Longitudinal Viral Load Clustering for People with HIV using Functional Principal Component Analysis

1038

Jiayang Xiao[1], Yunqing Ma[1], Xueying Yang[1], Bankole Olatosi[1], Xiaoming Li[1], Jiajia Zhang[1]
[1]University of South Carolina, Columbia, SC, USA

BACKGROUND

- It might not be accurate to measure the impact of viral load (VL) pattern using only timepoint observations because VL change is a dynamic pattern.
- Research on clustering historical/longitudinal VL measures particularly using the continuous scale is limited.
- Clustering people with HIV (PWH) into distinct groups only based on different VL indicators or aggregate measures of longitudinal VL measures might fail to correctly classify the PWH at risk.

STUDY OBJECTIVES

- Use functional data clustering to classify historical VL patterns and characterize each cluster by demographics, comorbidities, social behaviors, and CD4 count.

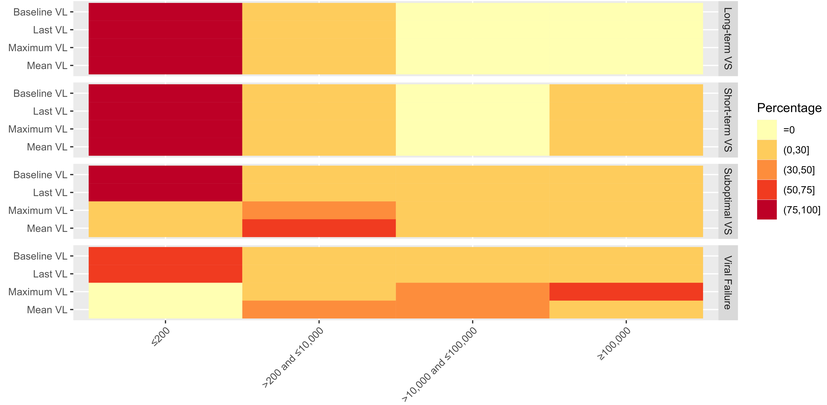
METHODS

- Analyzed adult PWH diagnosed from 2005 to 2020 in SC with a 5-year minimum follow-up from the first viral suppression (VS) to the last VL test using e-HARS data.
- Functional principal component analysis (FPCA) was used to classify PWH into different clusters based on their longitudinal historic VL patterns.
- ANOVA was used to test the difference in VL characteristics, demographics, comorbidities, substance uses, and longitudinal CD4 count during the follow-up period of each cluster.

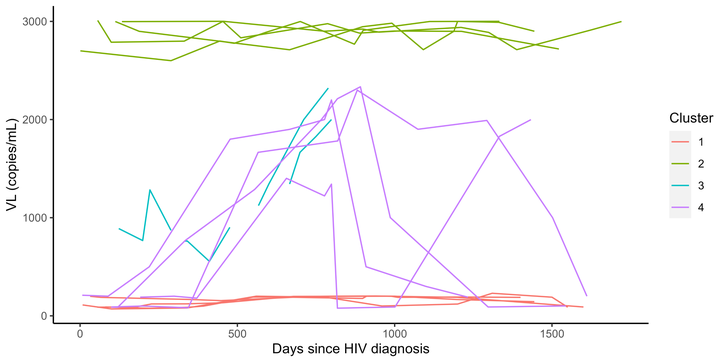
RESULTS

- **Long-term VS cluster (n=1,008):** Compared to other clusters, PWH were older age (≥ 50 years old, 17.5%), lower percentage of Black people (64.2%), minimum CD4 count ≥ 500 cells/μL (18.1%), and less alcohol use (20.8%), and were mostly 100% of time with VS (95.3%) with average of 11-year follow-up.
- **Short-term VS cluster (n=1,763):** Compared to other clusters, PWH were older age (≥ 50 years old, 19.5%), lower percentage of Black people (66.5%), higher percentage of individuals with comorbidity history (31.4%), and higher max CD4 count (>500 cells/μL, 93.8%). It is similar with long-term VS group except for shorter follow-up (6 years).
- **Suboptimal VS cluster (n=1,673):** PWH were mostly under 30 years old (44.8%) and Black (77.2%), with relatively lower percentage of individuals with mean VL <10,000 cells/mL (92.9%) and VR history (18.4%).
- **Viral failure cluster (n=1,472):** PWH were mostly under 30 years old (44.2%) and Black (78.6%), lower percentage of individuals with mean CD4 count >500 cells/μL (34.7%) and percentage of time with low CD4 count (never had low CD4 count, 41.7%).

VL characteristics heatmap plot by clusters



Historic VL patterns example plot



CD4 characteristics heatmap plot by clusters

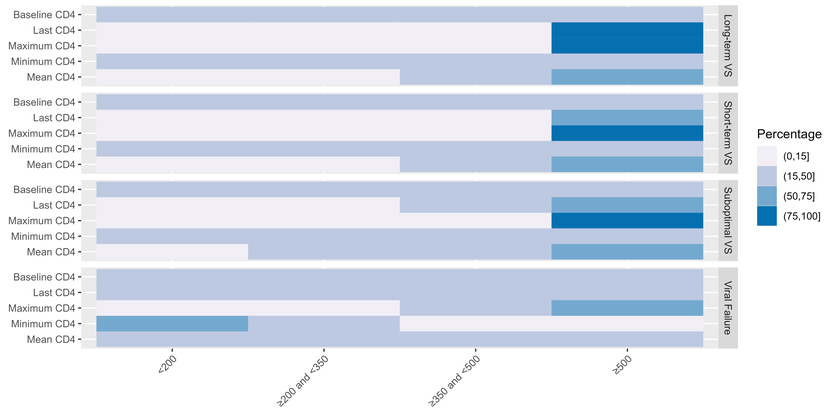


Table 1. Distributions for demographic count for four clusters from FPCA

Characteristics	Cluster 1 Long-term VS n=1,008 (17.3%)	Cluster 2 Short-term VS n=1,763 (29.8%)	Cluster 3 Suboptimal VS n=1,673 (28.3%)	Cluster 4 Viral Failure n=1,472 (24.9%)	P value
Age Group (Years)					<0.001
≥18 and <30	283 (28.1%)	678 (38.5%)	750 (44.8%)	651 (44.2%)	
≥30 and <40	251 (24.9%)	353 (20.0%)	389 (23.3%)	374 (25.4%)	
≥40 and <50	298 (29.6%)	389 (22.1%)	323 (19.3%)	317 (21.5%)	
≥50	176 (17.5%)	343 (19.5%)	211 (12.6%)	130 (8.8%)	
Sex					<0.001
Male	738 (73.2%)	1,363 (77.3%)	1,249 (74.7%)	1,030 (70.0%)	
Female	270 (26.8%)	400 (22.7%)	424 (25.3%)	442 (30.0%)	
Race/Ethnicity					<0.001
White	298 (29.6%)	484 (27.5%)	288 (17.2%)	246 (16.7%)	
Black	647 (64.2%)	1,173 (66.5%)	1,292 (77.2%)	1,157 (78.6%)	
Hispanic	48 (4.8%)	67 (3.8%)	63 (3.8%)	47 (3.2%)	
Others	15 (1.5%)	39 (2.2%)	30 (1.8%)	22 (1.5%)	
Transmission Risk					<0.001
Heterosexual	257 (25.5%)	301 (17.1%)	380 (22.7%)	406 (27.6%)	
MSM	48 (4.8%)	78 (4.4%)	100 (6.0%)	104 (7.1%)	
MSM/IDU	505 (50.1%)	1,000 (56.7%)	849 (50.7%)	683 (46.4%)	
Others	198 (19.6%)	384 (21.8%)	344 (20.6%)	279 (19.0%)	
Rurality or Residence					0.325
Urban	847 (84.0%)	1,481 (84.0%)	1,376 (82.2%)	1,209 (82.1%)	
Rural	161 (16.0%)	282 (16.0%)	297 (17.8%)	263 (17.9%)	

CONCLUSIONS

- FPCA has allowed for a more comprehensive understanding of the heterogeneity within the patient population, providing valuable insights into disease progression and treatment responses.
- Continuous clustering can identify unique patterns in viral load trajectories over time. This allows for the stratification of PWH into distinct groups based on their viral profiles, potentially reflecting different stages of infection, responses to treatment, or natural histories of the disease.
- Clusters may correlate with different risk profiles for the development of comorbidities or for the transmission of the virus.
- The findings from continuous clustering can inform the development of tailored treatment approaches and targeted intervention.
- This study enlightens that including confounders in clustering analysis for more comprehensive clustering results on VL trajectories and clinical understanding in further analysis.

ACKNOWLEDGEMENT

The research reported in this project was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI164947.

Jiayang Xiao. Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA;
Email: xiaoj@email.sc.edu

UNIVERSITY OF SOUTH CAROLINA

Big Data Health Science Center
University of South Carolina