

# A Joint Modeling Approach to Assess the Impact of CD4 Cell Count on the Risk of Loss to Follow up in HIV/AIDS Patients on Antiretroviral Therapy

Gurprit Grover<sup>1</sup>, Prafulla Kumar Swain<sup>1,\*</sup>, Vishal Deo<sup>2</sup>, Manoj Kumar Varshney<sup>3</sup>

<sup>1</sup>Department of Statistics, University of Delhi, India

<sup>2</sup>Department of Statistics, Ramjas College, University of Delhi, India

<sup>3</sup>Department of Statistics, Hindu College, University of Delhi, India

**Abstract** Lately, joint modeling has been reckoned as a very efficient technique for studying combinations of longitudinal and survival data generated from medical studies. In this paper we have developed a joint model to simultaneously study the longitudinal repeated measures on CD4 cell counts and the time to event (event being defined as loss to follow up) process of HIV/AIDS patients undergoing Anti-retroviral therapy (ART) treatment at Dr. Ram Manohar Lohia Hospital's ART centre, New Delhi, India. Apart from increasing the risk of the HIV infection to progress to AIDS, the event of loss to follow up in patients undergoing ART seriously interferes with the development, improvement, and validation of treatment techniques being used in the therapy. The fact that the problem of loss to follow up from ART has been understated in studies based on ART centers in India has motivated us to investigate the effect of various clinical, socioeconomic and demographic factors on the hazard of loss to follow up in patients undergoing ART treatment. The results of the joint model have been compared with those of the separate analyses of the longitudinal and the survival data. The parameter estimates of both methods are consistent; however, the joint analysis supports the dependence of the hazard of lost to follow up from ART treatment on the rate of change in CD4 counts, apart from the patient's baseline CD4 count. The estimated overall survival probability for HIV/AIDS patients retained on antiretroviral therapy was 0.81; 95%CI (0.76- 0.87).

**Keywords** HIV, AIDS, Antiretroviral Therapy, Joint modeling, Longitudinal data

## 1. Introduction

Longitudinal data and survival data often arise together in biomedical and clinical studies, where patients are followed up over time, and biomarkers are recorded repeatedly at different time points. There are various dynamic medical scenarios (viz. Liver cirrhosis studies in which our interest lies in the association of serum bilirubin and the time to death, Cancer studies in which we investigate the association between prostate specific antigen levels and the time to development of prostate cancer, HIV studies where interest lies in the association between CD4 cell counts (or viral load) and time to AIDS or death, or even Diabetics research where interest lies in the association of patients' hemoglobin level and time to death) where both the association structure of repeated biomarkers measurements and the primary survival endpoints are studied. Various standard methods are available for analyzing such data separately, viz. linear

mixed effects (LME) models for longitudinal data, and Cox proportional hazard (PH) models for survival data. LME model describes the process of the repeated measures on biomarkers over time while Cox PH model evaluates the effect of covariates on the hazard of an event. However, the use of separate models may be inappropriate and may provide biased inferences because they fail to take into account the association between the two components of the data [1]. One approach of studying the two components together is to incorporate the longitudinal measurements directly into the Cox model as time varying covariates, but this method does not account for the measurement errors in the time varying covariates and, therefore, can cause the estimated relative risk parameter in the time dependent Cox model to be biased towards the null [2]. Moreover, the time varying covariates are assumed to experience changes in their values at the times of follow up visits only and are assumed to remain constant during the time intervals in between these follow up visits, which is unrealistic in most of the situations.

In the circumstances discussed above, a joint model is desirable over separate models or even over the time dependent Cox model, to understand the association between

\* Corresponding author:

prafulla86@gmail.com (Prafulla Kumar Swain)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

the longitudinal and the time to event processes. Joint modeling can be perceived to be a sophisticated and complex approach in terms of estimation, however, its superiority comes from its ability to model the longitudinal repeated biomarkers measurements and the survival processes together while also taking into account the association between them. By including the random effects model for longitudinal data in the survival model, the patterns of a biomarker's performance and the relationship between its progression and survival time can be characterized. In this way, joint modeling provides less biased estimates and more efficient inferences than separate models [3, 4].

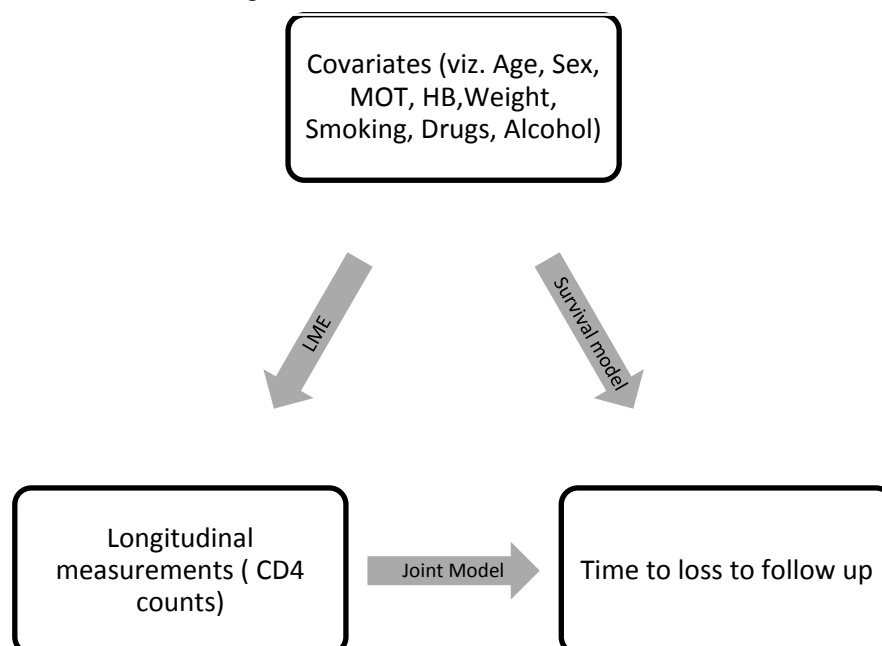
Joint modeling has been extensively studied by many authors in the last two decades. Tsiatis and Davidian (2004) [5] provided a very thorough overview of early work on joint models, including those of De Gruttola and Tu (1994) [6], Wulfsohn and Tsiatis (1997) [4], Henderson et al. (2000) [7], and Wang and Taylor (2001) [8], among others. More recent work, including Ding and Wang (2008) [9], Nathoo and Dean (2008) [10], Ye et al. (2008) [11], Albert and Shih (2010) [12], Jacqmin-Gadda et al. (2010) [13], Rizopoulos, (2012) [14], Wu et al. (2010) [15], and Huang et al. (2011) [16], have extensively discussed the development, estimation and applications of joint models.

Yi-Kuang Tseng et al., [17] used a joint modeling (graphic approach) to model survival time and time dependent CD4 count simultaneously for AIDS patients in Taiwan. Rui Martins et al. (2012) [18] developed a Bayesian hierarchical model for jointly modeling longitudinal and spatial survival data for a cohort of patients with HIV/AIDS in Brazil. Yi-Kuang Tseng et al. (2005) [19] proposed a viable joint modeling approach for accelerated failure time data and longitudinal and time-independent covariates. Thiebaut R. et al. (2005) [20], proposed a parametric joint modeling of bivariate longitudinal markers and a lognormal survival

model for the time to drop out of HIV 1 infected patients starting antiretroviral treatment in Europe.

Although many authors have applied joint modeling in HIV/AIDS epidemiology, most of their work has covered the joint modeling of the longitudinal CD4 cell counts and the survival times for the event death. In this paper we have made an attempt to develop a joint model of longitudinal CD4 cell counts and time to loss to follow up in HIV/AIDS patients undergoing antiretroviral therapy in an ART centre in Delhi, India.

The introduction of free ART services in India in 2004 has significantly reduced the mortality and morbidity in HIV populations. However, a successful treatment of individual patients and the monitoring and evaluation of ART programmes, both depend on regular and complete patient follow up [21]. Possible loss to follow up occurred after initiation of ART may be due to the regime side effects or due to their poor health conditions like having lower CD4 counts or due to migration or even due to some social and psychological causes. Patients who are lost from ART programme develop drug resistance virus which ultimately results in drug failure [22]. As a result, such patients are at a higher risk of illness and death because of AIDS related conditions. The severity of the problem of discontinuation of patients from the ART programme, leading to loss to follow up, makes it essential to conduct a proper investigation to identify the responsible factors. The factors leading to loss to follow up may also vary from country to country because of the dissimilarity in their socio-economic structures and also due to the difference in the level of awareness among the patients and their family members in those countries. However, there is a dearth of literature on investigating predictors of loss to follow up after initiation of ART in India, [23].



**Figure 1.** Schematic representation of Joint model for HIV/ AIDS patients on ART

Besides, accurate measures of loss to follow up are valuable for understanding processes and programmatic details, and they are critical to evaluating clinical outcomes. Therefore it is very important to study the loss to follow up in the presence of associated risk factors and biomarkers. The main objective of this paper is to build a joint model; simultaneous modeling of longitudinal CD4 counts and time to event (i.e loss to follow up) process and linking them using unobserved random effects through the use of a shared parameter model. The results are also compared with those obtained from separate analyses of the longitudinal and the survival data. The diagrammatical representation of joint modeling of longitudinal data and survival data has been depicted in the figure 1.

CD4 cell counts are a very important type of white blood cells that orchestrates the immune response of the body. CD4 cell count per cubic millimeter of blood is widely used as a biomarker for HIV disease progression and as an eligibility criterion to start antiretroviral therapy. Successful treatment for HIV patients depends on the therapy's ability to restore immune functions. Researchers have advocated that disease progression among patients with HIV infection is delayed substantially when CD4 counts increase in response to treatment [24, 25]. Identification of factors influencing disease progression is vital to effective care for patients and to improve their survival and quality of life. Apart from the assessment of clinical features and risk factors in patients, the detection of characteristics associated with a more rapid progression of CD4 counts can help identify patients who may benefit from closer and more frequent clinical follow up and earlier treatment interventions [24]. We anticipate that the path of changes in CD4 biomarker will have an impact on the major HIV complications.

The paper is organized as follows; section 2 describes material and methods used and section 3 presents the results and findings. Finally the discussions and concluding remarks are presented in sections 4 and 5 respectively.

## 2. Material and Methods

### 2.1. Study Population

We have considered 1259 adult (>18 years age) HIV/AIDS patients who were undergoing Antiretroviral Therapy in the ART centre of Dr. Ram Manohar Lohia Hospital, New Delhi, India, during the period April 2004 to November 2009, and were followed up through the ART routine register records till December 2010. Both longitudinal and survival data were recorded. The longitudinal response variable CD4 cell counts per mm<sup>3</sup> of blood were measured approximately every 6 months irrespective of their visit to ART centres. The event of interest was time to loss to follow up of the patients from ART treatment. As per National AIDS Control Programme (NACO) guidelines patients missing ART refills/appointments for more than three consecutive months at an ART centre are considered as lost to follow up. During the

study period 15.7% of the total patients were dead and 7.9% patients were lost to follow up from the treatment care. The baseline information such as age, sex, CD4 count, mode of transmission (MOT), weight, smoking habit, alcohol habit and hemoglobin were collected. These variables were entered into the model as categorical variables – Sex (male/female), MOT (sexual/Blood+IDU/Unknown), smoking (yes/no), alcohol (never/ever) and continuous variables-Age, Hemoglobin and Weight. The category 'sexual' of the covariate MOT included both Homosexual and Heterosexual transmissions. The software packages *survival* and *JM* in R have been used for data analysis.

### 2.2. Methodology

The study accommodates two different approaches of modeling the longitudinal CD4 data and the time to event data, viz. the separate modeling approach and the joint modeling approach (figure 1).

#### 2.2.1. Separate Modeling

A linear mixed model has been used for the longitudinal CD4 data and the time to event data has been modeled using a Cox proportional hazard model.

#### 2.2.2. Joint Modeling

Here our objective is to incorporate a time dependent covariate measured with error (longitudinal CD4 measurements of HIV/AIDS patients on ART) in a time to event model (loss to follow up being the event of our interest).

Suppose we have a set of  $n$  individuals followed over time. Let  $T_i$  ( $i=1,2,\dots,n$ ) be the time to loss to follow up of  $i^{\text{th}}$  HIV/AIDS patient on ART, and  $C_i$  be the corresponding censoring time. Let  $\delta_i = I(T_i \leq C_i)$  be the censoring indicator such that  $\delta_i = 0$  if the survival time of  $i^{\text{th}}$  individual is right censored and  $\delta_i = 1$ , otherwise. Thus the observed survival data are  $\{(t_i, \delta_i), i = 1, 2, \dots, n\}$ , where  $t_i = \min(T_i, C_i)$ .

A joint model is comprised of two sub-models, one for the longitudinal process and the other for the survival process. In this study, the joint distribution of the event times and longitudinal measurements is modeled via a set of random effects that are assumed to account for the associations between these two outcomes, as was earlier proposed by Gao et al. 2011 to analyze the direct impact of biomarkers variability on the survival outcome in a joint modeling set up.

#### The Submodels Specifications

In joint modeling framework, the longitudinal data are delineated by a conventional linear model assuming homogeneity within subject variance. However, such a homogeneity assumption automatically precludes the assessment of the research question "whether individuals with different levels of CD4 variability have different susceptibility to loss to follow from ART treatment". Consequently, the joint model proposed in this study combines both approaches; that is, it relates the variability of

longitudinal CD4 counts to time to loss to follow up of patients and also adds a frailty to the survival sub model. Frailty term in the survival sub model accounts for the heterogeneity among patients in terms of the CD4 trajectory affecting the time to event process.

### Longitudinal submodel

Linear mixed effect models are widely used to model continuous longitudinal data (i.e CD4 counts). The sequence of measurements  $y_{i1}, y_{i2}, \dots, y_{ini}$  for the  $i$ th subject at times  $t_{i1}, t_{i2}, \dots, t_{ini}$  is modeled as (Laird and Ware, 1982) [26]

$$y_{ij} = W_i(t_{ij}) + \epsilon_{ij} \quad (1)$$

$$W_i(t_{ij}) = X'_i(t_{ij})\beta + Z'_i(t_{ij})b_i + u_i\delta \quad (2)$$

$$b_i \sim N(0, \Psi), \epsilon_{ij} \sim N(0, \sigma^2)$$

Where  $y_{ij}$  is the observed longitudinal response for the  $i$ th patient measured at  $j$ th time point.  $X'_i(t_{ij})$  and  $Z'_i(t_{ij})$  are the design matrices for fixed ( $\beta$ ) and random ( $b_i$ ) effects respectively. Furthermore, we also have a vector of covariates (possibly time dependent),  $u_i \in U_i$ , and corresponding regression coefficient  $\delta$ . We assume that the  $\epsilon_{ij}$  is independent of the random effects, and that  $\text{cov}(\epsilon_{ij}, \epsilon_{ik}) = 0$  (where  $i \neq j$ ). The term  $W_i(t_{ij})$  can be viewed as the true individual level CD4 trajectories after they have been adjusted for the overall mean trajectory and other fixed effects.

### Survival Submodel

The Cox proportional hazard submodel is given by

$$h(t / b_i, v_i) = h_0(t) \exp(\alpha W_i(t_{ij}) + v_i \phi) \quad (3)$$

Where  $h_0(t)$  is the baseline hazard function,  $\alpha$  denotes the association parameter and  $\phi$  is a set of regression coefficients associated with a set of covariates (again possibly time dependent),  $v_i \in U_i$ . The value of the biomarker, as estimated by the longitudinal submodel, is included in the survival linear predictor as a time-varying covariate.

If the covariate is included in both submodels, we can obtain an overall effect on survival as the longitudinal trajectory of the marker, multiplied by the association parameter, plus the direct effect on survival.

### Maximum Likelihood Estimation

The likelihood method is a widely used approach for the parameter estimation in the joint model. Assuming that censoring and timing of longitudinal measurements are non-informative, the likelihood function  $L$  with the observed data for each subject ( $X_i, \delta_i, Z_i, t_i$ ) is given by Wulfson and Tsiatis (1997).

$$L = \prod_{i=1}^n \left[ \int_{-\infty}^{\infty} \left( \prod_{j=1}^{m_i} f(y_i(t_{ij}) | b_i, \theta) \right) f(b_i | \theta) f(T_i, d_i | b_i, \theta) db_i \right] \quad (4)$$

Where

$$f(y_i(t_{ij}) | b_i, \theta) = (2\pi\sigma_e^2)^{-1/2} \exp \left\{ -\frac{y_i(t_{ij}) - W_i(t_{ij})}{2\sigma_e^2} \right\}, \quad (5)$$

$$f(b_i | \theta) = (2\pi|V|)^{-1/2} \exp \left\{ -\frac{b_i' V^{-1} b_i}{2} \right\}, \quad (6)$$

and

$$f(T_i, d_i | b_i, \theta) = [h_0(t) \exp(\alpha W_i(t) + \phi v_i)]^{\delta_i} \times \exp \left\{ -\int_0^{T_i} h_0(u) \exp(\alpha W_i(u) + \phi v_i) du \right\} \quad (7)$$

The maximization of this function (4) with respect to  $\theta$  requires a combination of numerical integration and optimization algorithms, because both the integral with respect to the random effects in (5) and in the survival function given by (7) do not have an analytical solution. Despite some authors have employed standard numerical integration techniques, such as Monte Carlo or Gaussian quadrature, the Expectation-Maximization (EM) algorithm described by Wulfsohn and Tsiatis (1997) has been traditionally preferred. The intuitive idea behind the EM algorithm is to maximize the log-likelihood in two steps: the Expectation (E) step, where missing data is first completed, so we replace the log-likelihood of the observed data with a surrogate function, and then by maximizing the surrogate function with respect to  $\theta$  in the M-step. The closed form of maximum likelihood estimates are obtained by the Newton-Raphson method.

## 3. Results

Table 1 shows the baseline descriptive statistics of HIV/AIDS infected patients on ART. Out of the total 1259 patients 851 (67.6%) were males and the remaining 408 (32.6%) were females. Majority (64%) of the infected patients had reported sexual mode of HIV transmission, including both homosexual and hetro-sexual modes of transmission, followed by those with unknown mode of transmission who accounted for 26.3% of the total, and 9.7% of them had experienced blood+ IDU route of HIV infections. 37.01% and 34.4% of the infected patients had habits of smoking and drinking alcohol (ever) respectively. The median (IQR) age at the start of ART was 33 (29-39) years. The median weight at baseline was 50 kg (with IQR; 44-56). The mean (SD) hemoglobin level at the initiation of ART was 11.05 (1.73) and the mean (SD) CD4 count at baseline was 192.06 (105.28) cells/mm<sup>3</sup>. Figure 2 shows the profile of longitudinal CD4 cell count measurements over time for HIV/AIDS patients, separately for censored and event cases. The plotted profiles tend to generate a linearly increasing pattern which rationalizes the use of Linear Mixed Effect model to analyze the trajectory of CD4 cell count. Comparison of box plots (figures 4.1-4.4) constructed for

different categories of patients formed on the basis of the factors, sex, mode of transmission, smoking habit and alcohol consumption exhibits higher median CD4 counts for (a) female patients as compared to that of males, (b) patients with unknown mode of transmission than those with known modes of transmission, (c) regular smokers as compared to that of non smokers, and (d) alcohol consumers as compared to that of non consumers.

Figure 3 shows the Kaplan-Meier survival plot of HIV/AIDS patients retained on antiretroviral therapy. The estimated overall survival probability for HIV/AIDS patients retained on antiretroviral therapy was 0.81; 95%CI (0.76-0.87). Table 2 shows the estimates and corresponding standard error and p-value for the parameters of the two models, i.e. separate and joint models. In longitudinal submodel the predictors, time, sex, age, mode of transmission, and weight were found to be statistically significant at 5% level of significance.

The coefficient for the time effect has a positive sign indicating that on an average the CD4 cell count increases with time. The estimate average longitudinal regression coefficients for gender is positive and significantly different from zero, suggesting that the female patients had higher CD4 cell counts than males during the follow up. The regression coefficient for age is -1.369 (P-value < 0.003), suggesting that as age of patients increases there is significant decrease in CD4 count cells/mm<sup>3</sup>, level over the time. Similarly, the other predictors like weight and hemoglobin coefficients 2.791 (P-value < 0.001) and 4.187 (P-value < 0.05) respectively, indicates that CD4 count increases over the study period as unit increases in weight and hemoglobin level of patients.

In survival submodel, sex, age, hemoglobin and smoking status are not found to be significant predictor for loss to follow up event, whereas an alcoholic patient had [ $\exp(0.711) = 2.036$ ] times higher risk of lost to follow up as compared to

non-alcoholic patient. There is significantly higher hazard of loss to follow up from the treatment associated with rural patients as compared to the patients who belong to urban area. Patients with unknown mode of HIV transmission had a lower hazard for loss to follow up as compared to that of patients with sexual mode of transmission. Weight is positively associated with the trajectory of the CD4 count and also, significantly lower hazard for loss to follow up is associated with patients having higher weight after initiation of ART.

The estimates of the parameters of the separate and joint models are quite similar to each other but not identical, however the estimate of the association parameter in the joint analysis is significantly different from zero, providing strong evidence of association between the two submodels. The estimate of the association parameter due to the slope (trend) of CD4 count is negative (-0.213), indicates that the slope of CD4 count is negatively associated with the hazard of loss to follow up of patients from ART treatment. This indicates that an increasing trend in the CD4 count in patients undergoing ART treatment significantly reduces the risk of loss to follow up of those patients. Figure 5, shows the diagnostic plots for the fitted joint model for HIV/AIDS patients on ART. The top left panel depicts the subject specific residual for the longitudinal measurements on CD4 cell counts plotted against their corresponding fitted values and the residuals can be seen to be trending very close along the fitted line. The top right panel shows a Normal Q-Q plot of the standardized subject specific residuals for longitudinal process which is almost coinciding with the reference line passing through the origin and hence, is validating our assumption of normality of the error term in the longitudinal submodel. The bottom panels depict an estimate of the marginal survival and an estimate of the marginal cumulative risk function for the event process.

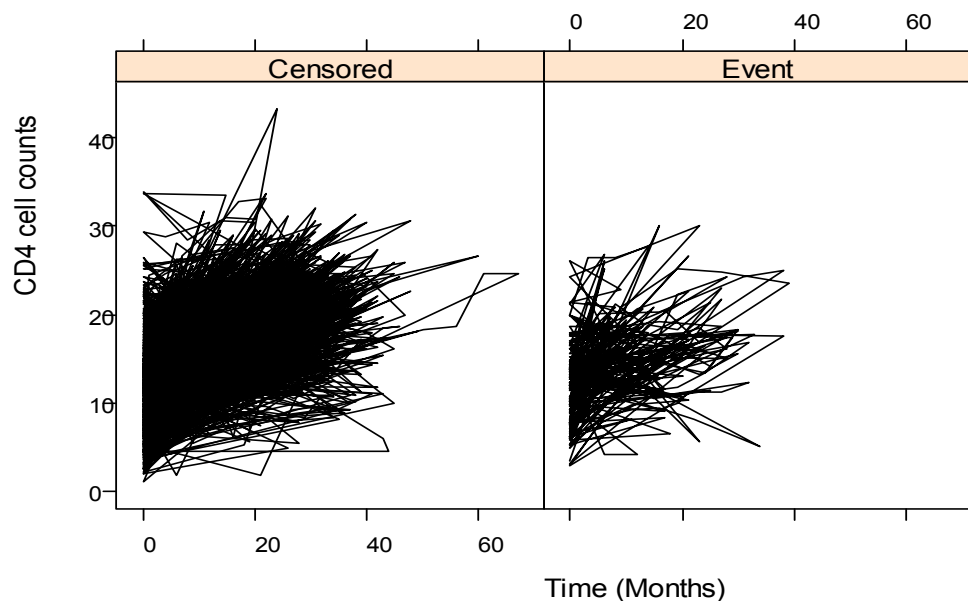


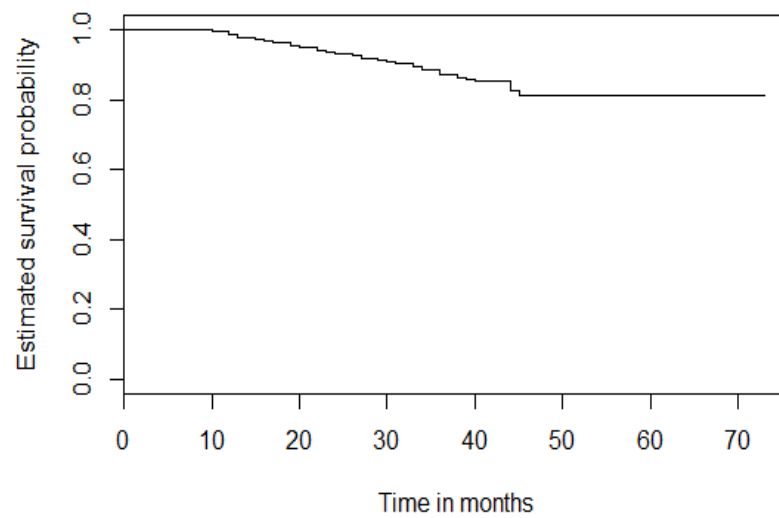
Figure 2. Longitudinal CD4 cell count measurements over time for HIV/AIDS patients, separately for censored/event

**Table 1.** Baseline characteristics of HIV/AIDS infected patients on ART

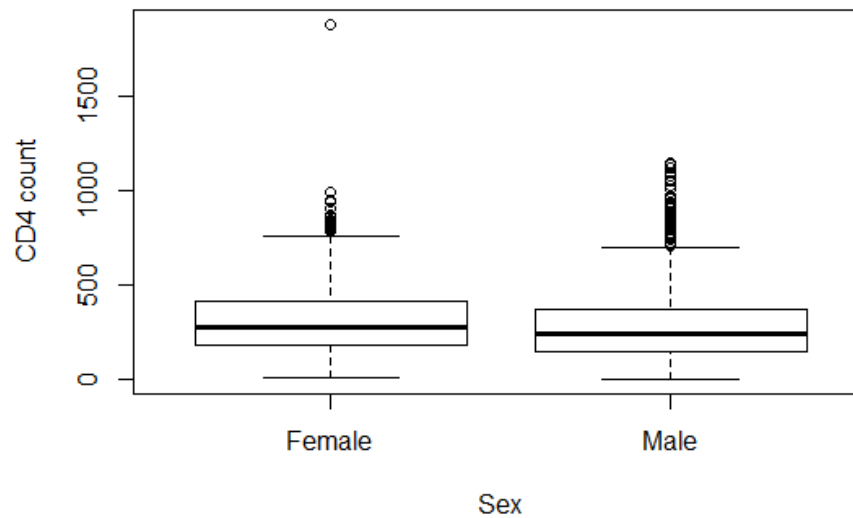
Characteristics	Category	No. of patients	Percent
Sex	Male	851	67.6
	Female	408	32.4
MOT	Sexual	806	64
	Blood+IDU	122	9.7
	Unknown	331	26.3
Smoking	No	793	62.99
	Yes	466	37.01
Alcohol	Never	826	65.61
	Ever	433	34.39
Location status	Urban	1056	83.87
	Rural	203	16.13
Status	Event	100	7.94
	Censored	1159	92.06
Median Age (IQR)			33 (29 -39)
Median Weight (IQR)			50 (44 - 56)
Mean HB ( SD)			11.05 (1.73)
Mean CD4 count at baseline (SD)			192.06 (105.38)

**Table 2.** Comparison of separate and joint models of longitudinal CD4 counts and time to loss to follow up from ART

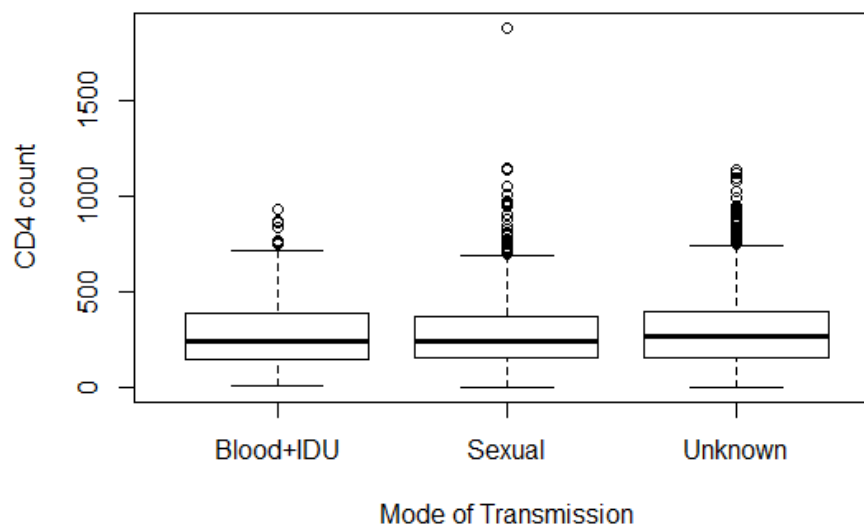
	Parameters	Separate Analysis			Joint Analysis		
		Est. value	Std.Error	P-value	Est.value	Std.Error	P-value
Longitudinal sub-model (CD4 count)	Intercept	42.191	32.343	0.192	42.359	32.225	0.189
	Time	8.688	0.171	0.000	8.684	0.172	0.000
	Sex (Female)	48.205	9.515	0.000	48.180	9.480	0.000
	Age	-1.368	0.462	0.003	-1.369	0.460	0.003
	Smoking (yes)	-16.930	10.681	0.113	-16.914	10.641	0.112
	MOT(Blood+IDU)	8.334	13.074	0.524	8.332	13.027	0.522
	MOT(Unknown)	-43.934	7.875	0.000	-43.893	7.846	0.000
	Weight	2.792	0.370	0.000	2.791	0.369	0.000
	Alcohol (Ever)	18.025	10.818	0.096	18.016	10.778	0.095
	Location (Rural)	-1.230	0.432	0.067	-1.243	0.421	0.066
	Hemoglobin	4.187	2.120	0.049	4.188	2.113	0.047
Survival sub-model (Time to loss to follow-up)	Intercept				-10.204	1.208	0.000
	Sex (Female)	-0.107	0.307	0.726	-0.086	0.312	0.784
	Age	0.006	0.013	0.655	0.006	0.013	0.643
	Smoking (yes)	0.086	0.293	0.769	0.087	0.295	0.769
	MOT (Blood+IDU)	0.519	0.328	0.113	0.541	0.328	0.099
	MOT (Unknown)	-0.789	0.247	0.001	-0.893	0.244	0.000
	Weight	-0.023	0.011	0.042	-0.023	0.012	0.055
	Alcohol (Ever)	0.683	0.305	0.025	0.711	0.306	0.020
	Location (Rural)	0.320	0.123	0.007	0.318	0.120	0.006
	Hemoglobin	0.046	0.061	0.450	0.050	0.061	0.414
	Association				-0.213	0.011	0.007
	log(shape)				0.904	0.099	0.000



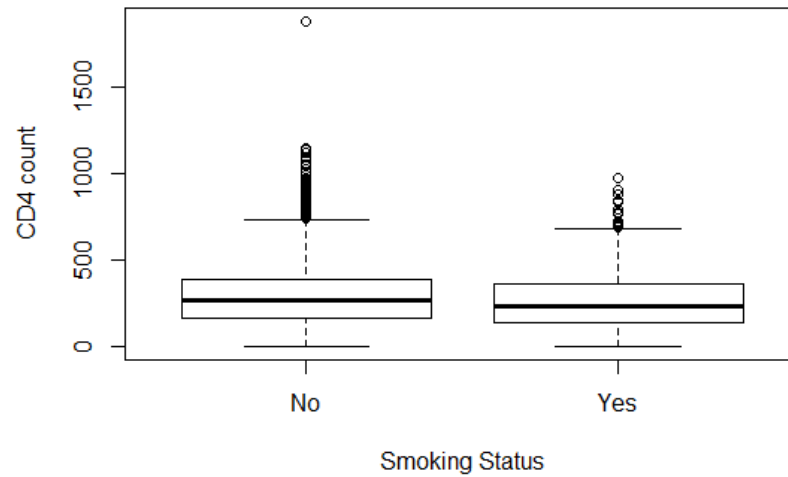
**Figure 3.** K-M Survival plot for HIV/AIDS patients retained on ART



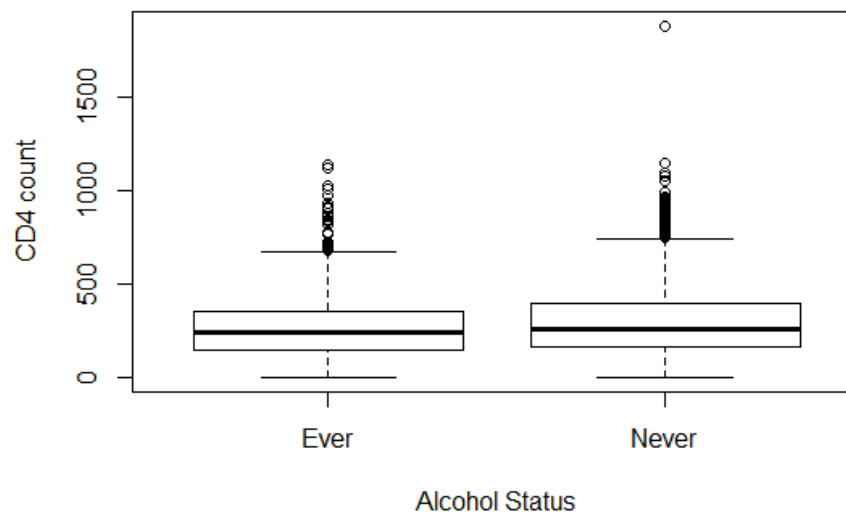
**Figure 4.1.** Box plot for CD4 counts by Sex



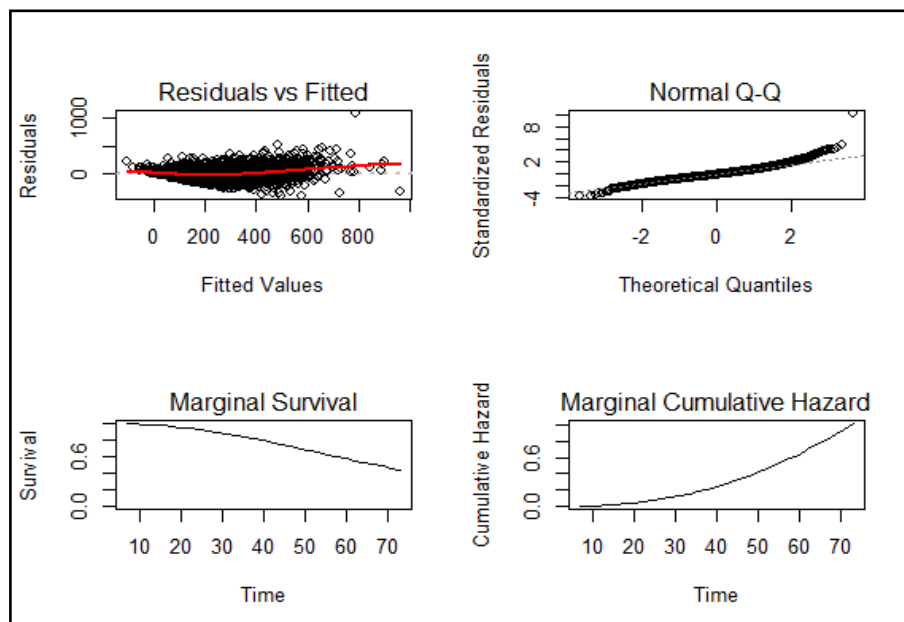
**Figure 4.2.** Box plot for CD4 counts by mode of transmission



**Figure 4.3.** Box plot for CD4 counts by smoking status



**Figure 4.4.** Box plot for CD4 counts by alcohol status



**Figure 5.** Diagnostic plots for the fitted joint model for HIV/AIDS patients on ART



## 4. Discussion

In this paper, we have studied the relationship between the trajectory of CD4 counts over time and the risk of loss to follow up of HIV/AIDS patients who started on ART using joint modeling with longitudinal random effects sub-model and a Cox proportional hazard sub-model. Also we have compared the results of the joint model with those obtained from separate analyses of the longitudinal and survival data.

The joint analysis has established significant dependence of the hazard of loss to follow up from ART treatment on the rate of change in CD4 counts, apart from the patient's baseline CD4 count. So, in clinical perspective, the trajectory of the CD4 count of patients undergoing ART treatment can prove to be a vital indicator about the risk of losing them to follow up. This knowledge becomes very crucial for devising methods to minimize morbidity and mortality rates due to HIV/AIDS condition as loss to follow up is expected to result in magnified exposure to serious risks in patients.

Although, the parameter estimates of both the separate analysis and the joint analysis are consistent, a considerable amount of difference is apparent between their values, possibly due to the model correction accounting for the correlation between the longitudinal CD4 counts and the time to loss to follow up from the ART treatment.

To our knowledge, there is a dearth of literature that highlights the problems of loss to follow up among HIV/AIDS infected patients started on ART treatment in India. Gerardo Alvarez-Uria et al. 2013 [23], have described the predictors associated with loss to follow up in HIV infected patients in a large cohort study in Anantapur, Andhra Pradesh, India. In this study, they have reported 10.3% loss to follow up during their study period, which is slightly higher than that of our data. Dalal et al. (2008) [27] and Geng et al. (2008) [28] have reported high proportion of patients lost to follow during the course of ART Treatment in resource-limited settings in South Africa and Africa respectively.

While predictors like mode of transmission, location status and weight at the start of ART, even being significant predictors of loss to follow up, cannot be controlled or improved upon during the course of ART, alcohol addiction can be dealt with by introducing a dedicated consultancy programme as a part of the ART. Since the risk of loss to follow up increases by around 2.036 times for alcoholic patients, curbing this addiction can significantly reduce the risk of losing the patients during the course of ART. Migration might also be one of the important factors that could increase the risk of loss to follow up from ART programme (Bygrave et al., 2010) [31].

There are many studies which have supported joint modeling of longitudinal data and survival time to the event process over separate modeling [1, 8], which has been again emphasized upon by the results of our study which has shown a very significant association between the longitudinal trajectory of the CD4 count and the time to event.

The previous studies have claimed that the patients having low socioeconomic status and those who are illiterate are significantly associated with loss to follow up from the treatment. In this study we were not able to capture these two important predictors in our model. Presence of opportunistic infections including TB and the nature of ART initial regimen are some other important factors which can have significant effect on the event of loss to follow up and can be included in the joint model set up for further analysis. The travel time and travel costs can also be major impediments for HIV patients to come to the clinics [29, 30], supporting the current policy of decentralization of ART centres by the Government of India.

Moreover, we have used only one ART centre data for analysis that may not be representative for the whole country. Considering the socioeconomic diversity of India, our results need to be substantiated by similar survival studies from other parts of India to conjure up a comprehensive picture of HIV/AIDS epidemiology in India.

## 5. Conclusions

In conclusion, the retention of patients in HIV-care is still a great challenge and the trajectory of CD4 cell count promises to be a crucial factor in predicting the event of loss to follow up. Also, joint models give a proper platform to study such situations where the association of longitudinal data with the time to event has utmost importance. So, joint model should be preferred over separate models for longitudinal and survival data analysis.

## REFERENCES

- [1] Lim et al. 2013, Joint modeling of longitudinal and event time data: application to HIV study, *Journal of Medical Statistics and Informatics*, 1:1. <http://dx.doi.org/10.7243/2053-7662-1-1>.
- [2] Prentice, R. L, (1982) Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika*, 69, 331-342.
- [3] Tsiatis AA, De Gruttola V and Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* 1995; 90:27-37.
- [4] Wulfsohn MS and Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330-9.
- [5] Tsiatis AA and Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*. 2004; 14:809-834.
- [6] De Gruttola V and Tu XM. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*. 1994; 50:1003-14.

- [7] Henderson R, Diggle P and Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1:465-80.
- [8] Wang Y and Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; 96:895-905.
- [9] Ding and Wang. Modeling Longitudinal Data with Nonparametric Multiplicative Random Effects Jointly with Survival Data, *Biometrics*. 2008; 64(2): 546-556. doi:10.1111/j.1541-0420.2007.00896.x.
- [10] Nathoo F. S and C. B. Dean, "Spatial multistate transitional models for longitudinal event data," *Biometrics*, vol. 64, no. 1, p. 271-279, 326, 2008.
- [11] W. Ye, X. Lin, and J. M. G. Taylor, "Semiparametric modeling of longitudinal measurements and time to- event data-a two-stage regression calibration approach," *Biometrics*, vol. 64, no. 4, pp. 1238-1246, 2008.
- [12] P. S. Albert and J. H. Shih, "On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure," *Biometrics*, vol. 66, no. 3, pp. 983-987, 2010.
- [13] H. Jacqmin-Gadda, C. Proust-Lima, J. M. G. Taylor, and D. Commenges, "Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model," *Biometrics*, vol. 66, no. 1, pp. 11-19, 2010.
- [14] Rizopoulos, D. (2012) *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. Boca Raton: Chapman and Hall/CRC.
- [15] Wu et al. (2010), L. Wu, W. Liu, and X. J. Hu, "Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors," *Biometrics*, vol. 66, no. 2, pp. 327-335, 2010.
- [16] Y. Huang, G. Dagne, and L. Wu, "Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors," *Statistics in Medicine*, vol. 30, no. 24, pp. 2930-2946, 2011.
- [17] Yi- Kuang Tseng et al. A joint modeling approach for evaluating the efficacy of HAART treatment for AIDS patients in Taiwan.
- [18] Rui Martins, Silva G.L, and Andreozzi. V. (2012), Spatial joint analysis of longitudinal and survival AIDS data in Brazil.
- [19] Y. K. Tseng, F. Hsieh, and J. L. Wang, "Joint modelling of accelerated failure time and longitudinal data," *Biometrika*, vol. 92, no. 3, pp. 587-603, 2005.
- [20] Thiebaut R, et al. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection, *Statist. Med.* 2005; 24:65-82.
- [21] Martin WG Brinkhof, François Dabis, Landon Myer et al., WHO: Early loss of HIV-infected patients on potent antiretroviral therapy programmes in lower income countries, <http://www.who.int/bulletin/volumes/86/7/07-044248/en/>.
- [22] Poppa et al. 2003, British HIV Association (BHIVA)/British Association for sexual health and HIV guidelines on provision of adherence support to individuals receiving antiretroviral therapy. *HIV Medicine*, 5, 46-60.
- [23] Gerardo Alvarez uria et al. 2013, Predictors of delayed antiretroviral therapy initiation, mortality and loss to follow up in HIV infected patients eligible for HIV treatment: Data from an HIV cohort study in India, *BioMed Research International*, 2013.
- [24] Nelson K and Williams CM. *Infectious Disease Epidemiology. Theory and Practice*. 2nd ed. 2007.
- [25] Langford SE, Ananworanich J and Cooper DA. Predictors of disease progression in HIV infection: a review. *AIDS Res Ther.* 2007; 4:11.
- [26] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963 - 974.
- [27] Dalal RP, Macphail C, Mqhayi M, Wing J, Feldman C, et al. (2008) Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa. *J Acquir Immune Defic Syndr* 47: 101-107.
- [28] Geng EH, Emenyonu N, Bwana MB, Glidden DV, Martin JN (2008) Sampling based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *JAMA* 300: 506-507.
- [29] S. M. Ingle, M. May, K. Uebel et al., "Outcomes in patients waiting for antiretroviral treatment in the Free State Province, South Africa: prospective linkage study," *AIDS*, vol. 24, no. 17, pp. 2717-2725, 2010.
- [30] E. H. Geng, M. B. Bwana, W. Muyindike et al., "Failure to initiate antiretroviral therapy, loss to follow-up and mortality among HIV-infected patients during the Pre-ART period in Uganda," *Journal of Acquired Immune Deficiency Syndromes*, vol. 63, no.2, pp. e64-e71, 2013.
- [31] Bygrave H, Kranzer K, Hilderbrand K, Whittall J, Jouquet G, et al. (2010) Trends in Loss to Follow-Up among Migrant Workers on Antiretroviral Therapy in a Community Cohort in Lesotho. *PLoS ONE* 5(10): e13198. doi:10.1371/journal.pon e.0013198.