

Declassifying eHARS: A Geospatial Simulation of the Enhanced HIV/AIDS Reporting System

John R. Bassler¹, Emily B. Levitan², Dustin M. Long¹

Background

Geographic, clinical, and demographic characteristics of persons with HIV (PWH) are collected by public health agencies and maintained in the CDC-developed Enhanced HIV/AIDS Reporting System (eHARS). Using this standardized document-based surveillance database, state health departments collect and manage case reports, lab reports, and other documentation about PWH.

Benefits

- **Better patient care:** more up-to-date information enables providers to make informed diagnoses and treatment
- **Enhanced data security:** eHARS uses state-of-the-art security measures to protect the confidentiality of HIV/AIDS data
- **Improved data analysis:** high quality data that allows for the identification of trends and patterns of the HIV epidemic (i.e. trends in Viral Suppression (VS))

Barriers

- **Access:** no access to entities outside of local and state health departments without contractual permission
- **Complexity:** eHARS is a diverse data source that includes hundreds of unique variables and specific organization

Results

This robust simulation of eHARS data allows for a myriad of HIV outcomes related to viral suppression to be examined. For example: **Figure 1**, shows the median time to VS for each County, demonstrating how geography to be included in analyses and visualizations. **Figure 2** represents the statistical analysis capabilities displaying an example of a Kaplan-Meier analysis of time to viral suppression.

Capabilities

- Simulation allows for the development of spatial analysis of HIV outcomes while maintaining confidentiality of the true eHARS database
- Simulated coordinates are randomly generated and do not correspond to any PWH's address
- Regression and time to event models, such as Cox-ph, or Frailty models are also possible



Figure 1. Choropleth of estimated median time to viral suppression by US County, HIV Diagnosis from 2011 - 2020

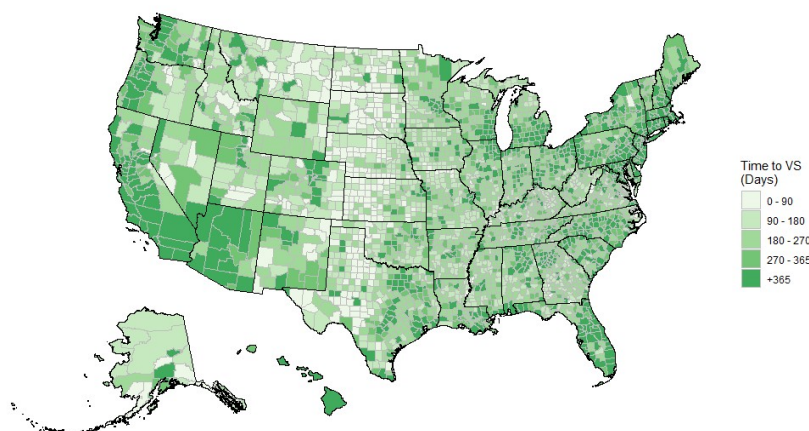
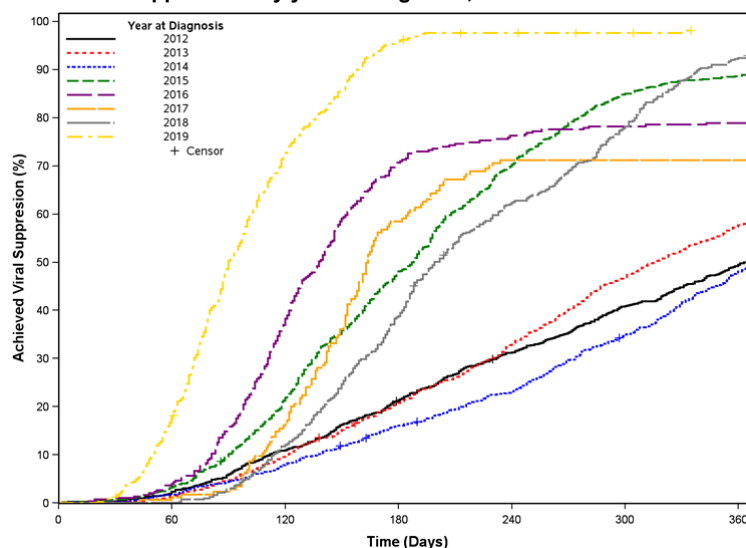


Figure 2. Kaplan-Meier analysis of estimated time to viral suppression by year of diagnosis, Alabama 2011-2020



Methods

Details of the format and description of eHARS specific variables were established with assistance from three state health departments, Louisiana, Mississippi, and Alabama. The distributional attributes of clinical variables (i.e., date of HIV diagnosis) and demographic variables (i.e., sex, race, ethnicity) were replicated to ensure continuity of the simulated data and the original eHARS database. Geographic data (latitude, longitude) within each state in the US and District of Columbia were simulated and included in the synthetic data set; including matching the simulated coordinates to each corresponding county and census tract. Furthermore, all date variables were given careful attention to ensure that the dates occurred in a natural progression and varied in time between events (HIV diagnosis, AIDS progression, etc.). While access to this data is limited to health departments to maintain the confidentiality of the data, we have built a synthetic eHARS data set, designed to replicate this system would provide researchers with the framework to develop models related to HIV, without needing to access a protected network.

Discussion

▪ Why make the simulation?

Understanding geographic variability of HIV outcomes is a critical component to Ending the HIV Epidemic – the simulated data provides many relevant geographic information and region-specific demographic data

▪ Who is this simulation for?

Partnerships between state health departments and research entities is mutually beneficial and under utilized – this publicly available data set helps reduce the barriers of establishing partnerships

▪ How can one use this?

All materials developed as a part of this project is freely available at:

<https://github.com/JRBassler/Geospatial-eHARS-Simulation>

or using the QR code

