(1) = \sum = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{pmatrix}$$

### Model Selection

It is one of a most series problem in data analysis. The variance-covariance structure with the lower AIC, BIC and likelihood ratio test (-2LnL) value can be an appropriate to select the model. Most commonly in epidemiological studies, investigators frequently attempt to construct the most desirable statistical model using the popular methods of forward, backward, and stepwise regressions [30].

### Intra-Cluster Correlation (ICC)

The covariance matrix of the random-effects is thought to summarize the ICC. If there are a large number of

random-effects components, then this leads to a complex covariance matrix and can increase computational burden [31].

### Multivariate Repeated Measurement Models

#### Kronecker Product Covariance Structure

The first approach considered is to fit a model with a kronecker product covariance structure. This model allows investigator to specify the variance-covariance matrix of the measurement error within each study unit (i.e., subject) and thus to examine the intra-and inter-subject correlations of the measurement errors. The data layout for this model has the following presentation, where the variable VISIT indicates that all the assessments are equally spaced while the variable TIME gives the exact measurement time (six month). Where, PHID stands for patient who have HIV infections identified card in the SAS program. To assure normality and homoscedasticity of the residual distribution, the response variable is defined as the change in value of subject at time  $t$  since the initial visit, i.e.,

$$X_i(t) = \log_e^D A_i(t) - \log_e^D A_i(0), Y_i(t) = \log_e^D S_i(t) - \log_e^D S_i(0), Z_i(t) = \log_e^D HMG_i(t) - \log_e^D HMG_i(0), i = 1, 2, \dots, 566.$$

For models with a kronecker product covariance, SAS only provides the possibility to fit bivariate models. The model is implemented by the REPEATED statement in the PROC MIXED and currently SAS only provides 3 alternative covariance structures: UN@CS, UN@AR(1) and UN@UN. For example, the following SAS codes fit a model with a covariance of UN@AR(1) for the relationship between CD4 and CD8 cells, where the COVTEST option requests significance tests for the covariance estimates (i.e., the random effects portion of the model), the NOCLPRINT option suppresses the display of the "Class Level Information" table, and the NOINT option in the MODEL statement suppresses the intercept terms because we have:  $X_i(0) = Y_i(0) = 0$  for each subject  $i$ . note that the variable VISIT rather than TIME is used in the MODEL statement, and thus assumes that all HIV infections have equally spaced assessments. The variable VISIT in the REPEATED statement takes exactly the same values as in VISIT, but VISIT is treated as a class variable. The syntax/ program in SAS version-9.4 was fitted the kronecker product covariance structure was presented below.

```
Proc mixed data=HIV_MIX covtest noclprint;
```

Title1 "model 1 mixed model with a kronoker product covariance";

Title2 "CD4 (X) versus CD8 (Y)";

Class PHID vname visit;

Model value=vname\*visit/s noint;

Repeated vname visit/ type=UN@AR(1) subject=PHID r rcorr;

Where vname="CD4" or vname="CD8";

The option TYPE=UN@AR(1) in the REPEATED statement specifies that the covariance matrix with in an individual has the following structure:

$$V \otimes \Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_x \sigma_y \\ \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 * \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} & \sigma_x * \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \\ \sigma_x * \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} & \sigma_y^2 * \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \end{bmatrix}$$

Thus, the above model assumes that (i) the two cells (CD4 & CD8) share a common intra-cells correlation as measured by:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}, \text{ and (ii) the inter-cells}$$

correlations (as measured  $\rho_y = \frac{\sigma_y}{\sqrt{\sigma_x^2 \sigma_y^2}}$ ) are the

same for two cells measured at the same time point. The two cells are independent if the matrix V has the form

of  $V = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$ , and the hypothesis  $H_0: \sigma_y = 0$

will be tested by the COVTEST option.

### Random Coefficient Mixed Models

Instead of modeling the variation with in the study unit as in the repeated measurement models, the random coefficient mixed models assume that the regression

coefficients are a random sample from some population of possible coefficient and allow one to model variations between study units [14]. In the presence of multiple response variables, a separate set of regression coefficients will be fitted for each response variable and the correlations among these random coefficients can be examined. In SAS this model is also implemented with PROC MIXED and the data layout of the model is exactly the same as in previous section. The following SAS codes fit a random coefficient mixed model for CD4 and CD8. We note that the variable TIME rather than VISIT is used in the model, indicating its capability to handle equally spaced measurements. The syntax/program in SAS version-9.4 was fitted the random coefficient mixed models were presented below (Table 1).

**Table 1:** The data layout presentation for multivariate longitudinal of HIV data.

PHID	Vname	Visit	Time	Value
1	CD4	1	0	217
1	CD4	2	5.866667	393
1	CD4	3	11.63333	2944
1	CD4	4	17.43333	545
1	HMG	1	0	13
1	HMG	2	5.866667	11.4
1	HMG	3	11.63333	11.2
1	HMG	4	17.43333	21
1	CD8	1	0	29
1	CD8	2	5.866667	43.9
1	CD8	3	11.63333	37.7
1	CD8	4	17.43333	29.4

Proc mixed data=HIV\_MIX covtest noclprint;

Title1 "model 2 mixed model with random coefficients";

Title2 "CD4 (X) versus CD8 (Y)";

Class PHID vname;

Model value=vname\*time/s noint;

Random vname\*time/ type=UN subject=PHID g gcorr;

Repeated / type=VC group=vname subject=PHID;

Where vname="CD4" or vname="CD8";

Run;

The RANDOM statement requests that two random slopes (one for CD4 and one for CD8) be fitted for each individual and the GROUP option in the REPEATED statement specifies that the variances of measurement errors are different for different cells. The G and GCORR

options in the RANDOM statement require the display of covariance and correlation matrix for the random slopes

respectively, with  $G = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$ , the two slopes are independent if  $\sigma_{xy} = 0$ .

## Results

Among the total of 566 per four visits (2264 individuals) HIV infections data at North Gondar Zone, Ethiopia for category of CD4 counts were 852(37.6%), while in CD8 counts were 644(28.4%) and the remaining hemoglobin levels were 768(33.9%) presented in Figure 2 and Table 2. However, Figure 1 revealed that the value of HIV infections data at North Gondar Zone, Ethiopia was approximately skewed to the right direction or decrease to the right direction.

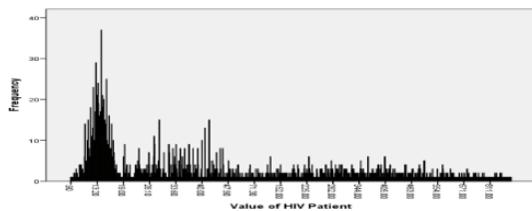


Figure 1: Bar-chart for value of HIV

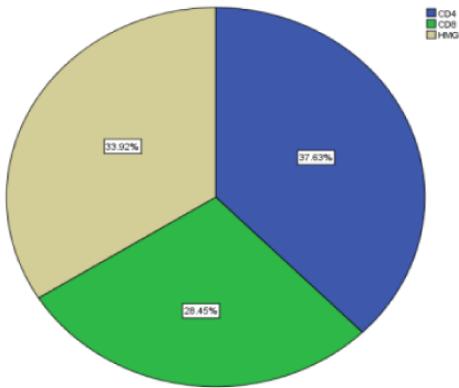


Figure 2: Pie-chart for variables name of HIV distribution

In Table 2 reveals that each visit of HIV infections per visits were 566(25%). According to this scenario in HIV infections of the cite Chilga have 1016(44.9%), Debank cite have 84(3.7%), Gondar cite have 456(20.1%) and the rest HIV infections have 708(31.3%) in Metema cite.

From the total of 566 HIV data the value of HIV infections that have 0.90 was the minimum and maximum amount of infections were 2944. Off the total 566 HIV infections

Table 2: Categorical variables of HIV infections data (n=566)

S. No.	Variables	Categories	Frequency (n)	Percentage (%)
1	Cite of Infections	Chilga	1016	44.9
		Debank	84	3.7
		Gondar	456	20.1
		Metema	708	31.3
2	Visit of HIV Infections	First	566	25.0
		Second	566	25.0
		Third	566	25.0
		Fourth	566	25.0
3	Variables name of Infections	CD4	852	37.6
		CD8	644	28.4
		HMG	768	33.9

Off the total HIV infections of 566 per four visits (2264 individuals) cases for cell counts of CD4, CD8 and hemoglobin levels were 213, 161 and 192 were presented respectively. In similar manner, the average value of the visit of HIV infections with in the variables name of HIV infections in CD8, hemoglobin and CD4 were 0.285, 0.339 and 0.376 presented respectively.

data the average value of HIV infections were 163.77 and with standard deviation of 235.81. In the same fashion for other variables minimum, maximum, mean and standard deviation were presented respectively Table 3,4.

Table 3: Cross-tabulation for visit on variables name of HIV infections data

Cross-tabulation		Variables name of HIV Infections			Total
		CD4	CD8	HMG	
Visit of HIV Infections	1	213	161	192	566
	2	213	161	192	566
	3	213	161	192	566
	4	213	161	192	566
Total		852	644	768	2264
Mean		0.376	0.285	0.339	-
Standard Deviation		14.577	12.676	13.841	-

Table 4: Continuous variables of HIV infections data at North Gondar Zone, Ethiopia (n=566)

	Variables	Minimum	Maximum	Mean	Std. Deviation
Dependent	Value of HIV Infections	.90	2944.00	163.77	235.81
	ART Number	1	137	69.46	40.32
Covariates	Time of HIV Infections	0.00	128.53	29.97	26.48

The results of the fitted mixed models are shown in table-5, where the slope parameter reflects the average six month change of the cell counts over time. Since our primary interest focuses on the inter-relationships among these cells, for simplicity, only the estimated correlation coefficients out of the random effect portion are listed in Table 5. Based on the fit statistics (smaller indicating a better fit), those models with UN@UN covariance structure provide a better fit to the data.

The results show that, patients with HIV infections, the values of CD4 and CD8 increase over time while

**Table 5:** Bivariate mixed models with a kronecker product covariance on HIV infections data

Features	CD4(X) and CD8(Y)		CD4(X) and HMG(Z)		CD8(Y) and HMG(Z)	
	UN@AR(1)	UN@UN	UN@AR(1)	UN@UN	UN@AR(1)	UN@UN
Fixed effects						
Slope <sub>x</sub>	0.0623**	0.0639**	0.0600**	0.0611**	---	---
Slope <sub>y</sub>	0.2244**	0.2239**	---	---	0.2266**	0.2235**
Slope <sub>z</sub>	---	---	-0.0279*	-0.0271*	-0.0282*	-0.0282*
Random effects						
$\rho_x$	0.466**	0.463**	---	---	---	---
$\rho_x$	---	---	0.029	0.069	---	---
$\rho_y$	---	---	---	---	-0.059	-0.055
Fit Statistics						
AIC	-3360	-3323	-1509	-1745	-880	-926
BIC	-3365	-3303	-1478	-1708	-867	-913
-2LnL	-3398	-3357	-1529	-1764	-898	-970

hemoglobin (HMG) decreases over time. All these changes are significantly different from zero (with  $P<0.001$  for CD4 and CD8,  $P<0.05$  for HMG respectively). The results also reveal that a strong positive correlation between CD4 and CD8 cells ( $\rho_x = 0.463$  with  $P<0.001$ ), but the correlation between CD4 and HMG ( $\rho_x = 0.069$ ,  $P=0.148$ ) as well as the correlation between CD8 and HMG ( $\rho_y = -0.055$ ,  $P=0.217$ ) are not statistically significant. We note that these correlations index had the extra associations among cell counts after removing the effect of involution process over time.

Statistically significant at both 95% & 99.9% level of confidence level (\* $P<0.05$ , \*\* $P<0.001$ ).

**Table 6:** Heterogeneous mixed models with random coefficients on HIV infections data

Features	(X, Y)	(X, Z)	(Y, Z)	(X, Y, Z)
Fixed effects				
Slope <sub>x</sub>	0.0053**	0.0053**	---	0.0053**
Slope <sub>y</sub>	0.0088**	---	0.0087**	0.0088**
Slope <sub>z</sub>	---	-0.0021*	-0.0021*	-0.0021*
Random effects				
$\rho_x$	0.753**	---	---	0.753**
$\rho_x$	---	-0.074	---	-0.074
$\rho_y$	---	---	-0.045	-0.034
Fit Statistics				
AIC	-3355	-1933	-823	-3423
BIC	-3343	-1906	-795	-3494
-2LnL	-3356	-1937	-833	-3443

Statistically significant at both 95% & 99.9% level of confidence level (\* $P<0.05$ , \*\* $P<0.001$ ).

The results of mixed models with random coefficients are listed in table-6 where the slope parameter represents the average change per six month for each cell counts over time. For simplicity, only the estimated correlation coefficients out of the random effect portion are listed in Table 6.

Results show that CD4 and CD8 increase over time ( $P<0.001$ ) while HMG decrease ( $P<0.05$ ). There exist a strong positive correlation between CD4 and CD8 ( $\rho_x = 0.753$  with  $P<0.001$ ), but the correlation between CD4 and HMG ( $\rho_x = -0.074$ ,  $P=0.536$ ) as well as the correlation between CD8 and HMG ( $\rho_y = -0.045$ ,  $P=0.764$ ) are not statistically significant. However, it should be pointed out that the correlation coefficients in table-6 represent the associations among random slopes (i.e., the trajectory of cell counts over time) rather than measurement errors. We also note this approach is easily extendable to multivariate longitudinal data with more than two response variables.

## Discussions

In this study was conducted comparison of the 2 approaches for multivariate longitudinal data modeling: the 2 different approaches considered in this study to explore different aspects of the joint evolution of multiple response variables, and thus each approach has its own advantages and limitations.

The first models with kronecker product covariance provide a convenient way to fit bivariate data and enable one to examine both inter-cell and intra-cell correlations of the measurement errors. However, this model possesses several apparent limitations, namely, a common intra-cell correlation for different cells, a constant inter-cell correlation for cells measured at the same time point, as well as the demanding for equally spaced measurements. Furthermore, SAS only provides the possibility to fit bivariate mixed models.

The second models with random coefficients provide an opportunity to examine correlations among the trajectories of cell counts over time and to capture the growth of these response variables. As exemplified by the HIV data, this approach is capable of handling equally spaced measurements and is easily extendable to multivariate models with more than two response variables. Another advantage of these mixed-model approaches is that they allow incomplete observations and thus can use information more efficiently.

## Conclusions

According to this study in the early stage of HIV infections, CD4 and CD8 steadily increase over time while the values of hemoglobin (HMG) decrease over time. There exists a strong positive correlation between the two cell counts of HIV infections. The correlations between the cell counts (CD4 and CD8) and hemoglobin levels are not significant in either measurement errors or random slopes, but CD4 and CD8 cells show a negative cross-lagged effect on hemoglobin levels, i.e., the higher the value of CD4 (CD8) at time  $t-1$ , the lower the hemoglobin measurement at time  $t$ .

## Declarations

### Ethical Approval and Consent to Participate

Ethical clearance had been obtained from Haramaya University, Dire Dawa, Ethiopia.

### Availability of Data and Materials

I used secondary data for this investigation obtained at University of Gonder Teaching and Specialized Hospital.

## Author Contributions

Author wrote the proposal, developed data collection format, supervised the data collection process and analyzed the data in consultation with Mr. Gashaw A. MSc in Public Health. He is also edited the document and gave critical comments. Author read and approved the final manuscript.

## Acknowledgments

In a special way, I wish to extend my sincere gratitude to Mrs. Elesabet G. for her support and guidance accorded me during this study. I would also like to acknowledge the contribution of my colleagues from whom I enjoyed fruitful discussion on challenging topics.

## Appendix

### Using SAS-Codes for Data Analysis.

### Kronecker Product Covariance Structure

Proc mixed data=HIV\_MIX covtest noclprint;

Title1 "model 1 mixed model with a kronecker product covariance";

Title2 "CD4 (X) versus CD8 (Y)";

Class PHID vname visit;

Model value=vname\*visit/s noint;

Repeated vname visit/ type=UN@AR(1) subject=PHID r rcorr;

Where vname="CD4" or vname="CD8";

### Random Coefficient Mixed Models

Proc mixed data=HIV\_MIX covtest noclprint;

Title1 "model 2 mixed model with random coefficients";

Title2 "CD4 (X) versus CD8 (Y)";

Class PHID vname;

Model value=vname\*time/s noint;

Random vname\*time/ type=UN subject=PHID g gcorr;

Repeated / type=VC group=vname subject=PHID;

Where vname="CD4" or vname="CD8";

## References

1. Hoefield RA, Kalra PA, Baker PG, Sousa I, Diggle PJ, Gibson MJ, et al. The use of eGFR and ACR to predict decline in renal function in people with diabetes, *Nephrology Dialysis Transplantation*. 2011; 26: 887-892.
2. Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen HV, Pressler T, Smyth RL, et al. Understanding the natural progression in % FEV1 decline in patients with cystic fibrosis: a longitudinal study, *Thorax*. 2012; 67: 860-866.
3. Yates JWT, Watson EM. Estimating insulin sensitivity from glucose levels only: use of a non-linear mixed effects approach and maximum a posteriori (MAP) estimation, *Computer Methods and Programs Biomedicine*. 2013; 109: 134-143.
4. Jang H, Conklin DJ, Kong M. Piecewise nonlinear mixed-effects models for modeling cardiac function and assessing treatment effects, *Computer Methods and Programs Biomedicine*. 2013; 110: 240-252.
5. Zeger SL, Irizarry R, Peng RD. On time series analysis of public health and biomedical data, *Annual Review of Public Health*. 2006; 27: 57-79.
6. Shelton BJ, Gilbert GH, Liu B, Fisher M. A SAS macro for the analysis of multivariate longitudinal binary outcomes, *Computer Methods and Programs in Biomedicine*. 2004; 76: 163-175.
7. Genolini C, Falissard B. Kml: A package to cluster longitudinal data, *Computer Methods and Programs in Biomedicine*. 2011; 104: 112-121.
8. Genolini C, Pingault JB, Driss T, Côté S, Tremblay RE, et al. A non-parametric algorithm for clustering joint trajectories, *Computer Methods and Programs Biomedicine*. 2013; 109: 104-111.
9. Weiss RE. *Modeling Longitudinal Data*, Springer-Verlag, New York. 2005.
10. Davis CS, Laird NM, Ware JH, Verbeke G, Molenberghs G. *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer Verlag. 2002.
11. Laird. Random-effects models for longitudinal data. *Biometrika*. 1982; 38: 963-974.
12. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer. 2009.

13. Singer JD. Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*. 1998; 23: 1998.
14. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc. 1996.
15. Thiebaut R, Jacqmin-Gadda H, Chene G, Leport C, Commenges D. Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*. 2002; 69: 249-256.
16. Newsom JT. A multilevel structural equation model for dyadic data. *Structural Equation Modeling*. 2002; 9: 431-447.
17. Galecki AT. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics-Theory and Methods*. 1994; 23: 3105-3119.
18. Chapman AB, Guay-Woodford LM, et al. Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): the Consortium for Radiology Imaging Study of Polycystic Kidney Disease (CRISP) cohort. *Kidney International*. 2003; 64: 1035-1045.
19. Henderson CR. *Applications of Linear Models in Animal Breeding*. Guelph Canada: University of Guelph Press. 1984.
20. Lyle C, Gurrin. Solving mixed model equations. New York: John Wiley and Sons. 2005; 120: 120-142.
21. MacCallum R, Austin JT. Application of Structural Equation Modeling in psychological research. *Annual Review of Psychology*. 2000; 51: 201-226.
22. Hatcher L. *A step-by step approach to using the SAS System for factor analysis and structural equation modeling*, Cary, NC: SAS Institute Inc. 1998.
23. Ormerod MG. Using the dispersion of laser light, that depends on the enzymes covering the cells of HIV infection. *Flow Cytometry* (Garland Pub).
24. Murray JD. *Mathematical Biology I. An Introduction to multivariate longitudinal data*. [Ed. Springer]. 2002.
25. Perelson AS, Weisbuch G. Immunology for physicists. *Rev Mod Phys*. 1997; 69: 1219-1267.
26. Perelson AS, Kirschner DE, De Boer R. Dynamics of HIV infection of CD4+ T cells. *Mathematical Biosciences*, 1993; 114, 81-125.
27. Perelson AS, Nelson PW. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM review*. 1999; 41: 3-44.
28. Perelson AS. Modeling viral and immune system dynamics *Nature Reviews Immunology*. 2002; 2: 28-36.
29. Diggle PJ. *Analysis of Longitudinal Data*, 2nd ed. New York: Oxford University Press. 2002.
30. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley and Sons, Inc, pp. 1989; 376-381.
31. Peng H, Lu Y. Model selection in linear mixed effect models. *Journal of Multivariate Analysis*. 2012; 109: 109-129.