International Journal of Health Sciences
March 2018, Vol. 6, No. 1, pp. 56-63
ISSN: 2372-5060 (Print), 2372-5079 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/ijhs.v6n1a7
URL: https://doi.org/10.15640/ijhs.v6n1a7

Joint Longitudinal and Survival Modeling of HIV in the Upper West Region of Ghana

Abdul Basit Abdul Rahaman[1] Albert Luguterah[2] Nurudeen Abdul-Rahaman[3] Suleman Nasiru[4] & Salma S. Abdul-Rahaman[5]

Abstract

The main aim of this research was to develop a joint longitudinal and survival model for HIV patients in the Upper West Region, Ghana. The study revealed that AZT/3TC/NVP drug regimen contributed better to patients' survival time and CD4 count growth using Weibull model and linear mixed effects model respectively. The two models further unveiled that factors such as preCD4 count, gender, and duration of treatment (months) significantly determined HIV patient's CD4 count, whilst drug regimen, age and preCD4 determined the survival of the patient. Results from the joint model of the longitudinal sub-model and survival sub-model indicated a significant association between the repeated CD4 count measurement and survival time of HIV patients.

Keywords: CD4 count, longitudinal modeling, survival modeling, HIV virus, Anti-Retroviral Therapy.

1. Introduction

In medical studies, periodically measured biomarkers are used to monitor the progression of a disease so as to enable health caregivers to prescribe the appropriate treatment for a patient at a particular period. One of such area broadly researched into is the dynamics of the CD4 count in Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome (HIV/AIDS) patient. HIV is a virus that attacks and destroys the infection-fighting CD4 cells of the body's immune system. The CD4 cells are therefore the important component of the human immune system which begins to deplete as the viral load progresses high. These cells are of two types; T-4 cells otherwise called CD4+ ('helper' cells) and T-8 cells normally referred to as the CD8+ cell ('suppressor' cells). The CD4+ leads the attack against infections whilst the CD8+ ends immune response and kills cancer cells and cells infected with the virus. The T-4 cells are normally referred to as CD4 because they have CD4 molecules on their surface. It is regarded as the fundamental indicator for prognostic information and a guide as to whether HIV infected individual is qualified for antiretroviral therapy. Moreover, the efficacy of antiretroviral therapy reflects on the initial viral decay rate and the increment of the CD4 count, hence the disease progression is delayed significantly (Ding & Wu, 2001; Nelson et al. 2007). The significant role of CD4+ cell in the survival of HIV/AIDS patients called for research into it dynamics. A multivariate regression conducted to evaluate predictors of CD4+ count and HIV-1 RNA levels in a multicenter AIDS cohort study, revealed that old men of 50yeasr and above on Highly Anti Active Retroviral Therapy (HAART) had their mean CD4+ adjusted upwards per microliter and similar pattern but quite higher adjustment of men CD4+ count was found in younger men (Xiuhong et al. 2011). Maintaining higher CD4 count and complete cessation of smoking may reduce the risk of non-AIDS-defining cancer among patients (Krishnan et al., 2011). In Ghana, Adams & Luguterah (2013) conducted a study in the Builsa District hospital and revealed that there is a strong association between treatment period and CD4 count increment. Meanwhile, the association between hazard rate of death and the longitudinal progression of CD4+ counts has received little attention and as such this research sought to jointly model the repeated measurement of the CD4 count and the hazard of death.

[1]Ghana Education Service, sfbasit@gmail.com, Nurudeen Abdul-Rahaman, University of Science and Technology of China, nurudeenkuberke@gmail.com
[2]University for Development Studies, Ghana, NasiruSuleman, University for Development Studies, Ghana

2. **Material and Method**

2.1 **Study area**

The Upper West Region is one of the ten regions of Ghana. It is located at the North Western corner of Ghana with latitude 9.8°- 11.0° North and longitude 1.6°- 3.0 west, bordered by Upper East region to the east, Northern region to the south, and Burkina Faso to the west and north. The region has a total population of 576,583, of whom 276,445 (47.9%) are males and 300,138 (52.1%) females (2010 census).

2.2 Data and Source

Secondary data were obtained at the public health departments specifically the HIV/AIDS unit at Jirapa District Hospital and the Regional Hospital, where enrolled clients CD4 counts were monitored regularly. Data collection was restricted to individuals infected with HIV patients on treatment, who had their therapy during January 2006 to December 2014 and are ten (10) years and above. Members under the study were censored if they dropped along the study period or failed to experience the event. Both longitudinal and survival data were obtained from individual patient's folders. The two outcome variables considered in this study included CD4 cell count per mm^3 of blood measured repeatedly for approximately every six months' interval visits and survival endpoint or time to event is death. A common measuring (observation) time limit is used for all patients where the maximum observation time considered to be on the 78th month (such that$n_i \leq 12$). The survival response in months was obtained by subtracting the date of entry for treatment from the date of the last visit. Nine (9) potential explanatory variables were considered in this study viz; Drugs, PreCD4, Alcohol, Smoking, Time, Education, Marital status, Gender, and Age.

2.3 **Methods of Data Analysis**

2.3.1 Longitudinal modeling

In a longitudinal or follow up studies, observations on individuals are repeatedly measured over time. As a result, modeling with simple linear regression will not be appropriate because of the assumption of independent observations. In that regards, a linear mixed effect is designed in which the repeated measurements using linear regression model are fitted where parameters vary across subjects. It takes into accounts within and between sources of variation; flexible enough to account for the natural heterogeneity population and can handle any degree of missing drop out data in the longitudinal data. In addition, each subject has a subject-specific mean response over time with each i^{th} subject measured at times $S_{i1} \ldots S_{in_i}$ model as;

$$y_i = \mu_i(s) + W_{1i}(s) \quad (2.1)$$

$$= X_{1i}^T(s)\beta_1 + Z_{1i}^T(s)b_i + \varepsilon_{i,} \quad (2.2)$$

where

$$\mu_i(s) = X_{1i}^T(s)\beta_1 \quad (2.4)$$

$$W_{1i}(s) = Z_{1i}^T(s)b_i \quad (2.5)$$

$$\mathrm{b_i} \sim \mathrm{N_q}(0, \psi) \varepsilon_\mathrm{i} \sim \mathrm{N}_{n_i}(0, \sigma^2 I)$$

y is an $n_i$dimensional vector of observed responses.

β_1 is a $p$dimensional vector of fixed effects.

b_i is a $q$dimensional vector of random effects.

$X_{1i}^T(s)$ a matrix of size $(n \times p)$ fixed effects possibly time-varying covariates is the mean response.

$Z_{1i}^T(s)$ is a matrix of size $(n \times q)$random effects covariates that incorporates random effects in the model.

2.3.2 Survival Data modeling

Both parametric and semi parametric models are available to model the survival data and the models used include; Weibull, exponential, lognormal and log-logistic including Cox proportional hazards models.

The survival model takes the form of a proportional hazard model,

$h_i(t) = h_0(t)exp(W_{2i}(s) + \emptyset \boldsymbol{v_i})$ (2.6)

where;

$h_0(t)$ Represent the base line hazard function.

$\boldsymbol{v_i} \in \cup_{\boldsymbol{i}}$is a vector of base line covariates with corresponding log hazard ratios $\emptyset$

$W_2(s)$ is similar to $W_1(s)$ in equation (2.5) .

Finally, the joint model consists of the two linked sub models obtained from equation (2.1) and (2.6)

$$h_i(t) = h_0(t)\exp\{\alpha W_1(s) + \emptyset \boldsymbol{v_i}) \quad (2.7)$$

The parameter α indicates the strength of association between the longitudinal biomarker and the time-to-event. If $\alpha = 0$ then the joint model is reduce to separate models and fitting a joint model will not yield any advantage.

2.3.3 Likelihood for the Joint model

Constructing the likelihood for the joint model,

$$\prod_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{m_i} f\left(y_i(t_{ij})|b_i,\theta\right)\right) f(b_i|\theta) f(T_i,d_i|b_i,\theta)db_i\right], \quad (2.8)$$

where

$$f\left(y_i(t_{ij})|b_i,\theta\right) = (2\pi\sigma_e^2)^{-1/2} exp\left\{-\frac{y_i(t_{ij})-W_i(s)}{2\sigma_e^2}\right\}, \quad (2.9)$$

$$f(b_i|\theta) = (2\pi|V|)^{-1/2} exp\left\{-\frac{b_i{}'V^{-1}b_i}{2}\right\}, \quad (2.10)$$

The likelihood component under the Weibull or exponential sub model can be express as;

$$f(T_i,d_i|b_i,\theta) = [h_o(T_i)exp(\alpha W_i(s) + \emptyset v_i)]^{d_i} exp\left\{-\int_0^{T_i} h_0(u)exp(\alpha W_i(u) + \emptyset v_i)du\right\} \quad (2.11)$$

2.3.4 Model selection criteria

Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) are indices of relative goodness-of-fit and was used to compare models with the same fixed effects but different covariance structures. Both of these criteria apply rather generally for purposes of model selection and hypothesis testing. Smaller AIC or BIC values show a better fit. However, the BIC has preferred if the distribution has a sufficiently large sample size because it penalizes more severely than the AIC does. In that regard, the two criteria will not always agree on the choice of best model or hypothesis since our objective is parsimony, hence BIC is more reliable than the AIC criterion.

$$AIC = L(\hat{\theta}) - q \quad (3.11.0)$$

$$BIC = L(\hat{\theta}) - (q+2)\,log(N^*) \quad (3.11.1)$$

where;

$L(\hat{\theta})$ is the maximized log-likelihood (ML) or restricted maximized log-likelihood (REML).

q is the number of parameters in the covariance matrix,

$P$is the number of fixed effect parameters.

N is the number of subjects.

3.0 Data Analysis and Discussion of Results

3.1 Separate analysis of longitudinal data

To ensure that the longitudinal data meet the normality assumption, a Q-Q plot of the CD4 count was plotted and it was realized that most of the values deviated from the normal line as indicated in Figure 3.1a. As a result, a square root transformation was performed on the CD4 count and after which it was observed that the data has attained normality as shown in Figure 3.1b

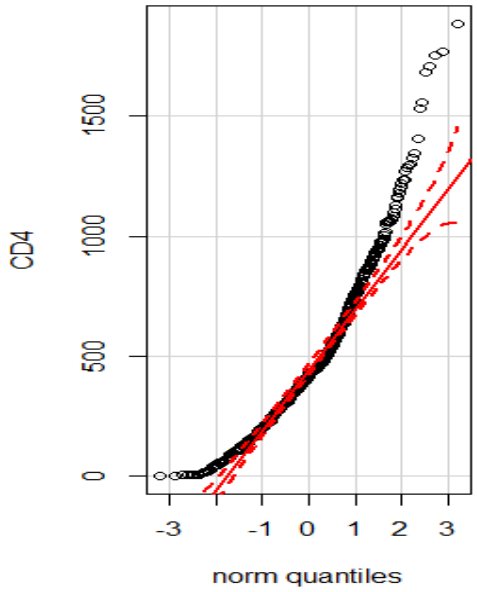

Figure 1aQ-Q plot of CD4 count

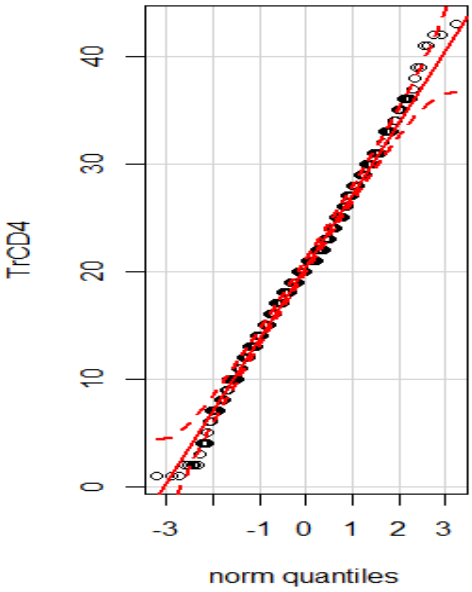

Figure 1b Q-Q plot of $\sqrt{\mathbf{CD4}}$ count

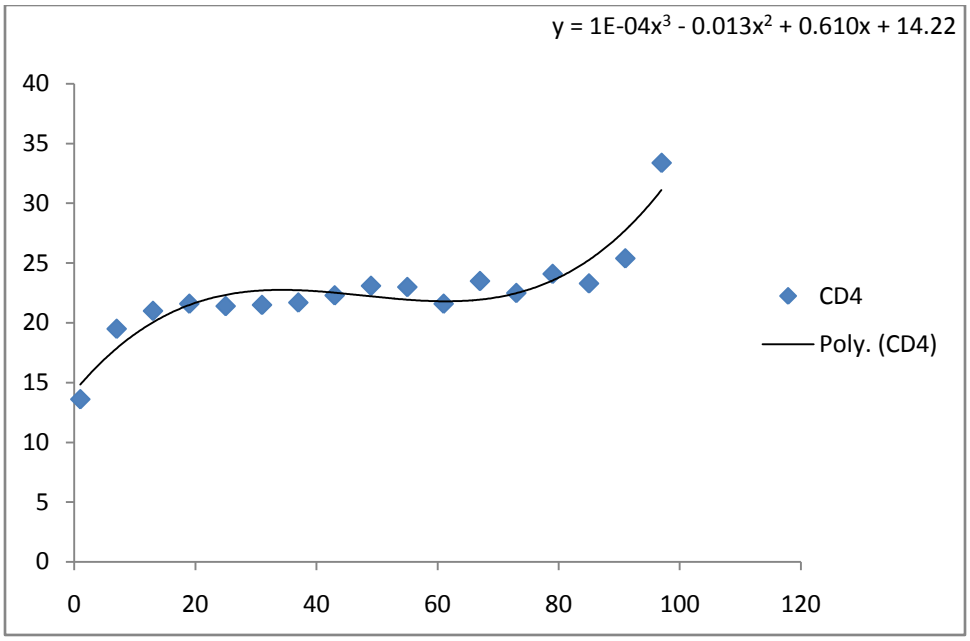

Figure 2 General pattern of CD4 count over time

3.1.1 Exploring the general mean structure of CD4 count

Following the data transformation, we needed to ascertain the possible relationship of the mean CD4 count over the space of time and as shown in Figure 3.2 with a fitted trend line. The appropriate trend line was determined using AIC and BIC values after fitting the various trend line. Hence, we considered the polynomial model of order 3 since it has the smallest AIC and BIC value as shown in Table 3.1. In addition, we determined the underlying covariance structure appropriate for the data by exploring some covariance structures. Using the AIC and BIC values as indicated in Table 3.3, the unstructured covariance structure has the least values of AIC and BIC and hence considered as the appropriate underlying covariance structure

Table 1. Information criteria for trend models

Model	AIC	BIC
Linear	82.47128	84.97092
Exponential	82.21864	84.71828
Logarithmic	81.77224	84.27188
Polynomial (order=2)	84.26567	87.59852
Polynomial (order=3)	**66.57995**	**70.74601**
Polynomial (order=4)	67.91535	72.91463

Table 2. AIC and BIC values of different linear mixed effect models

MODEL	AIC	BIC
Intercept	4680.9	4756.3
Intercept and Slope	**4670.9**	**4746.9**

Table 3. Statistic for the Covariance structures

Covariance Structure	AIC	BIC
Variance Component	3975.1	4055.7
First Autoregressive AR (1)	4089.1	4167.0
Compound Symmetry	4089.1	4167.0
Unstructured	**3973.0**	**4055.3**

Table 4. Parameter Estimate of Mixed Effect of full model

Parametres	Value	Std.Error	DF	t value	p value
Intercept	18.415667	2.0194988	553	9.118929	0.0000
Compared with EFV					
Drug (NVP)	0.105318	0.7207710	553	0.146119	0.8839
Compared with Female					
Gender (Male)	-2.216603	0.9605134	115	-2.307728	0.0228
Compared with (NO)					
Education (YES)	-0.348698	0.8054122	553	-0.432944	0.6652
Compared with (NO)					
Alcohol (YES)	1.081634	0.8386667	553	1.289706	0.1977
Compared with (NO)					
Smoking (YES)	0.523765	1.1737440	115	0.446235	0.6563
Time	0.047982	0.0122346	553	3.921831	0.0001
preCD4	0.016444	0.0026337	115	6.243499	0.0000
Married (YES)	-0.648003	0.8356824	553	-0.775417	0.4384
age	-0.028564	0.0380705	553	-0.750301	0.4534
Random effects	StdDev				
Intercept	3.9513331	(Intr)			
Time	0.0632366	-0.365			
Residual	3.6553211				

Table 5. Estimates of Reduced Linear Mixed-effect Model (Longitudinal Sub-model)

Pararmeters	Value	Std.Error	DF	t value	p-value
Intercept	17.242413	0.7444227	558	23.162127	0.0000
Gender (Male)	-2.405326	0.9201100	116	-2.614173	0.0101
Time	0.047306	0.0118133	558	4.004455	0.0001
PreCD4	0.016489	0.0026335	116	6.261213	0.0000
Random effects	StdDev				
Intercept	4.06727742	(Intr)			
Time	0.06143932	-0.425			
Residual	3.66074861				

Table 6. Survival model comparison

Criterion	Cox model	Weibull	Exponential	LlogisticLnormal	
AIC	7626.4	**573.05**	761.768	1455.896	901.871
BIC	7673.057	**690.55**	879.265	1568.875	1019.370

Table 7. Weibull model of all covariates

Parameter	Value	StdError	z	p
Intercept	3.24842	0.61861	5.25	1.51e-07
Compared with EFV				
Drug (NVP)	0.93058	0.27439	3.39	6.95e-04
Compared with (NO)				
Alcohol (YES)	-0.46628	0.26321	-1.77	7.65e-02
Compared with (NO)				
Smoking (YES)	-1.12953	0.201633	-5.6019	2.12e-01
Compared with (NO)				
Married (YES)	0.61088	0.28377	2.15	3.13e-01
Compared with (NO)				
Education (YES)	0.39391	0.26645	1.48	1.39e-01
Compared with Female				
Gender (Male)	-0.42460	0.30209	-1.41	1.60e-01
PreCD4	0.00396	0.00107	3.68	2.34e-04
Age	0.03295	0.01345	2.45	1.43e-02
Log(scale)	-0.62320	0.103190	-6.0394	
1.55e-09				
Scale= 0.536				

Table 8. Estimate of reduced Weibull model

	Value	Std Error	z	p
(Intercept)	3.90825	0.56445	6.92	4.39e-12
Drug (NVP)	0.84828	0.24895	3.41	6.56e-04
precd4	0.00316	0.00106	2.98	2.87e-03
Age	0.03015	0.01322	2.28	2.25e-02
Log(scale)	-0.49536	0.107221	-4.62	3.84e-06
Scale	0.609			

Table 9. Parameter Estimates of Joint Model

Parameters	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Longitudinal						
Time	0.0000103	2.95e-06	3.47	0.001	4.46e-06	0.000016
Gender	-2.873938	0.9356323	-3.07	0.002	-4.707743	-1.040132
PreCD4	0.0159314	0.0026117	6.10	0.000	0.0108126	0.0210502
Int	20.71174	1.342994	15.42	0.000	18.07952	23.34396
Survival						
Assoc: value	-0.3317109	0.1280213	-2.59	0.010	-0.5826281	-0.0807937
ln _ lambda						
Age	-0.0350959	0.0309258	-1.13	0.256	-0.0957094	0.0255176
preCD4	0.0040346	0.0031685	1.27	0.203	-0.0021755	0.0102447
Drug (NVP)	-1.16353	0.5333836	-2.18	0.029	-2.208942	-0.1181169
Int	3.273715	2.52749	1.30	0.195	-1.680074	8.227505

Random Effect	EStimates	Std. Err	[95% Conf. Interval]	
Time	0.0000144	4.34e-06	8.02e-06	0.000026
Int	4.104442	0.3520551	3.469309	4.855849
corr(Time*Int)	-0.6864479	0.1326006	-0.8699059	-0.3361284
Residual	3.620334	0.1211074	3.390582	3.865655

3.1.2 Linear mixed effect model type

A linear mixed effect model can be intercepted only model or intercept and slope model, as results, we compared the intercept model and intercept and slope model and it revealed that the model with intercept and slope has the least values of AIC and BIC as indicated in Table 3.6. Hence we proceeded with a model with random intercept and random slope.

3.1.3 Linear mixed effects model (random intercept and slope)

Table 3.3 shows that the initial CD4 count, gender and time of treatment (in months) were significant determinants of change in CD4 counts of patients on ART. Male patients have about 2.21 units decrease of CD4 counts less than their female counterparts for every CD4 count. The advantage of early diagnosis and treatment was shown in the PreCD4 count, thus when a patient starts treatment early, the average change gained in CD4 counts is expected to be 0.016 units for every CD4 count. The rate of change in the CD4 count is 0.048 counts per unit increase in time, suggesting the rate of change of CD4 count increase with time while a patient CD4 count decrease by about 0.03 units for every additional year of a patient's age holding other factors constant. Other covariates used in the study but were not statistically significant includes; alcohol drinking, cigarette smoking, education, married, and drug.
From Table 3.4 the full linear mixed effect model is given as;

$$\boldsymbol{E(CD4_{it}) = 18.42 + 0.11 * Drug(NVP) - 2.22 * Gender(Male)_i - 0.35 * Educated(yes) + 1.08 * Alcohol(yes) + 0.52 * Smoking(yes) + 0.05 * Time_{it} + 0.02 * PreCD4_i - 0.65 * Married(yes) - 0.03 * Age} \quad (3.0)$$

The stepwise method was used in the selection of the reduced model guided by AIC and BIC model selection criterion as indicated in Table 3.5. Therefore, the reduced linear mixed effect model is given as;

$$\boldsymbol{E(CD4_{it}) = 17.24 + 0.02 * preCD4_i + 0.05 * Time_{it} - 2.41 * Gender(male)_i} \quad (3.1)$$

3.2 Survival modeling

In other to choose the appropriate survival model for this research work; Cox PH models and four standard parametric models; Exponential, Log logistic, Weibull, and Lognormal were explored and then compared. Using AIC and BIC as shown in Table 3.5, Weibull model appeared the best since it has the least values of AIC and BIC.

3.2.1 The Weibull Model

The Weibull model as shown in Table 3.6 revealed that drug regiment NVP component will lower the hazard of a patient by 2.54 (p=6.95e-04) than EFV component of the drug regimen. However, the results revealed there was an increase in hazard with an increase in age by the hazard of 1.03 significant at p=1.43e-02. Patient with high preCD4 count lowers the hazards by 1.00 which was statistically significant at p=2.34e-04. However, the rest of the covariates were all non-statistical significant to the survival time of the patient. To determine the prognostic factors and the reduced model for prediction, the AIC stepwise selection criterion was used. The reduced model indicated that covariates such as PreCD4 count, Age and Drug are significant determinants of patient's survival time as shown in Table 3.7 which resulted in equation (3.3)
From Table 3.7 the full Weibull model is given as;

$$\boldsymbol{h(t) = \rho t^{(\rho-1)} \exp(0.93 * Drugs_i + 0.0032 * preCD4_i + 0.03 * Age_i - 1.13 * smoking + 0.61 * married(yes) - 0.47 * alcohol(yes) + 0.4 * education(yes) - 0.42 * gender(male))} \quad (3.2)$$

From Table 3.8 the reduced Weibull model is written as;

$$\boldsymbol{h(t) = \rho t^{(\rho-1)} \exp(0.85 * Drugs_i + 0.0032 * preCD4_i + 0.03 * Age_i)} \quad (3.3)$$

3.3 Joint Model

Because convergence failure prevents accurate computation of posterior model summaries, we used a Weibull model with r = 1 (i.e., an exponential) resulting in Table 3.8. The results revealed that the time of treatment has a statistically significant positive effect on the average CD4 and thus indicating that there is an increase of 0.0000103 units of CD4 count per units increase in time (95% CI: 4.46e-06, 0.000016). The results, in addition, revealed that PreCD4 count or initial CD4 count also positively affect the CD4 insignificantly, thus indicating that with early diagnosis and treatment the expected average increment in the CD4 count was 0.0159314 units (95% CI: 0.0108126, 0.0210502). But the male patient has a significant reduction in the average CD4 count of 2.873938 units (95% CI: -4.707743, -1.040132) as compared to their female counterparts.

Treatment NVP regimen component effect lowers the hazard of death by 1.16353 (95% CI: -2.208942, -0.1181169) than EFV regimen component. However, the other survival covariates that are non-statistically significant include age and preCD4 count. The estimated association parameter in the joint model was highly negative with a value of -0.3317109 (95% CI: -0.5826281, -0.0807937) and it is statistically significant in the Weibull model (p=0.010). This indicates that there is strong evidence of an association between the two sub-models.

Further investigation of this association showed that both initial level and slope of the CD4 count were negatively associated with the hazard of death. This finding is clinically predictable since patients with a more drastic decline in CD4 count would have poorer survival. Under this joint model, a patient's survival is associated with his/her longitudinal data pattern. The longitudinal sub-model and survival sub-model was obtained from Table 3.9; the longitudinal sub-model is written as,

$$\boldsymbol{E(CD4_{it}) = 20.71 + 0.02 * preCD4_i + 0.0000103 * Time_{it} - 2.87 * Gender(male)_i} \quad (3.4)$$

and the survival sub-model in the absence of random effects is given as,

$$\boldsymbol{\log\big(h(t)\big) = 3.27 - 1.2 * Drugs_i + 0.004 * preCD4_i - 0.04 * Age_i} \quad (3.5)$$

The incorporation of stochastic variable from the linear mixed effect model into the survival sub model gave us the joint model,

$$\boldsymbol{\log\big(h(t)\big) = -1.164 * Drugs_i + 0.004 * preCD4_i + -0.035 * Age_i - 0.332 * (4.104442 + 0.0000144 * Time} \quad (3.6)$$

4. Conclusion

We observed that there was a significant association between the repeated measured CD4 count and patients' survival time. Hence the rate of change of CD4 count has a significant effect on the survival time of HIV patient.

References

Adams, M., &Luguterah, A. (2013). Longitudinal Analysis of Change in CD4+ Cell Counts of HIV Patients on Antiretroviral Therapy (ART) in the Builsa District. *European Scientific Journal,* **9**(33**)**: 1857 – 7881.

Ding, A. & Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics,* **2**(1): 13–29.

Krishnan, S., Schouten, J. T., Jacobson, D. L., Benson, C. A., Collier, A. C., Koletar, S. L.,& Mitsuyasu, R. (2011). Incidence of Non-AIDS-Defining Cancer in Antiretroviral Treatment-Naïve Subjects after Antiretroviral Treatment Initiation: An ACTG Longitudinal Linked Randomized Trials Analysis. *Oncology*, **80**(1): 42–49.

Xiuhong, L., Margolick, J. B, Jamieson, B. D., Rinaldo, C. R., John, P. Phair, J. P., Lisa, P.,&Jacobson, L. P. (2011). CD4+ T-Cell Counts and Plasma HIV-1 RNA Levels Beyond 5 Years of Highly Active Antiretroviral Therapy. *Journal of Acquired Immune Deficiency Syndromes,* **57**:421–428.