

# Multilevel Modeling of the Progression of HIV/AIDS Disease Among Patients Under HAART Treatment

## Multilevel Ordinal Response Modeling

Awol Seid<sup>1</sup>

Received: 2 June 2015 / Revised: 21 July 2015 / Accepted: 28 September 2015 /  
Published online: 22 October 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Human immune deficiency virus results a noncurable disease acquired immuno deficiency syndrome (AIDS). After a person is infected with virus, the virus gradually destroys all the infection fighting cells called CD4 cells and makes the individual susceptible to opportunistic infections which cause severe or fatal health problems. The most effective treatment for the disease is the highly active antiretroviral therapy (HAART) which requires a lifelong commitment to adhere diligently to daily medications, dosing schedules and making frequent clinic visits. Several studies show that the CD4 cells count is the most determinant indicator of the effectiveness of the treatment or progression of the disease. The objective of this paper is to investigate the progression of the disease over time among patients under HAART treatment. Two main approaches of the generalized multilevel ordinal models; namely the proportional odds model and the nonproportional odds model have been applied to the HAART data. Also, the multilevel part of both models include random intercepts and random coefficients. In general, four models are explored in the analysis and then the models are compared using the deviance information criteria. Of these, the random coefficients nonproportional odds model is selected as the best model for the HAART data used as it has the smallest DIC value. This selected model shows that the progression of the disease increases as the time under the treatment increases. In addition it reveals that gender, baseline clinical stage and functional status of the patient have a significant association with the progression of the disease.

**Keywords** Nonproportional odds model · Proportional odds model · Random coefficients model · Random intercepts model

---

✉ Awol Seid  
es.awol@gmail.com; awol.seid@haramaya.edu.et

<sup>1</sup> Department of Statistics, Haramaya University, Dire Dawa, Ethiopia

## 1 Introduction

Human immune deficiency virus (HIV) is a virus that causes acquired immuno deficiency syndrome (AIDS). The virus attacks the body's immune system and destroys the CD4 cells that are crucial to the normal function of the immune system which defends the body against illness.

As HIV/AIDS is not a curable disease, patients are being treated by the combination of two or more classes of antiretroviral drugs, called highly active antiretroviral therapy (HAART), throughout their life time. The success of this lifelong HAART treatment in improving the health status of the patient depends on a number of factors. [5] indicated that CD4 count, viral load, total lymphocytes, body mass index and adherence to the treatment are significant predictors of mortality due to the disease. Of these and other factors, several cohort studies and clinical trials have shown that CD4 count is the strongest predictor of subsequent disease progression and survival [6, 13]. For example, CD4 counts below 200 cells per  $\text{mm}^3$  indicate occurrence of the greater risk of the disease, while the risk of progression increases substantially at CD4 counts  $< 350$  cells per  $\text{mm}^3$  and CD4 counts from 350 to 500 cells per  $\text{mm}^3$  are associated with risks of  $\leq 5\%$  across all ages [18]. Thus, lower CD4 counts are associated with greater risk of disease progression.

Many researchers have carried out studies on the change in CD4 cell counts of HIV patients. Several of these studies have look at change from the cross sectional point of view without considering the pattern of change over the period of study. For example, [11] used linear regression models to study some predictors of CD4 cell counts recovery in HIV-1 positive patients receiving Antiretroviral Therapy and showed that the change in CD4 cells is linear. [1] used a longitudinal approach and suggested that the pattern of growth in CD4 cell was not linear. In addition, [17] reveals that the mean square root CD4 cell counts of HIV/AIDS patients has a quadratic relationship with time. Therefore, several studies arrive at different relationship of the CD4 count over time.

The basis of this paper is, therefore, modeling of the progression of HIV/AIDS disease using random intercepts and random coefficients ordinal response models. The progression of the disease represent the seriousness of HIV/AIDS sickness which is determined by splitting the immunological indicator namely the *CD4* count into four stages as in [8]. These stages are CD4 count  $< 200$  cells per  $\text{mm}^3$ ,  $200 \leq \text{CD4}$  count  $< 350$  cells per  $\text{mm}^3$ ,  $350 \leq \text{CD4} < 500$  cells per  $\text{mm}^3$  and CD4 count  $\geq 500$  cells per  $\text{mm}^3$ . Thus, by considering changes over time, the main focus is to compare two approaches of the generalized multilevel ordinal models; namely the proportional odds model and the non-proportional odds model. The random part of both models include random intercepts and random coefficients.

Following parts of the paper are organized as follows. Section 2 describes the materials and methods used in the study. Section 3 presents the results and discussion. Finally, conclusions are provided in Sect. 4.

## 2 Materials and Methods

### 2.1 Description of the HAART Data

The HAART data used for this study is the data used by [17] which was obtained from Jimma University Specialized Hospital HIV Outpatient Clinic, South West of Ethiopia. It consists of 1464 HIV/AIDS patients who were 18 years old or older, and started the HAART treatment any time in between 1st January 2007 to 31st December 2011.

The response variable is CD4 representing the progression of the disease as described previously. Thus, it has four values; 1 if the CD4 count  $< 200$  cells per  $\text{mm}^3$ , 2 if  $200 \leq \text{CD4 count} < 350$  cells per  $\text{mm}^3$ , 3 if  $350 \leq \text{CD4} < 500$  cells per  $\text{mm}^3$  and 4 if the CD4 count  $\geq 500$  cells per  $\text{mm}^3$ .

This study considers 5 explanatory variables. These are CD4 observation time in months (Time), baseline age of the patient in years (Age), gender (Female, Male), baseline clinical stage of the patient (Stage I, Stage II, Stage III, Stage IV) and baseline functional status of the patient (Working, Ambulatory, Bedridden).

### 2.2 Multilevel Ordinal Models

Multilevel data are a commonly encountered phenomenon especially in the fields of medical, biological and social sciences. Statistical modeling of multilevel data has been in discussion for many years and many developments have been made [2, 9, 10]. In such studies, the interest lies in drawing inference about the regression parameters of a marginal model for correlated responses while the association structure between the responses is of secondary importance [3]. As most of the early developments are concentrated in the area of continuous response variables, the field of multilevel modeling for discrete categorical responses is a relatively new approach [7, 16].

In this study, a two-level analysis is employed. Consider a response variable  $Y$  having  $k$  categories ( $s = 1, 2, \dots, k$ ). Suppose category  $k$  is chosen as a reference category. Let  $j$  represent the higher level (level-2) units and let  $i$  represent the lower level (level-1) units (nested observations). In this particular study, the level-2 units are patients (identified by card numbers) and the level-1 units are the CD4 observation times. Thus, each observed response is identified by the combination of card numbers and observation times.

Assume that there are  $j = 1, 2, \dots, N$  level-2 units and  $i = 1, 2, \dots, n_j$  level-1 units within each patient (level-2 unit). The total number of level-1 observations across level-2 units is given by  $n = \sum_{j=1}^N n_j$ .

Let  $Y_{ij}$  be the value of the ordinal response associated with level-1 unit  $i$  nested within level-2 unit  $j$ . Thus,  $\pi_{ij}^{(s)}$  denotes the probability of patient  $j$  having a response variable value of  $s$  at the observation time  $i$ . The cumulative response probabilities are:

$$P(Y \leq s) = \gamma_{ij}^{(s)} = \sum_{h=1}^s \pi_{ij}^{(h)}, \quad s = 1, 2, \dots, k - 1 \quad (1)$$

where  $\gamma_{ij}^{(s)}$  represents the expected cumulative proportion (out of the total  $n_i$  observations) for the  $j^{\text{th}}$  patient at time  $i$ . Thus, expressing the category probabilities in terms of the cumulative probabilities:

$$\pi_{ij}^{(h)} = \gamma_{ij}^{(h)} - \gamma_{ij}^{(h-1)}; \quad 1 < h < k \quad (2)$$

but  $\pi_{ij}^{(1)} = \gamma_{ij}^{(1)}$  and  $\gamma_{ij}^{(k)} = 1$ .

The choice of the marginal model depends on the nature of the response scale. The basic ordinal response model often utilize the cumulative comparisons of the ordinal outcome with the logit link function.

$$\text{logit}(\gamma_{ij}^{(s)}) = \alpha^{(s)} + \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_j \quad (3)$$

where

$$\gamma_{ij}^{(s)} = \frac{\exp(\alpha^{(s)} + \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_j)}{1 + \exp(\alpha^{(s)} + \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_j)}.$$

In this model, the term  $\mathbf{u}_j$  are random effects specific to the second level. The  $\mathbf{x}$  represents the fixed effect variables and the  $\mathbf{z}$  (subset of the  $\mathbf{x}$  variables) are the variables associated with the random effects. Here, it is assumed that  $\mathbf{u}_j \sim N(\mathbf{0}, \boldsymbol{\Psi})$  where  $\boldsymbol{\Psi}$  is the variance covariance matrix of the random effects.

It is worth mentioning that the linear predictor differs in the marginal model (3) above. First, the category specific intercepts need to satisfy a monotonicity condition  $\alpha^{(1)} \leq \alpha^{(2)} \leq \dots \leq \alpha^{(k-1)}$  only when the family of cumulative link models are employed. Second, the regression parameter coefficients ( $\boldsymbol{\beta}$ ) of the covariates are category specific (do not vary across categories). Thus, the relationship between the explanatory variables and the cumulative logits does not depend on  $s$ . [14] calls this assumption of identical odds ratios across the  $k - 1$  cut-offs the proportional odds assumption. As written above, a positive coefficient for an explanatory variable indicates that as the values of the explanatory variable increase so do the odds that the response is greater than or equal to  $s$ .

Assuming an underlying multinomial distribution for the category probabilities, the covariance matrix of the the cumulative proportions is given by:

$$\text{cov}(y_{ij}^{(s)}, y_{ij}^{(r)}) = \frac{\gamma_{ij}^{(s)}(1 - \gamma_{ij}^{(r)})}{n_i}; \quad s \leq r \quad (4)$$

As noted by [15], violation of the proportional odds assumption is common. Thus, they described a (fixed-effects) partial proportional odds model in which the covariates are allowed to have differential effects on the  $t - 1$  cumulative logits. Hence, this non-proportional odds model is widely used in situations where there is no evidence to suggest the effect of certain variables do not behave proportionality across response categories. In such case, the cumulative proportions is modeled as:

$$\text{logit}(\gamma_{ij}^{(s)}) = \alpha^{(s)} + \mathbf{x}_{ij}\boldsymbol{\beta}^{(s)} + \mathbf{z}_{ij}\mathbf{u}_j \quad (5)$$

where the term  $\boldsymbol{\beta}^{(s)}$  depict the estimated coefficients that vary across the logits.

### 2.3 Bayesian Estimation and the Likelihood Function

Estimation of both proportional and nonproportional odds models are estimated under a Bayesian framework through Markov chain Monte Carlo (MCMC) methods using the specialized software package for fitting multilevel models MLwiN [4]. For this study, I used the Stata command `runmlwin` to fit the models in MLwiN 2.32 seamlessly from within Stata. The command allows to fit models by both the IGLS and MCMC algorithms and provides full control over all aspects of model specification and estimation [12].

Bayesian inference is based on the posterior distribution given the observed data. For the ordinal models, the probability of a response in category  $s$  for a given level-2 unit  $j$ , conditional on the random effects  $\boldsymbol{\theta}$  is equal to

$$\pi_{ij}^{(s)} = P(Y_{ij} = s | \boldsymbol{\theta}) = \gamma_{ij}^{(s)} - \gamma_{ij}^{(s-1)}.$$

In what follows, the general model allowing for nonproportional odds is considered, the more restrictive proportional odds model is just a special case.

Let  $\mathbf{Y}_j$  denote the vector of ordinal responses from level-2 unit  $j$  (for the  $n_j$  level-1 units nested within). The probability of any pattern  $\mathbf{Y}_j$  conditional on  $\boldsymbol{\theta}$  is equal to the product of the probabilities of the level-1 responses,

$$\ell(\mathbf{Y}_j | \boldsymbol{\theta}) = \prod_{i=1}^{n_j} \prod_{s=1}^k (\gamma_{ij}^{(s)} - \gamma_{ij}^{(s-1)})^{y_{ij}^{(s)}}$$

where  $y_{ij}^{(s)} = 1$  if  $Y_{ij} = s$  and 0 otherwise (i.e., for each  $ij$ -th observation  $y_{ij}^{(s)} = 1$  for only one of the  $k$  categories).

Since Bayesian analysis combines the prior distribution of parameters with the likelihood function of the observed data, the marginal density of  $\mathbf{Y}_j$  in the population is expressed as the integral of the likelihood  $\ell(\cdot)$  weighted by the prior density  $g(\cdot)$ ,

$$h(\mathbf{Y}_j) = \int_{\boldsymbol{\theta}} \ell(\mathbf{Y}_j | \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

where  $g(\boldsymbol{\theta})$  is prior distribution of the parameters. For this study, multivariate normal and inverse gamma priors are used for the fixed effects parameter vector and the variance of the random effects, respectively. Hence,  $g(\boldsymbol{\theta})$  represents the multivariate standard normal and inverse gamma densities.

The marginal loglikelihood from the  $N$  level-2 units is then given by,

$$\log L = \sum_{j=1}^N \log h(Y_j) = \sum_{j=1}^N \log \left( \int_{\theta} \ell(Y_j|\theta) g(\theta) d\theta \right).$$

which should be maximized to obtain the MCMC parameter estimates.

### 3 Results and Discussion

The HAART data used consists of  $N = 1464$  patients (level-2 units). Also the indicator of the progression of the disease, the CD4 count, was measured approximately every six months for a maximum of successive five years. Thus, there are 1–10 ( $n_j = 1, 2, \dots, 10$ ;  $j = 1, 2, \dots, 1464$ ) level-1 units (recoded responses) for each patient. Therefore, there are about  $n = 4655$  observations across all 1464 patients.

The four categories of the response variable (CD4) can be viewed as being increasingly good progression of the disease. The lowest value of the response variable ( $CD4_{ij} = 1$ ) represents the bad progression and the largest value ( $CD4_{ij} = 4$ ) represents good progression of the disease. The fourth category of the response ( $CD4_{ij} = 4$  or  $CD4 > 500$  cells per  $\text{mm}^3$ ) is taken as the reference category. Thus,  $CD4_{ij} \sim \text{Multinomial} \left( \pi_{ij}^{(1)}, \pi_{ij}^{(2)}, \pi_{ij}^{(3)} \right)$  where  $\pi_{ij}^{(1)}$ ,  $\pi_{ij}^{(2)}$  and  $\pi_{ij}^{(3)}$  denote the probabilities of having  $CD4 < 200$ ,  $200 \leq CD4 < 350$  and  $350 \leq CD4 < 500$  for the  $j$ th patient at the  $i$ th observation time, respectively.

To assess the progression of the HIV/AIDS disease, two ordinal multilevel models are fitted to the HAART data, the first assuming a proportional odds model and the second relaxing this assumption. For both analyses, the repeated ordinal responses (progression stages) are modeled in terms of the time-varying observation time effect and the other baseline covariates. In terms of the multilevel part of the model, both a random intercept (patient) effect and random coefficient (random intercept for each patient and random slope for the observation time) are included in both analyses.

Each model is, first, fitted by iteratively generalized least squares (IGLS) to obtain starting values for the model parameters for the Markov Chain Monte Carlo (MCMC) estimation. In all of the models, the MCMC estimation is based on a sampling of 50,000 iterations following a 25,000 iteration “burn-in” period.

#### 3.1 Random Intercepts Model

Random intercepts model (i.e.,  $\mathbf{z}_j = \mathbf{1}_{n_j}$ ) is used for both the proportional odds model and nonproportional odds model. The random intercepts proportional odds model is of the form:

$$\begin{aligned}
\log \left( \pi_{ij}^{(1)} / \pi_{ij}^{(2,3,4)} \right) &= \alpha_0^{(1)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j} \\
\log \left( \pi_{ij}^{(1,2)} / \pi_{ij}^{(3,4)} \right) &= \alpha_0^{(2)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j} \\
\log \left( \pi_{ij}^{(1,2,3)} / \pi_{ij}^{(4)} \right) &= \alpha_0^{(3)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j}
\end{aligned} \tag{6}$$

where  $u_{0j}$  is the random effect associated with each patient (card number). It is assumed that  $u_{0j} \sim N(0, \sigma_{u_0}^2)$ .

Since the response variable has four categories, the model is set up as three log-odds contrasts (logits). The first log odds contrast model is the log odds of  $CD4 < 200$  versus  $CD4 \geq 200$  which can be interpreted as the logit of the expected probability that patient  $j$  has CD4 count less than 200 at time  $i$ , given the values of the explanatory variables. The second log odds contrast model, the log odds of  $CD4 < 350$  versus  $CD4 \geq 350$  compares the log odds of that patient  $j$  has CD4 count less than 200 instead of having CD4 count 200 or more at time  $i$ , given the values of the explanatory variables. Similarly, the third log odds contrast model, the log odds of  $CD4 < 500$  versus  $CD4 \geq 500$  shows, for given the values of the explanatory variables, the logit of the expected probability that patient  $j$  has CD4 count less than 500 at time  $i$ . Also, separate intercepts (threshold parameters) are estimated in each log-odds contrast. The effects of all predictor variables are assumed constant across the log-odds contrasts (proportional odds assumption).

Similarly the nonproportional random intercepts model is of the form:

$$\begin{aligned}
\log \left( \pi_{ij}^{(1)} / \pi_{ij}^{(2,3,4)} \right) &= \alpha_0^{(1)} + \beta_1^{(1)} \text{Time}_{ij} + \beta_2^{(1)} \text{Age}_j + \beta_3^{(1)} \text{Gender}_j \\
&\quad + \beta_4^{(1)} \text{Stage II}_j + \beta_5^{(1)} \text{Stage III}_j + \beta_6^{(1)} \text{Stage IV}_j \\
&\quad + \beta_7^{(1)} \text{Ambulatory}_j + \beta_8^{(1)} \text{Bedridden}_j + u_{0j} \\
\log \left( \pi_{ij}^{(1,2)} / \pi_{ij}^{(3,4)} \right) &= \alpha_0^{(2)} + \beta_1^{(2)} \text{Time}_{ij} + \beta_2^{(2)} \text{Age}_j + \beta_3^{(2)} \text{Gender}_j \\
&\quad + \beta_4^{(2)} \text{Stage II}_j + \beta_5^{(2)} \text{Stage III}_j + \beta_6^{(2)} \text{Stage IV}_j \\
&\quad + \beta_7^{(2)} \text{Ambulatory}_j + \beta_8^{(2)} \text{Bedridden}_j + u_{0j} \\
\log \left( \pi_{ij}^{(1,2,3)} / \pi_{ij}^{(4)} \right) &= \alpha_0^{(3)} + \beta_1^{(3)} \text{Time}_{ij} + \beta_2^{(3)} \text{Age}_j + \beta_3^{(3)} \text{Gender}_j \\
&\quad + \beta_4^{(3)} \text{Stage II}_j + \beta_5^{(3)} \text{Stage III}_j + \beta_6^{(3)} \text{Stage IV}_j \\
&\quad + \beta_7^{(3)} \text{Ambulatory}_j + \beta_8^{(3)} \text{Bedridden}_j + u_{0j}
\end{aligned} \tag{7}$$

where  $u_{0j} \sim N(0, \sigma_{u_0}^2)$ .

The parameter estimates of Model (6) and Model (7) together with their 95 % credible interval are presented in Table 1. In both models, the result show that all explanatory variables, except age, are significant at 5 % level of significance in the proportional odds model. But in the nonproportional odds models, at least one of the design variables of the corresponding categorical variable is significant. Even, age is significant in the third contrast of the nonproportional odds model. Thus, all variables are significantly associated with the progression of the disease in the proportional odds model.

In terms of the random patient effects, the estimated between-patient variance for the proportional odds model is 2.3416 with the 95 % credible interval (1.9809, 2.7434) and for the nonproportional odds model is 2.4712 with its corresponding 95 % credible interval (2.1018, 2.8891). It can be easily observed that the estimated between patient variance is smaller in the proportional odds model than that of the nonproportional odds model. And also the credible interval in the proportional odds model is wider than that of the nonproportional odds model. Therefore, the credible intervals for the between patient variance in both models clearly indicate that the between patient variance is significant or the data are correlated within patients.

Generally, for a random intercepts model, it is often of interest to express the level-2 variance in terms of an intra-class correlation. For the ordinal multilevel model assuming normally distributed random effects, the estimated intra-class correlation is  $\hat{\sigma}_{u_0}^2 / (\hat{\sigma}_{u_0}^2 + \pi^2/3)$  where the term in the denominator represents the variance of the underlying latent response tendency.

As said before, the significance of the random patient effects in Table 1 shows that the data are correlated within subjects. Expressing this as an intra-class correlation, the attributable variance at the patient level equals 0.4158 (41.58 %) and 0.4289 (42.89 %) for proportional odds and nonproportional odds models respectively. This means that about 41 % of the variation in the response (progression of the disease) is explained by the difference among patients.

### 3.2 Random Coefficients Model

Similar to the random intercepts model, both the proportional and nonproportional odds models are also fitted in a random coefficients form. The random coefficients model fitted, here, includes a random slope for each observation time in addition to a random intercept for each patient. The proportional odds model analysis assumes the effects of the explanatory variables are the same across the three cumulative logits of the model, whereas the nonproportional odds model analysis estimates effects for each explanatory variable on each of the three cumulative logits.

The random coefficients proportional odds model has the form:



**Table 1** Bayesian estimates and the 95 % credible intervals for parameters of the random intercept models

Parameter	Proportional odds model (11)		Nonproportional odds model (12)		Nonproportional odds model (12)		Nonproportional odds model (12)	
	$\log\left(\frac{(\leq s)}{\pi_{ij}} / \pi_{ij}^{(> s)}\right); s = 1, 2, 3, 4$	95 % CI	$\log\left(\frac{(\leq s)}{\pi_{ij}} / \pi_{ij}^{(> s)}\right)$	95 % CI	$\log\left(\pi_{ij}^{(1,2)} / \pi_{ij}^{(3,4)}\right)$	95 % CI	$\log\left(\pi_{ij}^{(1,2,3)} / \pi_{ij}^{(4)}\right)$	95 % CI
$\hat{\alpha}^{(1)}$	-0.4543	(-0.8974, -0.0468)	-0.3575	(-0.8866, 0.1114)				
$\hat{\alpha}^{(2)}$	1.8104	(1.3631, 2.2190)			1.6351	(1.1620, 2.0877)		
$\hat{\alpha}^{(3)}$	3.4260	(2.9638, 3.8494)					2.7305	(2.1815, 3.2654)
Time	-0.1025	(-0.1100, -0.0952)	-0.1386	(-0.1504, -0.1269)	-0.0953	(-0.1041, -0.0865)	-0.0735	(-0.0838, -0.0634)
Age	-0.0007	(-0.0120, 0.0110)	-0.0029	(-0.0157, 0.0106)	0.0043	(-0.0076, 0.0164)	0.0169	(0.0026, 0.0318)
Female	Ref.							
Male	0.5174	(0.2880, 0.7464)	0.4966	(0.2427, 0.7514)	0.5173	(0.2579, 0.7892)	0.6655	(0.3544, 0.9851)
Stage I	Ref.							
Stage II	0.2296	(-0.0447, 0.5071)	0.4466	(0.1275, 0.7707)	0.1035	(-0.2125, 0.4289)	-0.1403	(-0.5076, 0.2313)
Stage III	0.5012	(0.2049, 0.8029)	0.6812	(0.3466, 1.0272)	0.4159	(0.0699, 0.7757)	0.1693	(-0.2463, 0.5721)
Stage IV	0.7597	(0.2564, 1.2736)	0.8689	(0.3190, 1.4338)	0.7832	(0.1681, 1.4208)	0.4450	(-0.2900, 1.2117)
Working	Ref.							
Ambulatory	0.2338	(-0.0286, 0.4877)	0.4008	(0.1102, 0.6865)	0.1005	(-0.2062, 0.3984)	-0.1080	(-0.4574, 0.2379)
Bedridden	0.3636	(-0.2439, 0.9585)	0.4541	(-0.2012, 1.1193)	0.1890	(-0.4934, 0.9071)	0.1381	(-0.6921, 1.0151)
Random effects								
$\hat{\sigma}_{\tau_0}^2$	2.3416	(1.9809, 2.7434)			2.4712	(2.1018, 2.8891)		

**Table 2** Bayesian estimates and the 95 % credible intervals for parameters of the random coefficient models

Parameter	Proportional odds model (11)		Nonproportional odds model (12)	
	Mean	95 % CI	Mean	95 % CI
$\hat{\alpha}^{(1)}$	-0.4028	(-0.9133, 0.1029)	-0.1277	(-0.6103, 0.3748)
$\hat{\alpha}^{(2)}$	2.2921	(1.7669, 2.8222)	2.0495	(1.5138, 2.5528)
$\hat{\alpha}^{(3)}$	4.3051	(3.7595, 4.8629)	-0.1311	(-0.1496, -0.1135)
Time	-0.1624	(-0.1788, -0.1465)	-0.0058	(-0.0189, 0.0083)
Age	-0.0048	(-0.0190, 0.0089)	0.5319	(0.2329, 0.8277)
Female	Ref.		0.7106	(0.3432, 1.0762)
Male	0.5929	(0.3291, 0.8616)	0.3880	(0.0421, 0.7364)
Stage I	Ref.		0.5841	(0.1944, 0.9721)
Stage II	0.4789	(0.1476, 0.8069)	0.9537	(0.2614, 1.6553)
Stage III	0.6955	(0.3334, 1.0682)	0.2674	(-0.0739, 0.6149)
Stage IV	0.9935	(0.3976, 1.6157)	0.1813	(-0.6019, 0.9851)
Working	Ref.		0.1782	(-0.8491, 1.2426)
Ambulatory	0.3855	(0.0675, 0.7098)	2.5515	(1.9321, 3.2277)
Bedridden	0.4674	(-0.2568, 1.2163)	0.0236	(0.0181, 0.0294)
Random effects			-0.0359	(-0.0748, 0.0013)
$\hat{\sigma}_{\mu_0}^2$	3.9518	(3.1992, 4.8077)		
$\hat{\sigma}_{\mu_1}^2$	0.0204	(0.0153, 0.0264)		
$\hat{\sigma}_{\mu_0, \mu_1}$	-0.1041	(-0.1532, -0.0614)		

$$\begin{aligned}
\log \left( \pi_{ij}^{(1)} / \pi_{ij}^{(2,3,4)} \right) &= \alpha_0^{(1)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij} \\
\log \left( \pi_{ij}^{(1,2)} / \pi_{ij}^{(3,4)} \right) &= \alpha_0^{(2)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij} \\
\log \left( \pi_{ij}^{(1,2,3)} / \pi_{ij}^{(4)} \right) &= \alpha_0^{(3)} + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_j + \beta_3 \text{Gender}_j \\
&\quad + \beta_4 \text{Stage II}_j + \beta_5 \text{Stage III}_j + \beta_6 \text{Stage IV}_j \\
&\quad + \beta_7 \text{Ambulatory}_j + \beta_8 \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij}
\end{aligned} \tag{8}$$

where  $u_{0j}$  and  $u_{1j}$  are the random intercept for each patient and random slope associated with the observation time. Here  $u_{0j} \sim N(0, \sigma_{u_0}^2)$ ,  $u_{1j} \sim N(0, \sigma_{u_1}^2)$  and  $\text{cov}(u_{0j}, u_{1j}) = \sigma_{01}$ .

Similar to the random coefficients proportional odds models, the random coefficients nonproportional odds model is expressed as:

$$\begin{aligned}
\log \left( \pi_{ij}^{(1)} / \pi_{ij}^{(2,3,4)} \right) &= \alpha_0^{(1)} + \beta_1^{(1)} \text{Time}_{ij} + \beta_2^{(1)} \text{Age}_j + \beta_3^{(1)} \text{Gender}_j \\
&\quad + \beta_4^{(1)} \text{Stage II}_j + \beta_5^{(1)} \text{Stage III}_j + \beta_6^{(1)} \text{Stage IV}_j \\
&\quad + \beta_7^{(1)} \text{Ambulatory}_j + \beta_8^{(1)} \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij} \\
\log \left( \pi_{ij}^{(1,2)} / \pi_{ij}^{(3,4)} \right) &= \alpha_0^{(2)} + \beta_1^{(2)} \text{Time}_{ij} + \beta_2^{(2)} \text{Age}_j + \beta_3^{(2)} \text{Gender}_j \\
&\quad + \beta_4^{(2)} \text{Stage II}_j + \beta_5^{(2)} \text{Stage III}_j + \beta_6^{(2)} \text{Stage IV}_j \\
&\quad + \beta_7^{(2)} \text{Ambulatory}_j + \beta_8^{(2)} \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij} \\
\log \left( \pi_{ij}^{(1,2,3)} / \pi_{ij}^{(4)} \right) &= \alpha_0^{(3)} + \beta_1^{(3)} \text{Time}_{ij} + \beta_2^{(3)} \text{Age}_j + \beta_3^{(3)} \text{Gender}_j \\
&\quad + \beta_4^{(3)} \text{Stage II}_j + \beta_5^{(3)} \text{Stage III}_j + \beta_6^{(3)} \text{Stage IV}_j \\
&\quad + \beta_7^{(3)} \text{Ambulatory}_j + \beta_8^{(3)} \text{Bedridden}_j + u_{0j} + u_{1j} \text{Time}_{ij}
\end{aligned} \tag{9}$$

where as usual  $u_{0j} \sim N(0, \sigma_{u_0}^2)$ ,  $u_{1j} \sim N(0, \sigma_{u_1}^2)$  and  $\text{cov}(u_{0j}, u_{1j}) = \sigma_{01}$ .

The parameter estimates of both of the models in a random coefficient terms are summarized in Table 2. Again when the random coefficient models are considered, the random patient effect is significant in both models. Also, the random linear time effect is significant in both analysis. The estimated covariance between the subject and time effects is negative but it is not significantly different from zero.

**Table 3** Comparison of the estimated models using DIC

Model	$\bar{D}$	$pD$	$DIC$
Random intercepts proportional odds (6)	9341.86	880.09	10221.95
Random intercepts nonproportional odds (7)	9192.82	906.10	10098.92
Random coefficients proportional odds (8)	8105.76	1383.61	9489.37
Random coefficients nonproportional odds (9)	8033.55	1304.46	9338.01

### 3.3 Model Comparison

Having fitted models by MCMC, then comparison of the models is done via the hierarchical modeling generalization of the Akaike Information Criterion ( $AIC$ ) called deviance information criterion ( $DIC$ ) [19]. The  $DIC$  is the sum of the posterior expectation (mean) of the deviance function ( $\bar{D}$ ) and the effective number of parameters ( $pD$ ). The term  $\bar{D}$  measures the goodness-of-fit of the model and the term  $pD$  measures the model complexity. Since a smaller  $\bar{D}$  indicates a better fit and a smaller  $pD$  indicates a parsimonious model, small value of the sum ( $DIC$ ) indicates preferred model.

Table 3 presents  $\bar{D}$ ,  $pD$  and  $DIC$  values for the four models fitted previously. When the posterior mean of the deviance function is examined, the value decreases from Model (6) to Model (9), it seems that the random coefficient nonproportional odds model fits the HAART data better. Turning to the effective number of parameters, the values increase from Model (6) to Model (9). But to select a model as best among other candidate models, both of the values should be small. Hence, when the sum of the two ( $DIC$ ) is considered, the random coefficient nonproportional odds model has the smallest value. Therefore, among the four candidate models, the nonproportional odds model in a random coefficients basis fits better for the HAART data used.

## 4 Conclusions

In this study, a novel application of Bayesian multilevel modeling for an ordinal categorical response has been discussed. Proportional odds and nonproportional odds models are considered for the analysis of the HAART data. In terms of the multilevel part of the model, both random intercepts models and models incorporating random patient intercepts and linear trends are investigated.

Specifically, four different two-level ordinal models are explored. Then, the models are compared using the Deviance Information Criteria. Of the four models, the random coefficients nonproportional odds model is the best fit for the HAART data as it has the smallest  $DIC$  value.

All the variables included in the model, except age, are found to be significantly associated with the progression of the HIV/AIDS disease. In particular, the progression of the disease increases as the time under HAART increases, assuming all the other variables constant. Also, male patients are more likely to have better progression as compared to female patients if all the other explanatory variables are assumed to be

fixed. Similarly the baseline clinical stage and functional status of the patients have differential effects on the progression of the disease.

In general, the modeling technique used in this study has not been widely used for modeling medical data with hierarchical structures. Hence the contribution of the paper would be invaluable for medical researchers and statisticians.

## Appendix

### Stata's `runmlwin` Codes

```
// Two-Level Proportional Odds Model - Random Intercepts Model
xi: runmlwin CD4 Cons (ObsTime Age Sex i.ClinStag i.FunStat, ///
    contrast(1/3)), level2(CardNum:(Cons, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) nopause

xi: runmlwin CD4 Cons (ObsTime Age Sex i.ClinStag i.FunStat, ///
    contrast(1/3)), level2(CardNum:(Cons, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) ///
    mcmc(burnin(25000) chain(50000)) initsprevious nopause

// Two Level Nonproportional Odds Model - Random Intercepts Model
xi: runmlwin CD4 Cons ObsTime Age Sex i.ClinStag i.FunStat, ///
    level2(CardNum:(Cons, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) nopause

xi: runmlwin CD4 Cons ObsTime Age Sex i.ClinStag i.FunStat, ///
    level2(CardNum:(Cons, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) ///
    mcmc(burnin(25000) chain(50000)) initsprevious nopause

// Two Level Proportional Odds Model - Random coefficients Model
xi: runmlwin CD4 Cons (ObsTime Age Sex i.ClinStag i.FunStat, ///
    contrast(1/3)), level2(CardNum:(Cons ObsTime, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) nopause

xi: runmlwin CD4 Cons (ObsTime Age Sex i.ClinStag i.FunStat, ///
    contrast(1/3)), level2(CardNum:(Cons ObsTime, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) ///
    mcmc(burnin(25000) chain(50000)) initsprevious nopause

// Two Level Nonproportional Odds Model - Random coefficients Model
xi: runmlwin CD4 Cons ObsTime Age Sex i.ClinStag i.FunStat, ///
    level2(CardNum:(Cons ObsTime, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
    denom(Cons) basecategory(3)) nopause

xi: runmlwin CD4 Cons ObsTime Age Sex i.ClinStag i.FunStat, ///
    level2(CardNum:(Cons ObsTime, contrast(1/3))) ///
    level1(Obs:) discrete(dist(multinomial) link(ologit) ///
```

```
denom(Cons) basecategory(3) ///
mcmc(burnin(25000) chain(50000)) initsprevious nopause
```

## References

1. Adams M, Luguterah A (2011) Longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on antiretroviral therapy (ART) in the Builsa District hospital. *Eur Sci J* 9(33):299–309
2. Aitkin M, Longford N (1986) Statistical modeling in school effectiveness studies. *J R Stat Soc A* 149:1–43 (with discussion)
3. Anestis T (2014) R Package multgee: a generalized estimating equations solver for multinomial responses. University of Cambridge, Cambridge
4. Browne WJ (2012) MCMC estimation in MLwiN: Version 2.26 Centre for Multilevel Modelling, University of Bristol, UK
5. Egger M (2007) Outcomes of antiretroviral treatment in resource limited and industrialized countries. In: Proceedings of the 14th conference on retroviruses and opportunistic infections
6. 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
7. Fielding A, Yang M (2005) Generalized linear mixed models for ordered responses in complex multilevel structures. *J R Stat Soc* 168:159–183
8. Giuseppe D, D'Amico G, Girolamo A, Janssen J, Iacobelli S, Tinari N, Manca R (2007) A stochastic model for the HIV/AIDS dynamic progression. *Math Probl Eng*. doi:[10.1155/2007/65636](https://doi.org/10.1155/2007/65636)
9. Goldstein H (1986) Multilevel covariance component models. *Biometrika* 74(2):430–431
10. Hedeker D, Gibbons RDA (1994) Random-effects ordinal regression model for multilevel analysis. *Biometrics* 50(4):933–944
11. Kulkarni H, Jason F, Greg G, Nancy F, Michael L, Braden H, Glenn W, Edmund T, Michael P, Matthew D, Alan R, Brian K, Sunil K, Vincent C (2011) Early post-seroconversion CD4+ cell counts independently predict CD4+ cell count recovery in HIV-1-positive subjects receiving antiretroviral therapy. *J Acquir Immune Defic Syndr* 57:387395
12. Leckie G, Charlton C (2012) Runmlwin: a program to Run the MLwiN multilevel modeling software from within stata. *J Stat Softw* 52(11):1–40
13. 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):946954
14. McCullagh P (1980) Regression models for ordinal data. *J R Stat Soc* 42:109–142
15. Peterson B, Harell FE (1990) Partial proportional odds model for ordinal response variables. *Appl Stat* 39:205–217
16. Rashbash J, Steele F, Browne WJ, Goldstein H (2004) A user's guide to MLwiN
17. Seid A, Getie M, Birlie B, Getachew Y (2014) Joint modeling of longitudinal CD4 cell counts and time-to-default from HAART treatment: a comparison of separate and joint models. *Electron J Appl Stat Anal* 07(02):292–314
18. Simone EL, Jintanat A, David AC (2007) Predictors of disease progression in HIV infection: a review. *AIDS Res Therapy* 4:11
19. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B* 64:583–616