Science of the Total Environment 754 (2021) 142396

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Analyzing the spatial determinants of local Covid-19 transmission in the United States

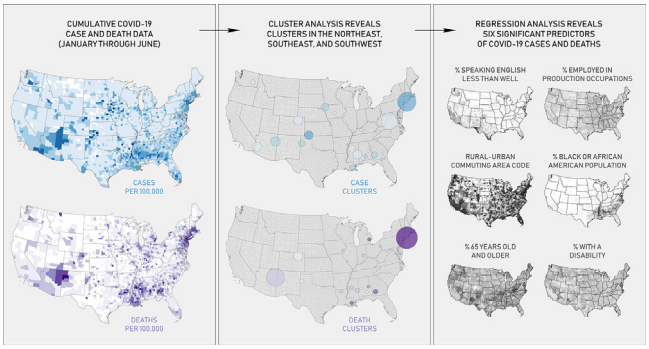
Lauren M. Andersen [a,*], Stella R. Harden [a], Margaret M. Sugg Ph.D. [a], Jennifer D. Runkle Ph.D. [b], Taylor E. Lundquist [a]

[a] Department of Geography and Planning, Appalachian State University, USA
[b] North Carolina Institute for Climate Studies, North Carolina State University, USA

HIGHLIGHTS

- Cluster analysis, three-stage regression used to assess COVID-19 determinants in U.S.
- COVID-19 hotspots predominantly in New England and the Southeast and Southwest
- Urban, Black population among significant predictors of COVID-19 cases and deaths
- Predictors significant at county, state levels despite behavior, policy differences

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 August 2020
Received in revised form 11 September 2020
Accepted 12 September 2020
Available online 18 September 2020

Editor: Jay Gan

Keywords:
COVID-19
Spatial determinants
Counties
Cluster analysis
Regression

ABSTRACT

The Coronavirus Disease 19 (COVID-19) has quickly spread across the United States (U.S.) since community transmission was first identified in January 2020. While a number of studies have examined individual-level risk factors for COVID-19, few studies have examined geographic hotspots and community drivers associated with spatial patterns in local transmission. The objective of the study is to understand the spatial determinants of the pandemic in counties across the U.S. by comparing socioeconomic variables to case and death data from January 22nd to June 30th 2020. A cluster analysis was performed to examine areas of high-risk, followed by a three-stage regression to examine contextual factors associated with elevated risk patterns for morbidity and mortality. The factors associated with community-level vulnerability included age, disability, language, race, occupation, and urban status. We recommend that cluster detection and spatial analysis be included in population-based surveillance strategies to better inform early case detection and prioritize healthcare resources.

Published by Elsevier B.V.

1. Background

Coronavirus Disease 19 (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), was first identified in December 2019 in the Wuhan prefecture, a rail and aviation hub in the Hubei Province of China. The outbreak was linked to a wet market in Wuhan (Mackenzie and Smith, 2020). The virus is believed to

* Corresponding author at: Department of Geography and Planning, Appalachian State University, P.O. Box 32066, Boone, NC 28608, USA.
E-mail address: andersenlm@appstate.edu (L.M. Andersen).

https://doi.org/10.1016/j.scitotenv.2020.142396
0048-9697/Published by Elsevier B.V.

originate from bats (Zhou et al., 2020), similar to the 2002 Severe Acute Respiratory Syndrome (SARS) and 2012 Middle East Respiratory Syndrome (MERS) outbreaks (Wu and McGoogan, 2020). The commonly-reported symptoms of the virus include a persistent cough, fever, fatigue, and loss of smell or taste (Grant et al., 2020; Menni et al., 2020).

Humans spread the virus to one another through contact and droplet, airborne, and fomite transmissions (WHO, 2020a). Groups considered high-risk include older adults and those with underlying medical conditions, such as obesity, diabetes, and heart disease (CDC, 2020a). Following the report of the first case in Wuhan, the virus spread outward to surrounding provinces in China. The first international case of COVID-19 was reported in Thailand on January 13th 2020, imported by an individual who recently visited Wuhan (WHO, 2020b). Since January, the virus has rapidly spread to other countries with significant outbreaks occurring in countries across the globe, including Italy, Iran, Brazil, and the United States (U.S.).

The earliest COVID-19 case in the U.S. was reported in Snohomish County, Washington on January 19th. The individual had recently returned from Wuhan (Holshue et al., 2020). On January 24th, a second case was reported in Cook County, Illinois, also by an individual who had recently returned from Wuhan. The individual's spouse later became the first known transmission in the U.S. (Ghinai et al., 2020). The earliest known COVID-19 death occurred in Santa Clara County, California on February 6th (CSCC, 2020).

The virus continued to spread with outbreaks in urban areas and locations where people live, work, and recreate closely, such as nursing homes, meatpacking plants, and ships (Kamp and Wilde Matthews, 2020; CDC, 2020b; Smith, 2020; Werner, 2020). Despite stay-at-home orders, the U.S. has maintained the highest case and death counts worldwide (WHO, 2020c). As of June 30th, there were cumulatively 2,618,817 cases and 126,623 deaths associated with COVID-19 in the U.S. (USA Facts, 2020a).

COVID-19 studies utilizing spatial analysis are valuable tools for community response. To date, over 63 studies worldwide have examined COVID-19 using geographic information systems (GIS), with studies focusing on spatiotemporal analysis, data mining, environmental variables, and health and social geography. Predictive modeling or studies that evaluate place-based risk factors have frequently focused on climatic variables (Franch-Pardo et al., 2020). In the U.S., researchers have begun to examine the spatial patterns and underlying risk factors of COVID-19. Mollalo et al. (2020) found that place-based factors like median household income, income inequality, percentage of nurse practitioners, and percentage of Black female population explain significant variation in COVID-19 incidence. Others have shown strong evidence of the spatial effects of COVID-19 with county-level socioeconomic factors in neighboring counties influencing incidence in the U.S. (Baum and Henry, 2020).

This study extends previous work by using geospatial methods to examine cumulative cases and deaths and the underlying place-based factors that increase a population's risk to COVID-19. The objective of the study is to understand the spatial determinants of the pandemic in U.S. counties by comparing socioeconomic variables to COVID-19 case and death data. Unlike previous studies (e.g., Mollalo et al., 2020), we utilize a longer time series and incorporate advanced hierarchical modeling approaches to understand how place affects transmission. As the pandemic continues, our study provides a methodological framework for future geographic research for understanding how place-based factors drive COVID-19 across the U.S.

2. Methods

2.1. COVID-19 outcomes

Cumulative county-level case and death counts between January 22nd and June 30th, 2020 were obtained from USA Facts (2020b, 2020c). As noted by Chen and Krieger (2020), USA Facts maintains separate reporting for individual counties (boroughs) in New York City, which is in contrast to the aggregated dataset provided by Johns Hopkins University (Dong et al., 2020). To minimize count differences based on county size, case and death rates per 100,000 were calculated based on population data from the 2014–8 American Community Survey (ACS) (USCB, 2020a) (Fig. 1).

2.2. Variable selection

A list of 34 possible risk factors was compiled based on previous studies and the authors' knowledge of factors that contribute to vulnerability (Table 1). Following Snyder and Parks (2020); Mollalo et al. (2020); and Chin et al. (2020), variables regarding race, age, gender, income, poverty, employment, and health were obtained from the 2014–8 ACS (USCB, 2020a). To evaluate potential rural-urban differences, Rural Urban Commuting Area (RUCA) codes were obtained from the U.S. Department of Agriculture (2019). RUCA codes are measures of population density, urbanization, and daily commuting that range from 1 (urban) to 10 (rural).

COVID-19 data was not available for the counties in American Samoa, Guam, Puerto Rico, the North Mariana Islands, and the Virgin Islands. These counties were removed from the analysis, resulting in a total of 3142 counties included in this study. To examine potential contextual variables for counties with versus without cases and deaths, descriptive statistics were calculated (Supplementary Table 1).

2.3. Analysis

To begin, separate cluster analyses were conducted in SaTScan, a mapping software that uses Kulldorff's spatial scan statistic, to identify significant clusters of cumulative COVID-19 cases and deaths across the contiguous U.S. at the county level (Kulldorff, 2018). SaTScan employs moving circular or elliptical windows across an area to locate significant spatial clusters of elevated risk during a specified period. Geographic hotspots of elevated risk were identified using a discrete Poisson model. COVID-19 cases were assumed to be Poisson distributed, as has been done in previous spatial COVID-19 surveillance studies (Desjardins et al., 2020; Hohl et al., 2020a). The Poisson model produces more conservative *p*-values than the alternative Bernoulli model (Root et al., 2009). SaTScan's user-specified window size determines the amount of the population for each county considered at-risk. Following a sensitivity analysis from 10% to 50%, a circular window scanning for a maximum of 20% of the total at-risk population was selected to avoid large cluster sizes and identify localized clustering of COVID-19. Adopting the approach used by Desjardins et al. (2020), a map of relative risk (RR) for COVID-19 cases in each county was created to show the variability of risk within clusters (Desjardins et al., 2020; Hohl et al., 2020b).

To determine significant predictors, a three-stage regression approach was applied separately for COVID-19 cases and deaths. First, variables were selected based on stepwise regression. The dependent variables were transformed using $\log(Y + 1)$ to ensure normality. Variables selected from this stage were RUCA, Black, English, Production, Disabled, ICU Beds, Vehicle, Asian, Older, Healthcare, Diabetes, Unemployed, Resources, Smokers, and Obesity for cases and RUCA, Black, Male, NH Beds, Distance, Income, ICU Beds, Older, Vehicle, Hawaiian, Population, Occupants, and Disabled for deaths. Tolerance and variance inflation factors were below 2.6 and below the recommended cutoff of 5 or 10, signifying that variance was not inflated due to collinearity among the independent variables (Craney and Surles, 2002).

Variable selection was based on subset regression, a preferable method to other explanatory regression methods for variable selection. This method is computational complex, so a smaller subset of variables from the stepwise regression was used (Field et al., 2012). Final subset regression results were limited to 5 variables to reduce heteroscedasticity among the residuals and to focus the analysis on

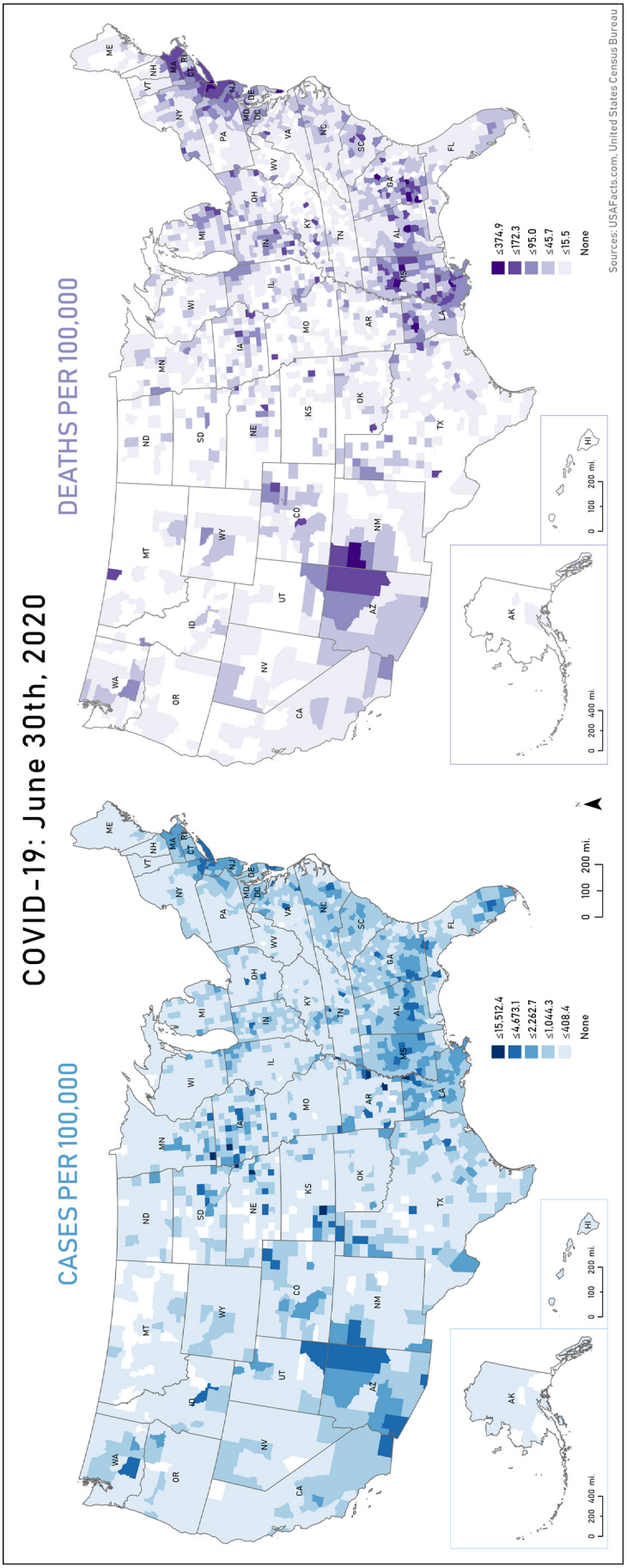


Fig. 1. Cumulative COVID-19 cases and deaths in the U.S. on June 30th 2020.

Table 1
The names, descriptions, and sources of the socioeconomic variables.

Name	Description	Source
Disabled	% civilian noninstitutionalized population, disabled	USCB (2020a)
English	% population ≥ 5 years, speak English "less than well"	
Unemployed	% civilian population ≥ 16 years, unemployed	
Income	Median household income for total households	
Insurance	% civilian noninstitutionalized population, no health insurance coverage	
Poverty	% families, income in past 12 months below poverty level	
Vehicle	% occupied housing units, no vehicles available	
Plumbing	% occupied housing units, lacking complete plumbing facilities	
Occupants	% occupied housing units, >1.01 occupants per room	
Population	Total population	
Density	Population per mi^2	
Male	% total population, male	
Female	% total population, female	
Older	% total population, ≥65 years	
Hispanic	% total population, Hispanic or Latino	
White	% total population, White alone	
Black	% total population, Black or African American alone	
Indian	% total population, American Indian and Alaska Native alone	
Asian	% total population, Asian alone	
Hawaiian	% total population, Native Hawaiian and Other Pacific Islander alone	
Healthcare	% civilian employed population ≥ 16 years, healthcare practitioners or technical occupations	
Service	% civilian employed population ≥ 16 years, service occupations	
Resources	% civilian employed population ≥ 16 years, natural resources/construction/maintenance occupations	
Production	% civilian employed population ≥ 16 years, production/transportation/material moving occupations	
Group	% total population, living in group quarters	
Vascular	2016–8 cardiovascular disease deaths per 100,000	CDC (2020c)
Diabetes	% total population, diabetic	
Obesity	% total population, obese	
Smokers	% persons ≥18 years current smokers	CHRRP (2020)
RUCA	2010 Rural-Urban Commuting Area (RUCA) codes	US Department of Agriculture (USDA), (2019)
Distance	Social distancing score	Unacast (2020)
ICU Beds	Hospital Intensive Care Unit (ICU) Beds per 100,000	Definitive Healthcare (2020)
NH Beds	Nursing Home (NH) Beds per 100,000	USDHS (2020)
UC Facilities	Urgent Care Facilities per 100,000	

the top predictors of COVID-19 cases and deaths. The final regression variables based on the results of the subset regression were RUCA, Black, English, Production, and ICU Beds for cases (Adjusted R^2: 0.4011, CP = 131.78, AIC = 7261.4, SBIC = 209.804) and RUCA, Black, ICU Beds, Older, and Disabled for deaths (Adjusted R^2: 0.3529, CP = 134.13, AIC = 7255.5, SBIC = 401.081).

Third, an Ordinary Least Square (OLS) regression was constructed to ensure proper model specification and significant predictors (Table 3). However, OLS regression resulted in spatial autocorrelation of the residuals for both deaths (Moran's I: 0.412, p-value<0.01) and cases (Moran's I: 0.352, p-value<0.001) violating the specification assumptions of OLS models that county-level cases and deaths and predictors are independent.

To address this violation, a spatial regression model that accounted for spatial autocorrelation in the residuals was applied using a queen's weight matrix. Spatial autocorrelation in OLS residuals may result from autocorrelation of the dependent variable (i.e., spatial lag) or autocorrelation in the error term (i.e, missing spatial autocorrelation predictors). The Lagrange Multiplier test and Robust Lagrange Multiplier (RLM) index were used to demonstrate the more appropriate spatial model. The RLM index indicated a better model fit with the spatial lag model compared to the spatial error, indicating that the spatial autocorrelation in the OLS's model residuals is likely the result of spatial autocorrelation for COVID-19 cases and deaths (e.g., the dependent variable) (Anselin, 2001). Due to lack of neighbors, seven counties were removed from the spatial lag model.

Lastly, Generalized Linear Mixed-Effect models were applied to determine the factors responsible for cumulative COVID-19 cases and deaths, separately. These models included a random effect term for each state to account for potential differences in behaviors and policies at the state level.

3. Results

The results of the cluster analysis indicated that areas of high relative risk (RR) for COVID-19 were predominantly in New England, the Southeast, and the Southwest (Fig. 2).

In total, 100 statistically significant (p-value<0.01) case clusters were identified and 78 statistically significant (p-value<0.01) death clusters were identified (Table 2). RR values greater than 1 indicate an increased risk within a cluster compared to outside of it (Chong et al., 2013). The ten most likely clusters had a RR value ranging between 1.87 and 22.6 for cases and 1.83 to 8.4 for deaths.

The OLS regression revealed that a higher percentage of Black individuals and fewer ICU beds were positively associated with COVID-19 cases and deaths. Additionally, high proportions of non-English speakers and persons working in production, transportation, and material moving occupations were positively associated with cumulative cases, while higher proportions of older and disabled individuals were positively associated with cumulative deaths (Supplementary Table 2). ICU Beds did not significantly impact the results. After excluding ICU Beds, RUCA 2 and 4–10, Black, English, and Production were significant predictors of COVID-19 cases and RUCA 5–10, Black, Older, and Disabled were significant predictors of COVID-19 deaths. For the following models, RUCA was reclassified from 10 values to 4 values with 1 representing urban areas, 2 representing micropolitan areas, 3 representing small towns, and 4 representing rural areas.

Results from the spatial lag models are summarized in Table 3. Findings showed that living in urban areas, micropolitan areas, or small towns and a higher proportion of Black individuals and non-English speakers were significant predictors of spatial patterns of COVID-19 cases. Notably, living in an urban area was associated with an increase in cases compared to a lower cumulative incidence in rural areas. For deaths, findings showed that a higher proportion of Black individuals and persons living with a disability were significant predictors. The spatial lag regression models explained 51% and 43% of the variance in cases and deaths, respectively.

Urban areas, micropolitan areas, and small towns were significant predictors across all models. This finding was also reflected by the cluster analysis. Of the 610 counties within case clusters, 265 (43%) were classified as urban areas, 178 (29%) as micropolitan areas, 129 (21%) as small towns, and 38 (6%) as rural areas. Of the 340 counties within death clusters, 178 (52%) were classified as urban areas, 95 (28%) as micropolitan areas, 57 (17%) as small towns, and 10 (3%) as rural areas.

Finally, Generalized Linear Mixed-Effect models were used to determine the predictors of cumulative COVID-19 cases and deaths at the state level. We observed that significant predictors of deaths and cases in the county-level spatial lag model remained significant at the state level despite variability in social distancing behaviors, stay-at-home mandates, and funding for assistance programs (e.g., Medicaid) (Table 4). The models explained 29% and 18% of the variance in cases and deaths, respectively.

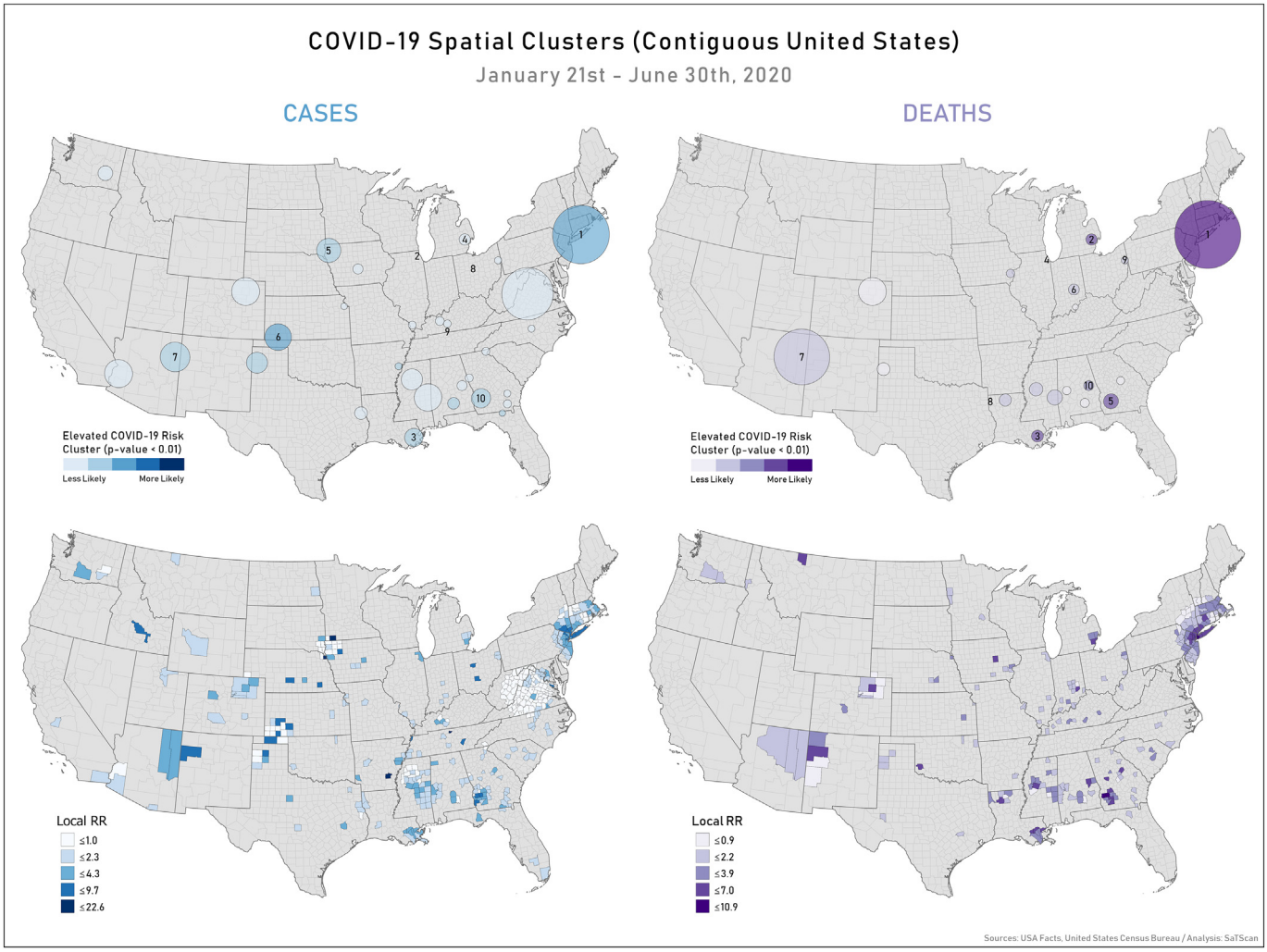


Fig. 2. The spatial clusters and relative risk per county for cases and deaths.

4. Discussion

In this study, cluster analysis and regression identified a subset of spatial determinants of COVID-19 at the county level across the U.S. from January 22nd to June 30th 2020 (Fig. 3). SaTScan analysis identified clustering of both cases and deaths, many of which could be attributed to outbreaks in cities, correctional facilities, meatpacking plants, and nursing homes. We observed significant geographic disparities in COVID-19 cases and deaths in urban areas and areas with higher percentages of Black individuals. Additionally, higher proportions of non-English speakers and persons working in production, transportation, and material moving occupations were significant determinants of cases and higher proportions of older and disabled individuals were significant determinants of deaths.

Table 2
COVID-19 spatial clusters from January 22nd though June 30th 2020 (RR = relative risk).

ID	Cases		Deaths	
	State (county)	RR	State (county)	RR
1	CT, NH, NJ, NY, MA, PA, RI, VT (*Many*)	5.15	CT, DE, NH, NJ, NY, MA, PA, RI, VT (*Many*)	8.35
2	IL (Cook)	2.46	MI (*Many*)	3.74
3	LA (*Many*)	3.42	LA (*Many*)	4.03
4	MI (*Many*)	1.87	IL (Cook)	2.09
5	IA, NE, SD, MN (*Many*)	2.90	GA (*Many*)	5.02
6	KS, OK, TX (*Many*)	4.69	IN (*Many*)	1.83
7	NM, AZ (*Many*)	3.76	AZ, NM, UT, CO (*Many*)	2.22
8	OH (Marion)	7.13	LA (Caddo)	2.37
9	TN (Trousdale)	22.59	OH, PA (*Many*)	2.09
10	GA, AL (*Many*)	3.05	AL (*Many*)	3.46

The cluster analysis identified likely clusters across the contiguous U.S. between January 22nd and June 30th 2020. The highest-risk case cluster and death cluster spanned multiple states in the Northeast, due in part to the COVID-19 outbreak in New York City (Watkins and Rashbaum, 2020). This finding corresponds to a primary cluster of cumulative COVID-19 cases also found in the Northeast by another published study between January 22nd and April 27th 2020 (Hohl et al., 2020a). The clusters of cases and deaths near Arizona was driven by the outbreak in Navajo Nation (Chappell, 2020).

Several of the highly-likely case clusters were associated with outbreaks in high-density locales, such as correctional facilities and meatpacking plants. For example, COVID-19 outbreaks occurred at the Trousdale-Turner Correctional Facility in Trousdale County, Tennessee and the Marion Correctional Institution in Marion County, Ohio (Hineman, 2020; Candisky, 2020). The fifth case cluster was due to an outbreak at the Smithfield Foods plant in Sioux Falls, South Dakota (AP, 2020) and the sixth case cluster was driven by outbreaks at several meatpacking plants in southwestern Kansas (Boyer, 2020).

Other high-risk clusters appeared near Albany, Georgia; Chicago, Illinois; Detroit, Michigan; and New Orleans, Louisiana, cities with higher poverty rates, higher percentages of Black, Indigenous, and People of Color (BIPOC), and below-average life expectancies compared to the

Table 3
The results of the spatial lag model.

Predictors	Cases				Deaths			
	Estimate	Std error	Z value	P value	Estimate	Std error	Z value	P value
RUCA [urban area]	1.62	0.71	2.26	0.024*	1.34	0.82	1.63	0.104
RUCA [micropolitan area]	1.55	0.71	2.17	0.030*	1.09	0.82	1.32	0.187
RUCA [small town]	1.40	0.71	1.97	0.050*	0.90	0.82	1.09	0.276
RUCA [rural area]	0.55	0.72	0.77	0.443	0.38	0.82	0.46	0.649
Black	0.02	0.00	12.38	<2e-16***	0.02	0.00	12.53	<2e-16***
Production	0.002	0.00	0.60	0.549	–	–	–	–
English	0.05	0.00	12.80	<2e-16***	–	–	–	–
Older	–	–	–	–	−0.01	0.01	−1.93	0.054
Disabled	–	–	–	–	−0.02	0.01	−4.19	2.8e-5***
	Rho: 0.57, R^2: 0.51, AIC: 9190				Rho: 0.52, R^2: 0.43, AIC: 10004			

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

United States (USCB, 2020b; CDC, 2020d). These cities have high numbers of individuals with preexisting medical conditions that put them at a higher risk of contracting COVID-19 (Porterfield, 2020; Brooks, 2020; Reyes et al., 2020). Additionally, many of the death clusters, including those in Columbiana County, Ohio and Tallapoosa County, Alabama, were the result of outbreaks in nursing homes (Johnson, 2020; Williams, 2020).

As COVID-19 incidence continues to escalate, it's likely that the numbers of cases and deaths in rural counties will increase. Some recent studies suggest that this transition may already be occurring (Paul et al., 2020). This is problematic because individuals in rural locales are more likely to be older, lack health insurance, and have preexisting medical conditions that increase their risk (Peters, 2020; Razzaghi et al., 2020). Due to fewer ventilators, beds, and employees, medical facilities in rural counties are less equipped to handle surges in cases (Kaufman et al., 2020). A subsequent analysis comparing the hospital and reclassified RUCA data revealed that hospitals in counties classified as rural had an average of 0.55 ventilators and 3.39 ICU beds per hospital, while counties classified as urban, micropolitan, and small town had an average of 3.66 ventilators and 14.42 ICU beds per hospital. Finally, some rural communities, such as Navajo Nation, dwell in homes that lack access to plumbing (Chappell, 2020). This makes basic preventative measures, such as handwashing, challenging.

The percentage of Black individuals was a significant determinant for both cases and deaths. This finding parallels other COVID-19 analyses documenting the disproportionate death rate among Black individuals (e.g., Thebault et al., 2020; Dyer, 2020). This can be attributed to systemic racism and the influence of adverse social determinants of health, including high poverty, high crime, and limited access to healthy foods and quality healthcare, in many urban communities where more BIPOC live (Yancy, 2020; Singh et al., 2017). Black individuals are also more likely than their white counterparts to be in frontline jobs that comprise the 'essential' service sector (Gould and Wilson, 2020). Finally, Black individuals are more likely to become seriously ill during pandemics due to a higher prevalence of underlying conditions and lower access to medical care (Hamidi et al., 2020; Quinn and Kumar, 2020). These findings build on previous work highlighting how communities with a higher proportion of Black individuals also experience more COVID-19 cases and deaths.

Similarly, the percentage of the population speaking English "less than well," was a significant predictor of cumulative cases. The higher number of cases in these communities can be attributed to communication challenges relating to receiving public announcements, as well as accessing and understanding medical care (Galvin, 2020). English was highly correlated with Hispanic ($r = 0.83$, $p < 0.01$), indicating that communities with a larger Hispanic population may be more likely to contract COVID-19. The disproportionate impact on Hispanics has also been documented in prior analyses of race and COVID-19 (Oppel et al., 2020).

These results suggest that BIPOC are more at-risk of contracting, and potentially dying, from COVID-19. Studies have demonstrated the negative impact that both direct and indirect exposure to police brutality has on the mental health of Black individuals (Bor et al., 2018). This impact is important to emphasize given recent nationwide protests against the murder of George Floyd, which have occurred parallel to the COVID-19 outbreak in the U.S. A recent federal survey revealed that these events have exacted a disproportionate emotional and mental toll on Black and Asian Americans compared to White Americans (Fowers and Wan, 2020). Additional work is needed to explore the impacts of these events on the mental health of vulnerable individuals.

Of the variables related to occupation, the percentage of persons working in production, transportation, and material moving occupations was the only significant predictor of cases. This sector includes employees that frequently travel or interact with others in close quarters,

Table 4
The results of the mixed-effects models at the state level.

Predictors	Cases			Deaths		
	Estimates	CI	p	Estimates	CI	p
RUCA [urban area]	1.49	1.34–1.63	<0.001***	1.10	0.93–1.28	<0.001***
RUCA [micropolitan area]	1.24	1.10–1.38	<0.001***	0.72	0.56–0.89	<0.001***
RUCA [small town]	0.99	0.85–1.13	<0.001***	0.52	0.36–0.67	<0.001***
Black	0.38	0.32–0.43	<0.001***	0.44	0.38–0.50	<0.001***
Production	0.09	0.04–0.14	0.001**	–	–	–
English	0.42	0.37–0.46	<0.001***	–	–	–
Older	–	–	–	−0.07	−0.13 to −0.01	0.015*
Disabled	–	–	–	−0.14	−0.20 to −0.08	<0.001***
Marginal R^2/conditional R^2	0.289/0.458			0.184/0.421		

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

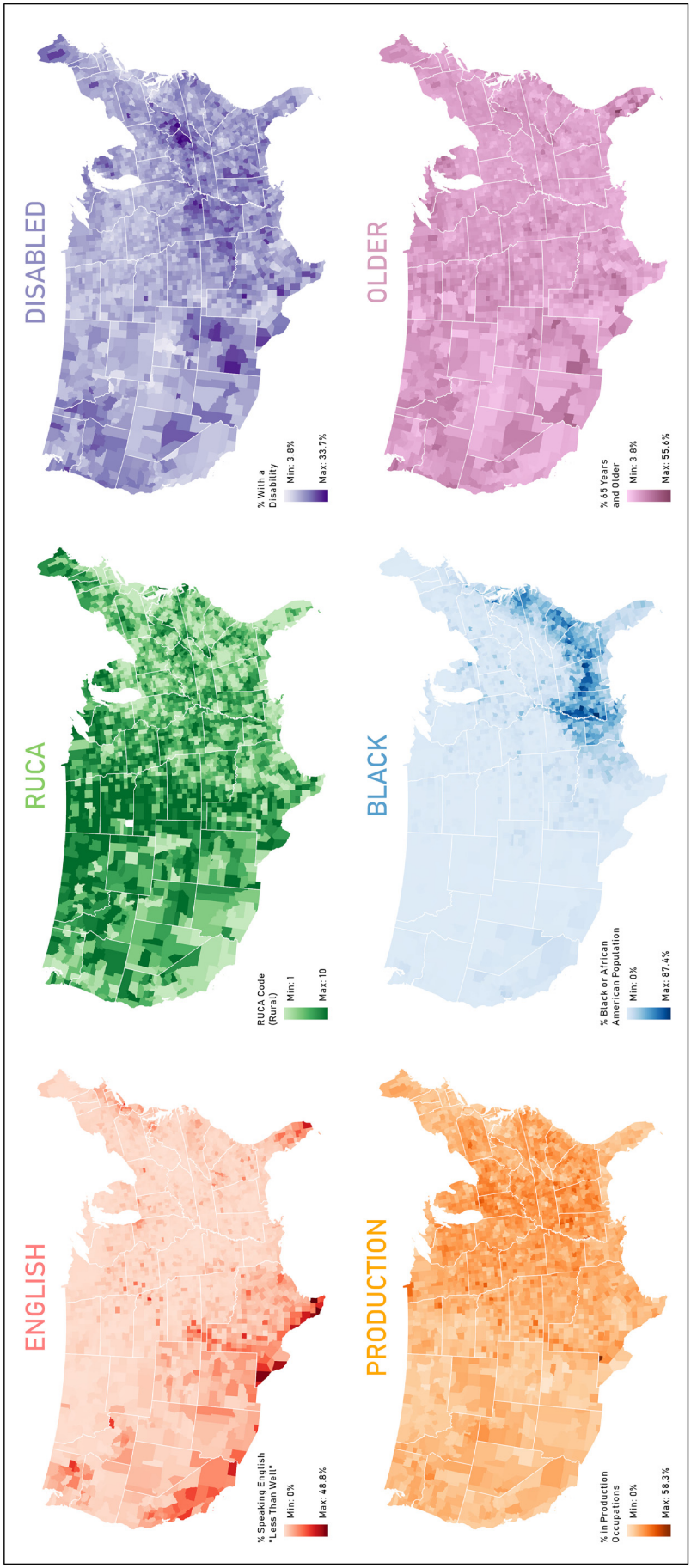


Fig. 3. The significant determinants of COVID-19.

including bus drivers, flight attendants, supermarket workers, and conveyor operators. Many of these occupations offer below-average salaries and lack paid sick leave, suggesting that employees may be unable to take time off work due to illness (Harrington, 2020).

Disability was a significant predictor for deaths. Similar to the older population, disabled individuals are considered to be high-risk for contracting COVID-19 because they may not be able to understand information, practice preventative measures, or communicate symptoms of illness (CDC, 2020a). People living with a disability may also be more likely to have a coexisting chronic condition and in many cases require in-home care which puts them at greater risk for infection. The regression found a negative relationship with COVID-19 deaths, which was unexpected and confirmed by examining Supplementary Table 1. As COVID-19 continues to spread geographically, this relationship should be re-evaluated at a smaller geographic scale (e.g., zip code) and for smaller geographic regions (e.g., the Southeast).

Age was a significant predictor for deaths with the percentage of individuals 65 years old and older predicting fewer deaths. Similar to disabled, this difference may be due to the location of early COVID-19 emergers. Older individuals are at a higher risk of contracting COVID-19 because they are more likely to have pre-existing medical conditions (CDC, 2020a). As revealed in the cluster analysis, many high death rates are due to outbreaks in nursing homes, which largely consist of older individuals. Additionally, many disabled individuals are also older, compounding vulnerability to COVID-19.

5. Strengths and limitations

There are notable limitations to the COVID-19 data. USA Facts confirms county-level data by referencing state and local agencies directly. Because the scale, timing, and style of reporting varies by county, the case and death data are subject to error. Additionally, USA Facts counts presumptive cases and confirmed cases, which is in line with reporting by the CDC (USA Facts, 2020d). Our study is also limited to the early period of the outbreak from January 22nd through June 30th, 2020, and the underlying spatial determinants of COVID-19 cases and deaths will likely change as the pandemic continues.

The main limitation of our study is the retrospective ecological study design, which limits the ability to determine causal relationships. Further, a prospective study measuring changes in morbidity and mortality over time, rather than cumulative cases and deaths, would contribute to the emerging findings on early cluster detection (i.e. Desjardins et al., 2020; Hohl et al., 2020a; Hohl et al., 2020b). However, our early analysis provides a framework for future research to identify regions with geographic disparities and the associated underlying causes of high COVID-19 cases and deaths. As the pandemic continues, we recommend future analysis that focuses on specific regions of the U.S. to fully understand the underlying risk factors and how they vary across geographic regions. We also recommend the use of a smaller spatial scale, which may avoid the potential for the modifiable areal unit problem and has been used in other COVID-19 studies to detect health disparities (Chen et al., 2020).

As testing increases, the number of confirmed cases is expected to increase. Future work should incorporate testing data into the analysis once these data are made publicly available. Additionally, future studies should analyze trends in cases and deaths by age, race, and gender at the individual level as this information becomes publicly available. Finally, the increasing availability of COVID-19 data at the zip code level provides opportunities for sub-county exploration of the determinants of COVID-19. The authors are particularly interested in using methods similar to Federgruen and Naha (2020) to explore determinants of COVID-19 across North Carolina zip codes. As COVID-19 cases and deaths continue to increase across the U.S., understanding the variables that contribute to incidence will be crucial for targeted interventions at the community level.

6. Conclusion

This study related social, economic, and demographic determinants to spatial variation in COVID-19 across the U.S. from January to June 2020. Findings revealed that urban areas and locations with a greater proportion of Black individuals were associated with higher numbers of cases and deaths. Future research should explore underlying risk factors for COVID-19 at a smaller geographic scale or the individual-level to confirm which environments and populations increase vulnerability. Ultimately, results from this study highlight locations where targeted public health interventions can be explored to mitigate outbreaks of COVID-19.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.142396.

CRediT authorship contribution statement

Lauren Andersen: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Stella Harden:** Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Margaret Sugg:** Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Supervision. **Jennifer Runkle:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision. **Taylor Lundquist:** Investigation, Writing - original draft, Writing - review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Appalachian State University Office of Research.

References

Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Basil Blackwell, Oxford, p. 310e30.

Associated Press (AP), 2020. South Dakota COVID-10 cases grow, additional death reported. https://apnews.com/79635961c8d3dbcf11bdd6cefac17248.

Baum, C.F., Henry, B., 2020. Socioeconomic factors influencing the spatial spread of COVID-19 in the United States. Boston College Working Papers in Economics 1009 https://ideas.repec.org/p/boc/bocoec/1009.html.

Bor, J., Venkataramani, A.S., Williams, D.R., Tsai, A.C., 2018. Police killings and their spillover effects on the mental health of Black Americans: a population-based, quasi-experimental study. Lancet 392 (10144), 302–310. https://doi.org/10.1016/S0140-6736(18)31130-9.

Boyer, C., 2020. Update: Coronavirus Clusters Grow Rapidly in Three Western Kansas Meatpacking Counties. High Plains Public Radio https://www.hppr.org/post/update-coronavirus-clusters-grow-rapidly-three-western-kansas-meatpacking-counties.

Brooks, B., 2020. Why Is New Orleans' Coronavirus Death Rate Twice New York's? Obesity is a Factor. Reuters https://www.reuters.com/article/us-health-coronavirus-new-orleans/why-is-new-orleans-coronavirus-death-rate-twice-new-yorks-obesity-is-a-factor-idUSKBN21K1B0.

Candisky, C., 2020. Coronavirus surges at Pickaway prison, now no. 2 hot spot in U.S. - behind Marion. Marion Star https://www.marionstar.com/story/news/local/2020/04/23/coronavirus-surges-pickaway-prison-now-no-2-u-s-behind-marion-ohio/3007507001.

Centers for Disease Control & Prevention (CDC), 2020a. People who are at increased risk for severe illness. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-increased-risk.html.

Centers for Disease Control & Prevention (CDC), 2020b. Meat and poultry processing workers and employers. https://www.cdc.gov/coronavirus/2019-ncov/community/organizations/meat-poultry-processing-workers-employers.html.

Centers for Disease Control & Prevention (CDC), 2020c. Interactive atlas of heart disease and stroke. https://nccd.cdc.gov/DHDSPAtlas/?state=County.

Centers for Disease Control & Prevention (CDC), 2020d. U.S. small-area life expectancy project. https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html.

Chappell, B., 2020. Coronavirus Cases Spike in Navajo Nation, Where Water Service Is Often Scarce. NPR https://www.npr.org/sections/coronavirus-live-updates/2020/03/26/822037719/coronavirus-cases-spike-in-navajo-nation-where-water-service-is-often-scarce.

Chen, J.T., Krieger, N., 2020. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US County vs. ZIP code analyses. Harvard Center for Population and Development Studies Working Paper Series 19(1) . https://www.hsph.harvard.edu/population-development/research/working-papers/harvard-pop-center-working-paper-series.

Chen, J.T., Waterman, P.D., Krieger, N., 2020. COVID-19 and the unequal surge in mortality rates in Massachusetts, by city/town and ZIP code measures of poverty, household crowding, race/ethnicity, and racialized economic segregation. Harvard Center for Population and Development Studies Working Paper Series 19(2) . https://www.hsph.harvard.edu/population-development/research/working-papers/harvard-pop-center-working-paper-series.

Chin, T., Khan, R., Li, R., Chen, J.T., Kriger, N., Buckee, C.O., Balsari, S., Kaing, M.V., 2020. U.S. county-level characteristics to inform equitable COVID-19 response. medRxiv 2020.04.08.20058248 https://doi.org/10.1101/2020.04.08.20058248.

Chong, S., Nelson, M., Byun, R., Eastwood, J., Jalaludin, B., 2013. Geospatial analyses to identify clusters of adverse antenatal factors for targeted interventions. Int. J. Health Geogr. 12 (46). https://doi.org/10.1186/1476-072X-12-46.

County Health Rankings and Roadmaps Program (CHRRP), 2020. Rankings data & documentation. https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation.

County of Santa Clara, California (CSCC), 2020. County of Santa Clara Identifies Three Additional Early COVID-19 Deaths. County of Santa Clara Emergency Operations Center https://www.sccgov.org/sites/covid19/Pages/press-release-04-21-20-early.aspx.

Craney, T.A., Surles, J.G., 2002. Model-dependent variance inflation factor cutoff values. Qual. Eng. 14 (3), 391–403. https://doi.org/10.1081/QEN-120001878.

Definitive Healthcare, 2020. USA hospital beds. ESRI COVID-19 GIS hub. https://coronavirus-resources.esri.com/datasets/1044bb19da8d4dbfb6a96eb1b4ebf629_0.

Desjardins, M.R., Hohl, A., Delmelle, E.M., 2020. Rapid Surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl. Geogr., 102202

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20 (5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

Dyer, O., 2020. COVID-19: Black people and other minorities are hardest hit in US. BMJ 369, m1483. https://doi.org/10.1136/bmj.m1483.

Federgruen, A., Naha, S.R., 2020. Variation in Covid-19 cases across New York City. medRxiv 2020.05.25.20112797 https://doi.org/10.1101/2020.05.25.20112797.

Field, A., Miles, J., Field, Z., 2012. Discovering Statistics Using R. Sage Publications.

Fowers, A., Wan, W., 2020. Depression and anxiety spiked among Black Americans after George Floyd's death. Wash. Post https://www.washingtonpost.com/health/2020/06/12/mental-health-george-floyd-census.

Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19: a review. Sci. Total Environ. 739, 140033. https://doi.org/10.1016/j.scitotenv.2020.140033.

Galvin, G., 2020. Language access issues a Barrier during COVID-19. US News & World Report https://www.usnews.com/news/healthiest-communities/articles/2020-04-16/language-access-problems-a-barrier-during-covid-19-pandemic.

Ghinai, I., McPherson, T.D., Hunter, J.C., Kirking, H.L., Christiansen, D., Joshi, K., Rubin, R., Morales-Estrada, S., Black, S.R., Pacilli, M., Fricchione, M.J., Chugh, R.K., Walblay, K.A., Ahmed, N.S., Stoecker, W.C., Hasan, N.F., Burdsall, D.P., Reese, H.E., Wallace, M., Wang, C., Moeller, D., Korpics, J., Novosad, S.A., Benowitz, I., Jacobs, M.W., Dasari, V.S., Patel, M.T., Kauerauf, J., Charles, E.M., Ezike, N.O., Chu, V., Midgley, C.M., Rolfes, M.A., Gerber, S.I., Lu, X., Lindstrom, S., Verani, J.R., Layden, J.E., 2020. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. Lancet 395 (10230), 1137–1144. https://doi.org/10.1016/S0140-6736(20)30607-3.

Gould, E., Wilson, V., 2020. Black Workers Face Two of the Most Lethal Preexisting Conditions for Coronavirus - Racism and Economic Inequality. Economic Policy Institute https://www.epi.org/publication/black-workers-covid.

Grant, M.C., Geoghegan, L., Arbyn, M., Mohammed, Z., McGuinness, L., Clarke, E.L., Wade, R.G., 2020. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): a systematic review and meta-analysis of 148 studies from 9 countries. (2020). PLoS One 15 (6), e0234765. https://doi.org/10.1371/journal.pone.0234765.

Hamidi, S., Sabouri, S., Ewing, R., 2020. Does density aggravate the COVID-19 pandemic? J. Am. Plan. Assoc. https://doi.org/10.1080/01944363.2020.1777891.

Harrington, J., 2020. Poor paid sick leave, proximity to people make these tough jobs in coronavirus crisis. USA Today https://www.usatoday.com/story/money/2020/04/18/american-jobs-with-the-biggest-sick-leave-problems-right-now/111535588.

Hineman, B., 2020. More than 1,800 prison inmates are infected with COVID-19 in Tennessee, TDOC Says. Nashville Tennessean https://www.tennessean.com/story/news/2020/05/11/tennessee-prison-inmates-coronavirus-trousdale-county-turner-correctional-facility/3113554001.

Hohl, A., Delmelle, E., Desjardins, M., 2020a. Rapid detection of COVID-19 clusters in the United States using a prospective space-time scan statistic: an update. SIGSPATIAL Special 12 (1), 27–33.

Hohl, A., Delmelle, E.M., Desjardins, M.R., Lan, Y., 2020b. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. Spat. Spatio-temporal Epidemiol. 34, 100354.

Holshue, M.L., DeBolt, C., Lindquist, S., Loft, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S.I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M.A., Weldon, W.C., Biggs, H.M., Uyeki, T.M., Pillai, S.K., 2020. First case of 2019 novel coronavirus in the United States. N. Engl. J. Med. 382 (10), 929–936. https://doi.org/10.1056/NEJMoa2001191.

Johnson, D., 2020. 54th COVID death recorded in Columbiana County. Salem News https://www.salemnews.net/news/local-news/2020/06/54th-covid%E2%80%88death-recorded-in-columbiana-county.

Kamp, J., Wilde Matthews, A., 2020. As U.S. nursing home deaths reach 50,000, states ease lockdowns. Wall Street J. https://www.wsj.com/articles/coronavirus-deaths-in-u-s-nursing-long-term-care-facilities-top-50-000-11592306919

Kaufman, B.G., Whitaker, R., Pink, G., Holmes, G.M., 2020. Half of rural residents at high risk of serious illness due to COVID-19, creating stress on rural hospitals. J. Rural. Health https://doi.org/10.1111/jrh.12481.

Kulldorff, M., 2018. SaTScan user guide. https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download.

Mackenzie, J.S., Smith, D.W., 2020. COVID-19 a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't. Microbiol. Australia, MA20013 https://doi.org/10.1071/MA20013.

Menni, C., Valdes, A.M., Freidin, M.B., Sudre, C.H., Nguyen, L.H., Drew, D.A., Ganesh, S., Varsavsky, T., Jorge Cardoso, M., El-Sayed Moustafa, J.S., Visconti, A., Hysi, P., Bowyer, R.C.E., Mangino, M., Falchi, M., Wold, J., Ourselin, S., Chan, A.T., Steves, C.J., Spector, T.D., 2020. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat. Med. https://doi.org/10.1038/s41591-020-0916-2.

Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. 728, 138884. https://doi.org/10.1016/j.scitotenv.2020.138884.

Oppel, R.A., Gebeloff, R., Rebecca Lai, K.K., Wright, W., Smith, M., 2020. The fullest look yet at the racial inequity of coronavirus. The New York Times https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html.

Paul, R., Arif, A.A., Adeyemi, O., Ghosh, S., Han, D., 2020. Progression of COVID-19 from urban to rural areas in the United States: a spatiotemporal analysis of prevalence rates. J. Rural. Health https://doi.org/10.1111/jrh.12486.

Peters, D.J., 2020. Community susceptibility and resiliency to COVID-19 across the rural-urban continuum in the United States. J. Rural. Health 36 (3), 446–456. https://doi.org/10.1111/jrh.12477.

Porterfield, C., 2020. Why Chicago, Detroit And New Orleans could become the next coronavirus 'hot spots'. Forbes https://www.forbes.com/sites/carlieporterfield/2020/03/27/why-chicago-detroit-and-new-orleans-could-become-the-next-coronavirus-hot-spots/#181c9f1c5df0.

Quinn, S.C., Kumar, S., 2020. Health inequalities and infectious disease epidemics: a challenge for global health security. Biosecur. Bioterrorism 12 (5), 263–273. https://doi.org/10.1089/bsp.2014.0032.

Razzaghi, H., Wang, Y., Lu, H., Marshall, K.E., Dowling, N.F., Paz-Bailey, G., Twentyman, E.R., Peacock, G., Greenlund, K.J., 2020. Prevalence of selected underlying medical conditions associated with increased risk for severe COVID-19 illness - United States, 2018. Morb. Mortal. Wkly Rep. 69, 945–950. https://doi.org/10.15585/mmwr.mm6929a1.

Reyes, C., Husain, N., Gutowski, C., St. Clair, S., Pratt, G., 2020. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. Chicago Tribune https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html.

Root, E.D., Meyer, R.E., Emch, M.E., 2009. Evidence of localized clustering of Hastrochises births in North Carolina, 1999-2004. Soc. Sci. Med. 68 (8), 1261–1267. https://doi.org/10.1016/j.socscimed.2009.01.034.

Singh, G.K., Daus, G.P., Allender, M., Ramey, C.T., Martin, E.K., Perry, C., De Los Reyes, A.A., Vedamuthu, I.P., 2017. Social determinants of health in the United States: addressing major health inequality trends for the nation, 1935-2016. Int. J. MCH AIDS 6 (2), 139–164. https://doi.org/10.21106/ijma.236.

Smith, R., 2020. Covid-19 hits cruise-ship crews hard. Wall Street J. https://www.wsj.com/articles/covid-19-hits-cruise-ship-crews-hard-11594486510

Snyder, B., Parks, V., 2020. Spatial variation in socio-ecological vulnerability to COVID-19 in the contiguous United States. SSRN https://doi.org/10.2139/ssrn.3587713.

Thebault, R., Ba Tran, A., Williams, V., 2020. The coronavirus is infecting and killing black americans at an alarmingly high rate. Wash. Post https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/?arc404=true.

Unacast, 2020. Unacast social distancing score (latest available). https://coronavirus-resources.esri.com/datasets/ab72fb3e9bf24d9594f0b942718bffeb_0.

US Census Bureau (USCB), 2020a. 2014–2018 American Community Survey 5-Year estimates. https://data.census.gov/cedsci.

US Census Bureau (USCB), 2020b. QuickFacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045219.

US Department of Agriculture (USDA), 2019. Rural-Urban Commuting Area Codes. USDA Economic Research Service https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.

US Department of Homeland Security (USDHS), 2020. Homeland infrastructure foundation-level data. https://hifld-geoplatform.opendata.arcgis.com.

USA Facts, 2020a. Coronavirus locations: COVID-19 map by county and state. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map.

USA Facts, 2020b. COVID-19 known cases dataset (July 1) [CSV]. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map.

USA Facts, 2020c. COVID-19 Deaths Dataset (July 1) [CSV]. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map.

USA Facts, 2020d. Detailed methodology and sources: COVID-19 data. https://usafacts.org/articles/detailed-methodology-covid-19-data.

Watkins, A., Rashbaum, W.K., 2020. How many people have actually died from coronavirus in New York? The New York Times https://www.nytimes.com/2020/04/10/nyregion/new-york-coronavirus-death-count.html.

Werner, B., 2020. 26 navy battle force ships have had COVID-19 cases. U.S. Naval Inst. News https://news.usni.org/2020/04/23/26-navy-battle-force-ships-have-had-covid-19-cases.

Williams, C., 2020. Many COVID-19 cases, deaths in Tallapoosa County attributed to long-term care facilities. The Outlook https://www.alexcityoutlook.com/news/many-covid-19-cases-deaths-in-tallapoosa-county-attributed-to-long-term-care-facilities/article_d155e288-9488-11ea-a348-1b1beaa656db.html.

World Health Organization (WHO), 2020a. Transmission of SARS-CoV-2: implications for infection prevention precautions. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

World Health Organization (WHO), 2020b. Novel coronavirus - Thailand (ex-China). https://www.who.int/csr/don/14-january-2020-novel-coronavirus-thailand-ex-china/en.

World Health Organization (WHO), 2020c. Coronavirus disease (COVID-2019) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

Wu, Z., McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China. JAMA 323 (13), 1239–1242. https://doi.org/10.1001/jama.2020.2648.

Yancy, C.W., 2020. COVID-19 and African Americans. JAMA 323 (19), 1891–1892. https://doi.org/10.1001/jama.2020.6548.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probably bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7.