Investigating the Association between Social Determinants of Health and Racial Disparities in HIV-related Mortality Rates Among non-Hispanic Black and non-Hispanic White Persons

United States, 2020

Hayley Hemme

5/5/23

# 1. Summary

This project aims to identify factors associated with racial disparities in HIV-related death rates in the United States. To assess this, we look at the ratio of Black-to-White HIV-related mortality rate (MRR) from 2020 in the United States. Additionally, we assess whether social determinants of health (SDOH) can explain variation in MRR at the state-level by employing spatial and non-spatial methods.

# 2. Introduction

The impact of human immunodeficiency virus (HIV) on marginalized communities in the United States continues to be a significant public health concern. Of particular concern is the significant disparity in HIV-related health outcomes between non-Hispanic Black and non-Hispanic White persons, including the rates of new HIV diagnoses, late HIV diagnoses, and HIV-related death. Research has shown a direct correlation between social determinants of health (SDOH) indicators and poor health outcomes for individuals with HIV. In this project, our aim is to investigate the factors contributing to the disparities in HIV-related death rates between Black and White individuals in the US. Specifically, we will examine whether the Black-to-White HIV-related mortality rate (MRR) from 2020 can be modeled by SDOH and late diagnoses. Additionally, we will explore whether a spatial model can explain variation in MRR at the state level after accounting for variables associated with SDOH or HIV-related outcomes. Our null hypothesis is that there is no spatial correlation between Black-to-White HIV-related MRR and SDOH or HIV-related outcomes, while our alternative hypothesis suggests that there is indeed a spatial correlation.

# 3. Background Information

Since the emergence of the epidemic in 1981 in the United States, more than 1.2 million people in the country have been diagnosed with human immunodeficiency virus (HIV), resulting in more than 700,000 HIV-related deaths (Jun 07 & 2021, 2021). The virus encodes three enzymes that allow it to escape host immunity and results in chronic infection that, without treatment, ultimately leads to acquired immunodeficiency syndrome (AIDS). Medical advancements have led to development effective antiretroviral therapies, many of which disrupt the function of these key enzymes, and are capable of not only treating HIV infection by reducing adverse HIV-related health outcomes, including slowing progression to AIDS, but are also capable of preventing infection before and after potential exposure for people who are HIV negative. Even still, more than 30,000 people in the United States continue to be diagnosed each year (Jun 07 & 2021, 2021). While the Ending the Epidemic initiative set up a goal of 90% reduction in new HIV infections by 2030 by upscaling efforts in HIV diagnoses treatment, and prevention, this is followed by nearly one decade of steady HIV diagnoses rates. Developing targeted intervention programs is essential to achieve this goal. To accomplish this, we need to gain a deeper understanding of the impact or contribution of complex factors such as social determinants of health or other barriers to poor HIV-related outcomes and disparities.

# 4. Materials and Methods

## 4.1 Data Source

The data used in this project was obtained from multiple publicly available source.

### 4.1.1 AIDSVu

Data was obtained from AIDSVu (*AIDSVu datasets. Includes*, n.d.) on counts and rates of HIV-related mortality by state in 2020. It is important to distinguish that these deaths are specifically the result of HIV infection, and do not represent all deaths that occurred in 2020 in the United States among persons with HIV. Social determinants of health (SDOH), were also obtained from AIDSVu by state from 2020. SDOH included a number of variables represented as a percentage particularly pertinent to health persons with HIV. AIDSVu is a public resource presented by Emory University’s Rollins School of Public Health in partnership with Gilead Sciences, Inc. and the Center for AIDS Research at Emory University (CFAR) and contains HIV surveillance data and other pertinent data to monitoring HIV-related outcomes in the United States.

### 4.1.2 US Census Bureau

Data on 2020 demographic composition by race/ethnicity and age by state was obtained from the [U.S. Census Bureau](#X6589fc6ab0dc82cf12099d1c2d40ab994e8410c) (“US Census Bureau,” n.d.). A five-year estimate of percent with access to broadband internet by state (released in 2020) was also obtained from the US Census Bureau.

### 4.1.3 Kids Count

Data on the percentage distribution of children under 14 years old was obtained from [Kids Count](https://datacenter.kidscount.org/) (*KIDS COUNT Data Center from the Annie E. Casey Foundation*, n.d.), an online resource that provides high-quality data on the well-being of children in the United States.

### 4.1.4 Spatial data

In order to assess spatial associations, this analysis was limited to the contiguous United States. Spatial object of the contiguous United States was obtained from library(maps) in R and a list of neighboring contiguous United States was obtained from a public repository on [GitHub](https://github.com/ubikuity/List-of-neighboring-states-for-each-US-state). (Ubikuity, 2022)

## 4.2 Data processing

### 4.2.1 Cleaning

In the data cleaning process, various tasks were performed to ensure the data were ready for analysis. Firstly, variables were reformatted and dataframes were merged as necessary. Additionally, several observations in response variables were suppressed due to small counts/ rates were represented by -9, and were re-coded to 0, rather than NA, as some analysis methods did not permit missing observations. Feature engineering was also employed, most notably for our primary outcome variable, the Black-to-White HIV-related death rate ratio (MRR).

### 4.2.2 Outcome Variables

The primary response variable in this analysis is the ratio of Black-to-White HIV-related death rate (MRR). To compute this variable, feature engineering was performed to create death rate (estimated) stratified by race (non-Hispanic Black and non-Hispanic White) using counts of HIV related deaths divided by the estimated population based on US Census data multiplied by 100,000 persons. Then, I computed MRR by obtaining the ratio of the estimated HIV-related death rate for non-Hispanic Black persons divided by the estimate HIV-related death rate for non-Hispanic White persons.

### 4.2.3 Predictors

Predictors used in this analysis were selected to serve as indicators for various aspects relating to SDOH, including access to healthcare, the built environment, and economic stability. A number of variables were initially considered, but ultimately five predictors were selected: the median household income (in US dollars); percent living in poverty (% of population); percent uninsured under 65 (% of population); percent without internet subscription(% of population); and percent living with food insecurity (% of population). Percentage diagnosed late is the percentage of persons diagnosed with HIV who progressed to AIDS within 12 months of diagnosis.

### 4.2.4 Statistical Analysis

The primary objective of this analysis was to examine the spatial distribution of the Black-to-White HIV-related MRR. Simple linear regression was conducted to explore possible predictors of MRR, selecting variables with a correlation coefficient of |-0.15| or more. Principal component analysis was used to reduce dimensionality and remove multicollinearity.

To test for spatial correlation, Moran’s I was employed. One assumption of using Moran’s I in hypothesis testing is that the data follow a normal distribution. Normality was assessed visually using a histogram and verified with a Shapiro-Wilke test. The analysis utilized the following neighborhood structure:

hij { 1, if counties i and j share a county line; 0, if otherwise.

Several regression models were employed, including multiple linear regression, negative binomial regression, and the conditional autoregressive model.

Multiple linear regression requires several key assumptions to be met, including linearity, normality of the response variable (MRR), and addressing the issue of multicollinearity. These assumptions were checked using scatterplots, visual assessment, Shapiro test, and PCA.

Given that the case data for HIV-related mortality tends to be heavily right-skewed, a Poisson regression model was evaluated, which assumes that the variance and mean are equal. However, when the conditional variance is greater than the conditional mean, overdispersion occurs. In this case, the negative binomial regression model, which can accommodate overdispersion, was considered.

Finally, the conditional autoregressive (CAR) model, a type of linear regression model that can account for spatial association between neighbors, was employed. The CAR model’s assumptions are conditional on the data at remaining site {*zi* ∶ *j* ≠*i*}, are normally distributed data Z(*si*), and there are only pairwise spatial interactions.

To compare models and assess relative goodness of fit, the Akaike information criterion (AIC) was used, where smaller values indicate better fit.

# 5. Results

## 5.1 Univariate Analysis

Our primary outcome of interest was the Black-to-White HIV-related MRR, which was computed for each state by dividing the estimate death rates for non-Hispanic Black persons divided by the estimate death for White persons. Table 1 shows the overall number of HIV-related deaths, population size older than 12 years, death rate and the MRR in the United States in 2020. The rate of HIV-related deaths among this Black persons was 7.2 times high than the rate of HIV-related deaths among White persons.

|  |
| --- |
| Table 1. Population Charactertistics |

To assess the distribution of Black-to-White HIV-related MRR in the US in 2020, we conducted a visual assessment ( [Figure 1](#fig-include-MRR) ) and a Shapiro-Wilks test. Our results showed that the distribution followed a normal distribution (p= 0.63), and no normalizing transformations were needed. [Figure 2](#fig-include-DR) shows the histogram of the raw HIV-related death rate by race, which highlights significant racial disparities in risk of HIV-related death.

|  |
| --- |
| Figure 1: Histogram of Black-to-White HIV-Related Mortality Rate Ratio |

|  |
| --- |
| Figure 2: Histogram of HIV-Related Death Rates |

To further explore the racial disparities in HIV-related mortality, we compared the Black-to-White HIV-related mortality rate ratio (MRR) by state ( [Figure 3](#fig-include-state) ) and by region ( [Figure 4](#fig-include-region) ). Four states, namely Idaho, Montana, Wyoming, and Vermont, did not have an estimate for MRR due to small or suppressed rates. The results from these figures reveal distinct disparities between the rate of HIV-related death by race and region. States in the Northeast region had the highest median Black-to-White HIV-related MRR, with a ratio of approximately 8.5, while states in the West region had the lowest median Black-to-White HIV-related MRR, at approximately 4.0.

|  |
| --- |
| Figure 3: Bargraph of Black-to-White HIV-Related MRR |

|  |
| --- |
| Figure 4: Boxplot of Black-to-White HIV-Related MRR, by Region |

## 5.2 Bivariate Analysis and Principal Component Analysis

Bivariate analysis and principal component analysis (PCA) were conducted to identify potential predictors associated with HIV-related MRR. Firstly, bivariate analysis was used to narrow down the range of predictors, where only variables with a correlation coefficient of |0.15| or more were included in the next step. [Figure 5](#fig-include-bivar) displays the results of the bivariate analysis, which includes the following predictors: Percent Living in Poverty, Median Household Income, Percent Uninsured Under 65 Years Old, Percent Living with Food Insecurity, Percent Without an Internet Subscription, and Percentage Diagnosed Late.

|  |
| --- |
| Figure 5: Simple Linear Regression of Outcome vs. Select Predictors |

Following the bivariate analysis, PCA was performed on the selected predictors. The PCA was based on a correlation matrix and resulted in a scaled and centered matrix. [Figure 6](#fig-include-pca) shows the results of the PCA, where only principal components (PCs) that individually explained more than 16.7% variation were included in the analysis. PC1 accounted for 56.6% of variation, while PC2 accounted for 17.1% of variance. All other PCs were excluded from the analysis. PC1 had strong positive associations with median household income and negative associations with percent living in poverty, percent without an internet subscription, and percent living with food insecurity. PC2 showed very strong negative associations with the percentage diagnosed late.

|  |
| --- |
| Figure 6: Results of Principal Component Analysis |

[Figure 7](#fig-include-map) provides a visual representation of the selected predictors using maps. The South generally had higher percentages of persons living in poverty, living with food insecurity, and persons uninsured under 65 years old, and had lower median household income. The percentage of persons diagnosed late was consistent throughout much of the United States. Interestingly, states with the highest percentage of persons with HIV who were diagnosed late were neighbored by states with the lowest rates. The percentage without an internet subscription was generally low throughout the United States, but was highest in Wyoming.

|  |
| --- |
| Figure 7: Maps of Selected Proxies for Social Determinants of Healths |

## 5.3 Full Models

### 5.3.1 Multivariate Linear Model

The results of the multivariate linear model indicate a statistically significant positive association between Black-to-White HIV-related MRR and PC1 (*p* < 0.005), as well as a statistically significant negative association between MRR and PC2 (*p* = 0.038).

|  |
| --- |
| Table 2. Summary of Multivariate Linear Model |

### 5.3.2 Negative Binomial Regression Model

The negative binomial regression model was selected over Poisson regression model due to overdispersion (*c* = 17.35). In order to use this regression method, I use the number of HIV-related deaths in each state using log(Population) as an offset, and race, PC1 and PC2 as predictors, using non-Hispanic White persons as the reference group. Regression coefficients were exponentiated, resulting in rate ratios. The HIV-related MRR was 6.46 times higher among non-Hispanic Black persons than non-Hispanic White persons (95% CI = (5.28, 7.93); z = 17.96; p < 0.005). The Black-to-White HIV-related MRR decreased with increasing components of PC1, while Black-to-White HIV-related MRR increased with increasing components of PC2, but these results were not statistically significant (*p* = 0.13 and *p* = 0.14 , respectively).

|  |
| --- |
| Table 3. Summary of Negative Binomial Model (exponentiated) |

### 5.3.3 Conditional Autoregressive Model (CAR)

The conditional autoregressive model (CAR) indicated a statistically significant positive association between Black-to-White HIV-related MRR and PC1, (*p* < 0.005), after accounting for spatial dependence, while evidence of a negative association between Black-to-White HIV-related MRR and PC2 was not statistically significant (*p* = 0.38).

|  |
| --- |
| Table 4. Summary of CAR Model |

## 5.4 Results

In 2020, a total of 13,173 HIV-related deaths occurred among non-Hispanic Black and non-Hispanic white populations in the United States, with a Black-to-White HIV-related MRR of 7.2 indicating significant racial disparities. [Figure 8](#fig-include-rate) displays the estimated HIV-related death rate for non-Hispanic Black and non-Hispanic White persons in the United States in 2020, highlighting significant regional differences, with the highest rates observed among non-Hispanic Black persons, particularly in the South region. On the other hand, non-coastal western states consistently exhibited the lowest rates of HIV-related deaths across the selected racial/ethnic groups. Additionally, no state in 2020 had an HIV-related death rate higher than ~10 per 100,000 among non-Hispanic White persons.

|  |
| --- |
| Figure 8: HIV-Related Mortality Rate, by Race/Ethnicity |

To evaluate the spatial dependence of the data, Moran’s test was performed on Black-to-White HIV-related MRR. The results showed statistically significant evidence of spatial dependence (z = 1.768; p = 0.039), with a Moran’s *I* value of 0.1507, indicating a degree of clustering of the Black-to-White HIV-related MRR. For uncorrelated data, the expected value of Moran’s *I* was *E(I)* = -0.0213 with a variance *var(I)* = 0.0095. [Figure 9](#fig-include-moran) shows a Moran’s scatter plot of spatially lagged Black-to-White HIV-related MRR and Black-to-White HIV-related MRR, depicting a positive slope indicative of clustering.

|  |
| --- |
| Figure 9: Moran’s Scatter Plot |

In the full analysis, PC1 and PC2, which represented multiple indicators of social determinants of health, along with the percentage of persons diagnosed late, were used to assess the associated racial disparity in HIV-related death rates. The CAR model, accounting for spatial variation in the data, performed marginally better than the non-spatial generalized linear model, with an AIC of 503.18. Despite the slight differences, considering the results of Moran’s test, the CAR model was selected as the best-performing model, as use of the spatial model was also supported by evidence of clustering from by the Moran’s test. Table 5 provides a comparison of AIC across models.

The CAR model showed statistically significant evidence of a positive association between Black-to-White HIV-related MRR and PC1; higher levels of component 1 are associated with higher levels of the Black-to-White HIV-related MRR. Since PC1 is positively associated with median household income and negatively associated with poverty, lack of internet subscription, and food insecurity, this suggests that higher levels of income and lower levels of poverty, lack of internet access, and food insecurity are associated with higher Black-to-White HIV-related MRR.

|  |
| --- |
| Table 5. Relative Comparison of Goodness of Fit |

[Figure 10](#fig-include-mrr) shows maps comparing the raw Black-to-White HIV-related MRR, fitted values from the CAR model, and the ratio of predicted over observed values. States with zero or suppressed counts for HIV-related deaths among non-Hispanic Black persons (Idaho, Montana, Wyoming, Vermont) are represented in dark green. [Figure 10](#fig-include-mrr) (c) shows that the model performed relatively well in the Northeast and South, as indicated by the light greens, which represent values closer to 1. The West coast states are represented in orange, indicating that the model over-predicted the disparity between HIV-related MRR between Black and White persons in these regions.

|  |
| --- |
| Figure 10: Maps of Black-to-White HIV-Related Mortality Rate Ratio |

# 6. Discussion

There are significant disparities in HIV-related mortality rates between non-Hispanic Black and non-Hispanic White populations in the United States. This study found spatial associations of this disparity with five indicators of SDOH: median household income, percent living in poverty, percent uninsured under 65, percent without internet subscription, and percent living with food insecurity. This analysis found evidence of a spatial association with higher Black-to-White HIV-related MRR and higher median household income and lower rates of poverty, Internet subscription, and food insecurity. While these findings were somewhat surprising, they helped reveal that HIV-related health disparities are complex, and other factors likely contribute to high HIV-related mortality rates among non-Hispanic Black persons.

This study has several limitations that should be addressed. Firstly, while the Black-to-White HIV-related mortality rate ratio can be used to identifying areas with significant racial disparities in HIV-related mortality rates, it may not accurately reflect areas with where HIV deaths are especially high. Additionally, this study is vulnerable to ecological fallacy, as the data was aggregated at the state-level, which can potential mask individual-level relationships between the predictors and outcome variable. Furthermore, state-level data may not capture the variation in HIV-related health disparities within states, such as differences in risk factors and outcomes between urban and rural populations or those between different racial and ethnic groups. Another limitation is that the neighborhood structure used in this study may not capture the true relationships between neighboring counties. In Georgia, this might particularly be an issue, since county sizes and shapes differ significantly. Including additional neighborhood structures in the analysis, such as one that defines neighbors based on those that lay within *x* distance of one another, may be a better approach. Finally, the analysis is limited to the 2020 and to non-Hispanic Black and non-Hispanic White persons, and may not reflect changes over time or across other racial and ethnic groups.

Strengths of this study included the use of multiple predictors to assess the relationship between SDOH and the Black-to-White HIV-related MRR. Additionally, use of various regression models, including non-spatial and spatial methods, allowed for a more comprehensive analysis of the data. The use of PCA helped to address multicollinearity issues among the predictors.

This analysis highlights the need for targeted interventions to address HIV-related health disparities. Future research would greatly benefit from analyzing individual-level data to better understand factors contributing to HIV-related racial disparities, as well as utilizing additional neighborhood structures, and a broader range of years.

# 7. Software

sessionInfo()

R version 4.2.1 (2022-06-23 ucrt)  
Platform: x86\_64-w64-mingw32/x64 (64-bit)  
Running under: Windows 10 x64 (build 22621)  
  
Matrix products: default  
  
locale:  
[1] LC\_COLLATE=English\_United States.utf8   
[2] LC\_CTYPE=English\_United States.utf8   
[3] LC\_MONETARY=English\_United States.utf8  
[4] LC\_NUMERIC=C   
[5] LC\_TIME=English\_United States.utf8   
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base   
  
other attached packages:  
 [1] reshape2\_1.4.4 readxl\_1.4.2 RColorBrewer\_1.1-3 AER\_1.2-10   
 [5] survival\_3.5-5 car\_3.1-1 carData\_3.0-5 lubridate\_1.9.2   
 [9] forcats\_1.0.0 stringr\_1.5.0 dplyr\_1.1.0 purrr\_1.0.1   
[13] readr\_2.1.4 tidyr\_1.3.0 tibble\_3.2.0 tidyverse\_2.0.0   
[17] spdep\_1.2-8 spatialreg\_1.2-8 sf\_1.0-11 Matrix\_1.5-3   
[21] spData\_2.2.2 mfx\_1.2-2 betareg\_3.1-4 MASS\_7.3-57   
[25] lmtest\_0.9-40 zoo\_1.8-11 sandwich\_3.0-2 broom\_1.0.4   
[29] factoextra\_1.0.7 FactoMineR\_2.7 ggcorrplot\_0.1.4 ggpubr\_0.6.0   
[33] kableExtra\_1.3.4 knitr\_1.42 ggfortify\_0.4.15 ggplot2\_3.4.1   
[37] here\_1.0.1   
  
loaded via a namespace (and not attached):  
 [1] TH.data\_1.1-1 colorspace\_2.1-0 ggsignif\_0.6.4   
 [4] deldir\_1.0-6 ellipsis\_0.3.2 class\_7.3-20   
 [7] modeltools\_0.2-23 rprojroot\_2.0.3 estimability\_1.4.1   
[10] rstudioapi\_0.14 proxy\_0.4-27 ggrepel\_0.9.3   
[13] flexmix\_2.3-19 DT\_0.27 fansi\_1.0.4   
[16] mvtnorm\_1.1-3 xml2\_1.3.3 codetools\_0.2-18   
[19] splines\_4.2.1 leaps\_3.1 Formula\_1.2-5   
[22] jsonlite\_1.8.4 cluster\_2.1.3 png\_0.1-8   
[25] compiler\_4.2.1 httr\_1.4.5 emmeans\_1.8.5   
[28] backports\_1.4.1 fastmap\_1.1.1 cli\_3.6.0   
[31] s2\_1.1.2 htmltools\_0.5.4 tools\_4.2.1   
[34] coda\_0.19-4 gtable\_0.3.1 glue\_1.6.2   
[37] wk\_0.7.1 Rcpp\_1.0.10 cellranger\_1.1.0   
[40] vctrs\_0.5.2 svglite\_2.1.1 nlme\_3.1-162   
[43] xfun\_0.37 rvest\_1.0.3 timechange\_0.2.0   
[46] lifecycle\_1.0.3 rstatix\_0.7.2 LearnBayes\_2.15.1   
[49] scales\_1.2.1 hms\_1.1.2 expm\_0.999-7   
[52] yaml\_2.3.7 gridExtra\_2.3 stringi\_1.7.12   
[55] e1071\_1.7-13 boot\_1.3-28.1 rlang\_1.1.0   
[58] pkgconfig\_2.0.3 systemfonts\_1.0.4 evaluate\_0.20   
[61] lattice\_0.20-45 htmlwidgets\_1.6.1 tidyselect\_1.2.0   
[64] plyr\_1.8.8 magrittr\_2.0.3 R6\_2.5.1   
[67] generics\_0.1.3 multcompView\_0.1-8 multcomp\_1.4-23   
[70] DBI\_1.1.3 pillar\_1.8.1 withr\_2.5.0   
[73] units\_0.8-1 scatterplot3d\_0.3-43 abind\_1.4-5   
[76] sp\_1.6-0 nnet\_7.3-17 KernSmooth\_2.23-20   
[79] utf8\_1.2.3 tzdb\_0.3.0 rmarkdown\_2.20   
[82] grid\_4.2.1 digest\_0.6.31 classInt\_0.4-9   
[85] flashClust\_1.01-2 webshot\_0.5.4 xtable\_1.8-4   
[88] stats4\_4.2.1 munsell\_0.5.0 viridisLite\_0.4.1

# 8. References

*AIDSVu datasets. Includes: 2020 State Mortality, 2020 State Prevalence, 2020 State SDOH, 2020 Viral Suppression, 2020 State New Diagnoses, 2020 State Late Diagnoses, and 2020 State Receipt of Care*. (n.d.). Emory University’s Rollins School of Public Health, Gilead Sciences, Inc. and the Center for AIDS Research at Emory University (CFAR).

Jun 07, P., & 2021. (2021). The HIV/AIDS Epidemic in the United States: The Basics.

*KIDS COUNT Data Center from the Annie E. Casey Foundation*. (n.d.). https://datacenter.kidscount.org/.

Ubikuity. (2022). *List-of-neighboring-states-for-each-US-state*. Retrieved from <https://github.com/ubikuity/List-of-neighboring-states-for-each-US-state>

US Census Bureau. (n.d.). https://data.census.gov/.