**Variables under investigation**

The longitudinal response variables for current study were CD4 cell count and BMI. The two response variables are different in nature. CD4 cell count is count/discrete in nature and BMI calculated using weight and height of a patient as BMI = and obtained from the hospital is categorized as normal (if 18.5 kg/m2< BMI< 24.9kg/m2) and abnormal (under-weight (if BMI<18.5 kg/m2)/ over-weight(if BMI > 25kg/m2)) which is dichotomous in nature.

On the other hand, the predictor variables for the two responses were age in years, gender (male, female), marital status (living with partner, living without partner), ownership of cell phone (yes, no), weight in kilogram, height in meters, baseline CD4 cell count in cells/mm3, disclosure of the disease (yes, no), educational status (no education, primary, secondary and tertiary), residential area (rural, urban), WHO stages (stage1, stage2, stage3 and stage4), medication adherence (adherent, non-adherent), level of income (low, middle and high), follow up times/visits (1,2….23) and dietary instruction adherence(adherent, non-adherent). Self-reported data were also employed to assess whether there is social support from the community around them and for the existence of mental depression of patients under HAART.

**Statistical models for current investigation**

The standard model for count data is Poisson distribution. It is, therefore, useful at the outset to review some fundamental properties and characterize results of the Poisson distribution. If the discrete random variable Y has Poisson distribution with intensity or rate parameter *μ*, *μ* > 0 and t is the exposure defined as the length of time which the event is recorded, then Y has the density[[23](#_ENREF_23)].

(Y=y) = , y = 0, 1, 2,… (1)

where E(y)= var(y)= . If the time period equals to one, then its density given in (1) equals to;

(Y=y) = , y = 0, 1, 2,… (2)

In over-dispersed Poisson model, an extra parameter is included and this helps to estimate how much larger the variance than the mean [[23](#_ENREF_23)]. In the over-dispersed distribution, one alternative approach to fit extra dispersion parameter is a quasi-Poisson model. It has two parameters; mean, μand over-dispersion parameter such that variance is a linear function of mean [[23](#_ENREF_23)]. Hence for random variable y that follows auasi-Poisson distribution, we have:

where is expected value or mean of count response variable and is over-dispersed parameter for quasi-Poisson model. The other potential count regression model used for over-dispersed count response data is negative-Binomial model. The first two moments of negative- Binomial regression model are [[23](#_ENREF_23)].

where is the expected value of count response variable and is over dispersed parameter for negative-Binomial model. If > 0, variance is greater than mean and becomes over dispersed[[23](#_ENREF_23)].

Since our CD4 cell count data was over-dispressed (variance > mean) for each follow up time/visit, the two models, quasi-Poisson and negative-Binomial regression models are potential candidates for fitting our data. Therefore, we compared the two models using the values of log-likelihood, AIC and BIC to assess goodness-of-fit for our data[[23](#_ENREF_23)]. Since, AIC and BIC were smaller for quasi-Poisson, it has been selected as a model of CD4 cell count marginal data analysis. Similarly, the binary logistic regression model was employeed for marginal data analysis of the diotoumus outcome response variable, BMI. In potential predictor variable selection, we considered all predictors in the model, and fitted each product term obtained from predictor variables one at a time.

To assess the association between CD4 cell count and BMI data, the joint generalized linear mixed model was fitted. In this model, the correlation between the two responses is specified through the random effect structure assuming separate random intercept for each outcome and combining them by imposing joint multivariate distribution on the random intercept. A direct specification of joint distribution for both outcomes using mixed effects model was developed [[18](#_ENREF_18)]. The conditional independence of random effects in joint model consideres the following components [[20](#_ENREF_20)]:

where log and logit are link function of qusi-Poisson and negative-Binimial models respectively. The formulation of a joint model by striking a joint multivariate distribution helped in developing a joint multivariate distribution in the random effects of the two separate models.