

Appendix 1 Introduction of the ongoing Big Data project

Summary of the ongoing Big Data project (R01AI164947). Applying Big Data analytics to integrated EHR datasets of people living with HIV (PLWH) in South Carolina (SC), the ongoing Big Data project proposed to examine the longitudinal dynamic pattern of viral suppression (VL), develop optimal predictive models of various viral suppression indicators, and translate the models to service-ready tools for clinical use. The ongoing project (R01AI164947) will extract longitudinal EHR and other relevant data of all PLWH in SC from 8 state agencies and then link the patient-level data with county-level data (e.g., healthcare infrastructure indicators) from multiple publicly available data sources (2005 to 2020) to address the following specific aims: Aim 1: Identify the longitudinal dynamics of VL among PLWH in SC using multiple indicators, including time to initial suppression, sustained suppression, viral rebound, viral blips, and other relevant VL measures. Aim 2: Determine the critical predictors of multiple VL indicators through AI-based modeling accounting for factors at the individual level, structural level, and socioenvironmental level. Aim 3: Develop a multifactorial decision system based on a risk prediction model to assist with identifying the risk of viral failure or viral rebound when patients present at clinical visits.

Data source and structure. Eight state agencies/systems that oversee and administer public data repositories in SC have contributed EHR and other relevant data to the existing Big Data project. They are SC Department of Health and Environmental Control (SC DHEC) Enhanced HIV/AIDS Reporting System (eHARS), Ryan White HIV/AIDS Program Data Report (RDR), SC Revenue and Fiscal Affairs Office (SC RFA), Health Sciences South Carolina (HSSC), SC Department of Mental Health (DMH), SC State Law Enforcement Division (SLED), SC Department of Alcohol and Other Drug Abuse Services (DAODAS), and Prisma. SC RFA serves

as the honest broker for the linkage of all identifiable data from various sources and removes all identifiable information from the linked data before releasing it to the research team, who enters into legal contract (s) (required for each new study or new analysis) with SC RFA.