Geospatial Health 2023; volume 18:1153

A spatiotemporal analysis of the social determinants of health for COVID-19

Claire Bonzani, Peter Scull, Daisaku Yamamoto

Department of Geography, Colgate University, Hamilton, New York, United States

Abstract

This research aims to uncover how the association between social determinants of health and COVID-19 cases and fatality rate have changed across time and space. To begin to understand these associations and show the benefits of analysing temporal and spatial variations in COVID-19, we utilized Geographically Weighted Regression (GWR). The results emphasize the advantages for using GWR in data with a spatial component, while showing the changing spatiotemporal magnitude of association between a given social determinant and cases or fatalities. While previous research has demonstrated the merits of GWR for spatial epidemiology, our study fills a gap in the literature, by examining a suite of variables across time to reveal how the pandemic unfolded across the US at a county-level spatial scale. The results speak to the importance of understanding the local effects that a social determinant may have on populations at the county level. From a public health perspective, these results can be used for an understanding of the disproportionate disease burden felt by different populations, while upholding and building upon trends observed in epidemiological literature.

Correspondence: Claire Bonzani, Department of Geography, Colgate University, 13 Oak Drive, Hamilton, New York, United States.
Tel.: 914.215.4779
E-mail: cbonzani@colgate.edu

Key words: COVID-19; social determinants; GWR.

Acknowledgments: we would like to thank the Colgate Research Council for their support of this research and Colgate Geography Department for their guidance on the project. Additionally, we would like to thank Professor Rebecca Upton for her help uncovering and explaining the trends found in these results. Finally, we would like to highlight the contributions of Jasmin Lopez, particularly her assistance generating the final maps.

Conflict of interest: the Authors declare no conflict of interest.

Received for publication: 18 September 2022.
Revision received: 22 January 2023.
Accepted for publication: 24 January 2023.

Geospatial Health 2023; 18:1153
doi:10.4081/gh.2023.1153

Introduction

The virus responsible for the COVID-19 pandemic is known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The first outbreak of COVID-19 was traced to Wuhan, China, but has since spread to every corner of the world, with severe outbreaks in most countries (CDC, 2021). As of July 2022, there have been roughly 1.02 million deaths and over 50% of America (including 75% of children) have been infected (Mandavilli, 2022).

The disease primarily spreads through exposure to infectious respiratory fluids (CDC, 2021). However, the likelihood of infection is often dependent on duration of exposure, proximity to the infected individual and the environmental conditions where an individual encounters the virus (CDC, 2021). With the emergence of a vaccination plan in April 2021, unvaccinated individuals are more likely to experience symptomatic and extreme infections, while vaccinated individuals are often protected from infection (CDC, 2021). When COVID-19 first emerged at the epidemic scale, little was known about the social factors that contribute to an individual's likelihood to be infected or die from complications. However, as the spread of the virus escalated to a global scale, the social factors associated with virulence became more apparent.

Early literature on COVID-19 was primarily focused on initially determined co-morbidities linked to individual cases of severe infection and death including chronic lung diseases, heart conditions, obesity, and smoking (Mouliou *et al.*, 2021). Place-based studies of co-morbidities highlighted spatial clustering of counties with high mortality together with high chronic disease prevalence and high social vulnerability, indicating a positive association not only between co-morbidities and COVID-19 mortality, but social factors as well (Islam *et al.*, 2021). As the pandemic progressed, observers began to identify and focus on the more place-based social determinants of COVID-19. Social Determinants of Health (SDH) are "the non-medical factors that influence health outcomes" (WHO, 2022) and include the conditions in which people live and work, as well as the wider set of systems surrounding them.

Beyond co-morbidities, the determinants for infection and death by COVID-19 have been further connected to sociodemographic and socioeconomic indicators of health. For example, Fielding-Miller *et al.* (2020) found that a higher percentage of non-English-speaking individuals in the United States (US), individuals older than 65 years, and those living at or below the poverty level in non-urban areas are at a heightened risk for mortality due to COVID-19. Perhaps one of the most notable relationships between COVID-19 infection and SDH has to do with ethnic affiliation that can act as an indicator, not only for severe COVID-19 infection, but also for the burden of co-morbidities, crowded living conditions in urban areas and increased likelihood to work in public-facing careers (Hooper *et al.*, 2020). Paul *et al.* (2021) expand upon these results, revealing that across the US, the non-white population is strongly correlated with the rate of both COVID-19 infection and mortality.

Vaccines became available to the US adult population in 2021, and they were hugely effective at reducing COVID-19 infection rates – reducing infection rates by 80-88% compared to unvaccinated individuals (Bernal *et al.*, 2021). However, the US vaccination campaign was divisive, leading to further stratifications in disease spread and virulence, as many demographic groups chose to remain unvaccinated. According to the Kaiser Family Foundation's research into people's willingness to be vaccinated, roughly one in five adults still affirm that they will either never get the vaccine or will only get the vaccine if required. Shmueli (2021), in a study into the Health Belief Model for COVID-19, found that certain demographic factors, such as age, ethnicity, gender, educational level, suffering from chronic disease, and receiving previous vaccinations, were all behaviours associated with the likelihood of an individual accepting to be vaccinated. With respect to the wave of COVID-19 *Delta*-variant, literature has made it apparent that social factors contribute to an individual's willingness to receive the vaccine, with Republicans and those who mainly consume "conservative news media" being less likely to become vaccinated (Viswanath *et al.,* 2021). Furthermore, the Kaiser Family Foundation (2021) reveals that there are large gaps in vaccination rates between urban (75%), rural (58%), and suburban (73%) areas. As a result of these findings, there has been increased interest in the social determinants of COVID-19 vaccination, whereas the previous focus was on the interplay of social determinants and mortality rates.

Some of the social determinants that correlate with disproportionate disease burden (e.g., individual access to food, population density, air pollution indices and housing type) relate to the structure of the physical environment (Singu *et al.*, 2020). For instance, COVID-19 spreads more rapidly in areas with higher population density. As reported by Sy *et al.* (2021) doubling of the population density increases the index of pathogen's contagiousness and transmissibility (R_0) by 0.11. Social structures, such as access to quality medical care stemming from health insurance are another factor contributing to differences in disease burden in diverse populations (Grey II et al., 2020). Health insurance gaps were associated with 44% of COVID-19 infections and 32% of COVID-19 deaths between January 2020 and the end of August 2020 in the US (Garber, 2021). Further, community economic indicators, such as local unemployment rate, are associated both with disproportionately high case rates and CFRs, something which might be due to the fact that many individuals rely on their employers for health insurance coverage (Singu *et al.*, 2020). Conversely, increased median income is associated with increased COVID-19 infection risk in the US, showing that community economic indicators affect mortality and infection differently (Abedi *et al.*, 2021).

Since the spread of COVID-19 is influenced and mediated by social determinants of health, which have inherent spatial dimensions, many studies have utilized spatial statistical methods to explore the geography of COVID-19 in relation to existing social, medical, and environmental conditions. Mollalo, Rivera and Vahabi (2021), for example, used Local Moran's *I* classes to identify spatial clustering of U.S. counties with high and low pre-existing mortality rates and COVID-19 fatalities. Mollalo, Vahedi and Rivera (2020) explored the relationship between various environmental and socio-demographic variables and COVID-19 incidence at the US county scale, using Geographically Weighted Regression (GWR) and multiscale GWR (MGWR). This approach has also been used to model the spatial variation in COVID-19 vaccine hesitancy, again at the county level in the US (Mollalo and Tatar, 2021). While these studies demonstrate the advantage of local models, over global models (e.g., ordinary least square (OLS) models and spatial lag models), they rely on point-in-time COVID-19 data and do not account for temporal variation.

In part due to the COVID-19 outbreak being very recent, studies using GWR to explore both spatial and temporal infection trends in association with SDH are still limited in number and scope. However, Wu and Zhang (2021), identified seasonal patterns in the relationship between disease spread and selected SDH in the US state of Texas finding that winter was the most sensitive season for virus spread in certain counties. Chen *et al.* (2022) utilized GWR to analyse how risk for COVID-19 infection changed from January to mid-September 2020 at the US county level. Their results show that even in the first year of the pandemic, there were temporal variations in the risk for COVID-19 infection. Nevertheless, there are still few studies to date that have systematically analysed both the temporal and spatial relationships between COVID-19 and SDH, especially across the major waves of infection. Our study aimed to fill this gap by analysing the relationship between COVID-19 (numbers of cases and fatality) and SDH, using GWR across all counties in the contiguous US (*i.e.*, excluding Alaska and Hawaii) between January 2020 and January 2022. We examined whether the existing understanding of the pandemic's mobility invariably holds over space and time.

Materials and Methods

Data

Four different COVID-19 waves were chosen for analysis, based on changes in policy response, disease virology or both. COVID-19 data was pulled from The New York Times GitHub page, which updates the COVID-19 cases and deaths at the county level on a daily basis (New York Times, 2022).

The waves were defined as follows: i) Wave-1: January 1, 2020 – September 30. The end was determined by the US epidemic curve, by choosing the date where the nation-wide epidemic curve had fully bottomed; ii) Wave-2: October 1, 2020 – April 19th, 2021. The end was determined by the date where vaccines became available to all adults; iii) Wave-3: April 20, 2021 – November 30, 2021.The end was determined by the first known case of the Omicron variant in the US; and iv) Wave-4: December 1, 2021 – January 18, 2022. The end was determined at the date of data collection. For the case measure, the cumulative number was transformed into cases per 100,000 population and the death rate was measured by the CFR, which equals the number of deaths of all confirmed cases. Data for SDH were identified based on existing accounts of socioeconomic and demographic factors deemed influential in the COVID-19 outbreak. Initially 16 candidate variables were collected from various sources including the US Census and the US Economic Service. By following several variable reduction strategies (described below), nine variables were eventually selected: non-white population (Hooper *et al.*, 2020; Paul, 2021), less than high school education (Shmueli *et al.*, 2020), uninsured population (Gray *et al.*, 2020), age 65+ (Neumann-Podczaska *et al.*, 2020), population density (Sy *et al.*, 2021), median income (Abedi *et al.*, 2021; Hawkins, Charles, and Mehaffey, 2020; Quan *et al.*, 2021; Whittle and Diaz-Artiles 2020), unemployment rate (Singu *et al.*, 2020), republican population (Kaiser Family Foundation, 2021; Viswanath *et al.*, 2021) and vaccination rate (Bernal *et al.*, 2021; Table 1). The geometric data of county boundaries were obtained as shapefiles from the US Census TIGER/Line website.

Models

We used both OLS and GWR modelling in this study. The standard OLS model specifies the relationship between a set of explanatory variables and a dependent variable as:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \qquad \text{Eq. 1}$$

where in county i, y_i is the COVID-19 case rate (alternatively CFR); β_0 the intercept; x_{ik} the i^{th} observation of the k^{th} explanatory variable; β_k the regression coefficient for the k^{th} independent variable; and ε_i a random error term. The OLS model implicitly assumes spatial stationarity in the relationship between the explanatory and dependent variables, even if this assumption does not always hold in reality.

GWR extends the OLS model by allowing for locally variable coefficient estimates (Fotheringham *et al.* 2003). The difference between GWR and traditional regressions can be observed by adding the geographical coordinates (u_i, v_i) to the model:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \qquad \text{Eq. 2}$$

GWR fits a local model of the dependent variable to every feature in the dataset. These local models are constructed by incorporating the explanatory variables of features that fall within a defined neighbourhood (Fotheringham *et al.*, 2003). In our study all OLS and GWR models were generated using ArcGIS Pro 2.9 software (ESRI, Redlands, CA, USA).

Model calibration

To construct the final models, we used the following strategy to transform variables and to reduce the number of explanatory variables. First, explanatory variables that had a skewed distribution were log-transformed. These variables included population density, uninsured rate, age 65+ and non-white population. Furthermore, following Mollalo and Tatar (2021), all dependent and explanatory variables were converted to standardized z-score (mean = 0, standard deviation = 1) for both the GWR and the OLS models. This standardization enabled the comparison of models across different waves and helped to deal with the issues of local multicollinearity, as measured in local condition numbers.

In terms of the variable reduction strategy, from the initial 16 candidate variables, we first excluded variables that were highly correlated (using a 0.7 Pearson correlation coefficient as threshold). An initial OLS regression based on this subset of variables was run for each of the COVID-19-dependent variables (cases per 100,000 population and CFR). Variance Inflation Factor (VIF) values were calculated for all explanatory variables, none of which exceeded 3.0 (with the commonly recommended threshold of 7.5). Moreover, we examined the local condition number to check for local multicollinearity in the GWR models and removed additional variables, resulting in the inclusion of only nine in the final models as shown in Table 1. As the COVID-19 data for New York City were not aggregated at the county-level, but at the city level, all the other social determinants had to be adjusted to the city level for New York City, resulting in a dataset that combined all social determinants and COVID-19 data from the following counties: Bronx, Kings, New York, Richmond, and Queens.

For the neighbourhood definition (bandwidth) of the GWR models, we used the “golden search” method where the number of neighbouring counties used for each model was selected to obtain the lowest corrected Akaike’s information criterion (AICc) for each model, which is currently a common practice in GWR modelling (Fotheringham *et al.,* 2003). Finally, of the total 3,104 coun-

Table 1. Variables used.

Social determinant of health	Source	Date
COVID-19 Case and CFR	New York Times	2022
Population density	US Census	2020
Unemployment rate	US Economic Research Service	2019
Median income	US Economic Research Service	2019
Population with less than high school education	US Economic Research Service	2019
Percent Republicans (voted for Donald Trump in the 2020 election)	MIT Election Labs	2020
Uninsured population	US Census	2019
Age ≥65	Administration for Community Living	2019
Non-white population	US Census	2020
Average percent fully vaccinated people	CDC	2020-2021

MIT, Massachusetts Institute of Technology; CDC, Center of Disease Control; US Economic Research Service, https://www.ers.usda.gov/; Administration for Community Living, https://acl.gov/

ties, 10 were omitted from the analysis for the case rate and 16 for the CFR in the fourth wave. This is because the source data for these counties reported cumulative cases in Wave-4 that were lower than the cumulative case counts in Wave-3. Thus, the interpretation of these results must be done with care.

Results

The results from each OLS and GWR model are shown in Table 2 and the list of significant variables from the OLS regression in Table 3. As expected, the GWR models were able to account for much more variance compared to the OLS results indicating that there is a spatial component to the relationship between SDH and COVID-19 infection or death. This suggests that the SDH variable is more strongly associated with the rates of COVID-19 infection, while other factors that were not included in the model (*e.g.*, pre-existing health conditions) strongly affect the CFR (Tables 2 and 3).

The per capita case GWR models were able to account for a significant amount of variance as shown in the local adjusted R^2 values (Figure 1) in all parts of the country across all waves, except for parts of central US states and Texas. Although the GWR models dealing with CFR were able to account for much less variance than the case models, the GWR model was able to account for variance in the Northeast, Midwest and along parts of the west coast (Figure 1). Local condition numbers were also mapped (Figure 1). Counties that have high local R^2 values and low condition numbers (<30) represent areas where relationships were both meaningful and reliable. Areas where condition numbers exceeded 30 indicate potential presence of local multi-collinearity (Brundson *et al.*, 2012). GWR also computes a separate regression coefficient for each county based on a function-calculated number of neighbouring counties. These visualized coefficients showed great spatiotemporal variation in the relationship between SDH and COVID-19 per capita case or CFR (Figures 2 and 3). The variation told us that a particular SDH may be positively associated with COVID-19 cases or deaths in some areas of the country (red gradation in the maps), while negatively associated in other areas (blue gradation in the maps). These associations can be tracked from wave to wave, illustrating change throughout the pandemic. In some cases, the direction of the relationship change sign between waves in the same regions, revealing some of the spatiotemporal complexity.

Discussion

The GWR results revealed a few noteworthy trends and patterns that we wish to highlight with special reference to age, ethnicity, vaccination status, education level, economy, and general social factors.

Age

Older adults are more likely to have one (or more) of co-morbidities, such as cardiac disease, diabetes, chronic lung disease and hypertension, which are particularly linked with CFR due to COVID-19 (Neumann-Podczaska *et al.*, 2020). Across many coun-

Table 2. Model comparisons.

	OLS		GWR		
Dependent Variable	AIC	Adj. R^2	AIC	Adj. R^2	Number of neighbours
First Wave Cases	7404.59	0.366	5941.52	0.653	137
First Wave CFR	8497.57	0.095	7927.56	0.285	286
Second Wave Cases	8449.68	0.112	6790.90	0.547	131
Second Wave CFR	8224.11	0.169	7716.23	0.346	213
Third Wave Cases	7727.14	0.297	6049.59	0.597	150
Third Wave CFR	7959.35	0.238	7297.60	0.426	246
Fourth Wave Cases	5938.55	0.234	4089.11	0.632	147
Fourth Wave CFR	8128.62	0.073	7578.19	0.286	218

OLS, ordinary least squares; GWR, geographically weighted regression; CFR, case fatality rate; AIC, Akaike's information criterion.

Table 3. Significant OLS variables by wave.

	Case significant variables (p<0.05)	Fatality significant variables (p<0.05)
Wave-1	Population density, Unemployment rate, Median income, Age 65+, Non-white, Less than High School education.	Population density, Unemployment rate, Median income, Uninsured population, Age 65+, Non-white, Less than High School education.
Wave-2	Unemployment rate, Median income, Republican, Uninsured, Age 65+	Median income, Republican, Age 65+, Non-white, Less than High School education.
Wave-3	Population Density, Unemployment Rate, Median Income, Republican, Uninsured, Age 65+, Vaccination.	Population density, Unemployment rate, Median income, Republican, Uninsured, Age 65+, Non-white, Vaccination.
Wave-4	Population density, Unemployment rate, Median income, Republican, Uninsured, Age 65+, Non-white, High School education, Vaccination.	Unemployment Rate, Republican, Uninsured, Age 65+, Non-white.

ties, there is a clear positive association between age and COVID-19 fatality, and predominantly negative association between age and case rates (Figures 2 and 3). A closer examination of the maps shows some interesting local trends and patterns. For example, there is a reversal of association between older populations and fatality along the south-eastern states, including Florida, from positive (Waves 1 through 3) to negative (Wave-4) associations. This may be related to the fact that Wave-4 covered only winter months (December to January), and older people may engage in outdoor activities more actively in warmer regions of the country.

In terms of case rates, while negative associations between age and COVID-19 infection dominated the country during Wave-1, there was a positive relationship in much of the Northeast, including New York City (Figure 2). Much of the early attempts to control COVID-19 were focused on disease outbreaks among elderly communities, particularly in nursing homes. New York City, for

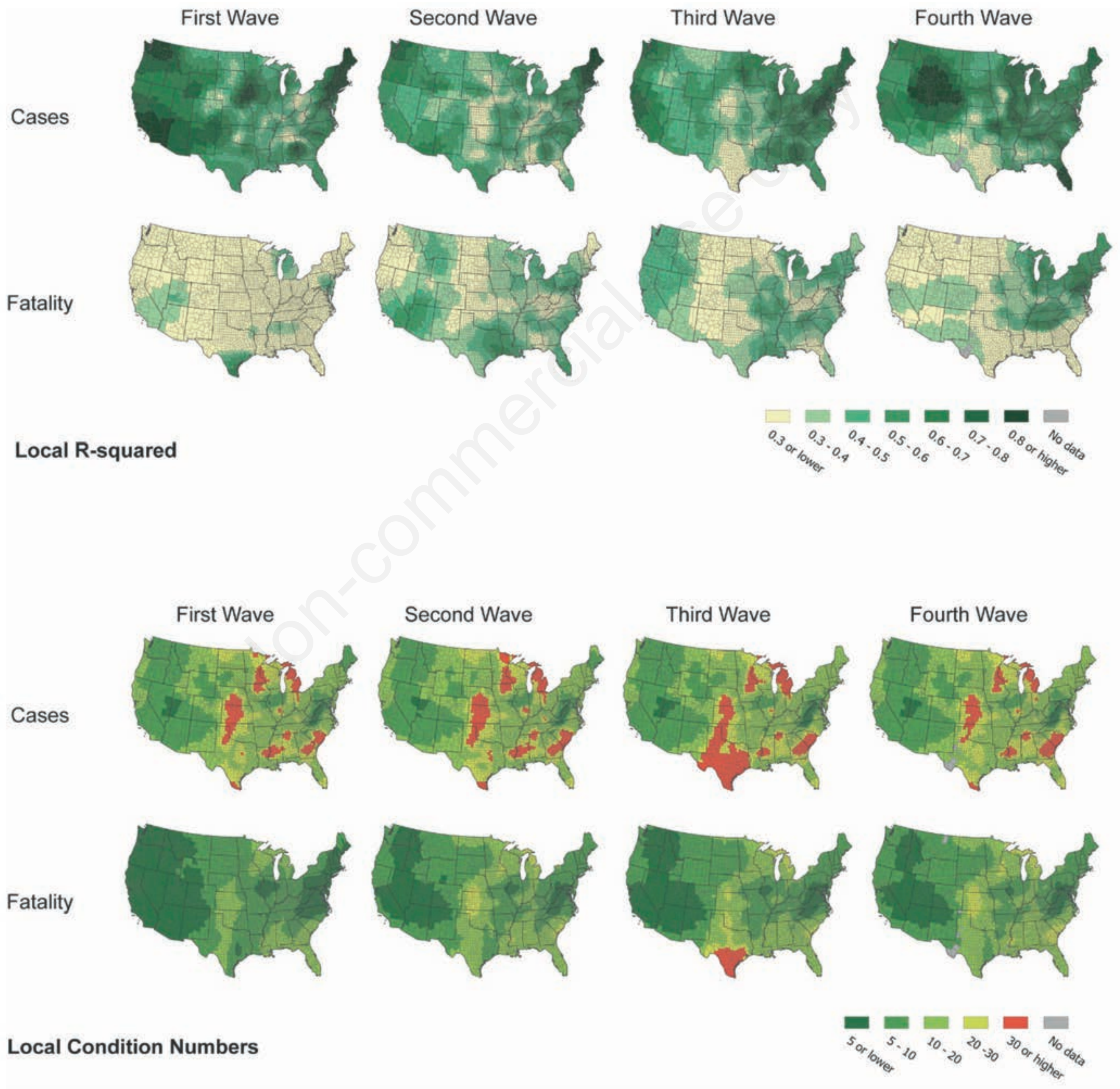

Figure 1. Geographically weighted regression model results – cases and fatalities by wave.

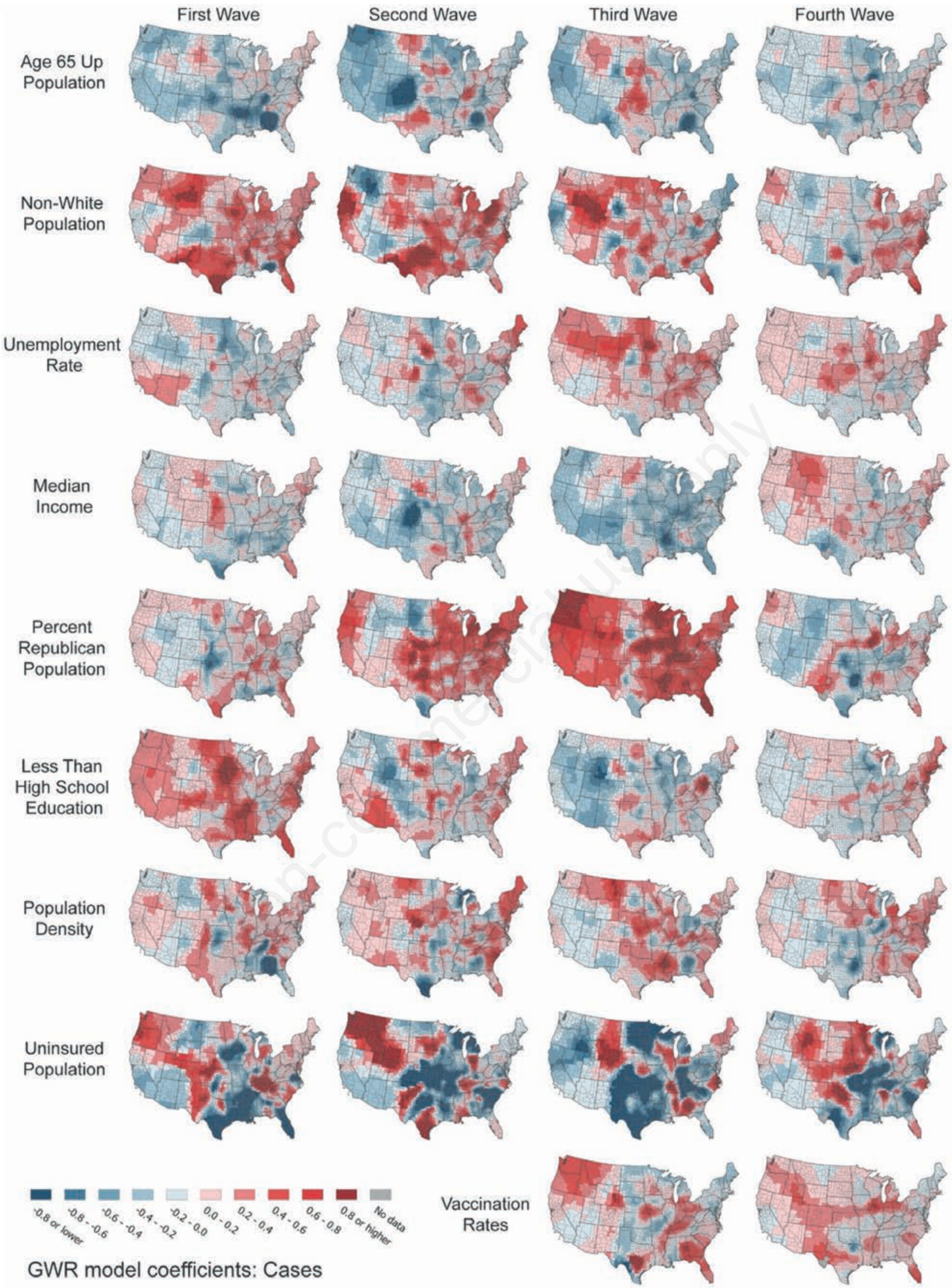

Figure 2. Weighted regression model coefficients by SDH – cases.

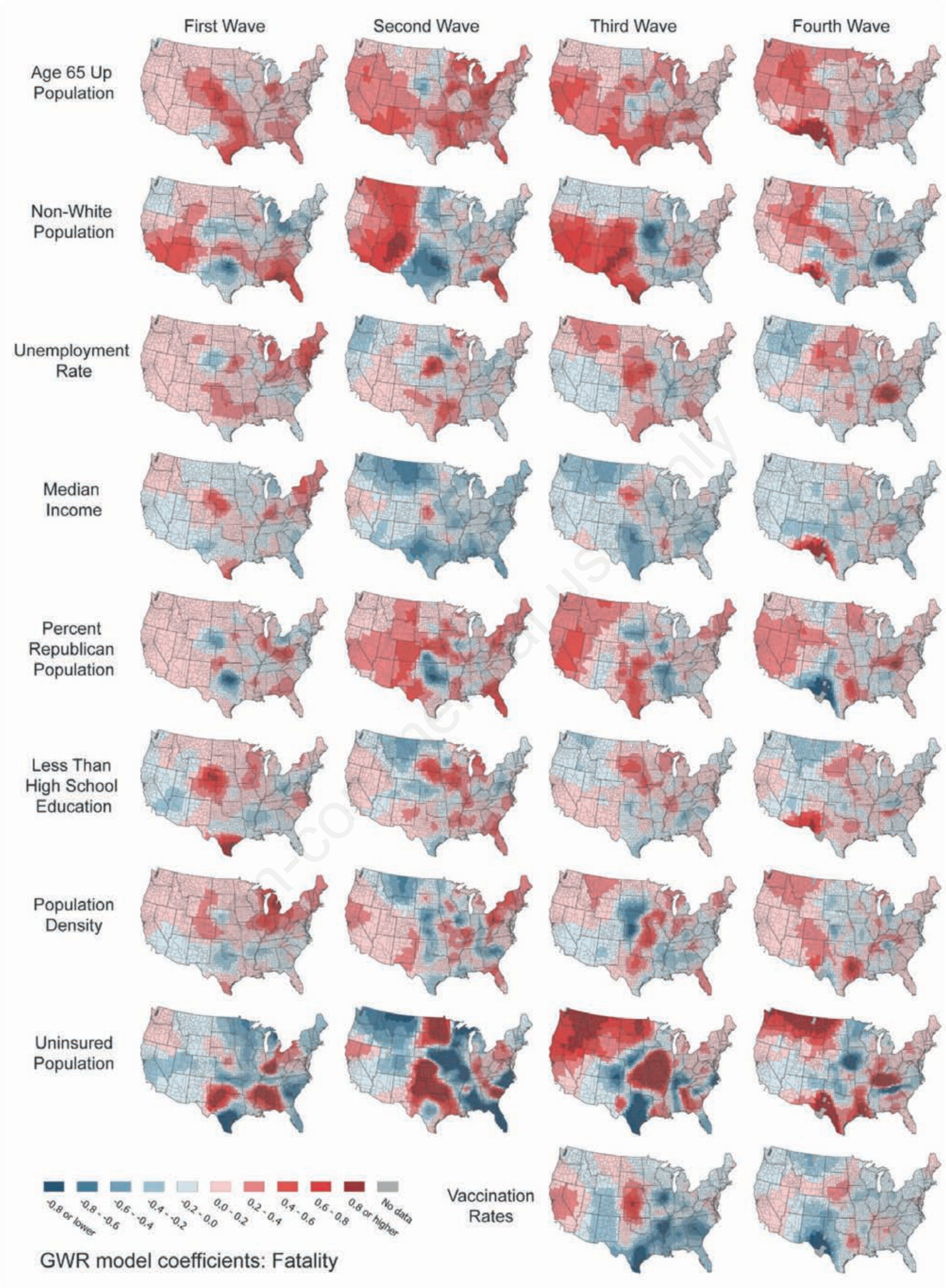

Figure 3. Weighted regression model coefficients by SDH – fatalities.

example, struggled to protect residents of its nursing homes, as cases skyrocketed under Governor Cuomo's nursing home policy (Ruiz, 2020). These "local stories" may be reflected in the Wave-1 results, as there was a positive relationship between age and COVID-19 infection in north-eastern counties, including those belonging to New York City (Figure 2). On the other hand, areas around the Florida Panhandle showed a strong and consistent negative relationship throughout the first three waves. Such "outliers" offer baseline information and spark productive questions for further inquiries and local investigations.

Ethnicity

Many studies have analysed the relationship between non-white population and COVID-19 infection. In the US state of California, Hsu and Hayes-Bautista (2021) found that case rates in individuals across all age groups were 1.5 – 6 times higher in non-white populations compared to the white non-Hispanic population. In major US metropolitan areas, counties with high non-white populations had COVID-19 infection rates approximately eight times higher than urban counties with majority white populations (Adhikari *et al.*, 2020). Our findings showed a strong positive association between non-white population and infection in large areas of the country, especially during Wave-1 (Figure 2). However, we wish point out that there are areas with a negative relationship between COVID-19 infection and non-white ethnicity (*e.g.*, the Pacific Northwest during Wave-2, and northern California during Wave-3); the positive association becomes generally weaker over time and a strong positive association is not necessarily prominent in urban counties (Figure 2). These results show that although existing research has highlighted significant associations between non-white population and COVID-19 infection, these associations continue to change across both time and space, revealing some of the immense complexity of the pandemic and the mitigation policies instituted.

In terms of COVID-19-related CFR, Golestaneh *et al.* (2020), explain that African American/Black populations have experienced a disproportionate impact of COVID-19 disease severity potentially because of disparities in co-morbidities, such as asthma, diabetes, hypertension, and obesity in addition to unequal access to healthcare services. In a similar vein, Paul *et al.* (2021) uncovered a strong positive relationship between non-white population and COVID-19 mortality at the national scale. Our results nevertheless indicate strong regional patterns in this assumed association. Most broadly in this connection, a strong positive association appears to hold in the western United States, including the coastal states, the Rocky Mountain Region, and much of the Southwest, though Florida was an exception to the pattern during the first two waves (Figure 3). Few previous studies addressed the potential presence of such regional contrasts in terms of ethnicity and COVID-19 infection. Furthermore, this positive association increased as the pandemic evolved and spread to more counties in the western US. The fact that many of these western US counties (California in particular) showed a negative association between COVID-19 infection and the non-white population, but a positive one between CFR and ethnicity, is a cause for alarm, as non-white populations may be taking measures to protect themselves from the infection, but once infected experience more severe symptoms and a disproportionately high death burden.

Economic status

Various previous studies examined associations between COVID-19 infections and personal economic conditions such as employment status (Fielding-Miller 2020), employer-based health insurance (Dragano *et al.*, 2021) and income (Abedi *et al.*, 2021; Hawkins, Charles, and Mehaffey, 2020; Quan *et al.*, 2021; Whittle and Diaz-Artiles 2020). Broadly, our results showed pre-pandemic local unemployment to be positively associated with a higher number of infections and CFR with higher income levels negatively associated these variables, yet distinct spatial and temporal variations were also visible (Figures 2 and 3). For example, Wave-1 revealed a positive association between COVID-19 CFR and unemployment rate in most areas of the country (Figure 3), while this pattern reversed when the fourth wave struck, with the Northeast showing a negative association between unemployment rate and COVID-19 CFR (Figure 3). Compared to the other waves, most counties were negatively associated with the unemployment rate during Wave-4, making the pre-pandemic unemployment rate a less clear SDH predictor of CFR as the pandemic progressed.

Looking at income levels, during Wave-2 and Wave-3, higher median income was related to higher CFR associated with COVID-19 in southern California, the Rocky Mountain States, the Southwest, and parts of the southern US (Figure 3). However, there was a geographic variation of these trends, as in earlier waves, with a negative association between increased median income and CFR observed along the northern and southern border states (Figure 3). Another interesting insight can be gained by comparing case and CFR within the same region over time. For example, in much of the Northeast, New York City included, there was a strong positive association between increasing median income and an increasing number of cases, across all the pandemic waves (Figure 2). This upholds the finding that increased median income was associated with increased risk of COVID-19 infection (Abedi *et al.,* 2021). In Wave-1, this positive association between increased median income and increased cases was also seen with respect to CFR, meaning that wealthier populations were both more likely not only to become infected by COVID-19, but also to become severely infected. However, in later waves, our analysis revealed a reversal of this trend, as wealthier populations were still more likely to be infected with COVID-19, but increased income became associated with decreased CFR (Figure 3). This may be indicative of the ability of wealthy populations to access emerging COVID-19 treatment, thus decreasing likelihood for severe, fatal infection.

Political orientation

It was reported in October 2021 that only 61% of Republicans had received one dose of a COVID-19 vaccine, compared to 90% of Democrats (Kaiser Family Foundation, 2021). For the Delta variant wave (Wave-3), vaccination was clearly effective at reducing the number of COVID-19 infections (80% - 88%) compared to unvaccinated individuals (Bernal *et al.*, 2021). When vaccines were available (during Wave-3 and Wave-4), there was a significant positive association between the percent republican population and COVID-19 infection in much of the Midwest, the Southern States, and the Pacific Northwest (Figure 2). During Wave-1, when vaccination was not an option and there was less polarizing rhetoric surrounding COVID-19, there was a negative association between the percent republican population and the number of cases in parts of the Midwest and the Southern States. Therefore, the increasingly positive relationship between percent republican population and COVID-19 cases in these areas may be due to the decreased likelihood of Republican populations to complete a full COVID-19 vaccination series or adhere to COVID-19 mitigation guidelines. Albrecht (2022) reported that people living in counties with a high percentage of "Trump-voters" were more likely to die from COVID-19. Our findings do indicate a largely positive relationship between COVID-19 fatality rate and the republican population, with some regional exceptions, especially during Wave-2 and Wave-3 (Figure 3).

Education level

Existing literature highlights a strong connection between education attainment and lower COVID-19 mortality (Albrecht, 2022). Shmueli *et al.* (2020), for example, reason that low education level is correlated with low COVID-19 vaccine uptake. Our Wave-1 data confirm a generally positive relationship between COVID-19 case rates and low education level in central and northeastern US, along with parts of North Carolina and Virginia (Figure 2). Importantly, this is at the time when vaccines were not yet available, which hints at the possibility of also other factors linking low education levels with high infection rates. Moreover, as the pandemic evolved (e.g., during Wave-2), this relationship held less strongly and even began to reverse in some areas of the country. Reaching Wave-4, many of the areas with a previous, positive association, now either had a weak positive, sometimes negative association between individuals with less than high school education and COVID-19 infection (Figure 2). These findings tell us to pay particular attention when declaring the relationship between socio-demographic attributes and disease outcomes in such a rapidly changing pandemic setting as that of COVID-19.

Vaccination

A completed COVID-19 vaccination series is correlated with decreased likelihood for infection (Bernal *et al.*, 2021). With regard to Wave-3, our findings support existing research, particularly in the Northeast (where vaccination rates were high), the Southwest, and the Central States. However, the results for the Gulf States during Wave-3 and the Rocky Mountain region during Wave-4, show areas where vaccination had a surprising positive association with the number of COVID-19 infections (Figure 2). However, these areas had a negative association between vaccination and CFR due to COVID-19 highlighting the vaccine's efficacy in preventing severe infection, but this might also have been the result of better prevention of infection in these areas (Figure 3).

Other social factors

Literature on the early part of the pandemic often pointed out that areas with higher population densities had higher infection rates (Sy *et al.*, 2020). In New York City, for example, Whittle and Diaz-Artiles (2020) found that increasing the population density by 10,000 people per km^2 was associated with a 2.4% increase in the COVID-19 rate. Our results confirm this result by finding a strong positive association between the number of COVID-19 cases and the population density in New York City and much of the Northeast during Wave-1 (Figure 2). In addition, existing research has highlighted the correlation between lack of health insurance on the one hand and increased disease burdens, infections, and deaths on the other (Grey II *et al.*, 2020; Garber 2021). Our study results agree as we found a positive association between lack of insurance and the number of infections in the Northeast and Rocky Mountain states during both Wave-1 and Wave-2. However, in these same regions, the relationship between uninsured and CFR was negative during Wave-2 (Figure 3). Furthermore, by Wave-3, most counties had a negative relationship between uninsured and COVID-19 infection (Figure 2). These findings are surprising, as they point to a reversal in the previous trends reported by existing research.

Our study provides some context with respect to the spatial variation with COVID-19 disease burden but did not tease out all the nuances in the spatial relationships between SDHs and the COVID-19 pandemic. GWR can reveal spatial patterns where diverse populations have felt a disproportionate impact of disease occurrence and severity, in turn leading to more accurate and specific public health policies that can target the most severely impacted populations as they change with time. While the GWR models used in this study significantly improved upon the OLS models, future studies could aim to replicate these results using MGWR, which has seen improved accuracy in COVID-19 studies, compared to standard GWR (Mollalo, Rivera & Vahabi, 2021).

Conclusions

This analysis was focused on unearthing the broad spatial patterns, considering not only existing literature relating SDH with cases or fatality rate, but also the cultural and political context surrounding COVID-19 at local levels. From the onset of the pandemic, the on-going spread of COVID-19 was discussed relative to its various SDHs. However, as the pandemic evolved, it became apparent that there were regional and local variations in the way that COVID-19 affected diverse populations. To understand these variations, applying GWR showed an immense geographic variation for each variable, across each disease wave and between the number of cases and the CFR. Further research may be conducted to understand why certain counties experienced a reversal of their GWR coefficients between waves, perhaps incorporating changes in the regional political landscape or changes in the values of independent variables not included in the present study. Future studies may utilize GWR and/or MGWR to analyse the impact of emerging policy and disease response trends. For example, these approaches could be used to see how the rollout of the "under-5" vaccination campaign affects the severity and infectivity for this younger age group, or how the removal of vaccine mandates can impact infection rates after the *Omicron* wave. Considering shifting pandemic response and policies across various geo-political scales, regression analyses can continue to provide information on the changing spatiotemporal dimensions of COVID-19 infection and severity as related to SDH.

References

Abedi V, Olulana O, Avula V, Chaudhary D, Khan A, Shahjouei S, Li J, Zand R, 2021. Racial, Economic, and Health Inequity and COVID-19 Infection in the United States. J Racial Ethn Health Disparities 8:732-42.

Adhikari S, Pantaleo N, Feldman J, Ogedegbe O, Thorpe L, Troxel A, 2020. Assessment of Community-Level Disparities in Coronavirus Disease 2019 (COVID-19) Infections and Deaths in Large US Metropolitan Areas. JAMA Netw Open 3:e2016938.

Albrecht D, 2022. Vaccination, politics, and COVID-19 impacts. BMC Public Health 22:317.

Bernal J, Andrews N, Gower C, Phil D, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Debrera G, Myers R, Campbell C, Amirthalingam G, Edmunds M, Zambon M, Brown

K, Hopkins S, Chand M, Ramsay M, 2021. Effectiveness of COVID-19 vaccines against the B.1.617.2 (Delta) Variant. N Engl J Med 7:585-94.

Brunsdon C, Charlton M, Harris P, 2012. Living with collinearity in local regression models. In Proceedings of the Tenth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 6. Available from: https://www.geos.ed.ac.uk/~gisteac/proceedingsonline/GIS-RUK2012/Papers/presentation-9.pdf

Centers for Disease Control, 2021. About COVID-19. Last updated January 24. Available form: https://www.cdc.gov/ coronavirus/2019-ncov/your-health/about-covid-19.html

Chen V, Park K, Sun F, Yang T, 2021. Assessing COVID-19 risk with temporal indices and geographically weighted ordinal logistic regression in US counties. PLoS One 17. DOI: https://doi.org/10.1371/journal.pone.0265673

Dragano N, Rupprecht C, Dortmann O, Schieder M, Wahrendorf M, 2020. Higher risk of COVID-19 hospitalization for unemployed: an analysis of 1,298,416 health insured individuals in Germany. medRxiv https://doi.org/10.1101/2020.06.17.20133918

Fielding-Miller R, Sundaram M, Brouwer K, 2020. Social determinants of COVID-19 mortality at the country level. PLoS One, 15. DOI: https://doi.org/10.1371/journal.pone.0240151

Fotheringham, A S, Brunsdon C, Charlton M, 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester, West Sussex: Wiley.

Garber J. 2021. One in three COVID-19 deaths may be linked to health insurance gaps. Lown Institute. March 22, 2021. Available from: https://lowninstitute.org/one-in-three-covid-19-deaths-linked-to-gaps-in-health-insurance-coverage/

Golestaneh L, Neugarten J, Fisher M, Billett H, Reyes Gil M, Johns T, Yunes M, Mokrzycki M, Coco M, Norris K, Perez H, Scott S, Kim R, Bellin E, 2020. The association of race and COVID-19 mortality. E Clin Med 25:100455.

Gray II D, Anyane-Yeboa A, Balzora S, Issaka R, May F, 2020. COVID-19 and the other pandemic: populations made vulnerable by systemic inequity. Nat Rev Gastroenterol Hepatol 17:520-22.

Hawkins R, Charles E, Mehaffey J, 2020. Socio-economic status and COVID-19 -related cases and fatalities. Public Health 189:129-34.

Hooper M, Napoles A, Perez-Stable E, 2020. COVID-19 and Racial/Ethnic Disparities. JAMA 323:2466 67.

Hsu P, Hayes-Bautista D, 2021. The epidemiology of diversity: COVID-19 case rate patterns in California. J Immigr Minor Health 23:857-62.

Islam N, Lacey B, Shabnam S, Erzurumluoglu A, Dambha-Miller H, Chowell G, Kawachi I, Marmot M, 2021. Social inequality and the syndemic of chronic disease and COVID-19: County-level analysis in the USA. J Epidemiol Community Health 75:496-500.

Kaiser Family Foundation. 2021. KFF COVID-19 Vaccine Monitor: October 2021. Last updated October 2021. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-october-2021/

Mandavilli A. The coronavirus has infected more than half of Americans, the C.D.C. reports. New York Times, April 27, 2022. Available from: https://www.nytimes.com/2022/04/26 /health/coronavirus-antibodies-americans-cdc.html_

MIT, 2020. Election Data and Science Lab; U.S. President 1976-2020; https://electionlab.mit.edu/data

Mollalo A, Vahedi B, Rivera K, 2020. GIS-based Spatial Modeling of COVID-19 Incidence Rate in the Continental United States. Sci Total Environ 728:138884.

Mollalo A, Tatar M, 2021. Spatial Modeling of COVID-19 Vaccine Hesitancy in the United States. Int J Environ Res Public Health 18:9488.

Mollalo A, Rivera K, Vahabi N, 2021. Spatial statistical analysis of pre-existing mortalities of 20 diseases with COVID-19 mortalities in the continental United States. Sustain Cities Soc 67:102738.

Mouliou D, Kotsiou O, Gourgoulianis K, 2021. Estimates of COVID-19 risk factorsamong social strata and predictors for a vulnerability to the infection. Int J Environ Res Public Health 18:8701.

Neumann-Podczaska A, Al-Saad S, Wieczorowska-Tobis K, 2020. COVID-19 - clinical picture in the elderly population: A qualitative systematic review. Aging Dis 11:998-1008.

New York Times. 2021. COVID-19 Tracker. Distributed by GitHub. Available from: https://github.com/nytimes/covid-19-data/blob/master/us-counties.csv

Paul A, Englert P, Varga M, 2021. Socio-economic disparities and COVID-19 in the USA. J Phys Complex 2:035017.

Quan D, Wong L, Shallal A, Madan R, Hamdan A, Ahdi H, Daneshvar A, Mahajan M, Nasereldin M, Van Harn M, Opara I, Zervos M, 2021. Impact of race and socioeconomic status on outcomes in patients hospitalized with COVID-19. J Gen Intern Med 36:1302-09.

Ruiz J. Cuomo says New York followed federal guidelines when sending coronavirus patients to nursing homes. CNN, May 24 2020. Available from: https://edition.cnn.com/2020/05/23/politics/cuomo-new-york-nursing-homes-coronavirus-patients/index.html

Shmueli L, 2021. Predicting intention to receive COVID-19 vaccine among the general population using the health belief model and the theory of planned behavior model. BMC Public Health 21:804.

Singu S, Acharya A, Challagundla K, Byrareddy, S, 2020. Impact of social determinants of health on the emerging COVID-19 pandemic in the United States. Front Public Health 8. DOI: https://doi.org/10.3389/fpubh.2020.00406

Sy K, White L, Nichols B, 2021. Population density and basic reproductive number of COVID-19 across United States counties. PLoS One, 16. DOI: https://doi.org/10.1371/ journal.pone.0249271

U.S. Census Bureau, 2021. 2020 Census; generated by Claire Bonzani; using data.census.gov; https://data.census.gov

U.S. Census Bureau, 2021. TIGER/Line Shapefiles; https://www.census.gov/cgi-bin/geo/shapefiles/index.php

USDA, 2021. Economic research service, unemployment and median household income and education tables; https://www.ers.usda.gov/data-products/county-level-data-sets/

Viswanath K, Bekalu M, Dhawan D, Pinnamaneni R, Lang J, and McLoud R, 2021. Individual and social determinants of COVID-19 vaccine uptake. BMC Public Health 21:818.

Whittle R, Diaz-Artiles A, 2020. An ecological study of socioeconomic predictors in detection of COVID-19 cases across neighborhoods in New York City. BMC Med 18:271.

WHO, 2022. Social determinants of health. Available from: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

Wu X, Zhang J, 2021. Exploration of spatial-temporal varying impacts on COVID-19 cumulative cases in texas using geographically weighted regression (GWR). Environ Sci Pollut Res Int 28:43732-46.