Strong Effects of Population Density and Social Characteristics on Distribution of Covid-19 Infections in the United States

Kumar B. Rajan, PhD,[1] Klodian Dhana, MD PhD,[2] Lisa L. Barnes, PhD,[3] Neelum T. Aggarwal MD,[3] Laura E. Evans, MD,[4] Elizabeth A. McAninch, MD,[2] Jennifer Weuve, ScD,[5] Robert S. Wilson, PhD,[4] Denis A. Evans, MD[2]

[1]Department of Public Health Sciences, University of California, Davis, CA

[2]Department of Internal Medicine, Rush University Medical Center, Chicago IL

[3]Department of Neurological Sciences, Rush University Medical Center, Chicago IL

[4]Department of Emergency Medicine, University of Washington, Seattle, WA

[5]Department of Epidemiology, Boston University, Boston, MA

Corresponding author: Kumar B. Rajan, PhD, One Medical Sciences 1C, Davis, CA 95616. Tel: 916-734-9658.

ABSTRACT

Coronavirus disease 2019 (Covid-19) has devastated global populations and has had a large impact in the United States with the number of infections and deaths growing exponentially. Using a smooth generalized additive model with quasipoisson counts for total infections and deaths, we developed a county-level predictive model that included population demographics, social characteristics, social distancing, and testing data. This model strongly predicted the actual US distribution of Covid-19, accounting for 94.8% of spatial-temporal variation in total infections and 99.3% in Covid-19 related fatalities from March 15, 2020. US counties with higher population density, poverty index, civilian population, and minorities, especially African Americans had a higher number of confirmed infections adjusted for county population. Social distancing measured by the change in the rate of human encounter per km^2 relative to pre-covid-19 national average was associated with slower rate of Covid-19 infections, whereas higher testing was associated with higher number of infections. The number of people infected was increasing, however, the rate of increase in new infections was starting to show signs of plateauing starting from the second week of April. Our model projects 2.11 million people to test positive for Covid-19 and 122,951 fatalities by June 1, 2020. Importantly, our model suggests strong social differences in the infections and deaths across US communities, and inequities in areas with larger African American minorities and higher poverty index expected to show higher rates of Covid-19 infections and deaths. Preventive steps including social distancing and community closures have been a cornerstone in stopping the transmission and potentially reducing the spread of the disease. Crucial knowledge of the role of social characteristics in the disease transmission is essential to understand current disease distribution, predict future distribution, and plan additional preventive steps.

INTRODUCTION

The Covid-19 is a global pandemic affecting 187 countries with over 3.84 million confirmed infections, 269,000 deaths, and a staggering fatality rate of 7.0%.[1] In the US alone, there are over 1.25 million infections with 76,000 deaths and a fatality rate of 6.0% that has remained steady. However, there are considerable variations in the Covid-19 infection and death rates across US communities over time. Hence, understanding the geospatial and temporal variation in Covid-19 infections and deaths needs serious and urgent attention. Many of these US communities show large variations in chronic health conditions, population density, and socio-economic status with poor access to essentials and lower social distancing, all of which could lead to higher rates of infections and deaths.

The objective of this research manuscript is to develop a social transmission model to study the geospatial temporal variation in infections and deaths across US counties. The social transmission model will utilize county-level population demographics, focusing on population density, number of minorities, and age distribution, and social characteristics,[2] such as, poverty index[3] and number of non-professional civilian population, and social distancing using rate of unique human encounters per Km^2 relative to US national pre-COVID baseline. This social transmission model will allow us to study the contribution of population density and social characteristics on the distribution of Covid-19 in US communities. Importantly, studying Covid-19 infections and deaths using our social transmission model will allow us to better understand the predictors of current disease distribution across US communities, the ability to predict future distribution across US communities and develop a national-level estimate, and most significantly identify the US communities that require the most resources to slow the infection and death rates.

METHODS

The data for this project comes from several compiled sources for testing data, daily infections, and daily deaths, 2010 US census data, and data on social distancing. More details for each of these sources are provided below.

Testing Statistics: The source for total number of tests for Covid-19 came from the COVID tracking project[4] and the US CDC.[5] The COVID tracking project aggregates the testing data by individual states and reports the number of people tested, including private labs. However, not all states report their figures, and this data should be considered as a general indication of testing output. The CDC provides the specimens tested in the CDC labs and public health labs in 49 states, New York City, Puerto Rico, USAF, and 15 California Counties. With these two sources, we would be able to obtain a general count of total tests performed in the US, with the counts having up to 7 days of lag when specimens are accessioned, testing is performed and summarized.

Test Cases and Deaths: Several Covid-19 data have been made available for research purposes. We use county-level epidemiological data on confirmed cases and deaths starting from March 1, 2020, which is available from Johns Hopkins University that is updated on a daily time series pattern.[6] Other epidemiological data including WHO situational reports and Atlantic Covid-19 tracking project were also be considered to check the accuracy and reports from these three sources. Data downloads from the source were automatic and a daily update was performed to get the most recent data.

US Census Data: The U.S. Census Bureau is the leading source of statistical information about the people living in the US in the form of a decennial census, which count the entire U.S. population every ten years (combination of long and short forms), along with several other

surveys.[7] The US census bureau collects several pieces of information from the population and has several hundred identified population and housing tables down to the block level. The 2010 US census data is available, which has been downloaded, curated, and integrated with county-level infections and deaths.

American Community Survey (ACS): The ACS is an ongoing monthly survey sent to 3.5 million addresses to produce detailed population and housing estimates each year.[8] The ACS is designed to produce critical information on small geographic areas and releases annual estimates for over 35,000 communities. The ACS collected several pieces of economic and community data that are relevant to this project. The ACS is also performed through the census bureau, but more detailed data was only collected starting from 2000. We use economic data from 2008 ACS survey on poverty index and non-professional civilian population for each county.

Social Distancing: According to the CDC and WHO, social distancing is currently the most effective way to slow the spread of Covid-19 through US communities. Unacast has developed a social distancing data program that consists of daily encounter, daily visitation, and daily non-essential visits compared to pre-COVID and averaged for the US population.[9] We used encounters rate since it provides the most appropriate to study the change in human encounters per square Km of residents in each US county.

Statistical Analysis

Descriptive plots for infections and deaths summarized over all US counties provided information on the cumulative infections and deaths. New daily infections and deaths were estimated as a lagged difference of cumulative infections and deaths between current and previous days. Similar characteristics were estimated for testing, hospitalization, and encounter rates across all US counties.

For our social transmission model, we used a smooth generalized additive model[10] with quasipoisson counts for total infections and deaths that included population density, poverty index, proportion of non-Hispanic Whites, Blacks, and Hispanics, proportion of females and non-professional civilians, age distributions (below 20, 20-40, 40-60, above 60) and social distancing for each US county. A county-level model was developed in several steps; the first step using time since March 1, 2020 and latitude and longitudinal coordinates for counties explained about 16% variation in the rate of confirmed infections. The addition of population demographics, social characteristics, and social distancing explained around 98% of variation in Covid-19 infections.[11] This model also included splines for time since March 1, 2020, population density, and latitude and longitude. In a separate model, we included testing characteristics and found the predictive models to be unstable due to large county-level missing data and underreporting and severe lags; these additional variables were therefore, excluded from our final model. A second social transmission model for Covid-19 deaths including all the variables in the infections model and the number of infections in each US county. All plots and statistical models were performed in Microsoft R Open Version 3.5.3 x86_64-pc-linux-gnu (64-bit) and Intel MKL for parallel mathematical computing using 18 cores.[12]

RESULTS

According to the most recent estimate, 1.25 million US residents are infected with Covid-19, around 3,454 per million residents. Of those infected, 75,543 US residents had died, 207.3 per million US residents. The cumulative number of infections and deaths show continual increase, however, the rate of increase in new Covid-19 infections (Figure 1A) show dramatic increase until April 8, 2020, and since then the new infection rates show signs of plateauing with a slow

downward trend. A similar increasing pattern in cumulative deaths was observed, however, the rate of deaths peaked on April 15, 2020, with a steady downward trend since then (Figure 1B). Thus, the figures for new Covid-19 infections and deaths may show signs of reaching the peak for the US outbreak.

Testing for Covid-19 has also steadily increased in the US starting from 41,191 total tests by March 15, 2020 and increasing to 6,817,925 by May 3, 2020. The daily increase in testing had risen to 105,423 by April 1, 2020, and to 218,465 by April 30, 2020. The highest daily increase in testing was 303,982 on May 1, 2020. Figure 1(C) shows the number of new daily tests in the US, which had increased over time with 25,000-30,000 tests performed daily starting from May 1, 2020. In terms of US states, New York had performed the most tests, close to 2,000,000 with California testing half as often as New York, followed by Florida, Texas, and Washington.

According to the COVID tracking project, there were 125,796 hospitalizations with Covid-19, with overall cumulative Covid-19-associated hospitalizations at 42.3 per 100,000 with the highest rates in people over the age of 65. Daily new hospitalization in the US had been around 2,000 new hospitalization cumulative over the states (Figure 1D). However, on May 1st, new Covid-19 related hospitalizations showed a dramatic increase to around 10,000 on that day, which fell back to 2,000 new hospitalizations. Reporting on hospitalizations has been uneven by the states. New York had the most hospitalizations, and, of the remaining US states, Connecticut, Massachusetts, and Florida reported the highest numbers of hospitalizations.

We developed a social contagion model to predict the distribution, within the US, of Covid-19 infections and deaths from population demographics, social characteristics, and social distancing. The model for infections accounted for 94.8% variation in the data with 92.2%

deviance, and 99.2% variation with 96.8% deviance in deaths across 3,364 US counties. The population demographics and social characteristics were also strongly associated with the rate of increase in confirmed Covid-19 infections. According to the social contagion model, we predict that 2.11 million US residents will have confirmed Covid-19 infections and 122,951 deaths by June 1, 2020. The actual (red line), estimated (blue line), and predicted (brown line) for Covid-19 infections is shown in Figure 2A and for deaths in Figure 2B.

Geospatial variation in confirmed Covid-19 infections (Figure 3) and deaths (Figure 4) with population demographics and social characteristics from March 1, 2020 to June 1, 2020 is evident in time lapsed US maps. Most of the eastern US and large counties in the western US show considerable increases in number of cases. Many counties in the central US are not reporting Covid-19 cases both due to incomplete data and lack of Covid-19 cases, but the exact reasons are harder to discern. Geospatial variations in confirmed Covid-19 deaths from March 1, 2020 to June 1, 2020 are starker with much of the North East and Midwest and South East, South West, and North West showing higher death rates.

In US counties with higher proportions of African Americans, the rate of Covid-19 infections increased by 5.6% for one-unit increase in percentage Blacks (Figure 5), whereas, the rate of increase was 2.6% in Whites and 4.9% in Hispanics. Additionally, in US counties with a higher poverty index, the rate of infections was 4.8% for one-unit increase in US census poverty index. In areas with higher non-professional civilian population, the rate of infection was also higher. In US counties with larger young population (20-40) and older population (60-80), the rate of infections was higher, and rates of infection in counties with large young population (below 20) and older population (above 80) the rate of infections were lower. The rate of death increased by 0.6% for each unit increase in poverty index, and 1% for each percent higher

proportion of non-Hispanic Blacks (Figure 6), whereas, the rates decreased by 1% for each percent higher Non-Hispanic whites. Ages over 40 were associated with higher death rates, whereas below 20 was associated with lower death rates. These findings suggest strong effects of population demographics and social characteristics on confirmed Covid-19 infections.

Population density, social distancing, time, and geospatial variation were also associated with the number of confirmed Covid-19 infections (Figure 5) and deaths (Figure 6). In US counties with higher population density, the rate of increase in Covid-19 confirmed infections was exponentially higher. About 10% of US counties (N=311) have population density higher than 500 residents per square mile and accounted for 80% total infections and deaths in the US. As social distancing became higher, the rate of Covid-19 infections was lower. The time trend showed a steep increase in the rate of infections from March 15, 2020 to about April 15, 2020, with the rate of infections leveling off and then slowing between April 15, 2020 and May 5th, 2020.

Social distancing has a strong and consistent association with Covid-19 infections and deaths in the US communities. Our social transmission model predicts that by June 1, 2020, the US will have 2,113,073 confirmed Covid-19 cases if the social distancing across all counties remains the same. If we see a 20% decline in social distancing, we project 46,433 additional COVID infections and 66,764 additional infections if social distancing were to decrease by 30%. Social distancing will also have substantial influence on We project that by June 2020, the US will have 122,951 deaths attributed to Covid-19, if the social distancing across all counties remains the same. If we see a 20% decline in social distancing, we project 2,785 additional deaths, and 4,006 additional deaths if social distancing were to decrease by 40%.

DISCUSSION

The social transmission model shows the high relevance and significance of population and social characteristics on Covid-19 infections and deaths in the US communities. According to our model, US counties with high population density have high rates of Covid-19 infections and deaths, with rates of infections and deaths increasing exponentially. A small fraction (10%) of densely populated US counties account for 80% of all confirmed infections and deaths across the US. These densely populated communities had twice the number of non-Hispanic Blacks than communities that were less densely populated as is also reflected in our social contagion model, where rates of infections and deaths were dramatically higher in US counties with large number of non-Hispanic Blacks.

Community-level transmission was slower in communities with higher social distancing. As social distancing increased, the rate of increase in confirmed infections and deaths started to decline, suggesting substantial increase in confirmed infections and deaths may be attributable to reduction in social distancing. The high Covid-19 infections and deaths in the densely populated areas were seen despite higher social distancing. Also, of significance is that communities with high poverty index and social characteristics in general had lower social distancing compared to geographical areas with low poverty index and similar social distancing. If social distancing restrictions were to be reduced in these densely populated lower socio-economic areas, we may be more likely to see higher number of confirmed infections and deaths in these communities. If social distancing can be improved in densely populated areas with high poverty indices, we may be likely to see substantial reductions in confirmed infections and deaths. However, areas with higher proportion of non-Hispanic Blacks shows significantly higher rate of infection and deaths, and these effects were larger than the social distancing effects, which in general was more

protective in areas with higher proportion of non-Hispanic Whites. These findings suggest strong effects of race/ethnicity on infections and deaths in the US communities.

Age distribution plays a significant role on the rate of increase in Covid-19 confirmed infections and deaths. It is noteworthy that areas with high infections and deaths, also had larger number of younger residents 20-40 years old, and larger number of older residents, 60-80 years old over the age of 60. This population dynamic suggests that young residents may be more likely to be asymptomatic carriers of the coronavirus. In areas with higher number of middle aged adults, those 40-60 years old had lower rate of infection, perhaps suggesting that more social distancing is being maintained in the middle age groups than the younger age groups with those of older ages showing a high susceptibility for infections due to the higher number of chronic health conditions associated with age.

The rate of change in number of infections and deaths increased exponentially in late March and early to middle of April. However, the rate of new infections has stabilized over time, reaching a plateau, where it continues to remain steady. The new infection and death rates across the US communities have started to decline, perhaps mostly due to social distancing, however, the evidence for continued decline over 14 days as mandated by the US government is yet to be observed in most of the densely populated areas. The stable rate of new infections, and lack of data on Covid-19 deaths from many counties are troublesome, since even a phased re-opening in the densely populated US communities may cause a large increase in infections and deaths, unless more precautions and preventive measures are put in place.

The social transmission model provides a framework for incorporating population demographics and social characteristics in addition to temporal and geospatial patterns as predictors of Covid-19 infections and deaths in the US communities. Even if testing were to be

dramatically increased, this approach alone does not address the highly infectious character of Covid-19. The major roles of population demographics and social characteristics may be more effectively reduced through social distancing. Focusing our preventive efforts on population centers with higher number of non-Hispanic Blacks, poverty and low-socioeconomic areas and improving both social distancing and testing in those areas might offer a better chance of reducing the spread of Covid-19 and deaths associated with Covid-19 across US communities.

REFERENCES

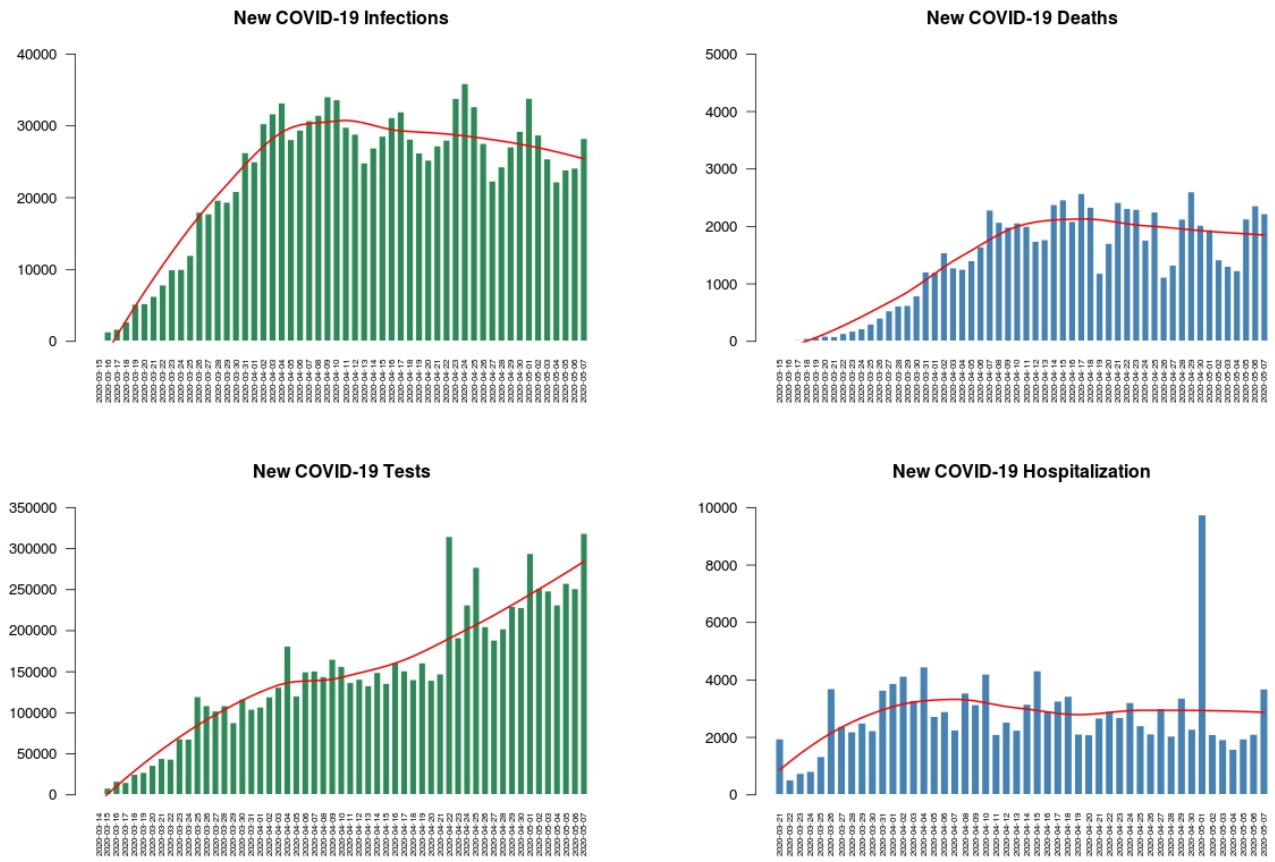
1. WHO Situation Reports, 2020. Source website: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed daily.
2. Braveman PA, Cubbin C, Egerter S, Williams DR, Pamuk E. Socioeconomic disparities in health in the United States: What the patterns tell us. American Journal of Public Health, 2010;100(S1):S186-S196. doi:10.2105/AJPH.2009.166082
3. Carter RT. Racism and psychological and emotional injury: Recognizing and assessing race-based traumatic stress. The Counseling Psychologist, 2007;35(1):13-105. doi:10.1177/0011000006292033
4. COVID tracking project. https://covidtracking.com/data/us-daily. Accessed daily.
5. CDC. The public health laboratory testing for COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html . Accessed daily.
6. Johns Hopkins data repository daily reports. https://coronavirus.jhu.edu/data/new-cases. Accessed daily.
7. US Census Data. https://www.census.gov/data.html. Accessed May 1, 2020.
8. American Community Survey. https://www.census.gov/programs-surveys/acs/data.html. Accessed May 1, 2020.
9. Unacast social distancing data. https://www.unacast.com/covid19/social-distancing-scoreboard. Accessed daily.

10. Wood SN. Modelling and smoothing parameter estimation with multiple quadratic penalties. Journal of the Royal Statistical Society. Series B. 2000;62 (2):413–428. doi:10.1111/1467-9868.00240.

11. Wood SN, Pya N, Saefken B. Smoothing parameter and model selection for general smooth models (with discussion). Journal of the American Statistical Association. 2016;111 (516): 1548–1575.

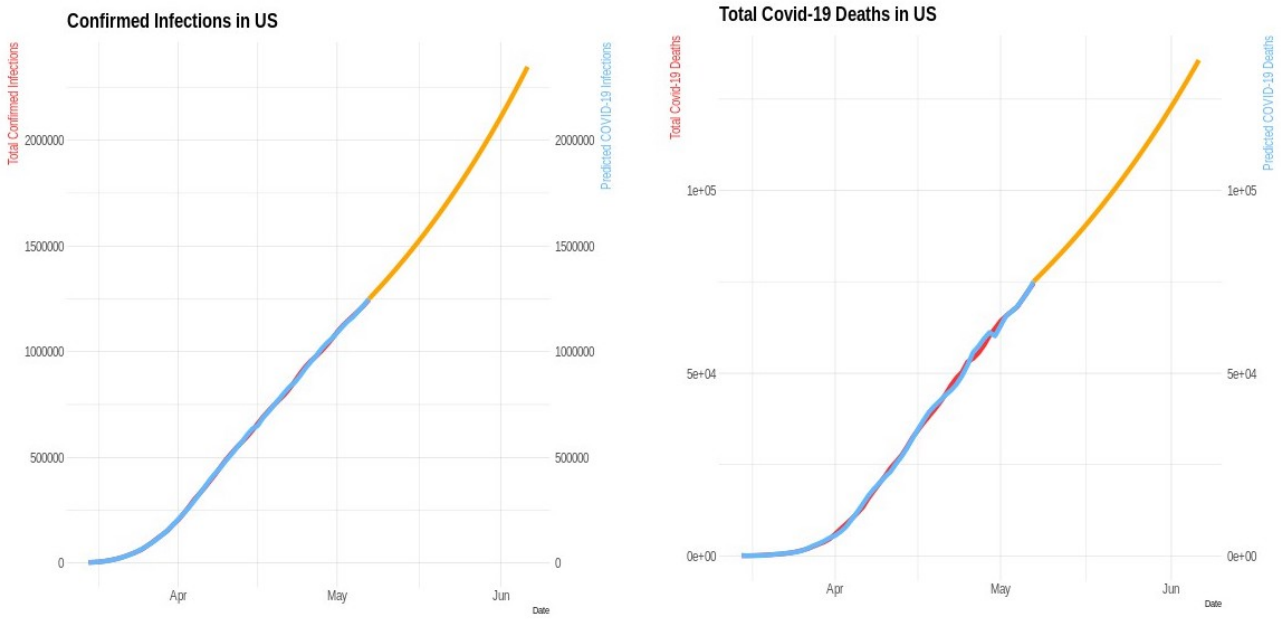
12. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Figure 1. Daily New Covid-19 Infections, Deaths, Testing, and Hospitalization in the United States from March 15, 2020



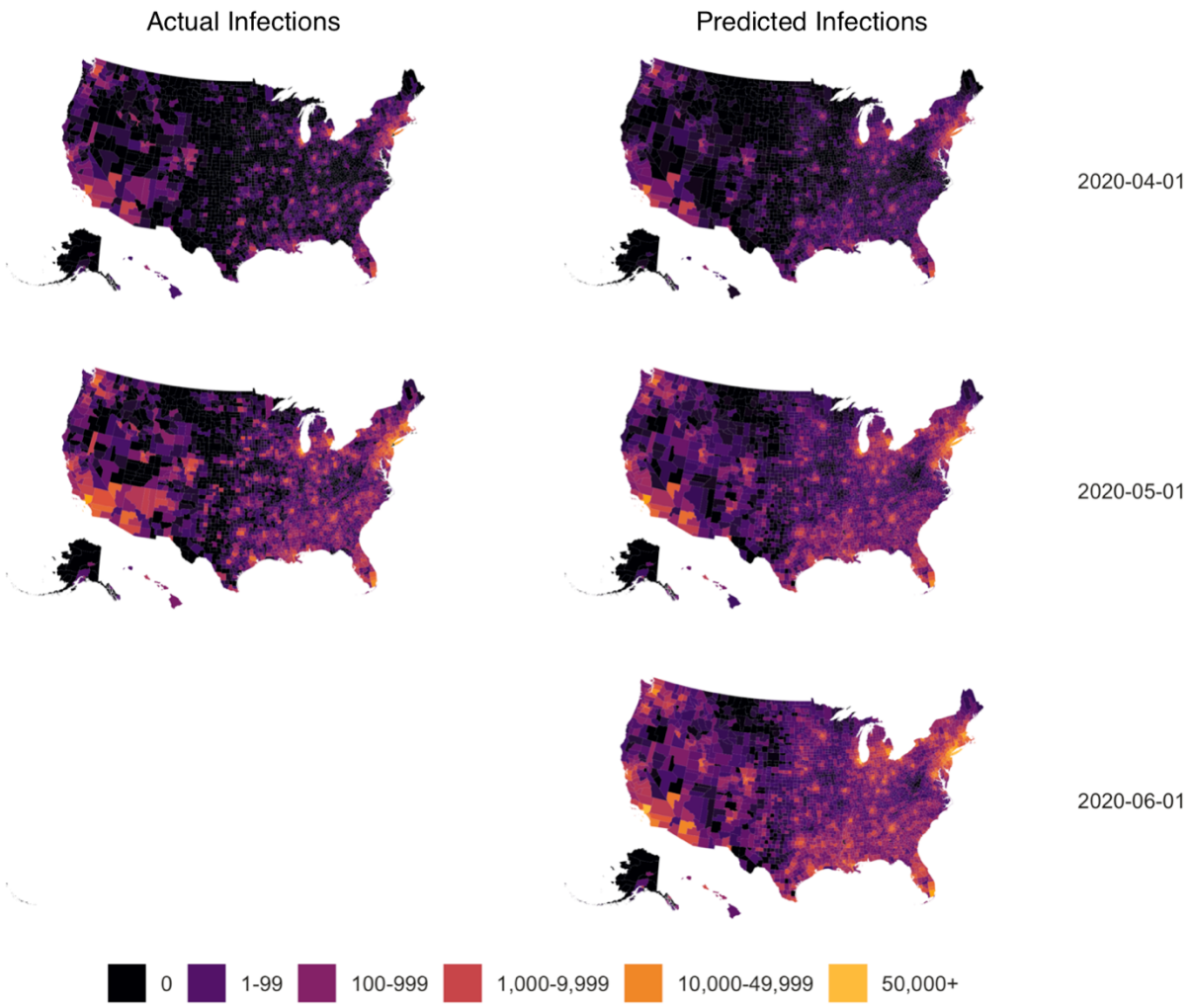
Legend: The top left panel (green) shows the daily new Covid-19 infections and the top right panel shows the daily new Covid-19 deaths in the US. The bottom left panel (green) shows the daily new Covid-19 tests and the bottom right panel shows the daily new Covid-19 hospitalizations in the US. A loess smoother (red line) is fit to the daily new infections and deaths in the US.

Figure 2. Cumulative Confirmed Infections and Deaths in the US from March 15, 2020 to June 1, 2020 with Actual and Predicted Values using the Social Transmission Model



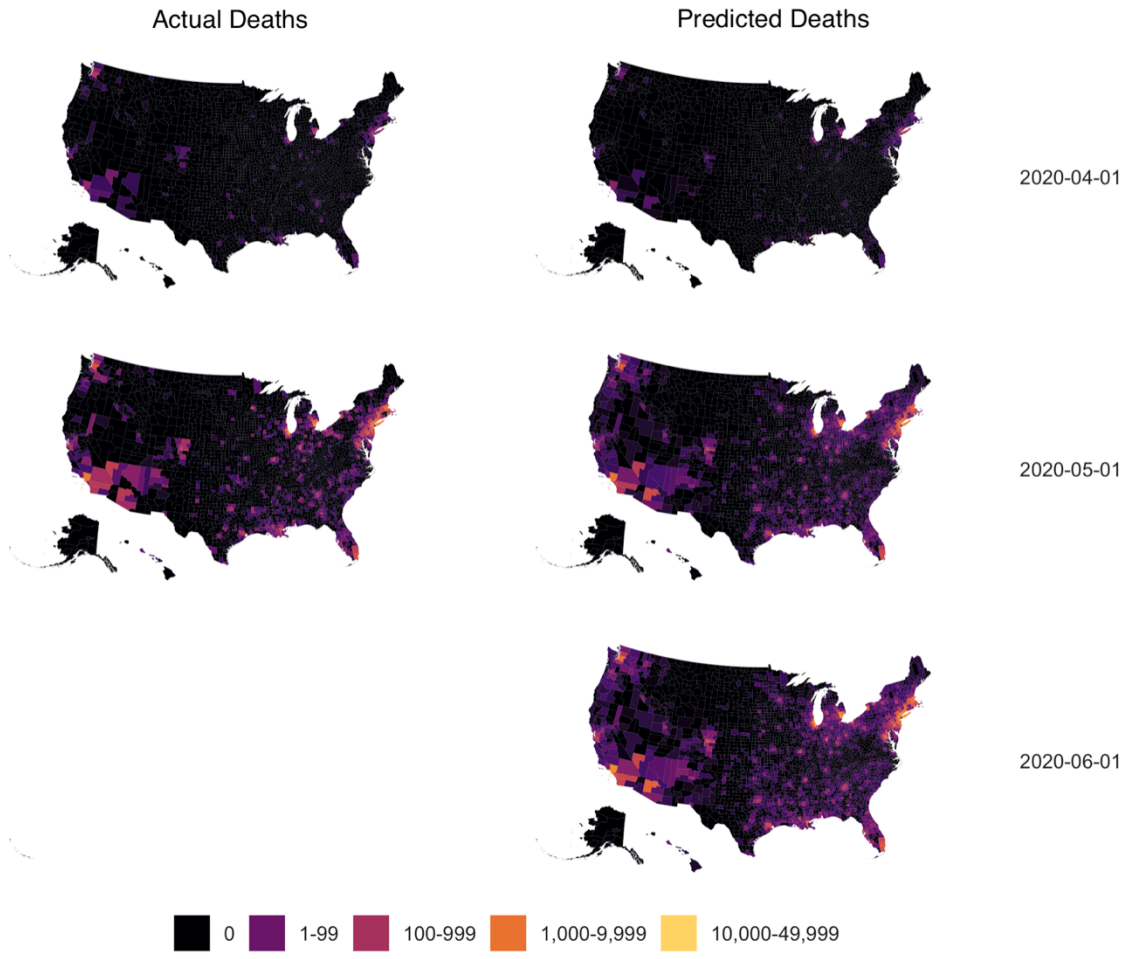
Legend: The left panel shows the confirmed infections and right panel the reported deaths in the US. In the figures, red line shows the actual confirmed Covid-19 infections and deaths, the blue line shows the estimated Covid-19 infections and deaths, and the brown line shows the predicted Covid-19 infections and deaths until June 1, 2020.

Figure 3. Distribution of Actual and Predicted Covid-19 Infections on Selected Dates from March 15 to June 1, 2020 in US Counties



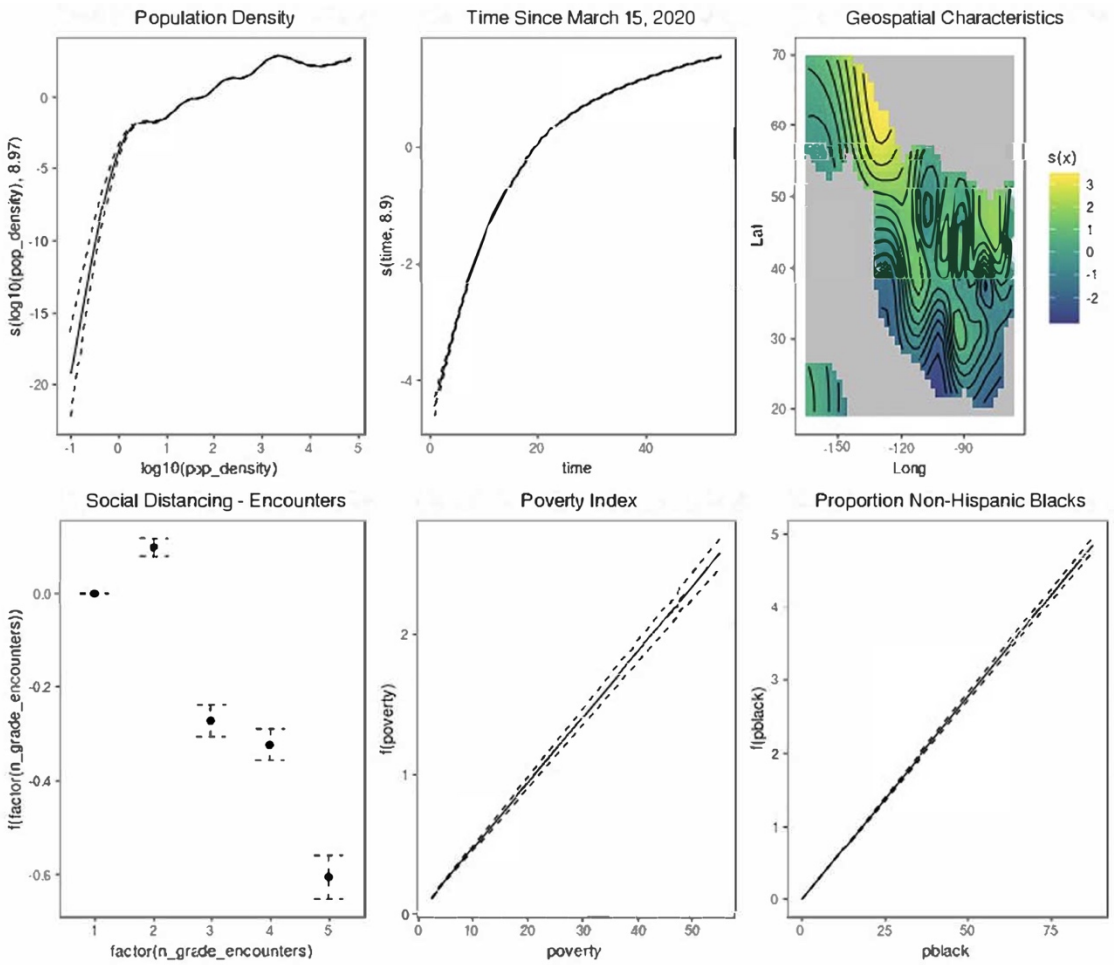
Legend: The left side shows the actual Covid-19 infections in US counties on April 1 and May 1, 2020. The right panel shows the predicted Covid-19 infections in US counties on April 1, May1, and June 1, 2020.

Figure 4. Distribution of Actual and Predicted Covid-19 Deaths on Selected Dates from March 15 to June 1, 2020 in US Counties



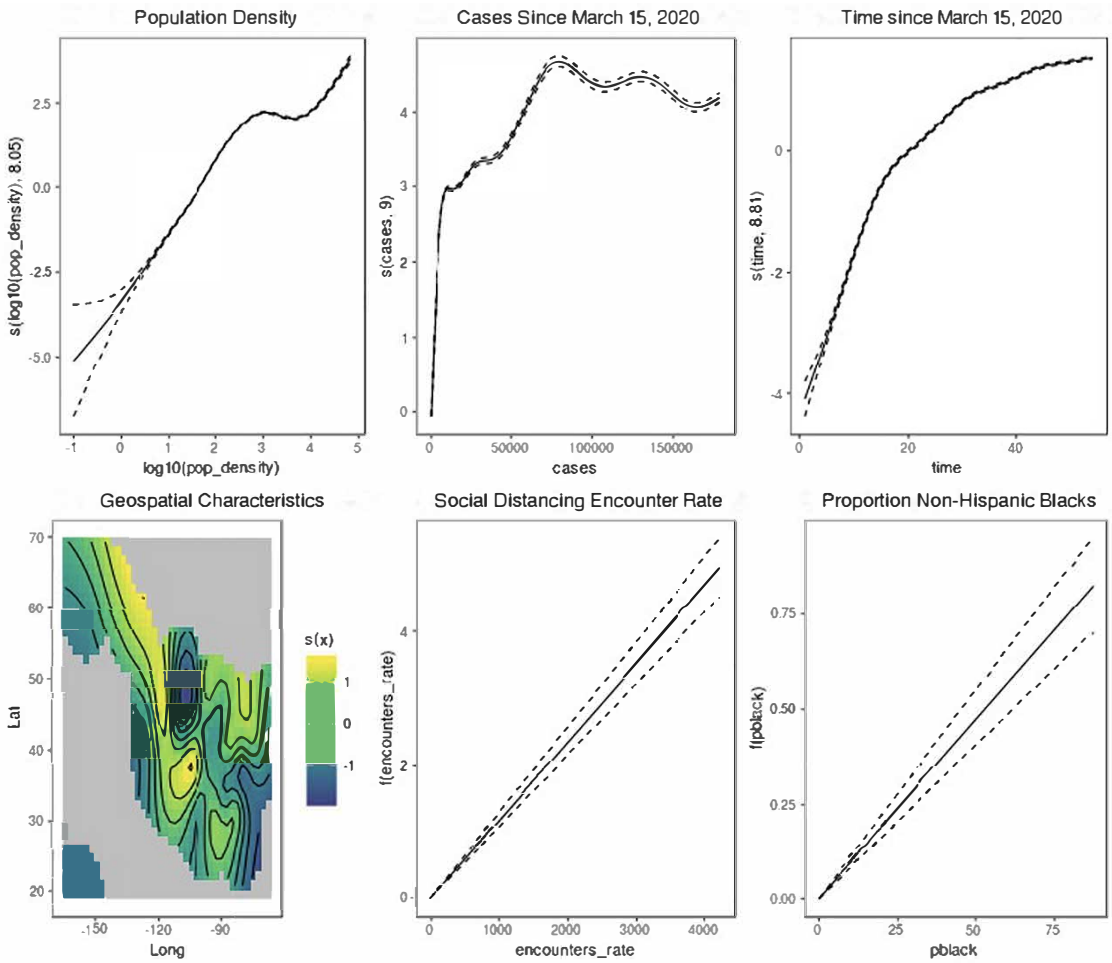
Legend: The left side shows the actual Covid-19 deaths in US counties on April 1 and May 1, 2020. The right panel shows the predicted Covid-19 deaths in US counties on April 1, May1, and June 1, 2020.

Figure 5. Relation of Population Density, Time, Geographical, Social Distancing, Poverty, and Percentage Blacks on Rate of Covid-19 Infections and Deaths in US Counties



Legend: The rate of Covid-19 infections uses smooth terms for log10 transformed population density, time since March 15, 2020, and geospatial characteristics, and additive terms for social distancing using encounter rates, poverty index, and proportion of non-Hispanic Blacks.

Figure 6. Relation of Population Density, Time, Geographical, Social Distancing, Poverty, and Percentage Blacks on Rate of Covid-19 Infections and Deaths in US Counties



Legend: The rate of Covid-19 infections uses smooth terms for log10 transformed population density, number of cases, time since March 15, 2020, and geospatial characteristics, and additive terms for social distancing using encounter rates, and proportion of non-Hispanic Blacks.