Retrospective HIV/AIDS Data Analysis - New York City (2011-2016)

Introduction

According to Dr. Anthony Fauci, Director of the National Institute of Allergy and Infectious Disease, HIV and Coronavirus share a common denominator. Both of these viruses "started off as a new virus that the human species had never had any experience with. It jumped, as many of these new emerging infections do, from an animal reservoir to a human." As COVID spreads throughout the United States, we saw that this virus also, like AIDS, disproportionately affected communities of color. According to the COVID Racial Tracker, "nationally, African-American deaths from COVID-19 are **nearly two times greater** than would be expected based on their share of the population. In four states, the rate is three or more times greater."

Understanding and focusing on these disparities is crucial for helping communities respond to COVID effectively by developing better public health protocols, community outreach, testing and treatment. By understanding a previous epidemic spread, we can glean knowledge of how to tackle this current one.

Data Source

We analyzed HIV/AIDS diagnosis data related to deaths by borough, age, sex, United Hospital Fund (UHF) in New York City from 2011-2015. Data was provided by the Department of Health and Mental Hygiene as reported to the HIV Epidemiology and Field Services Program in 2016. Last update was February 2020. All data shown are for people ages 13 and older.

Methods

Using python and a number of libraries (available on Jupyter notebook), we performed basic statistical analysis and visualization. We began with **data wrangling**. We handed outliers and aggregate values that may skew the data when represented using Pandas. Next, we **explored our data** to interpret what it represents by the categories of Sex, Race, Gender, Diagnosis, Linkage to Care, Viral Load Suppression Rate and Date Rate. Lastly, we performed various **data visualizations** with Seaborn and Matplotlib to stratify multiple levels of analysis to gain insights.

Data Summary

According to our analysis, we found that black males, between the age range of 20-29 years in Brooklyn, were the highest reported individuals diagnosed with HIV. Latino/Hispanic and white communities followed respectively. This trend decreases yearly across all boroughs. We also found that older individuals and women were more likely to be linked up to care within the first three months of diagnoses which also helped viral suppression percentage.

Interestingly, we also found that non-HIV related death rate and HIV related death rate were highest in Staten Island, although it had the lowest HIV and AIDS incidence.

Data Impact

Our findings find similarities between the AIDS epidemic and COVID pandemic – single racial communities and genders are disproportionately affected with both outbreaks. Both viruses had a higher prevalence with younger groups, yet mortality was higher in older individuals. Linkage to care and also suppression were key factors that affected this group, leading to current questions of how COVID is being tested, treated and managed in affected communities. Politicization, swinging-pendulum, stigma and lack of vaccination/clear treatment are additional underlying parallels that highlight similar struggles throughout the years.

Ultimately, this research proposal questions how resources can be allocated to prevent the contraction and spread of HIV to heavily affected boroughs and associated cohorts. The group will provide a statistical and business analysis of the data.