

Social Determinants Associated with COVID-19 Mortality in the United States: Supplemental Document

1. INTRODUCTION

This supplement is an expansion of the methods section. It contains the data sources used in the model, statistical variations discussed in the methods, time-series analysis of the high-risk model, and the list of all factors used in the social determinant association study.

A. Code Replication and Data Frames

Our code can be found through the COVID-Notebooks Github, on our [Social Determinants branch](#).

B. Data Sources

Table S1 shows the data sources used for the negative binomial model analysis.

Table S1. Publicly Available Data Sources Used in Negative Binomial

	Source	Data
COVID-19	John Hopkins University the Center for Systems Science and Engineering (JHU-CSSE) Coronavirus Resource Center (https://coronavirus.jhu.edu)	County-level COVID-19 deaths, date since first infection (Github)
Health and Social Determinants	United States Census and American Community Survey (https://www.census.gov/programs-surveys/acs/data.html)	County-level race percentages, population density, and percent with less than high school education
	2020 County Health Rankings per the Robert Wood Johnson Foundation (https://www.countyhealthrankings.org)	County-level data on: median household income, obesity prevalence, diabetes prevalence, population 65 years and older
	American Lung Association, estimated from the 2017 BRFSS (https://www.lung.org/research/trends-in-lung-disease/prevalence-incidence-lung-disease)	County-level prevalence in: lung cancer, COPD, adult asthma
	Homeland Infrastructure Foundation-Level Data (HIFLD) (Hospitals)	Hospital beds per county
Death Rates	CDC WONDER (https://wonder.cdc.gov/mcd.html)	County-level death counts for all cause, cancer, deaths of despair, and cardiovascular disease
State Policies	Raifman et al, Boston University School of Public Health, COVID-19 United States state policy database (www.tinyurl.com/statepolicies)	State-level data on: days since first stay at home order, days since reopening, days since reclosure of bars, mask requirements

C. Statistical Variations

To test performance and fit, the negative binomial mixed model for the high-risk factors was compared to zero inflated, fixed negative binomial, and spatial correlation variants. These models were compared using AIC and BIC. The main high-risk model was also analyzed at 16 different time points for robustness. Additionally, variants of the model that excluded New York city data and excluded counties with less than 10 cases were analyzed for robustness.

Similar results yield from comparing the negative binomial mixed model to the zero inflated, fixed negative binomial, and spatial correlation variants. The negative binomial mixed model had an AIC of 15806.4 and a BIC of 15969.4. Although the spatial correlation model had a slightly improved AIC of 15797.0, suggesting slightly better fit, the BIC is slightly higher than the main model at 15972.0. Overall, none of the statistical variants had performed significantly better, and the main negative binomial mixed model is robust to analysis.

The negative binomial mixed models that excluded New York City data and excluded counties with less than 10 infections had lower AIC scores, as they used lower numbers of counties, but similar coefficients and levels of significance to the main model. However, we did find that excluding counties with less than 10 infections reduced the MRR for days since first infection to 33.46%, compared to 124.94% reported in the main model. While the main model found cardiovascular death rate to be statistically insignificant, this model reports cardiovascular death rate to have a p-value of 4.68e-02 and an MRR of 8.93%.

Notebooks to reproduce all results can be found on the Rensselaer IDEA Github¹.

D. Temporal Analysis

To test the robustness of the high-risk model, data from 16 time points were analyzed. Figure S1 shows the Mortality Rate Ratio (MRR) of each explanatory variable at each time point and the p-value of each explanatory variable at each time point.

The MRRs of the explanatory variables have little change with respect to time. The only non-constant variable is days since first infection. This temporal variable decreases in impact as more data is included.

The p-values of most of the explanatory variables do not have significant variance. Population density, median household income, and days since social isolation in a county show trends of increasing p-values. However, the other explanatory variables are consistent in their statistical significance.

Overall, the MRR and p-value trends show that the model is robustness with respect to time.

¹<https://github.com/TheRensselaerIDEA/COVIDMINDER>

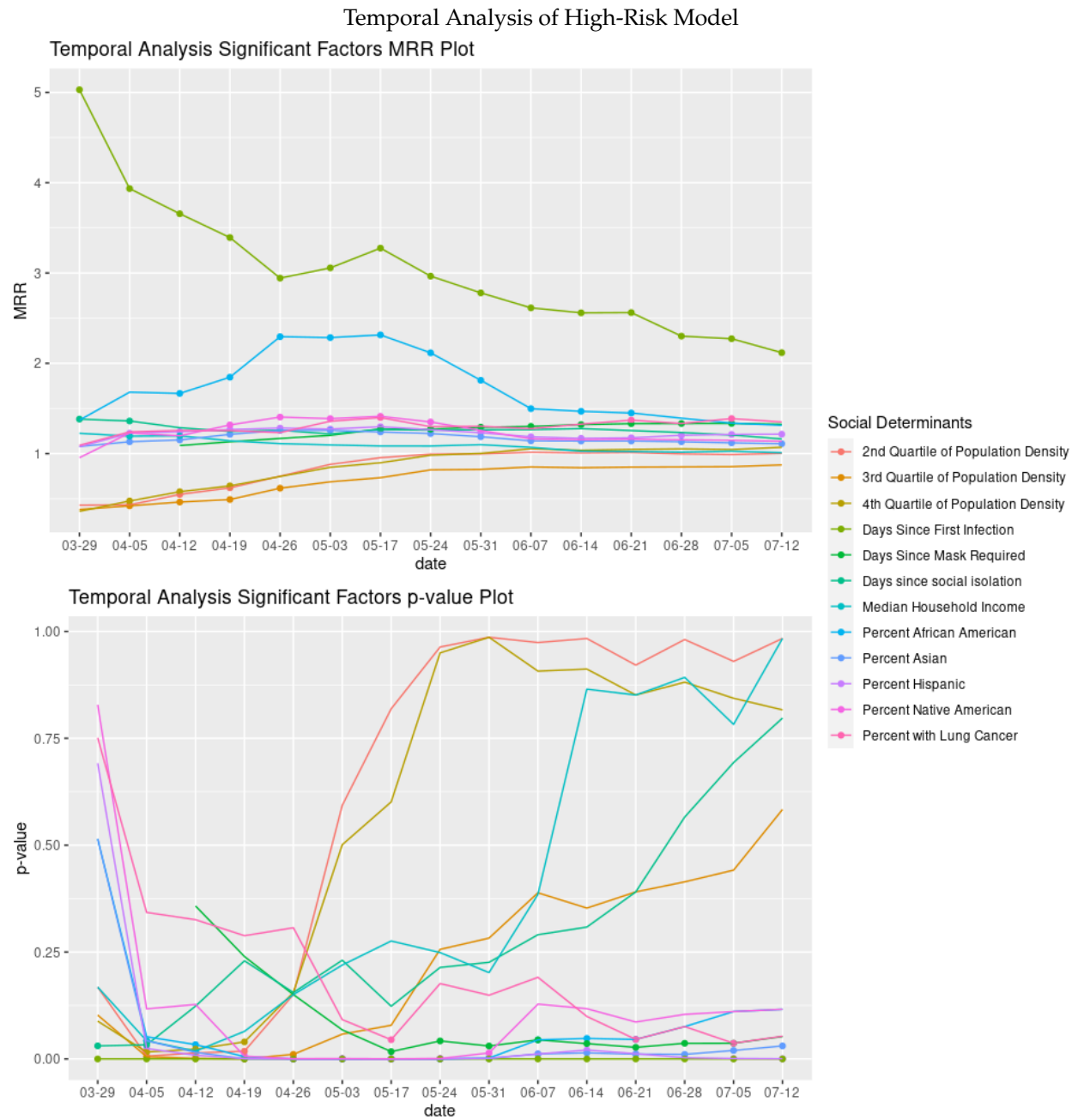


Fig. S1. Time series analysis of MRR and p-value of high-risk model. The legend above the graphs identify each coefficient with a plot color. The top graph depicts MRR vs time. The bottom graph depicts p-value vs time.

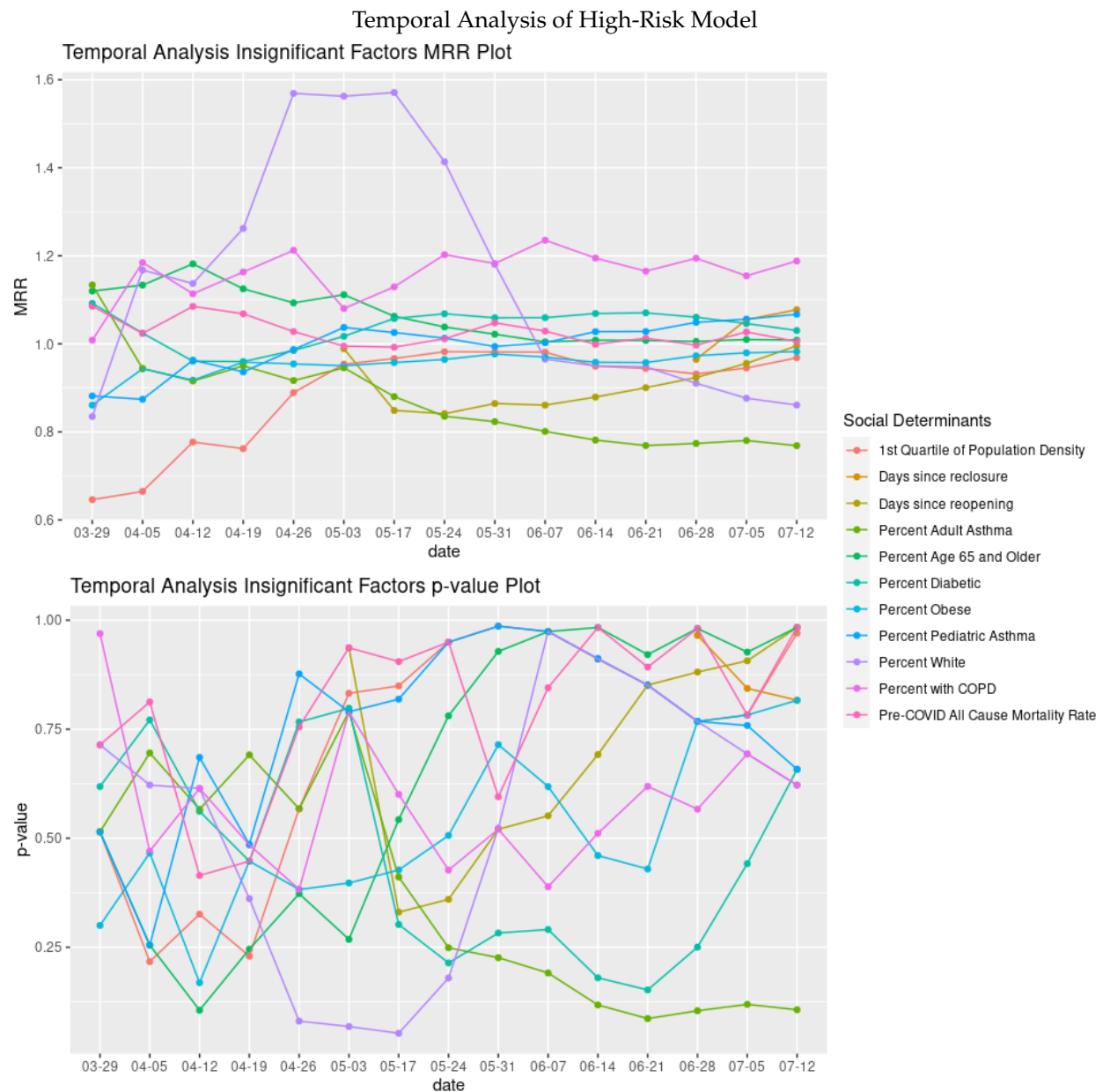


Fig. S2. Time series analysis of MRR and p-value of high-risk model. The legend above the graphs identify each coefficient with a plot color. The top graph depicts MRR vs time. The bottom graph depicts p-value vs time.

E. Social Determinants

In the study, we screened for the following social determinants:

- Death Rate Metrics: [CDC WONDER](#)
 - Infant Death Rate
 - Cancer Death Rate
 - Suicide Death Rate
 - Assault Death Rate
 - Cardiovascular Death Rate
 - Motor Vehicle Death Rate
 - Drug Overdose Death Rate
 - All Cause Death Rate
 - Death of Despair Rate
- Lung Disease Metrics: [American Lung Association](#)
 - Pediatric Asthma Prevalence
 - All Adult Chronic Lung Disease Prevalence
- Composition and Lifestyle Metrics: [County Health Rankings](#)
 - % Not Proficient in English
 - % Insufficient Sleep
 - % Unemployed
 - % Workers who Drive Alone
 - % With Long Commute and Drives Alone
 - % With Food Insecurities
 - % With Access to Exercise Opportunities
 - % Frequent Mental Distress
 - % Smokers
 - % of County With % Excessive Drinking Habits
 - % of County with Overcrowding
 - % Less than 18 years of age
 - % Homeowners
 - % of Households with Severe Cost Burden
 - Average number of physically unhealthy days
 - Social Association Index
 - Segregation Index
 - Average Daily PM 2.5
 - Presence of Water Violation
 - Average 3rd Grade Reading Performance
 - High School Graduation Rate
- Healthcare Metrics: [County Health Rankings](#)
 - Preventable Hospitalization Rate

- Primary Care Physicians Rate
- Other Primary Care Provider Rate
- % Screened with Mammogram
- % Uninsured- Overall Rate
- % Uninsured- African American
- % Uninsured- White
- % Uninsured- Asian
- % Uninsured- Native American
- % Uninsured- Hispanic
- % of County With Poor or Fair Health
- % Medicare FFS Users with Influenza Vaccination
- Chlamydia Prevalence Rate
- Mental Health Provider Rate
- HIV Prevalence Rate

F. Variable List

This is a comprehensive list of any social determinant, health factor, or COVID-19 variables. Please note that all variables are on the county level in the United States unless stated otherwise. Descriptions are taken from sources from which the data comes from.

Data source links:

- Atmospheric Composition Analysis Group (<https://sites.wustl.edu/acag/>)
- Gridmet via Google Earth ([University of Idaho Gridded Surface Meteorological Dataset](https://climate.geog.udel.edu/gridmet/))
- JHU CSSE COVID-19 Database (<https://coronavirus.jhu.edu>)
- US Census / American Community Survey (<https://www.census.gov/programs-surveys/acs/data.html>)
- County Health Rankings (<https://www.countyhealthrankings.org>)
- The COVID Tracking Project (<https://covidtracking.com/>)
- COVID-19 U.S. state policy database (www.tinyurl.com/statepolicies)
- Homeland Infrastructure Foundation-Level Data ([HIFLD](https://hifld.org))
- CDC WONDER (<https://wonder.cdc.gov/mcd.html>)
- American Lung Association (<https://www.lung.org/research/trends-in-lung-disease/prevalence-incidence-lung-disease>)

Variable	Source	Definition
mean_pm25	Atmospheric Composition Analysis Group	Average PM2.5 (air quality), 2000-2016
mean_winter_temp	Gridmet via Google Earth	Average winter temperature, 2000-2016
mean_summer_temp	Gridmet via Google Earth	Average summer temperature, 2000-2016
mean_winter_rm	Gridmet via Google Earth	Average winter relative humidity, 2000-2016
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
mean_summer_rm	Gridmet via Google Earth	Average summer relative humidity, 2000-2016
Lat	JHU CSSE COVID-19 Database	Latitude of county, based on center
Long_	JHU CSSE COVID-19 Database	Longitude of county, based on center
Confirmed	JHU CSSE COVID-19 Database	Number of confirmed COVID-19 cases
Deaths	JHU CSSE COVID-19 Database	Number of COVID-19 deaths
Recovered	JHU CSSE COVID-19 Database	Number of people recovered from COVID-19 illness
Active	JHU CSSE COVID-19 Database	Number of active cases, defined by total cases minus recovered and deaths
Incidence_Rate	JHU CSSE COVID-19 Database	COVID-19 cases per 100,000 people
Case.Fatality_Ratio	JHU CSSE COVID-19 Database	Number of recorded deaths divided by cases
poverty	US Census/American Community Survey	Percent living in poverty
popdensity	US Census/American Community Survey	Population density (person per sq. mile)
medianhousevalue	US Census/American Community Survey	Median house value
pct_owner_occ	US Census/American Community Survey	Percent owner-occupied housing
education	US Census/American Community Survey	Percent of people with less than a high school education
population	US Census/American Community Survey	Overall population
pct_blk	US Census/American Community Survey	Percent black
hispanic	US Census/American Community Survey	Percent Hispanic
pct_asian	US Census/American Community Survey	Percent Asian
pct_native	US Census/American Community Survey	Percent Native American
pct_white	US Census/American Community Survey	Percent white
q_popdensity	US Census/American Community Survey	Population density, broken into quartiles to represent urban/rural. These are values 2-5, with the smallest numbers being more urban.
older_pcent	US Census/American Community Survey	Percent of population age 65+
young_pcent	US Census/American Community Survey	Percent of population ages 0-14
prime_pcent	US Census/American Community Survey	Percent of population ages 15-44
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
mid_pcent	US Census/American Community Survey	Percent of population ages 45-64
State	County Health Rankings	As listed
County	County Health Rankings	As listed
Years of Potential Life Lost Rate	County Health Rankings	Premature death measure; years of potential life lost before age 75 per 100,000 people
% Fair or Poor Health	County Health Rankings	As listed
Average Number of Physically Unhealthy Days	County Health Rankings	As listed
Average Number of Mentally Unhealthy Days	County Health Rankings	As listed
% Low Birthweight	County Health Rankings	As listed
% Smokers	County Health Rankings	As listed
pct_obesity	County Health Rankings	Percent of adults with obesity
Food Environment Index	County Health Rankings	Index from 0 to 10 which indicates a healthy food environment
% Physically Inactive	County Health Rankings	As listed
% With Access to Exercise Opportunities	County Health Rankings	As listed
% Excessive Drinking	County Health Rankings	Percent of adults who participate in heavy or binge drinking
# Alcohol-Impaired Driving Deaths	County Health Rankings	As listed
# Driving Deaths	County Health Rankings	As listed
% Driving Deaths with Alcohol Involvement	County Health Rankings	As listed
# Chlamydia Cases	County Health Rankings	As listed
Chlamydia Rate	County Health Rankings	As listed
Teen Birth Rate	County Health Rankings	As listed
# Uninsured	County Health Rankings	As listed
% Uninsured	County Health Rankings	As listed
# Primary Care Physicians	County Health Rankings	As listed
Primary Care Physicians Rate	County Health Rankings	Physicians per 100,000 population
Primary Care Physicians Ratio	County Health Rankings	Ratio of total population to number of physicians in a county
# Dentists	County Health Rankings	As listed
Dentist Rate	County Health Rankings	Dentists per 100,000 population
Dentist Ratio	County Health Rankings	Ratio of total population to number of dentists in a county
# Mental Health Providers	County Health Rankings	As listed
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
Mental Health Provider Rate	County Health Rankings	Ratio of total population to number of mental health providers in a county
Mental Health Provider Ratio	County Health Rankings	Providers per 100,000 population
Preventable Hospitalization Rate	County Health Rankings	Hospitalization stays for conditions treatable in outpatient settings, per 100,000 Medicare patients
% With Annual Mammogram	County Health Rankings	Percent of females age 65-75 who receive annual mammograms, per 100,000 Medicare patients
% Vaccinated	County Health Rankings	As listed
Cohort Size	County Health Rankings	9th grade cohort
High School Graduation Rate	County Health Rankings	Percentage of 9th grade cohort that graduates
# Some College	County Health Rankings	Number of adults, age 25-44, with some college education
% Some College	County Health Rankings	Percent of adults, age 25-44, with some college education
# Unemployed	County Health Rankings	As listed
Labor Force	County Health Rankings	As listed
% Unemployed	County Health Rankings	As listed
% Children in Poverty	County Health Rankings	As listed
80th Percentile Income	County Health Rankings	For median household income
20th Percentile Income	County Health Rankings	For median household income
Income Ratio	County Health Rankings	Ratio of household income at the 80th percentile to the 20th percentile
# Single-Parent Households	County Health Rankings	As listed
# Households	County Health Rankings	As listed
% Single-Parent Households	County Health Rankings	As listed
# Associations	County Health Rankings	Number of social associations
Social Association Rate	County Health Rankings	Associations per 10,000 population
Annual Average Violent Crimes	County Health Rankings	As listed
Violent Crime Rate	County Health Rankings	As listed
# Injury Deaths	County Health Rankings	As listed
Injury Death Rate	County Health Rankings	As listed; per 100,000 population
Average Daily PM2.5	County Health Rankings	As listed
Presence of Water Violation	County Health Rankings	As listed
% Severe Housing Problems	County Health Rankings	As listed
Severe Housing Cost Burden	County Health Rankings	Percent housing problems, such as: overcrowding, high housing costs, lack of kitchen or plumbing facilities
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
Overcrowding	County Health Rankings	Percent of households with overcrowding
Inadequate Facilities	County Health Rankings	Percent of households without kitchen or plumbing facilities
% Drive Alone to Work	County Health Rankings	As listed
# Workers who Drive Alone	County Health Rankings	As listed
% Long Commute - Drives Alone	County Health Rankings	As listed
Life Expectancy	County Health Rankings	As listed
pre_covid_deaths	County Health Rankings	Deaths predating COVID-19 infection in the U.S.
pre_covid_death_rate	County Health Rankings	Death rate predating COVID-19 infection in the U.S.
child_deaths	County Health Rankings	Number of deaths of children under the age of 18
Child Mortality Rate	County Health Rankings	As listed; per 100,000 population
infant_deaths	County Health Rankings	Number of infant deaths
Infant Mortality Rate	County Health Rankings	As listed; per 100,000 population
% Frequent Physical Distress	County Health Rankings	Percent of adults with poor physical health 14 or more days per month
% Frequent Mental Distress	County Health Rankings	Percent of adults with poor mental health 14 or more days per month
pct_diabetes	County Health Rankings	Percent of adults with diabetes
# HIV Cases	County Health Rankings	As listed
HIV Prevalence Rate	County Health Rankings	As listed
# Food Insecure	County Health Rankings	Number of people with inadequate access to food
% Food Insecure	County Health Rankings	Percent of people with inadequate access to food
# Limited Access	County Health Rankings	Number of people who are low income and do not live close to a grocery store
% Limited Access to Healthy Foods	County Health Rankings	Percent of people who are low income and do not live close to a grocery store
# Drug Overdose Deaths	County Health Rankings	As listed
Drug Overdose Mortality Rate	County Health Rankings	As listed
# Motor Vehicle Deaths	County Health Rankings	As listed
Motor Vehicle Mortality Rate	County Health Rankings	As listed; MV is short for Motor Vehicle
% Insufficient Sleep	County Health Rankings	Percent of adults who report less than 7 hours of sleep
# Uninsured_1	County Health Rankings	Number of adults under 65 who do not have health insurance
% Uninsured_1	County Health Rankings	Percent of adults under 65 who do not have health insurance
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
# Uninsured_2	County Health Rankings	Number of children who do not have health insurance
% Uninsured_2	County Health Rankings	Percent of children who do not have health insurance
Other Primary Care Provider Rate	County Health Rankings	Primary care providers other than physicians per 100,000 population
Other Primary Care Provider Ratio	County Health Rankings	Population to primary care providers other than physicians
% Disconnected Youth	County Health Rankings	Percent of teens 16-19 who are not working or in school
Average Grade Performance	County Health Rankings	Average reading test scores for 3rd graders
Average Grade Performance_1	County Health Rankings	Average math test scores for 3rd graders
Median Household Income	County Health Rankings	As listed
% Enrolled in Free or Reduced Lunch	County Health Rankings	As listed
Segregation index	County Health Rankings	Black/White; higher values indicate more residential segregation
Segregation Index	County Health Rankings	non-White/White; higher values indicate more residential segregation
Homicide Rate	County Health Rankings	As listed
suicide_deaths	County Health Rankings	Number of suicide deaths
Suicide Rate (Age-Adjusted)	County Health Rankings	As listed; per 100,000 population
Crude Rate	County Health Rankings	Suicide rate, without adjustment per 100,000 population
# Firearm Fatalities	County Health Rankings	As listed
Firearm Fatalities Rate	County Health Rankings	As listed
Non-Petitioned Cases	County Health Rankings	Number of non-petitioned juvenile cases
Petitioned Cases	County Health Rankings	Number of petitioned juvenile cases
Denominator	County Health Rankings	Estimated number of juveniles; population age 10 to upper age of jurisdiction, rounded to the nearest hundred
Juvenile Arrest Rate	County Health Rankings	Delinquency cases per 1,000 juveniles
Average Traffic Volume per Meter of Major Roadways	County Health Rankings	As listed
# Homeowners	County Health Rankings	As listed
% Homeowners	County Health Rankings	As listed
# Households with Severe Cost Burden	County Health Rankings	Number of households which spend 50 percent or more of their income on housing expenses
% Severe Housing Cost Burden	County Health Rankings	Percent of households which spend 50 percent or more of their income on housing expenses
# less than 18 years of age	County Health Rankings	As listed
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
% less than 18 years of age	County Health Rankings	As listed
pct_age65	County Health Rankings	Percent of adults age 65 or older
# Black	County Health Rankings	As listed
% Black	County Health Rankings	As listed
# American Indian & Alaska Native	County Health Rankings	As listed
% American Indian & Alaska Native	County Health Rankings	As listed
# Asian	County Health Rankings	As listed
% Asian	County Health Rankings	As listed
# Native Hawaiian/Other Pacific Islander	County Health Rankings	As listed
% Native Hawaiian/Other Pacific Islander	County Health Rankings	As listed
# Hispanic	County Health Rankings	As listed
% Hispanic	County Health Rankings	As listed
# Non-Hispanic White	County Health Rankings	As listed
% Non-Hispanic White	County Health Rankings	As listed
# Not Proficient in English	County Health Rankings	As listed
% Not Proficient in English	County Health Rankings	As listed
% Female	County Health Rankings	As listed
# Rural	County Health Rankings	As listed
% Rural	County Health Rankings	As listed
state	The COVID Tracking Project	As listed
positive	The COVID Tracking Project	Number of people with confirmed or probable infection, state level
negative	The COVID Tracking Project	Number of negative tests, state level
pending	The COVID Tracking Project	Number of tests pending/being processed, state level
hospitalizedCurrently	The COVID Tracking Project	As listed, state level
hospitalizedCumulative	The COVID Tracking Project	As listed, state level
inIcuCurrently	The COVID Tracking Project	Number of people currently in the ICU, state level
inIcuCumulative	The COVID Tracking Project	Number of people who have been in the ICU, total on state level
onVentilatorCurrently	The COVID Tracking Project	As listed, state level
onVentilatorCumulative	The COVID Tracking Project	As listed, state level
recovered	The COVID Tracking Project	Number of people recovered from illness, state level
state_deaths	The COVID Tracking Project	Deaths in each state confirmed or probable due to COVID-19, state level
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
hospitalized	The COVID Tracking Project	Old label for hospitalizedCumulative
positiveTestsViral	The COVID Tracking Project	Completed tests with positive results, state level
negativeTestsViral	The COVID Tracking Project	Completed tests with negative results, state level
positiveCasesViral	The COVID Tracking Project	Number of people with completed positive tests, state level
deathConfirmed	The COVID Tracking Project	Deaths which have confirmed COVID-19 diagnosis, state level
deathProbable	The COVID Tracking Project	Deaths which have a probable COVID-19 diagnosis, where COVID-19 is listed as an underlying cause, state level
positiveIncrease	The COVID Tracking Project	Viral tests completed with positive results, state level
negativeIncrease	The COVID Tracking Project	Old label
total	The COVID Tracking Project	Old label
totalTestResults	The COVID Tracking Project	Total number of people tested (not per specimen), state level
totalTestResultsIncrease	The COVID Tracking Project	Daily increased in totalTestResults
posNeg	The COVID Tracking Project	Old label
deathIncrease	The COVID Tracking Project	Daily increase in death
hospitalizedIncrease	The COVID Tracking Project	Daily difference in hospitalized
stay_at_home	COVID-19 U.S. state policy database	Date since stay at home order issued, state level
reopen	COVID-19 U.S. state policy database	Imported reopening data
reclosure	COVID-19 U.S. state policy database	Imported reclosure data
mask	COVID-19 U.S. state policy database	Imported mask data
date_since_social	COVID-19 U.S. state policy database	Date since original stay at home order started, state level
date_since_reopen	COVID-19 U.S. state policy database	Date since stay at home order ended, state level
date_since_reclosure	COVID-19 U.S. state policy database	Date since bars were require to close again, state level
date_since_mask	COVID-19 U.S. state policy database	Date since mask required in public spaces, state level
beds	Homeland Infrastructure Foundation-Level Data	Total number of hospital beds in each county
date_since	JHU CSSE COVID-19 Database	Date since first COVID-19 infection
All.Cause.death_rate	CDC WONDER	All cause death rate
Assault.death_rate	CDC WONDER	Assault death rate
Cancer.death_rate	CDC WONDER	Cancer death rate
Cardiovascular.death_rate	CDC WONDER	Cardiovascular disease death rate
Continued on next page		

Table S2 – continued from previous page

Variable	Source	Definition
Despair.death_rate	CDC WONDER	Suicide and overdose death rate
PediatricAsthma	American Lung Association	Prevalence of pediatric asthma
AdultAsthma	American Lung Association	Prevalence of adult asthma
COPD	American Lung Association	Prevalence of COPD
AdultChronicLungDisease	American Lung Association	Prevalence of adult chronic lung disease
LungCancer	American Lung Association	Prevalence of lung cancer
2013 code	NCHS Urban-Rural Classification Scheme for Counties (CDC)	Counties are assigned a number 1-6, with 1 being the most urban and 6 being the most rural