

Application of Poisson Mixed Combined Models for Identifying Correlations of CD4 Count Progression in HIV Infected TB Patients During ART Treatment Period

Aboma Temesgen¹

¹ Department of Statistics, Haramaya University, Dire Dawa, Ethiopia

Correspondence: Aboma Temesgen, Department of Statistics, Haramaya University, Dire Dawa, Ethiopia.
E-mail: aboma.temesgen@gmail.com

Received: April 24, 2017 Accepted: May 8, 2017 Online Published: July 25, 2017

doi:10.5539/ijsp.v6n5p42 URL: <https://doi.org/10.5539/ijsp.v6n5p42>

Abstract

CD4 count is used to measure the number of CD4 cells in the blood mostly during ART treatment to know the risk progression of HIV in the HIV infected patients. This continuously measured CD4 count during the treatment period results in longitudinal data having correlation and over dispersion effects. While modeling such data to identify associated factors of change in CD4 count to monitor the progression of HIV most of the study did not consider these two main effects. The main aim of this study was also to consider these two main effects to identify the risk factors of CD4 count progression based on 239 HIV infected TB patients who were 18 years old and above taking ART treatment from 1st September 2009 to 1st July 2014 at Jimma University Specialized Hospital. The result of the study showed that the Poisson normal Gamma combined model which handles correlation and over dispersion effects of CD4 count simultaneously was an appropriate fit of the data among different Poisson mixed combined models considered for the study based on Akaike information criteria (AIC) comparisons. The estimated model depicts linear time and its interaction effect with functional status category group of the patients have a positive effect whereas quadratic time has a negative effect on the progression of CD4 count. The model also showed that baseline bedridden and ambulatory functional status group patients have lower average CD4 count measurements in comparison with working functional status group patients' counterparts. Therefore, while modeling CD4 count correlation and over dispersion should be taken into consideration since the CD4 count value was correlated due to repeated measurement and its variance larger than mean leading to over dispersion. Being at bedridden, ambulatory functional status at baseline in comparison with working functional status group and having quadratic time effects were also the associated risk factors that lower the CD4 count measurements of the patients during the ART treatment period at the study area.

Keywords: Correlation, Over dispersion, Combined model, Poisson Normal Gamma

1. Background of the Study

CD4 count measures the number of CD4 T lymphocytes (CD4 cells) in a sample of your blood. It is the most important laboratory indicator of how well the immune system is working, the strongest predictor of HIV progression and survival. It is also one of the key factors in determining both the urgency of antiretroviral therapy (ART) and survival according to findings from clinical trials and cohort studies (Mellors, JW. & et al., 1997; Mellors, JW., et al., 2002)

CD4 count was considered to hold predictive value for no more than the subsequent 6 month period, with individual patients contributing multiple 6 month periods of follow up. Lower CD4 counts are associated with greater risk of disease progression. This risk of progression to AIDS increases substantially at CD4 counts less than 350 cells/mm³, the greatest risk increase occurring as CD4 counts fall below 200 cells/mm³. The risk of disease progression at 200 cells/mm³, the threshold for ART initiation in resource limited settings (CASCADE, 2004)

Human Immunodeficiency Virus (HIV) infected patients may have TB infection either latent or active TB disease. HIV infection speeding up the progression from latent to active TB when the CD4 count of the patient was lowered and TB bacteria also accelerate the progress of HIV infection (Mayer, K., 2010). In 2013 of the estimated 9 million people who developed TB an estimated 1.1 million (13%) were HIV positive. There were also in 2013 360,000 deaths from HIV associated TB equivalent to 25% of all TB deaths, and around 25% of the estimated 1.5 million deaths from HIV/AIDS (Global Tuberculosis Control, 2014)

Human Immunodeficiency Virus increased risk of developing tuberculosis (TB) with losing of cell mediated immunity, along with a quantitative decline in circulating CD4 lymphocytes and tuberculosis occurs sooner than other opportunistic

infections. TB additionally contributes to reduction in CD4 count in HIV/TB co-infected patients and leads to greater improvement in count following treatment as compared to CD4 matched TB uninfected individuals (Wanchu, A. & et al., 2014)

In many medical and biomedical areas measuring the count outcome such as CD4 count during ART treatment is very common. When such data are collected longitudinally from a given subject repeatedly over time it results in repeated measurement of the observations within subject. However, Statistical modeling such data have several challenges because of this observations are correlated as the results of repeated measurements and over dispersed as the result of the variance of such data exceeds it's mean (Breslow, N. & Clayton, D., 1993; Wolfinger, R. & OConnell, M., 1993). Therefore, while statistical modeling such data one should have to consider these two effects in order to relate independent covariates with change in outcome variable during the measurement time.

CD4 count measurement is also among the count outcome measured for the HIV infected patients in order to know the progression of disease during treatment period. Various studies also reported factors related to change CD4 count measurement for HIV infected TB patients during their treatment period. But, there was no more studies considered the effects of correlation and over dispersion for this longitudinal measured CD4 count while reporting different factors related to change in CD4 count during the treatment period. An appropriate model, which is able to indicate related factors to change in CD4 during ART treatment, with an appropriate methodology, would provide support for possible therapeutic interventions and, consequently, a better quality of life and survival for patients. The main aim of this study was also to identify factors related to progression of CD4 count measurement of HIV infected TB patients using an appropriate Poisson mixed model with the consideration over dispersion which happens due to variance exceeds the mean of Poisson distribution and correlation effects which happens due to repeated measures of CD4 count within subject.

2. Materials and Methods

2.1 Data Source

Data for the study was obtained from HIV infected TB patients treated at Jimma University Specialized Hospital HIV outpatient Clinics, South West of Ethiopia. The study population consists of all HIV infected TB patients who were 18 years or older, and who were on ART treatment at any time between 1st September 2009 and 1st July 2014. Among 550 patients having the case at the Hospital, 239 patients who have at least one CD4 count measurement during the treatment period and having full record data were considered in the study. All the patients' epidemiological, laboratory and clinical information were collected from the patients chart of ART follow up retrospectively.

2.1.1 Variables of the Study

The outcome variable of this study was longitudinally measured CD4 count of HIV infected TB patients during ART treatment period. CD4 counts the number of cells per mm^3 of blood, which is an indication for the progression of HIV and measured approximately within 6 months interval during the treatment period. The 12 independent variables which was extracted from patients chart were listed with their categories on the Table 1 below.

Table 1. List of Covariates Considered in the Study

Variable name	Categories
Age	Years (Baseline)
Weight	Kilogram(Baseline)
Marital status	Single, Married,Others
Residence	Rural, Urban
Educational level	Not educated, Primary, Secondary, Tertiary
Use of alcohol	Yes, No
Smoking	Yes, No
Clinical stage	Stage-I, Stage-II,Stage-III, Stage-IV
Functional status	Working, Ambulatory, Bedridden
Religion	Muslim, Orthodox, Protestant
Sex	Female, Male
Current status	Active, Missed, Died, transferred

Notice that WHO Clinical Stage which is classified into four; I, II, III and IV; where Stage I indicates asymptomatic disease, Stage II indicates mild disease, Stage III indicates advanced disease and Stage IV indicates severe disease. Hence

disease severity increases from Stage I to Stage IV. Functional Status of the patients is also categorical covariate with three categories: Working, Ambulatory and Bedridden. Working patients are those patients who can able to work day to day while ambulatory patients are those patients who can able to work some time but bedridden patients cannot able to work due to the infectious disease. The marital status group also have three categories; married, single and others which includes separated, windowed and divorced individuals.

2.2 Ethical Consideration

Ethical clearance was obtained from Department of Statistics of Jimma University. Personal information was kept confidentially without disclosing to others during data collection from patient cards.

2.3 Methods of Data Analysis

2.3.1 Exploratory Data Analysis

Exploratory analysis of longitudinal data seeks to discover patterns of systematic variation across groups of patients, as well as aspects of random variation that distinguish individual patients. This study also considered individual profile plots to explore within and between subject variability and mean structure plots to explore the average progression of CD4 count measurement over the measurement time which helps as input for modeling.

2.3.2 Generalized Linear Mixed Model

Non-Gaussian repeated count data such as longitudinally measured diastolic and CD4 count to know the progression of patients during treatment period is most frequently modeled using generalized linear mixed model (GLMM) which allows the inclusion of subject specific random effects in the model (Breslow, N. & Clayton, D., 1993; Wolfinger, R. & OConnell, M., 1993; Engel, B. & Keen, A., 1992).

For such non Gaussian repeated measurement data let Y_{ij} be the outcome variable on the i^{th} subject measured at j^{th} time point and the random effect to be in the model for subject specific variation is \mathbf{b}_i having normally distribution with mean $\mathbf{0}$ and variance covariance matrices \mathbf{D} , then it is assumed that the conditional distribution of the outcome variable Y_{ij}/\mathbf{b}_i is independent and it belongs to the exponential family of distributions which is expressed as:

$$f_i(y_{ij}|\beta, \mathbf{b}_i, \phi) = \exp\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\} \tag{1}$$

Where the mean of the distribution is given by:

$$\eta(\psi'(\lambda_{ij})) = \eta(\mu_{ij}) = E(Y_{ij}|\beta, \mathbf{b}_i) = \mathbf{X}_{ij}^T\beta + \mathbf{Z}_{ij}^T\mathbf{b}_i \tag{2}$$

Where, $\eta(\cdot)$ is the known link function, \mathbf{X}_{ij} is a p-dimensional design matrix of the fixed effect parameters β , \mathbf{Z}_{ij} is a q-dimensional design matrix of the random effects \mathbf{b}_i and ϕ is scale (over dispersion) parameter. In this setting the likelihood contribution of i^{th} subject is given by:

$$f_i(y_{ij}|\beta, \mathbf{D}, \phi) = \int \prod_{j=1}^{ni} f_i(y_{ij}|\beta, \mathbf{b}_i, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \tag{3}$$

Based on the equation (3) the general likelihood which is the function of β, ϕ and \mathbf{D} is expressed as:

$$L(\beta, \mathbf{D}, \phi) = \prod_{i=1}^N \int \prod_{j=1}^{ni} f_i(y_{ij}|\beta, \mathbf{b}_i, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \tag{4}$$

This likelihood function in equation (4) does not have an analytical solution, and hence numerical approximations are needed and an extensive overview of different approximations for this likelihood is available in Molenberghs and Verbeke (1993) and Skrondal and Rabe-Hesketh (2004).

In case of CD4 count measurement which is specific to this study, let Y_{ij} be the value of CD4 count measurement measured on i^{th} patient at time point j^{th} has a Poisson distribution with parameter λ_{ij} . Then, the conditional of the mean is modeled as:

$$\lambda_{ij} = \exp(\mathbf{X}_{ij}^T\beta + \mathbf{Z}_{ij}^T\mathbf{b}_i) \tag{5}$$

2.3.3 Models Combining Over Dispersion with Random Effects

In most practical cases both over dispersion and correlation can happen together, and this led Molenberghs et al. (2010) to formulate a flexible and unified modeling framework which termed as combined model. This combined model simultaneously captures over dispersion and correlation for a wide range of data's including longitudinal count data's. In

this modeling framework the normally distributed subject-specific random effects capture the correlation due to repeated measurement while the conjugate measurement-specific random effects on the natural parameter, is used to accommodate over dispersion.

According to Molenberghs et al. (2010) the combined distribution model which accommodates both the over dispersion and normal random effects can be expressed in the form of:

$$f_i(y_{ij}|\beta, \mathbf{b}_i, \theta_{ij}, \phi) = \exp\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\} \tag{6}$$

This model expression has similar notation to that equation (1) except inclusion of θ_{ij} parameter which is included in this model to accommodate over dispersion effect. The conditional mean for this combine distribution is expressed as:

$$E(Y_{ij}|\beta, \mathbf{b}_i, \theta_{ij}) = \mu_{ij}^c = \theta_{ij}k_{ij} \tag{7}$$

Where; the random variable $\theta_{ij} = G_{ij}(\vartheta_{ij}, \sigma_{ij}^2)$, $k_{ij} = \exp(\mathbf{X}_{ij}^T\beta + \mathbf{Z}_{ij}^T\mathbf{b}_i)$, ϑ_{ij} and σ_{ij}^2 are the mean and the variance of θ_{ij} respectively which accommodate over dispersion in the model. As in GLMM \mathbf{b}_i has normal distribution and the link function is expressed as $\eta_{ij} = \mathbf{X}_{ij}^T\beta + \mathbf{Z}_{ij}^T\mathbf{b}_i$. The two different notations η_{ij} and λ_{ij} refer to the linear predictor and/or the natural parameter and to encompasses the random variables θ_{ij} respectively. Based on conditional mean parametrization of equation (7) which allows for the random effect θ_{ij} for capturing over dispersion the likelihood contribution of i^{th} subject is expressed as:

$$f_i(y_i|\beta, \mathbf{D}, \vartheta_i, \Sigma_i) = \int \prod_{i=1}^{n_i} f_{ij}(y_{ij}|\beta, \mathbf{b}_i, \theta_i) f(\mathbf{b}_i|\mathbf{D}) f(\theta_i|\vartheta_i, \Sigma_i) d\mathbf{b}_i d\theta_i \tag{8}$$

From this the general likelihood function is given by:

$$\begin{aligned} L(\beta, \mathbf{D}, \vartheta_i, \Sigma_i) &= \prod_{i=1}^N f_i(y_i|\beta, \mathbf{D}, \vartheta_i, \Sigma_i) \\ &= \prod_{i=1}^N \int \prod_{i=1}^{n_i} f_{ij}(y_{ij}|\beta, \mathbf{b}_i, \theta_i) f(\mathbf{b}_i|\mathbf{D}) f(\theta_i|\vartheta_i, \Sigma_i) d\mathbf{b}_i d\theta_i \end{aligned} \tag{9}$$

Where; $E(\theta_i) = E([\theta_{i1}, \theta_{i2}, \dots, \theta_{ini}]^T) = \vartheta_i$ and $\text{Var}(\theta_i) = \Sigma_i$

Specific to this study, let Y_{ij} be the CD4 count measurements for subject $i=1, 2, \dots, N$ at time point $j=1, 2, \dots, n_i$ has a Poisson distribution where the normal random effect is introduced to accommodate correlation due to repeated measurement of CD4 count within subject and gamma random effect is included to accommodate over dispersion due to the variance CD4 count was large than its mean. Accordingly, Poisson model with normal and gamma random effects can be specified as follows:

$$Y_{ij} \sim \text{Poi}(\lambda_{ij} = \theta_{ij}k_{ij}) \tag{10}$$

The conditional mean $\lambda_{ij} = \theta_{ij}k_{ij}$ is modeled by:

$$k_{ij} = \exp(\mathbf{X}_{ij}^T\beta + \mathbf{Z}_{ij}^T\mathbf{b}_i) \tag{11}$$

Where; $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ and $\theta_{ij} \sim \text{Gamma}(\alpha, \nu)$, \mathbf{X}_{ij} and \mathbf{Z}_{ij} p-dimensional and q dimensional matrices of known covariate for values β and \mathbf{b}_i a unknown fixed and random vector coefficients respectively.

For the distribution of the outcome variable in equation (10) if θ_{ij} over dispersion random effects is omitted the resulting model become Poisson Normal. Similarly, if we omit the normal random effect in this equation the resulting model becomes Poisson Gamma model. To estimate the all parameters of the model SAS NLMIXED procedure which is used to fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects using adaptive Gaussian quadrature was considered.

2.3.4 Model Comparisons

Before proceeding to model comparisons back ward automatic variable selection techniques was employed to exclude non-significant variable from the given generalize linear mixed model. The automatic variable selection was done to fit all models with the same variable for the sake of comparisons among the models.

To select an appropriate model among candidate models considered in the study that associates predictor variables with change in CD4 count over time -2log-likelihood and AIC (Sakamoto, Y. & et al., 1986) of the models were considered. Among these candidate models the model having minimum value of -2log-likelihood and AIC was considered as an appropriate fit of the data where the AIC value is given by:

$$AIC = -2\log - \text{lik} + 2p$$

Where; $\log\text{-lik}$ and p are the log likelihood and number of estimated parameters respectively.

3. Results

3.1 Descriptive Results

As indicated on Table 2 of the baseline characteristics of the patients with baseline and over all average CD4 count during treatment period the average age of the patients were 32.11 years with standard deviation 8.88 years at baseline whereas; the average weight of the patients was 48.45 kilogram with standard deviation 10.78 at baseline. The distribution of the patients with their sex also indicated 134(56.07%) and 105(43.93%) of them were male and female respectively whereas larger average CD4 count (184.94) was observed in male category group in comparison with females category.

The residence of the patients also indicates that 203(84.94%) of them were from urban areas whereas about 36(15.06%) of the were from the rural areas of the Jimma town. The educational level distribution of the patients depicts 53(22.18%), 104(43.51%), 67(28.03%) and 15(6.28%) of the patients were non educated, primary, secondary and tertiary educated individuals respectively. The lower average baseline CD4 count(96.733) with standard deviation value of 75.029 was observed in tertiary educated patients category group.

The baseline functional status of the patients also indicated 120(50.21%), 97(40.59%), 22(9.21%) of them were at ambulatory, working and bedridden functional status group respectively. Furthermore, the average baseline CD4 count by functional status also depicts lower average CD4 count (74.64) with standard deviation (78.15) was observed in bedridden functional status group in comparison with ambulatory and working functional status category groups.

The WHO clinical stages categories which shows the level of severity of the disease of the patient describes that 29(12.14%), 115(48.12%) and 95(39.75%) of the patients were at clinical stage I and II, III and IV respectively. The lower average CD4 count (155.58) with standard deviation(118.55) was observed in clinical stage III patients in comparison with clinical stage patients at baseline.

The average CD4 count during treatment period also helps to know over all average value of CD4 count by each category group which was from when the patient started the treatment to the end of the study or during 48 months visit. Accordingly, the all possible average CD4 count during the treatment period lied between 122.36 with standard deviation 121.10 and 320.32 with standard deviation 188.81 which was observed in bedridden functional status category group and protestant religious category group respectively.

Table 2. Baseline Demographic and Clinical Characteristics of the Patients

Covariates	Categories	Count (%)	Baseline CD4 count	Over all CD4 count
			Mean (SD)	Mean (SD)
Sex	Female	105(43.93)	157.83(153.32)	314.82(223.49)
	Male	134(56.07)	184.94(145.11)	284.63(183.37)
Functional Status	Ambulatory	120(50.21)	155.78(147.84)	284.21(194.38)
	Bedridden	22(9.21)	74.636(78.15)	122.36(121.10)
	Working	97(40.59)	216.69(148.81)	334.08(206.97)
Clinical stage	I and II	29(12.14)	236.10(155.64)	331.99(179.27)
	III	115(48.12)	155.58(118.55)	280.15(193.28)
	IV	95(39.75)	174.895(174.35)	306.11(223.22)
Patient status	Active	124(51.88)	174.79(139.61)	316.75(209.28)
	Died	29(12.13)	148.45(150.95)	244.55(202.10)
	Missed	47(19.67)	136.19(137.63)	261.49(192.31)
	Transferred out	39(16.32)	230.10(175.71)	281.58(160.18)
Religion	Muslim	87(36.40)	155.83(132.38)	306.33(195.83)
	Orthodox	136(56.90)	172.16(146.585)	290.26(209.32)
	Protestant	16(6.69)	273.938(214.77)	320.32(188.81)
Educational level	Non educated	53(22.18)	193.43(135.64)	303.32(194.74)
	Primary	104(43.51)	201.01(171.16)	325.22(219.12)
	Secondary	67(28.03)	130.54(118.32)	263.19(182.64)
	Tertiary	15(6.28)	96.73(75.03)	240.15(155.79)
Marital status	Married	105(43.93)	176.68(152.97)	309.85(198.29)
	Single	73(30.54)	184.69(153.03)	295.09(212.55)
	Others	61(25.52)	152.08(136.38)	282.03(197.30)
Residence	Rural	36(15.06)	152.19(152.87)	253.04(155.95)
	Urban	203(84.94)	176.72(148.46)	305.66(208.79)
Alcohol use	No	146(61.09)	182.39(155.80)	317.08(209.87)
	Yes	93(38.91)	158.33(137.37)	263.57(185.01)
Smoking	No	181(75.73)	180.40(154.10)	310.24(207.01)
	Yes	58(24.27)	150.03(130.71)	254.24(181.28)
		Mean		SD
Age		32.11		8.88
Weight		48.45		10.78

SD = Standard Deviation

3.1.1 Average CD4 Count Measurement During the Treatment Period

As shown on Table 3 of average CD4 count and its standard deviation during the measurement period for all measurement period the variance which is square of standard deviation was larger than the mean of CD4 count measurement. But, in theoretic for count data assuming Poisson distribution the mean must be equal to the variance and the description this data indication the existence of over dispersion.

The trend of average CD4 count indicates the average CD4 count has increasing value from the baseline time (173.03) up to 24th months (401.91) and has decreasing value after 24th month to end of the follow up period. Similarly, standard deviation value which describes the variation of CD4 count has also the increasing trend from baseline (149.91) to 18th months (208.33) and has decreasing trend from 30th months to the last of the follow up period.

Table 3. Mean and Standard Deviation of CD4 Count During Measurement Time

Time	0	6	12	18	24	30	36	42	48
Mean	173.03	287.27	350.77	401.90	401.91	423.23	383.44	293.33	220.67
SD	149.07	178.76	191.11	208.33	204.67	211.26	198.83	122.76	35.23

SD = Standard Deviation

3.1.2 Exploratory Data Analysis

Before proceeding to model building an exploratory data analysis is essential in longitudinal data to identify variance structure, the mean structure, fixed and random effects to be included in the model. Accordingly, the CD4 count measurement of the patients was explored using individual profile plots and mean structure plots which was described as follows:

The individual profile plot of Figure 1 was plotted by considering CD4 count some of the patients to have clear visualization for exploration. This plot clearly depicts the existence within and between variability in CD4 count measurements. This is, therefore, an indication for inclusion of random intercept and slope of the CD4 count to capture subject specific variability of the CD4 count since the plot indicated the existence of within and between variability of CD4 count measurements.

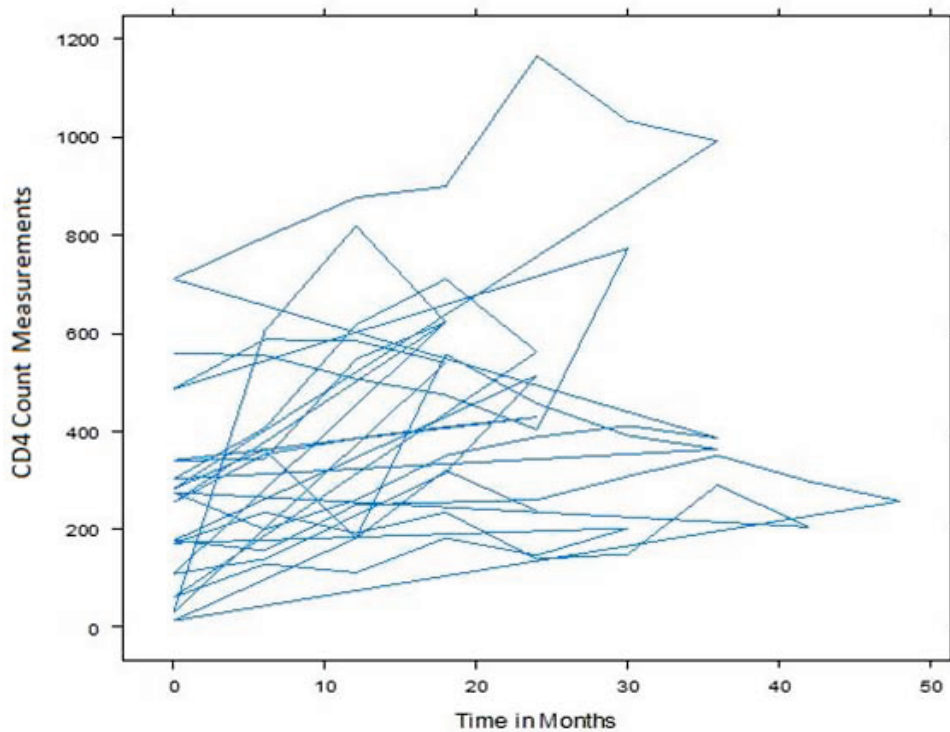


Figure 1. Individual profile plot

The mean structure plots of Figure 2 below also shows the mean progression of CD4 count during treatment period and helps to explore an appropriate fixed time effects to be included in the model. As observed from the plot the red line plot was the progression of CD4 count and the black line was the progression of CD4 count with loess smoothing techniques.

The average CD4 count progression plot depicts the quadratic change of CD4 count over time with some linear effects whereas the average CD4 count progression with loess smoothing technique plots depicts the linear change of CD4 count with some quadratic effects. Hence, the data is unbalanced longitudinal data mean structure plot with loess smoothing techniques was preferable to determine the mean structure of CD4 count evolution. Accordingly, since the plot with smoothing techniques shows linear and some quadratic change of CD4 count over time linear and quadratic time effects were considered in the model.

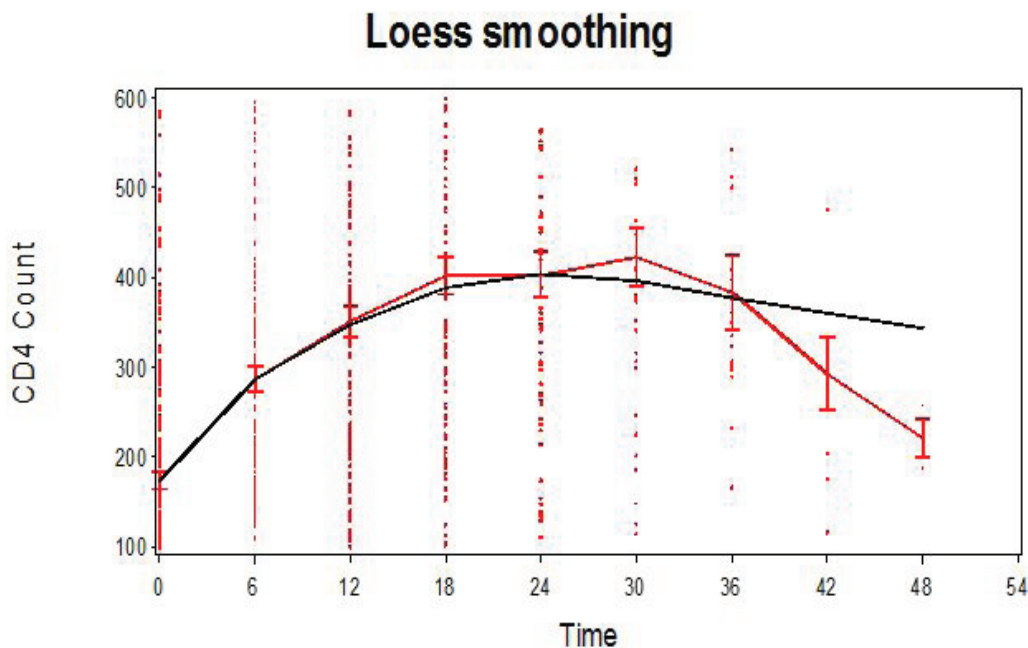


Figure 2. Mean profile plot

3.2 Models Comparison

To select an appropriate model as observed in Table 4 different models were fitted with the consideration of different random effects and different combined models were also considered. This model comparison was made to come up with an appropriate model that handles the progression of CD4 count measurement over time with associated factors. Among these candidate models the estimated parameters of Poisson mixed model has smaller estimated values than the parameters of the PN combined model. But, PM model standard error of estimated parameters were somewhat larger than that of the PN combined model. The AIC value of these two models indicates PM has a smaller AIC value than the PN combined model showing PM model is preferable than the PN combined model in fitting the data but these two models do not consider the over-dispersion effect.

Similarly, the comparison of the Poisson combined model with the gamma (PG) and gamma and normal random effects (PNG) depicts the PG model has larger estimated values of the fixed effects coefficient and standard errors than the PNG combined model parameters. This PG combined model only considered the over-dispersion effects with exclusion of correlation effects which arise due to repeated measurements of CD4 count. The estimated AIC value of this PG was also larger than the PNG combined model showing the PG combined model was not an appropriate fit of the data than the PNG.

The AIC and $-2\log$ -likelihood comparison of these four models indicated the PNG combined model has the minimum AIC among the remaining candidate models and is considered as an appropriate fit of the data. This result was an indication that the PNG combined model was preferable because this model takes into account correlation due to the repeated measurements of CD4 count and over-dispersion which is due to the variance of the CD4 count measurement being larger than its mean value simultaneously. Kassahun et al. (2014) and Molenberghs et al. (2010) also used this proposed flexible and unified modeling structure of the PNG model to simultaneously capture over-dispersion and correlation for a wide range of longitudinal data including count, binary and time-to-event responses.

Finally, to look for the improvement of the PNG combined model the model was refitted with the consideration of different random effects for the correlation effects due to the repeated measurements. Among these refitted PNG candidate models the PNG model which was fitted with a random intercept having the minimum AIC values was considered as an appropriate model than the PNG combined models which were refitted with consideration of a random time slope and both a random time slope and intercept.

Table 4. Comparison of Candidate Models

Effects	Combined Models			
	PM	PN	PNG	PG
	Est.(SE, p-value)	Est.(SE,P-Value)	Est.(SE, P-value)	Est.(SE,P-value)
Intercept	5.27(0.090,<.001)	5.30(0.010,<.001)	5.33(0.117,<.001)	5.47(0.080,<.001)
Time	0.04(0.001,.0001)	0.04(0.001,<.0001)	0.05(0.006,<.0001)	0.06(0.008,<.0001)
Functional status				
Working (ref.)				
Ambulatory	-0.36(0.106, 0.001)	-0.44(0.009, <.0001)	-0.33(0.097,0.001)	-0.26(0.072,0.0004)
Bedridden	-1.21(0.184,.0001)	-1.36(0.022,<.0001)	-1.13(0.168,<.0001)	-0.98(0.134,<.0001)
Marital status				
Single				
Married	-0.17(0.057,0.003)	-0.20(0.010,<.0001)	-0.05(0.108,<.0001)	-0.09(0.081,0.267)
Others	0.26(0.085,0.002)	0.22(0.011,<.0001)	-0.18(0.122,0.647)	-0.23(0.089,0.011)
Time*Ambulatory	0.01(0.001,<.0001)	0.01(0.001,<.0001)	0.01(0.004,0.134)	0.01(0.005,0.104)
Time*Bedridden	0.04(0.002,<.0001)	0.04(0.0015,<.0001)	0.04(0.011,0.001)	0.01(0.012,0.225)
Time*Married	0.004(0.001,<.0001)	0.004(0.001,<.0001)	0.004(0.005,0.001)	0.009(0.005, 0.091)
Time*Others	0.010(0.001,<.0001)	0.010(0.001,<.0001)	0.011(0.005,0.024)	0.014(0.006,0.011)
Time**2	-0.001(0.001,<.000)	-0.001(0.0001,<.0001)	-0.001(0.001,<.0001)	-0.001(0.0002,<.0001)
Random effects				
σ^2_{b0}	0.59(0.058,<.0001)	0.11(0.0013,<.0001)	0.31(0.05,<.0001)	-
Over dispersion(α)	-	-	4.73(0.315,<.0001)	2.37(0.115,<.0001)
$\frac{1}{\text{Overdispersion}}(v)$	-	-	0.21(0.014,<.0001)	0.42(0.021,<.0001)
Comparison with AIC and log likelihood				
Criteria	PM	PN	PNG	PG
-2loglik	20038.16	29712	9849.3	10040
AIC	20062.16	29736	9875.3	10064
Comparison of PNG with different random effects				
Criteria	Random intercept only	Random intercept and time slope	Linear time slope only	
-2loglik	9849.3		10403	10354
AIC	9875.3		10433	10380

PM = Poisson mixed, PN= Poisson normal, SE= Standard Error
 PNG = Poisson Normal Gamma, PG= Poisson Gamma, Est.= Estimate

3.2.1 Factors Associated with Progression of CD4 Count Measurement

As can be observed from Table 5 linear time have positive effects whereas the quadratic time has negative effects on the progression of CD4 count measurements. The estimated coefficients of for the functional status category group which was negative depicts that ambulatory and bedridden functional status category patients had lower CD4 count measurements in comparison with working functional status category counterparts. Aboma T. and Teshome K. (2016) on their joint modeling of CD4 count and measurement in HIV/TB co-infected patients also revealed that linear time positive effect whereas quadratic time and being at bedridden functional status at baseline has negative effects on the square root of CD4 count measurements.

The negative coefficient for the linear time interaction with functional status of the patients also depicts both ambulatory and bedridden functional status group patients had larger CD4 progression in comparison working functional status category patients. The interaction of linear time with marital status also indicates that others marital status category patients had larger CD4 count progression than single marital status category group patients.

The estimated coefficient for the linear time effect(0.05, p-value=0.0001) depicts that the long of expected CD4 count of the patient increased with 0.05 with unit increment of time in moth whereas for the quadratic time (-0.001, p-value =0.0001) depicts the log of expected CD4 count was decreased by 0.001 with unit increment in quadratic time effect holding other variables constant. Similarly, the estimated coefficients of functional status categories for ambulatory group (-0.33, p-value =0.0006) and bedridden group (-1.13, p-value= 0.0001)depicts that the log of expected CD4 count in ambulatory and bedridden functional status category groups were 0.33 and 1.13 lower than working functional status category group patients respectively holding other variables constant. Furthermore, the linear time interaction with ambulatory functional status coefficient (0.01, p-value=0.0014) depict the log of expected CD4 count was 0.01 larger than working functional status patients category group during the treatment period holding other variables constant.

The estimated Poisson model coefficients can be also interpreted as the expected incidence rate ratios by exponentiation of the estimated model coefficients. This incidence rate ratio depicts how the expected rate CD4 count increased or decreased with unit increment holding other variables in the model constant if the estimated coefficient was a continuous variable.In the case of categorical variable the estimated incidence rate ratio depicts how the rate of expected CD4 count for the estimated category group was lower or larger in comparison with the reference group counterpart.

Accordingly, the estimated coefficient for the linear time effect (0.05) depicts the expected rate of CD4 count measurement of the patients was increased with rate of 1.05(exp(0.05)) with unit increment of time month whereas the estimated coefficient for quadratic time effect (-0.001) depicts the expected of CD4 count was decreased with the rate of 0.10(exp(-0.001)) with unit increment of quadratic time effects holding other variables constant. Similarly, the estimated coefficient for bedridden functional status group (-1.13) depicts the expected rate of CD4 count in bedridden functional status group was 0.32(exp (-1.13)) lower than that of working functional status group holding other variables constant.

Table 5. The Selected PNG Combined Model With Estimated IRR

Parameter	Estimate(SE)	P-Value	95% Confidence Interval	IRR
Intercept	5.33(0.117)	0.0001	[5.10, 5.56]	206.56
Time	0.05(0.006)	0.0001	[0.04, 0.06]	1.05
Functional status				
Working(ref.)				
Ambulatory	-0.33(0.097)	0.0006	[-0.52, -0.14]	0.72
Bedridden	-1.13(0.168)	0.0001	[-1.46, -0.801]	0.32
Marital Status				
Single(ref.)				
Married	-0.05(0.108)	0.647	[-0.26,0.16]	0.95
Others	-0.18(0.122)	0.1342	[-0.43, 0.06]	0.83
Time* Ambulatory	0.01(0.004)	0.001	[0.01, 0.020]	1.01
Time*Bedridden	0.04(0.011)	0.001	[0.02, 0.06]	1.04
Time*Married	0.01(0.005)	0.296	[-0.004, 0.014]	1.01
Time*Others	0.01(0.005)	0.0238	[0.002, 0.021]	1.01
Time**2	-0.001(0.0001)	0.0001	[-0.0014, -0.0009]	0.10
Random effects				
σ_{b0}^2	0.31 (0.053)	0.0001	[0.20, 0.41]	1.36
Over dispersion(α)	4.73(0.315)	0.0001	[4.11,5.35]	112.91
$\frac{1}{Overdispersion}(\nu)$	0.21(0.014)	<0.0001	[0.18,0.24]	1.24

SE = Standard Error, IRR= Incidence Rate Ratio

4. Conclusions

The Poisson normal gamma (PNG) combined model which considers simultaneously correlation effects of CD4 count due repeated measurement and over dispersion effect due the variance of the CD4 count measurement was larger than it's mean of the distribution was considered as an appropriate fit of the data. Being at bedridden and ambulatory functional status group at base line in comparison with working functional status group patients at baseline and quadratic time effects were the risk factors that lowers the CD4 count in HIV infected TB patients at the study area.

References

- Aboma, T., & Teshome, K. (2016). Joint Modeling of Longitudinal CD4 Count and Weight Measurements of HIV/Tuberculosis Co-infected Patients at Jimma University Specialized Hospital. *Springer, Ann. Data. Sci*, 3(3), 321-338.
- Breslow, N., & Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25. <https://doi.org/10.1080/01621459.1993.10594284>
- CASCADE (2004). Collaboration: Short-term risk of AIDS according to current CD4 cell count and viral load in anti retro viral drug naive individuals and those treated in the mono therapy era. *AIDS*, 18(1), 51-58. <https://doi.org/10.1080/01621459.1993.10594284>
- Egger, M., May, M., Chene, G., & et al. (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: A collaborative analysis of prospective studies. *Lancet.*, 360(9327), 119-129. [https://doi.org/10.1016/S0140-6736\(02\)09411-4](https://doi.org/10.1016/S0140-6736(02)09411-4)
- Engel, B., & Keen, A. (1992). A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, 48, 1-22.
- Global Tuberculosis Control. (2014).
- Kassahun, W., Neyens, T., ..., & Verbeke, G. (2014). Modeling Hierarchical Data, Allowing for Over-dispersion and Zero

Inflation, in Particular Excess Zeros. *Stat Med*.

- Mayer, K. (2010). Synergistic Pandemic: Confronting the Global HIV and Tuberculosis Epidemics. *Journal of Clinical Infectious Diseases*, 50(3), S67.
- Mellors, JW., Munoz, A., Giorgi, JV., & et al. (1997). Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med*. 126(12), 946-954. <https://doi.org/10.7326/0003-4819-126-12-199706150-00003>
- Molenberghs, G., & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Molenberghs, G., Verbeke, G., Demtrio, C., & Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statist Sci*, 25(3), 325-347. <https://doi.org/10.1214/10-STS328>
- Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. D. Reidel Publishing Company, Dordrecht.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling*. London: Chapman and Hall/CRC. <https://doi.org/10.4103/0377-4929.72070>
- Wanchu, A., Kuttiatt, V., ..., & Varma, S. (2014). CD4 cell count recovery in HIV/TB co-infected patients versus TB uninfected HIV patients. *Indian Journal of Pathology and Microbiology*, 53(4), 745-49. <https://doi.org/10.4103/0377-4929.72070>
- Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: a pseudo likelihood approach. *Journal of Statistical Computation and Simulation*, 48, 233-243. <https://doi.org/10.1080/00949659308811554>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).