Characterizing groups using latent class mixed models: antiretroviral treatment adherence analysis

Ruth Fuentes-García[1], Lizbeth Naranjo[1,*], Yanink Caro-Vega[1,2]

[1] Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico
* lizbethna@ciencias.unam.mx
[2] Departamento de Infectología, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico

Abstract. Latent class mixed models are used to identify families of trajectories for antiretroviral treatment adherence profiles of people with HIV. The demographic characteristics associated to these adherence classes could help pinpoint patients more likely to show decreasing adherence profiles. Additionally, within each of the adherence groups a further latent class analysis with a bivariate response is used to characterize the joint behavior of CD4-CD8 ratio and CD4+T. These could lead to better understanding of how different profiles of the outcome could be linked to different adherence trajectories. This is of interest since low CD4-CD8 ratios have been associated to increased risks both for developing AIDS and death.

Keywords: Antiretroviral treatment adherence, Classification, Latent class mixture models

1 Introduction

Current Antiretroviral treatment (ART) regimes allow people with HIV (PWH) to achieve and maintain viral suppression up to 3 to 5 years since first ART initiation [6, 7, 20]. Adherence to the treatment, that is, opportune and complete intake of the prescribed dosage, is of crucial importance for the patient's evolution. Incomplete adherence has been identified as one of the potential factors that could drive heightened inflammation, immune activation and coagulopathy in adult people with HIV receiving ART [8–10]. Hence, understanding the adherence dynamics across time for PWH is relevant for the adequate surveillance of their condition.

The linear mixed model is frequently used to describe change over time of a dependent Gaussian variable, often a set of covariates is considered to determine their possible effect over such change. A particular challenge in our problem at hand is the presence of heterogeneity in the population, where subpopulations with different growth curve shapes could fit better. Namely we are interested in

identifying different patterns in the response variable, defining subpopulations, which could be linked to a difference in behavior among the observations. Latent class mixed models (LCMM), provide a useful proposal to tackle the problem. These models combine the characteristics of a linear mixed model with a latent variable, which divides the population into subsets of homogeneous observations. To estimate the parameters in these models we used the library lcmm available in R [15]. Muthén [14, 13] proposed a finite mixture model with random effects known in the literature as Growth Mixture Models, as a tool to represent unobserved heterogeneity between observations, hence capturing latent trajectory classes with different growth curve shapes. His approach is widely used and is implemented in the licensed software Mplus, it uses a continuous latent variable, defined as a function of the covariables of interest with class specific parameters and individual random effects. Given the class membership, this model leads to different multivariate Normal distributions that those defined by the LCMM, however, in general, it is similar in purpose and construction.

The HIV viral load and CD4+T lymphocytes counts are the most used biomarkers to monitor HIV-positive patients. The HIV infection progression is characterized by a progressive decrease of CD4+T counts and an increase of CD8+T counts, and therefore a simultaneous decrease of CD4/CD8 ratio. The CD4/CD8 ratio has been used as a biomarker of disease stage in PWH, as a marker of immune activation and immune senescence. Persistently low CD4/CD8 ratio has been associated with residual inflammation and identified as a good predictor of increased risk of death [5]. The CD4/CD8 ratio might represent a good predictor of AIDS and non-AIDS events, it is a more widely available measurement than other inflammatory biomarkers and is part of routine clinical care when monitoring CD4+T in many clinical settings. However, low CD4/CD8 ratios could be observed for different combinations of values of the numerator and denominator of the ratio. It would therefore be useful to explore the joint behavior of CD4+T and CD4/CD8 ratio. Antiretroviral adherence in people with HIV and virological suppression is associated with a decreased CD4-CD8 ratio [3].

The outline of this paper is organized as follows. In Section 2 we describe the population under study and the adherence to treatment. Section 3 describes the general setting for the latent class mixed models. In Section 4 we show the results and the classes identified through the LCMM; and finally a discussion and conclusions are given in Section 5.

2 Materials

In this section we describe the population, the data sources and data collection used for this study, and the ART adherence definition.

2.1 Population

A subset of PWH receiving medical care at the HIV / AIDS Clinic at the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ) is

considered, which is a national referral center located in Mexico City. Patients included in this analysis were those diagnosed with HIV-infection, that started their first antiretroviral treatment (ART) regimen between January 2005 and September 2017, were followed up for at least 6 months, and achieved viral suppression. The viral suppression is defined as having less than 200 copies of HIV per milliliter of blood (copies/μL). These patients were followed only if they showed uninterrupted periods of HIV suppression and if adherence information was available. Therefore, time in follow-up started the date of the first HIV RNA measurement $<$ 200 copies/μL after having started ART. Follow-up ended the date of a confirmed HIV RNA measurement $\geq$ 400 copies/μL, a period longer that 6 months without ART provision, last medical visit, death, or cohort administrative censoring (September 28th, 2017). For this analysis the subset of the population we considered corresponds to 988 patients in their first five years of their follow-up period.

2.2 Data collection

Data from the HIV / AIDS clinic and pharmacy databases was retrieved. Additionally, prospective information on all PWH receiving care in clinic was collected. The database contains patients' demographic, clinical and laboratory characteristics. Data are collected at each visit and entered immediately into the database. As a clinical HIV-care center receiving funds for ART from Federal Governments Ministry of Health, the clinic provides free-of-charge ART prescribed by the physicians in the clinic [18]. Patients refill their medications monthly, or bimonthly, and the pharmacy record electronically the ART provision. The pharmacy registers the refill date and when 30 or 60-day provision of ART is delivered to each patient.

The CD4+T and CD8+T lymphocyte counts (cells/μL) are measured by flow cytometry, at baseline (at enrollment) and routinely two or three times a year, during follow-up visits [2]. The CD4/CD8 ratio is computed from the division of the CD4+T lymphocyte counts (cells/μL)over the CD8+T lymphocyte counts (cells/μL).

2.3 ART adherence definition

We measured adherence (ADH) using the medication possession ratio (MPR) [12] based on pharmacy registries. We divided the 30-day doses of ART provided for a month period by the total number of days that the patient took to refill their next prescription to calculate the percentage of days that patients have ART on their possession. All surplus ART were taken into account in case the patients had picked-up a previous prescription prior to 30 days. For example, if a patient filled his/her prescription at day 25, she/he was considered to have a 100% adherence for that month, and we carried forward the extra 5-day provision to the next month estimated MPR [12]. The moving average of the last 4 months was taken as the measure of adherence for the statistical analysis [4]. Adherence was capped at 100%.

3 Latent class mixture models

The latent class mixed model considers a heterogeneous population composed by K latent classes of subjects characterized by K mean profiles of trajectories. Each subject is assigned to one and only one latent class. The number of classes should be established a priori choosing the best fit comparing the performance of models with different number of subpopulations, commonly through likelihood based criteria. The model for subject i with n_i repeated measurements is given as follows:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\lambda}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i,$$

where $\mathbf{Y}_i$ is the response vector for subject i with his/her n_i observations along time, $\mathbf{X}_i$ is the design matrix of fixed effects and $\mathbf{Z}_i$ is the design matrix for random effects; $\boldsymbol{\beta}$ is the vector of regression coefficients for the fixed effects; $\mathbf{b}_i$ is the random effects vector; and random error terms corresponds to $\mathbf{e}_i$; the latter two terms are assumed mutually independent with mean $\mathbf{0}$ with $\text{Var}(\mathbf{b}_i) = \mathbf{B}$ and $\text{Var}(\mathbf{e}_i) = \mathbf{R}_i$. Finally, class specific fixed effects are included in the model through the term $\mathbf{W}_i\boldsymbol{\lambda}_k$, where $\mathbf{W}_i$ and $\boldsymbol{\lambda}_k$ are the class-specific fixed effects and parameter vector for class k, respectively. Note that $\boldsymbol{\beta}$ will apply to all values of the corresponding column in $\mathbf{X}_i$, while $\boldsymbol{\lambda}_k$ will only affect observations in class k. For subject i, dummy variables are defined:

$$c_{ik} = \begin{cases} 1 & \text{if subject } i \text{ belongs to class } k, \\ 0 & \text{otherwise,} \end{cases}$$
$$(c_{i1}, \ldots, c_{iK}) \sim \text{Multinomial}\left(1, (\pi_{i1}, \ldots, \pi_{iK})\right),$$

where $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iK})'$. Probabilities π_{ik} are given by:

$$\pi_{ik} = \mathbb{P}(c_{ik} = 1|\mathbf{v}_i) = \frac{\exp(\mathbf{v}_i'\boldsymbol{\alpha}_k)}{\sum_{j=1}^{K}\exp(\mathbf{v}_i'\boldsymbol{\alpha}_j)},$$

$\mathbf{v}_i$ are variables used to determine the class membership for subject i and $\boldsymbol{\alpha}_j$'s are the unknown parameters vectors associated to class assignations, with constraint $\boldsymbol{\alpha}_1 = \mathbf{0}$.

The likelihood given the observed design matrices could be written as:

$$\log L(\theta) = \sum_{i=1}^{n}\log\sum_{k=1}^{K}\pi_{ik}f\left(y_{ik}|c_{ik}=1\right),$$

where $f(y_{ik}|c_{ik} = 1)$ is the density of a $\text{N}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\lambda}_k\ ,\ \mathbf{Z}_i\mathbf{B}\mathbf{Z}_i' + \mathbf{R}_i)$.

This model could be extended to describe multiple observed longitudinal markers. In this case an underlying latent process $\Lambda_i(t)|c_{ik}$ is assumed, generated by the d longitudinal marker,

$$\Lambda_i(t)|c_{ik} = \mathbf{X}_i(t)'\boldsymbol{\beta} + \mathbf{W}_i(t)'\boldsymbol{\lambda}_k + \mathbf{Z}_i(t)'\boldsymbol{b}_i + \omega_i(t),$$

where $\{\omega_i(t)\}$ is a zero mean Gaussian stochastic process. Let Y_{dij} be the measure of marker d, $d = 1, \ldots, D$, for subject i, $i = 1, \ldots, N$, at occasion j, $j = 1, \ldots, n_{di}$. The corresponding measurement of time is denoted t_{dij}. Here

$$Y_{dij} = H_d(\tilde{Y}_{dij}; \boldsymbol{\eta}_d), \tag{1}$$

where H_d is a parameterized marker-specific link function, $\boldsymbol{\eta}_d$ are the parameters for H_d; in this work linear link functions were used, i.e., $H^{-1}(Y) = \frac{Y-\eta_1}{\eta_2}$, where η_1 and η_2 are location and scale parameters, respectively. And $\tilde{Y}_{dij}$ is the noisy latent process at time t_{dij},

$$\tilde{Y}_{dij} = \Lambda_i(t_{dij})|c_{ik} + \mathbf{U}_i(t_{dij})'\boldsymbol{\gamma}_d + \delta_{di} + \epsilon_{dij},$$

where $\mathbf{U}_i$ are covariates with a marker-specific effect γ_d called contrast since $\sum_{d=1}^{D} \gamma_d = 0$, δ_{di} is a random intercept, and $\epsilon_{dij} \sim N(0, \sigma^2_{\epsilon_d})$ is the independent random measurement error. In this case the individual contribution to the likelihood would consider the Jacobian determinant of the link function and the density of a multivariate normal variable.

Optimizing the log-likelihood for mixture models could be done using EM algorithms or Newton-Raphson type algorithms. For this analysis we used the lcmm library available in R [15] which estimation uses maximum likelihood theory with a powerful modified Marquardt iterative algorithm, and a Newton-Raphson like algorithm. These choices were based in previous analyses regarding the speed and convergence of the method. A set of initial values is required, and the choice could affect the convergence, hence to ensure the global maximum is reached they suggest using several values. The choice of such values could be made through a grid or using random draws from one class estimated model. It is worth mentioning that several values should be tried to ensure there are no convergence concerns.

Once the models are fitted using the lcmm library, a quick comparison between them could be done using library LCTMtools. The common criteria used to compare models are Akaike information criteria (AIC) [1], Bayesian information criteria (BIC) [17] and entropy [11] (a criterion to assess the quality of the classification), where the best model is that having the lowest AIC and BIC, and the highest entropy.

4 Results

In this section we show the results of the LCMM. In the first stage the outcome was the ART adherence over time, and LCMM is used to find groups of ART adherence trajectories. Then, in the second stage, within each adherence group we considered a bivariate outcome, namely the CD4/CD8 ratio and the CD4+T counts in their logarithmic scale, and the LCMM is used to explain the biomarkers behavior by using some characteristics of the patients as explanatory variables.

4.1 Latent classes from ART adherence

In this section, we describe the population characteristics at baseline and the ART adherence classes. The analyzed data has 988 patients (894 male and 94 female) with a median follow up of 4.13 (interquartile range, IQR, 1.87-7.18) years; their median ART adherence was 99.17% and IQR 95.83%-100%, having an asymmetric distribution; their median age at the beginning of follow-up was 33 and IQR 28-41; 287 (29%) patients were diagnosed with AIDS-related conditions indicator previous to enrollment in care (if the patient was diagnosed with any of the opportunistic infections or other specific illnesses marked as AIDS-defining events by the CDC[3] at an earlier date or at the time of enrollment); the CD4+T counts at the beginning of follow-up were categorized as $\leq$200, 201-500 and >500 cells/mm^3 (341, 469 and 178 patients, respectively); their median CD4+T counts at the beginning of follow-up was 288.5 (IQR 158.0-443.8) and at the end of follow-up 484.0 (IQR 337.0-647.2); their median CD8+T counts at the beginning of follow-up was 847.0 (IQR 574.0-1175.2) and at the end of follow-up 727.0 (IQR 545.8-990.0); their median CD4/CD8 ratio at the beginning of follow-up was 0.329 (IQR 0.189-0.534) and at the end of follow-up 0.641 (IQR 0.431-0.914).

LCMM models were fitted for 2, 3 and 4 classes, where the dependent variable corresponds to ART adherence. The random effects we included are the intercept and time changes, expressed through a longitudinal behavior. The fixed effects include only a third-grade polynomial of time. Patients' sex and age were initially considered, but there were no significant changes in the estimated adherence trajectories. The proposed model was chosen based on the AIC, BIC, entropy and relative entropy, shown in Table 1, the best results were found under LCMM with 2 latent classes of ART adherence, having the lowest BIC and the highest entropy. We used a grid of initial points, and the results show convergence.

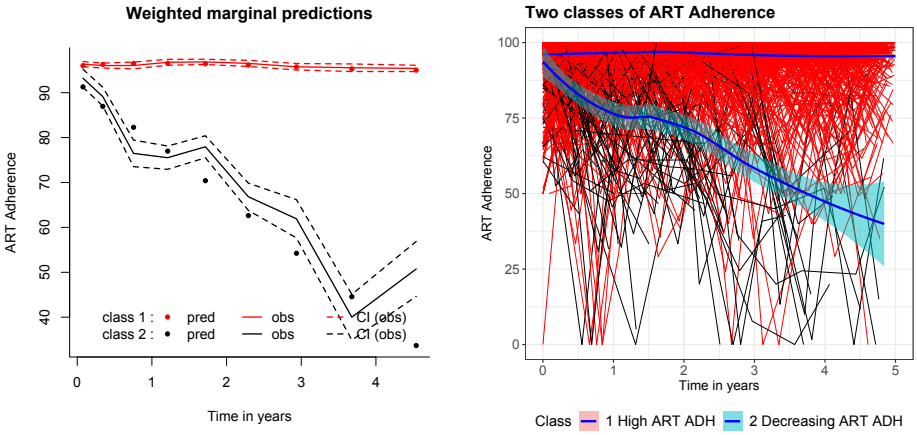
Table 1: AIC, BIC, entropy and relative entropy for LCMM with different number of latent classes for ART adherence.

Number of classes	AIC	BIC	Entropy	Relative entropy
1	66931.8	66975.8		
2	66433.5	66492.2	39.08	0.94
3	66439.5	66512.2	629.16	0.42
4	66445.5	66533.6	1030.25	0.25

The best model estimates two types of trajectories for the ART adherence with trajectories shown in Figure 1. The first class has 924 (93.52% of the total population) patients who have high levels of ART adherence for a long time, in

[3] https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5710a2.htm

average 95%. The second class is a very small class of 64 (6.48% of the total population) patients whose adherence decreases along time. This class started with a 95% average as well but decreased with time, reaching an average of 40% in the last observed time.



(A) (B)

Fig. 1: (A) Weighted marginal predictions for the ART adherence estimated by the LCMM with two classes. (B) Observed profiles of the two classes for the ART adherence; blue solid lines show the mean profiles computed with the loess smoothing method.

Table 2 shows the categorical variables describing the patient's characteristics per class of adherence. It summarizes the number of patients for the following variables; sex (male, female), previous AIDS defining event (conditions indicator previous to enrollment in care; no, yes), and CD4+T counts at the beginning of follow-up (in three categories: $\leq$ 200, 201-500, $>$ 500 cells/mm^3). In order to explore if there are differences between classes we used the Pearson's Chi-squared test with Yates' continuity correction. All p-value are greater than the significance level $\alpha = 0.05$, showing there are no differences between classes.

Table 3 summarizes continuous variables per class of patients, it shows the median and interquartile range for age, CD4/CD8 ratio, CD4+T count and CD8+T count and ART adherence, both at the beginning and at the end of the follow-up. Differences in location for the distributions between classes of latent classes were tested using Wilcoxon's rank sum test with continuity correction. All p-values are greater than the significance level $\alpha = 0.05$, showing there are no differences between classes except for the ART adherence and time. In addition, the time at the end of follow-up, which varies for individuals, is summarized per class, people who have taken medications for a longer period (class 1, high levels of ART adherence) have more adherence in comparison to people who have taken medications for a shorter period (class 2, decreasing ART adherence).

Table 2: Patient's characteristics by class of ART adherence computed by LCMM (cases and percentage), and Pearson's Chi-squared test.

Characteristics		Classes of ART ADH		Chi-squared test	
		1 High ($n = 924$)	**2** Decreasing ($n = 64$)	χ-squared	p-value
Sex	Male	838 (90.7%)	56 (87.5%)	0.3863	0.5342
	Female	86 (9.3%)	8 (12.5%)		
Previous AIDS	No	649 (70.2%)	52 (81.2%)	3.0075	0.0829
defining event[1]	Yes	275 (29.8%)	12 (18.8%)		
Baseline CD4+T	≤ 200	320 (34.6%)	21 (32.8%)	3.5071	0.1732
counts	201-500	443 (48.0%)	26 (40.6%)		
(cells/μL)[1]	> 500	161 (17.4%)	17 (26.6%)		

[1]Condition previous to enrollment in care.

Table 3: Patient's characteristics by class of ART adherence computed by LCMM (median and interquartile range), and Wilcoxon's rank sum test.

Characteristics		Classes of ART ADH		Wilcoxon test	
		1 High ($n = 924$)	**2** Decreasing ($n = 64$)	W	p-value
Age in years[2]		33 (28-41)	32 (28-37)	33590	0.0685
Time[3] at the end of follow-up		4.33 (1.96-4.75)	2.23 (1.37-3.46)	17780	<0.001
CD4/CD8	Beginning[2]	0.328 (0.188-0.530)	0.367 (0.210-0.577)	27107	0.2651
ratio	End[3]	0.647 (0.445-0.915)	0.602 (0.331-0.914)	32934	0.1275
CD4+T counts	Beginning[2]	287 (158-438)	322 (162-507)	26493	0.1637
	End[3]	486 (338-646)	461 (329-667)	30038	0.8318
CD8+T counts	Beginning[2]	845 (575-1167)	921 (549-1241)	28028	0.4856
	End[3]	723 (550-971)	907 (501-1145)	25884	0.0953
Percentage of	Beginning[2]	100 (97.5-100)	97.8 (86.0-100)	36529	0.0006
ART Adherence	End[3]	98.9 (95.6-100)	52.2 (37.5-70.8)	57197	< 0.001

[2]At the beginning of follow-up.
[3]At the end of follow-up (varies for individuals).

4.2 Latent classes for bivariate response of CD4/CD8 ratio and CD4+T within groups of adherence

CD4/CD8 ratio and CD4+T counts have been used as biomarkers of disease stage in PWH (p.e. [16]). Despite their relationship, they are usually studied sometimes using linear mixed models (LMM), but not jointly in a bivariate response model (p.e. [19]). We notice that a CD4/CD8 ratio equal to 1 could be obtained from CD4+T counts of 900 and CD8+T counts of 900, but also from 200, or any other quantity with the same CD4+T and CD8+T values. However,

having CD4+T of 900 is a normal patient's condition, but having CD4+T of 200 indicates a high risk of opportunistic infections.

Figure 2 shows the density of these variables per ART adherence class at the beginning and end of the follow-up, computed by kernel density estimates. We believe there are different features of patients within each class and a joint analysis could provide a partition that could help to identify characteristics linked to consistently low ratios. In this section we show the joint analysis of the CD4/CD8 ratio and the CD4+T counts, using the latent process mixed model for these multivariate longitudinal markers (multLCMM) [15].

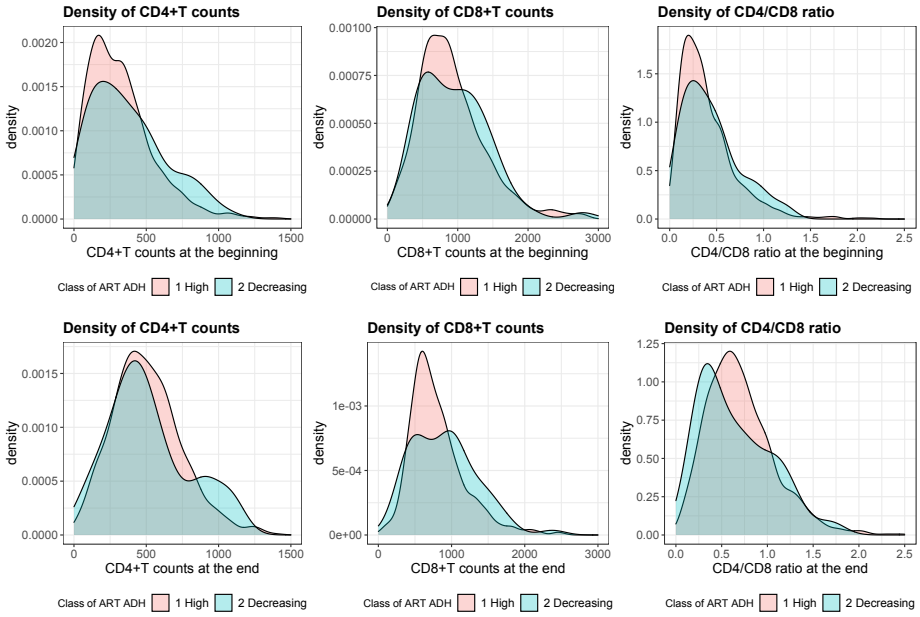


Fig. 2: Density of CD4+T counts, CD8+T counts and CD4/CD8 ratio by ART adherence latent class. Top: at the beginning of the follow-up. Bottom: at the end of the follow-up.

The dependent variable is the bivariate vector of biomarkers CD4/CD8 ratio and CD4+T in logarithm scale, i.e. (log CD4/CD8 ratio, log CD4+T). Random effects for the intercept and time were included, using a Brownian motion for its variance-covariance structure. The independent covariates are the ART adherence (ADH) measured in a continuous range of 0-100% and a 4-order polynomial for the time. In order to explore other factors associated to the biomarkers, we have also included sex (male vs. female), age in years at ART initiation, and AIDS-related conditions indicator previous to enrollment in care (yes or no) as fixed effects. Lastly, the class effects are the time and the ART adherence.

After fitting the LCMM for the ART adherence in subsection 4.1, the two classes obtained will be labeled as high (class 1) and decreasing (class 2) levels. The joint models described above were fitted for each ART adherence class separately. For both (high and decreasing) ART adherence classes 3 and 2 latent classes were used in the multLCMM, respectively. In each ART adherence class, the number of classes was chosen by BIC and entropy. Results are described in the next subsections.

High levels ART adherence class

Figure 3 displays the weighted subject-specific predictions of the latent process of the three latent classes estimated by the multLCMM, and the observed profiles for CD4/CD8 ratio, CD4+T counts and CD8+T counts computed by the loess smoothing method. Estimated parameters are summarized in Table 4. From the negative parameter estimates for age, sex, and AIDS, we observe that older male patients with AIDS show the least increase of the joint values of the response.

Table 4: High levels of ART adherence class: estimates of the multLCMM with bivariate response (log CD4/CD8 ratio, log CD4+T).

Variable	Estimate	SE	Wald	p-value
Fixed effects in the longitudinal model				
ART ADH class1	0.00051	0.00053	0.948	0.34303
ART ADH class2	0.00710	0.00228	3.109	0.00187
ART ADH class3	-0.03314	0.00474	-6.989	0.00000
AIDS: Yes	-0.20878	0.05515	-3.786	0.00015
Age	-0.00904	0.00253	-3.786	0.00035
Sex: Female	0.24589	0.08206	2.996	0.00273
Parameters of the link functions H_d *in* (1)				
log CD4/CD8 ratio -Linear η_1	-0.89597	0.07395	-12.116	0.00000
log CD4/CD8 ratio -Linear η_2	0.66828	0.02058	32.468	0.00000
log CD4 -Linear η_1	5.75555	0.05869	98.070	0.00000
log CD4 -Linear η_2	0.52911	0.01694	31.231	0.00000

Table 5 shows patient's characteristics for the subclasses in the high levels of ART adherence class (cases, or median and interquartile range). In order to test differences by latent classes we used the Pearson's Chi-squared test with Yates' continuity correction (for categorical variables) and Kruskal-Wallis test (for continuous variables). Results show there are no significant differences for sex, age and CD8+T at the beginning of follow-up. There are significant differences for AIDS, CD4+T counts and CD4/CD8 ratio. This means that latent classes are characterized mainly by the AIDS condition and levels of CD4+T counts and CD4/CD8 ratio at baseline.

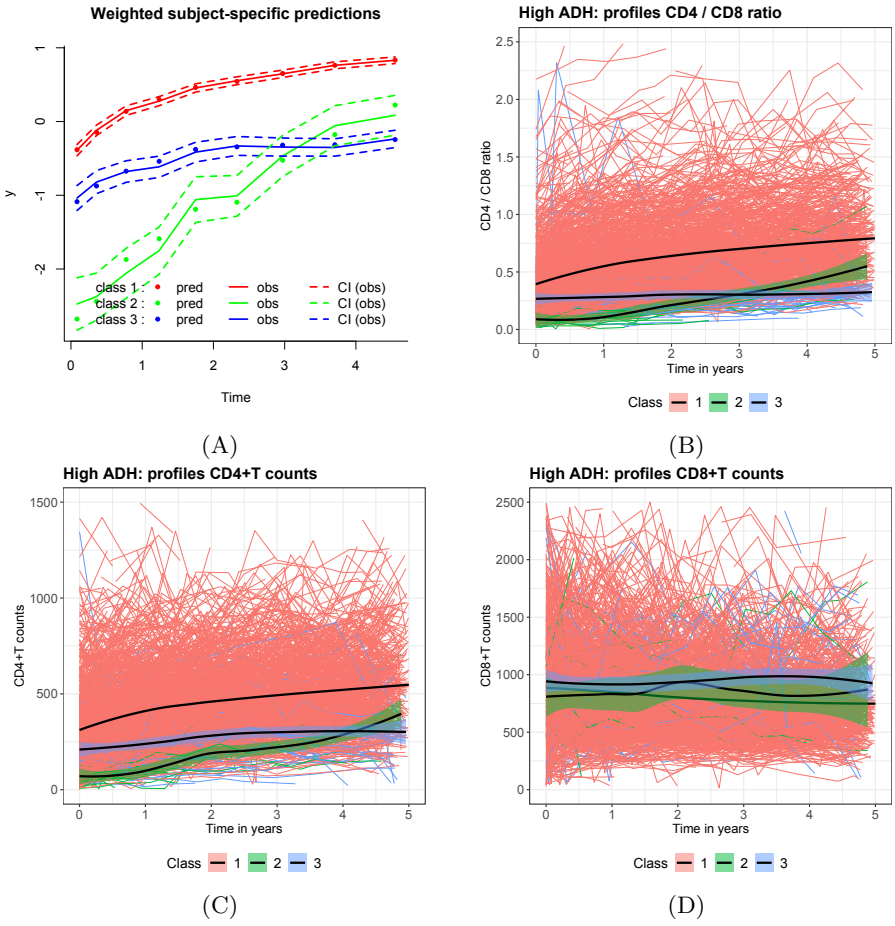


Fig. 3: Hight levels of ART adherence class: three classes from the multLCMM with bivariate response (log CD4/CD8 ratio, log CD4+T). (A) Weighted subject-specific predictions estimated by the multLCMM with three classes. Observed profiles of the three classes for (B) CD4/CD8 ratio, (C) CD4+T counts and (D) CD8+T counts; solid lines show the mean profiles computed with the loess smoothing method.

From Figure 3 we note that class 1 encompasses most of the patients; the CD4/CD8 ratio has different trajectory behavior for classes 2 and 3; class 2 has almost parallel trajectories to those of class 1, the latter in a lower level. Class 3 has a constant trajectory along time. Since these are all patients who have high levels of ART adherence, this could imply that the patient's condition is not only determined by adherence to treatment. There are also patients whose specific characteristics will lead to poor CD4+T and CD4 / CD8 ratio conditions, and that this condition will prevail, only among those with the worst initial condition.

Table 5: High levels of ART adherence class: characteristics (cases, or median and interquartile range) of patients from LCMM with bivariate response (CD4/CD8 ratio,CD4+T), Pearson's Chi-squared test, and Kruskal-Wallis test.

High ART Adherence		Classes from multLCMM			Chi-squared test	
		1	**2**	**3**	χ-squared	p-value
Characteristics		($n = 861$)	($n = 16$)	($n = 47$)		
Sex	Male	783	13	42	1.851	0.3962
	Female	78	3	5		
Previous AIDS	No	612	6	31	8.907	0.0116
defining event[1]	Yes	249	10	16		
Baseline CD4+T	≤ 200	271	15	34	58.088	<0.001
counts	201-500	432	1	10		
(cells/μL)[1]	> 500	158	0	3		

Hight ART Adherence		Classes from multLCMM			Kruskal-Wallis test	
Characteristics		**1**	**2**	**3**	χ	p-value
Age		33	37	36	5.540	0.0626
in years[2]		(28-41)	(34-46)	(30-45)		
CD4/CD8	Beginning[2]	0.341	0.074	0.188	58.974	<0.001
ratio		(0.203-0.539)	(0.045-0.094)	(0.127-0.288)		
	End[3]	0.676	0.229	0.280	128.460	<0.001
		(0.492-0.931)	(0.095-0.329)	(0.188-0.323)		
CD4+T	Beginning[2]	301	37	158	53.134	<0.001
counts		(169-446)	(19-89)	(112-212)		
	End[3]	502	169	242	91.362	<0.001
		(361-658)	(87-314)	(140-327)		
CD8+T	Beginning[2]	845	734	896	2.529	0.2824
counts		(584-1170)	(460-931)	(540-1262)		
	End[3]	717	885	857	5.956	0.0509
		(549-952)	(457-1254)	(592-1324)		

[1]Condition previous to enrollment in care.
[2]At the beginning of follow-up.
[3]At the end of follow-up (varies for individuals).

Decreasing ART adherence class

Figure 4 displays the weighted subject-specific predictions of the latent process of the two latent classes estimated by the multLCMM, and the observed profiles for CD4/CD8 ratio, CD4+T counts and CD8+T counts, computed by the loess smoothing method. Estimated parameters are summarized in Table 6. In this case, from the negative parameter estimates for age, sex, and AIDS, we observed that the lower joint increase would again correspond to older male patients with AIDS.

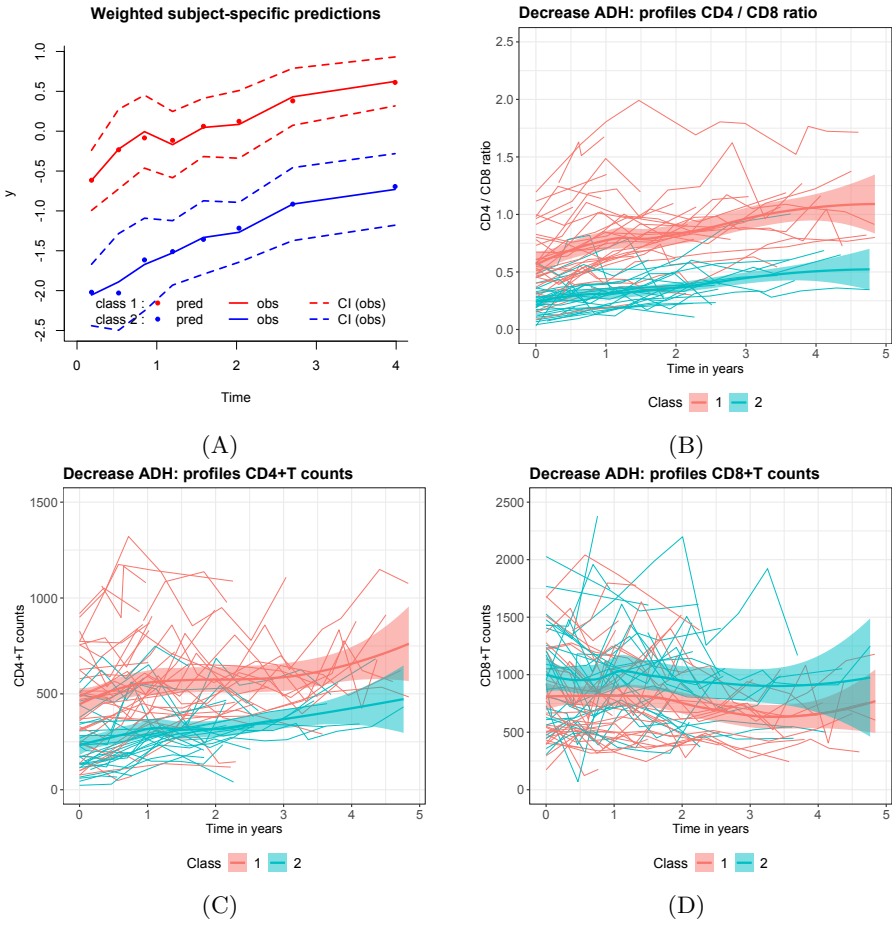


Fig. 4: Decreasing ART adherence class: two classes from a multLCMM with bivariate response (log CD4/CD8 ratio, log CD4+T). (A) Weighted subject-specific predictions estimated by the multLCMM with two classes. Observed profiles of the three classes for (B) CD4/CD8 ratio, (C) CD4+T counts and (D) CD8+T counts; solid lines show the mean profiles computed with the loess smoothing method.

Table 7 shows patient's characteristics for classes of patients within the decreasing ART adherence class (cases, or median and interquartile range). In order to test differences by latent classes we use the Pearson's Chi-squared test with Yates' continuity correction (for categorical variables) and Wilcoxon's rank sum test (for continuous variables).

The results in the decreasing ART adherence class analysis show there are no significant differences for sex, age, AIDS and CD8+T, however differences for CD4+T counts and CD4/CD8 ratio are significant. This suggests latent classes are characterized by the CD4+T counts and CD4/CD8 ratio at baseline, but it

Table 6: Decreasing ART adherence class: estimates of the multLCMM with bivariate response (log CD4/CD8 ratio, log CD4+T).

Variable	Estimate	SE	Wald	p-value
Fixed effects in the longitudinal model				
ADH class1	0.00051	0.00125	0.408	0.68305
ADH class2	0.00365	0.00145	2.513	0.01197
AIDS: Yes	-1.32212	0.20963	-6.307	0.00000
Age	-0.01434	0.00676	-2.122	0.03382
Sex: Female	0.60809	0.19330	3.146	0.00166
Parameters of the link functions H_d *in* (1)				
log CD4/CD8 ratio -Linear η_1	-0.34856	0.17443	-1.998	0.04569
log CD4/CD8 ratio -Linear η_2	0.57959	0.06167	9.399	0.00000
log CD4 -Linear η_1	6.20719	0.13867	44.762	0.00000
log CD4 -Linear η_2	0.45373	0.05166	8.782	0.00000

is about the decreasing ART adherence class. Note that once again patients who started with low levels of CD4+T counts and CD4/CD8 ratio will continue in similar conditions over time, regardless of their lower levels of ART adherence.

5 Discussion and Conclusions

Using LCMM to model ART adherence we found two groups of trajectories, within each group a joint model for CD4+T counts and CD4/CD8 ratio distinguished heterogenous subpopulations.

The majority of PWH receiving care at the HIV/AIDS Clinic at the INCMNSZ initiate and maintain a good level of ART adherence throughout time. However, there is a small subset whose adherence decreases to levels lower than 50% (class 2 Decreasing ART adherence in Section 4.1). These patients have a lower proportion of AIDS at enrolment in care (see Table 2) and show a lower increase in their CD4+T counts and CD4/CD8 ratio between the beginning and the end of the follow-up (see Table 3). Within this decreasing ART adherence class we found a class of patients who started with low CD4+T counts and CD4/CD8 ratio lower than 0.5 (see Figure 4 and Table 7); for these patients the initial ART adherence level shows a higher effect on the bivariate response in the joint model (see Table 6). Therefore, age, sex, AIDS and initial CD4+T counts seem to be crucial to maintain good CD4/CD8 ratios for those who start with adverse conditions, good ART adherence will provide a desirable positive effect in CD4+T counts and CD4/CD8 ratio. Additionally, further study of these classes could include specific analysis for clinical outcomes.

Latent class mixed models enable a useful interpretation of longitudinal heterogeneity in the population of interest. Identifying and characterizing different families of trajectories of a specific outcome, could focus the results of a model, allowing to detect covariates related to the difference and getting better and

Table 7: Decreasing ART adherence class: characteristics (cases, or median and interquartile range) of patients from multLCMM with bivariate response (log CD4/CD8 ratio, log CD4+T), Pearson's Chi-squared test, and Wilcoxon rank sum test.

Decrease ART Adherence		Classes from multLCMM		Chi-squared test	
		1	**2**	χ-squared	p-value
Characteristics		($n = 36$)	($n = 28$)		
Sex	Male	32	24	0.0001	1
	Female	4	4		
Previous AIDS	No	28	24	0.2344	0.6283
defining event[1]	Yes	8	4		
Baseline CD4+T	≤ 200	7	14	11.61	0.0030
counts	201-500	14	12		
(cells/μL)[1]	> 500	15	2		

Decrease ART Adherence		Classes from multLCMM		Wilcoxon test	
Characteristics		**1**	**2**	W	p-value
Age in years[2]		31 (27-35)	32 (28-38)	472	0.6720
CD4/CD8 ratio	Beginning[2]	0.560 (0.394-0.808)	0.218 (0.130-0.313)	875	<0.0001
	End[3]	0.847 (0.612-1.143)	0.347 (0.238-0.406)	869	<0.0001
CD4+T counts	Beginning[2]	478 (293-711)	201 (127-344)	773	0.0003
	End[3]	567 (441-886)	345 (291-460)	785	0.0002
CD8+T counts	Beginning[2]	799 (501-1156)	1071 (687-1429)	373	0.0774
	End[3]	721 (450-1083)	990 (735-1213)	366	0.0628

[1]Condition previous to enrollment in care.
[2]At the beginning of follow-up.
[3]At the end of follow-up (varies for individuals).

more specific answers to research questions. The mean profiles of the classes are defined according to time and covariates through latent class-specific mixed models. In this study, both fixed and random effects can be class specific. This is an advantage in comparison to the standard linear mixed models where the population is considered homogeneous and only one mean profile of response is needed. Moreover, the lcmm allows the study of multivariate longitudinal markers instead of a unique one.

Acknowledgments

This work was supported by the NIH-funded Caribbean, Central and South America network for HIV epidemiology (CCASAnet), a member cohort of the International Epidemiologic Databases to Evaluate AIDS (leDEA) (U01AI069923). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This award is funded by the following institutes: National Institute Of Allergy And Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Heart, Lung, And Blood Institute (NHLBI), National Institute Of Diabetes And Digestive And Kidney Diseases (NIDDK), National Institute On Drug Abuse (NIDA), National Institute On Alcohol Abuse And Alcoholism (NIAAA), Fogarty International Center (FIC), and National Cancer Institute (NCI).

LN and RFG have been in part supported by UNAM-DGAPA-PAPIIT, Mexico (Project IN118720). Thanks to Dra. Brenda Crabtree-Ramírez and Dr. Juan G. Sierra-Madero from Departamento de Infectología, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, for data, clinical information and discussion. We are also grateful to two anonymous reviewers who provided insightful comments and suggestions that have greatly improved both the readability and the content of this article.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov, F. Csaki (eds.) Proceedings of 2nd International Symposium on Information Theory, pp. 267–281. Académiai Kiado, Budapest, Hungary (1973)
2. Belaunzarán-Zamudio, P., Caro-Vega, Y.N., Shepherd, B.E., *et al.*: Monitoring of HIV treatment in seven countries in the WHO Region of the Americas. Bull. World Health Organ. **93**(8), 529–39 (2015)
3. Belaunzarán-Zamudio, P.F., Naranjo, L., Caro-Vega, Y., Castillo-Mancilla, J.R., Camiro-Zuñiga, A., Fuentes-García, R., Crabtree-Ramírez, B.E., Sierra-Madero, J.G.: Incomplete ART adherence is associated with lower CD4-CD8 ratio in virally suppressed patients with HIV-infection in Mexico. Submitted (2021)
4. Boussari, O., Subtil, F., Genolini, C., *et al.*: Impact of variability in adherence to HIV antiretroviral therapy on the immunovirological response and mortality. BMC Medical Research Methodology **15**, 10 (2015)
5. Bruno, G., Saracino, A., Monno, L., Angarano, G.: The revival of an "old" marker: CD4/CD8 ratio. AIDS Rev. **19**(2), 81–88 (2017)
6. Calmy, A., Tovar Sanchez, T., Kouanfack, C., *et al.*: Dolutegravir-based and low-dose efavirenz-based regimen for the initial treatment of HIV-1 infection (NAMSAL): week 96 results from a two-group, multicentre, randomised, open label, phase 3 non-inferiority trial in Cameroon. Lancet HIV **7**(10), e677–e687 (2020)
7. Caro-Vega, Y., Belaunzarán-Zamudio, P.F., Crabtree-Ramírez, B.E., *et al.*: Durability of efavirenz compared with boosted protease inhibitor-based regimens in antiretroviral-naïve patients in the Caribbean and Central and South America. Open Forum Infectious Diseases **5**(3), ofy004 (2018)

8. Castillo-Mancilla, J.R., Brown, T.T., Erlandson, K.M., *et al.*: Suboptimal adherence to combination antiretroviral therapy is associated with higher levels of inflammation despite HIV suppression. Clinical Infectious Diseases: an official publication of the Infectious Diseases Society of America **63**(12), 1661–1667 (2016)
9. Castillo-Mancilla, J.R., Brown, T.T., Palella, F.J.J., *et al.*: Partial normalization of biomarkers of inflammation and immune activation among virally suppressed men with HIV infection and high ART adherence. Open Forum Infectious Diseases **7**(4), ofaa099 (2020)
10. Castillo-Mancilla, J.R., Phillips, A.N., Neaton, J.D., *et al.*: Incomplete ART adherence is associated with higher inflammation in individuals who achieved virologic suppression in the START study. Journal of the International AIDS Society **22**(6), e25,297 (2019)
11. Celeux, G., Soromenho, G.: An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification **13**(2), 195–212 (1996)
12. McMahon, J.H., Jordan, M.R., Kelley, K., *et al.*: Pharmacy adherence measures to assess adherence to antiretroviral therapy: review of the literature and implications for treatment monitoring. Clinical Infectious Diseases **52**(4), 493–506 (2011)
13. Muthén, B.: Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: D. Kaplan (ed.) Handbook of Quantitative Methodology for the Social Sciences, pp. 345–368. Sage, Newbury Parc, CA (2004)
14. Muthén, B.O.: Beyond SEM: General latent variable modeling. Behaviormetrika **29**(1), 81–117 (2002)
15. Proust-Lima, C., Philipps, V., Liquet, B.: Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. Journal of Statistical Software **78**(2), 1–56 (2017)
16. Sauter, R., Huang, R., Ledergerber, B., *et al.*: CD4/CD8 ratio and CD8 counts predict CD4 response in HIV-1-infected drug naive and in patients on cART. Medicine **95**(42), e5094 (2016)
17. Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics **6**(2), 461–464 (1978)
18. Sierra-Madero, J.G., Belaunzarán-Zamudio, P.F., Crabtree-Ramírez, B., Magis-Rodriguez, C.: Mexico's fragmented health system as a barrier to HIV care. Lancet HIV **6**(2), e74–e75 (2019)
19. Wang, W.L., Fan, T.H.: Bayesian analysis of multivariate t linear mixed models using a combination of IBF and Gibbs samplers. Journal of Multivariate Analysis **105**(1), 300–310 (2012)
20. Wohl, D.A., Yazdanpanah, Y., Baumgarten, A., *et al.*: Bictegravir combined with emtricitabine and tenofovir alafenamide versus dolutegravir, abacavir, and lamivudine for initial treatment of HIV-1 infection: week 96 results from a randomised, double-blind, multicentre, phase 3, non-inferiority trial. Lancet HIV **6**(6), e355–e363 (2019)