The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships

Ibraheem M. Karaye, MD, DrPH, Jennifer A. Horney, PhD, MPH

Introduction: Because of their inability to access adequate medical care, transportation, and nutrition, socially vulnerable populations are at an increased risk of health challenges during disasters. This study estimates the association between case counts of COVID-19 infection and social vulnerability in the U.S., identifying counties at increased vulnerability to the pandemic.

Methods: Using Social Vulnerability Index and COVID-19 case count data, an ordinary least squares regression model was fitted to assess the global relationship between COVID-19 case counts and social vulnerability. Local relationships were assessed using a geographically weighted regression model, which is effective in exploring spatial nonstationarity.

Results: As of May 12, 2020, a total of 1,320,909 people had been diagnosed with COVID-19 in the U.S. Of the counties included in this study (91.5%, 2,844 of 3,108), the highest case count was recorded in Trousdale, Tennessee (16,525.22 per 100,000) and the lowest in Tehama, California (1.54 per 100,000). At the global level, overall Social Vulnerability Index ($e^b=1.65$, $p=0.03$) and minority status and language ($e^b=6.69$, $p<0.001$) were associated with increased COVID-19 case counts. However, on the basis of the local geographically weighted model, the association between social vulnerability and COVID-19 varied among counties. Overall, minority status and language, household composition and transportation, and housing and disability predicted COVID-19 infection.

Conclusions: Large-scale disasters differentially affect the health of marginalized communities. In this study, minority status and language, household composition and transportation, and housing and disability predicted COVID-19 case counts in the U.S. Addressing the social factors that create poor health is essential to reducing inequities in the health impacts of disasters.

INTRODUCTION

When disasters of any type occur, the socially vulnerable are at the greatest risk. Recent reports on morbidity and mortality of coronavirus disease 2019 (COVID-19) have indicated a higher burden among racial/ethnic minorities, the elderly, the poor, and people with the lowest educational attainment. The disproportionate burden of the pandemic among socially vulnerable communities has been explained in part by the endemic inequities in these populations, including income, education, nutrition, transportation, housing, jobs, environment, psychosocial stresses, and health care, which contribute to poor health. For example, African Americans are more likely to live in environmentally polluted neighborhoods than whites, which increases their propensity to developing environmental health problems like asthma, as well as their subsequent risk of COVID-19 morbidity. According to a report published by the National Association for the Advancement of Colored People and Clean Air Task Force, African
Americans are 75% more likely to reside in proximity to a polluting facility like a factory or refinery than other Americans and are exposed to air that is 38% more polluted compared with whites. Poor housing and crowded neighborhoods, where minorities are over-represented, make the type of physical distancing that can be protective against COVID-19 infection less feasible, increasing the risk of becoming infected. 

Other chronic diseases associated with worsening prognosis of COVID-19 include hypertension and diabetes. Unfortunately, the opportunity to improve diet and increase physical exercise are substantially associated with material factors—household income, financial assets, and the local built environment. 

Lower median household incomes among African Americans, along with lower rates of medical insurance and institutional racism, diminish access to health care. According to the Institute of Medicine (now known as the National Academy of Medicine), regardless of income, insurance status, age, or severity of medical condition, racial and ethnic minorities are more likely to be provided inferior medical care than is provided to whites. A review of the billing information on COVID-19 by Rubix Life Sciences also revealed that physicians are less likely to refer African Americans for screening when they present with symptoms.

Despite the increasing concerns about the disproportionate burden of the pandemic on minority communities, race-stratified data are yet to be released for most U.S. states; only 7 states—California, Illinois, Louisiana, Michigan, Minnesota, New York, and North Carolina—had released race-disaggregated data by April 12, 2020. Through April 9, 2020, a total of 16,231 deaths have occurred, but only 35% of the mortality records included information on race and ethnicity. Owing to the insufficiency of race-stratified data, few primary studies on social vulnerability to COVID-19 have been undertaken. Using publicly available secondary spatial data on social vulnerability and COVID-19 case counts, this study assesses the association between social vulnerability and the pandemic’s case counts in the U.S. and identifies hot spots where social vulnerability was positively associated with case counts.

**METHODS**

**Study Sample**

County-level publicly available data on social vulnerability for the lower 48 U.S. states were obtained from the Centers for Disease Control and Prevention (CDC’s Agency for Toxic Substances and Disease Registry (2020). Shapefiles for COVID-19–confirmed case counts per county were derived from the Johns Hopkins University’s Center for Systems Science and Engineering (current as of May 12, 2020). U.S. county-level data on average daily sunlight, precipitation, air temperature, heat index, and fine particulate matter (PM2.5) were retrieved from the CDC Wide-ranging Online Data for Epidemiologic Research. Data on the number of people tested for COVID-19 were derived from the COVID Tracking Project. Finally, regional boundary data on states and counties were downloaded from data.gov.

**Measures**

In the event of disasters of all types, including pandemics like COVID-19, social factors like the percentage of people living in poverty or without access to transportation can contribute to the degree to which communities withstand disaster impacts and minimize human suffering and financial loss. Factors that describe a community’s social vulnerability were compiled by CDC into a Social Vulnerability Index (SVI) to enable the identification of populations most likely to need support before, during, and after a disaster. CDC’s SVI estimates the relative vulnerability of each census tract or county in the U.S. by ranking 15 variables in 4 major themes: SES, household composition and disability, minority status and language, and housing and transportation.

The socioeconomic theme comprises 4 variables: percentage below poverty, percentage unemployed, per capita income, and percentage with no high school diploma. The household composition and disability theme comprises the percentage aged ≥65 years, percentage aged ≤17 years, percentage civilian with a disability, and percentage single-parent households. The minority status and language theme comprises 2 variables: percentage minority and the percentage who speak English less than well. Finally, the housing and transportation theme constitutes 5 social variables: percentage multiunit structures, percentage mobile homes, percentage crowding, percentage no vehicle, and percentage group quarters. The overall index is scored from 0 to 1 with higher values denoting higher vulnerability. Additional detail on CDC’s SVI variable selection and methods is available elsewhere. For each county, the authors derived the outcome variable by dividing the cumulative counts (between January 21, 2020, and May 12, 2020) of confirmed COVID-19 cases by county total population multiplied by 100,000. To satisfy the parametric requirement for normality, the authors log-transformed the outcome variable and exponentiated the model coefficients for ease of interpretation. Overall, 8.5% of U.S. counties (264 of 3,108; 3.5% of the national population) were excluded from the study because of missing data.

**Statistical Analysis**

The global association between COVID-19 case counts and social vulnerability was assessed by fitting a multiple linear regression model using ordinary least squares (OLS) regression. The relationship of the outcome variable was assessed with SVI and its 4 themes, including SES, household composition and disability, minority status and language, and housing and transportation. Covariates were included to adjust for the potential confounding effect of population size, population density, number of people tested, average daily sunlight, precipitation, air temperature, heat index, and PM2.5. OLS regression is considered global as it estimates a single parameter for each variable and assumes that the
relationship between the outcome and predictor variables holds the same everywhere in the study area.

However, because vulnerability to COVID-19 may actually be nonstationary and vary between U.S. counties, the authors also fitted a geographically weighted regression model (GWR) using the variables previously included in the OLS model. Unlike the global OLS regression, GWR estimates a local model and computes parameters for all sample points considering the spatial variation in relationships. The model was configured with a fixed (Gaussian) Kernel type, corrected Akaike Information Criterion’s bandwidth, and a cross-validated optimal number of 30 neighbors. To identify which SVI themes predict COVID-19 case counts across the spatially varying regions of the U.S., coefficient raster maps were generated for the predictor variables using hot-/cold-rendering color schemes.

Model performance was assessed using the coefficient of determination, adjusted coefficient of determination, Joint F and Wald statistics, Koenker (Bruesch–Pagan) statistic, and corrected Akaike Information Criterion. The impact of missing data on model performance was assessed by filling missing values with an estimated average, derived from 30 contiguous edged-spatial neighbors, and conducting a sensitivity analysis on the complete data. Similarly, the potential effect of testing rate on model performance was assessed with a sensitivity analysis of data from 6 states (Massachusetts, Rhode Island, Connecticut, Delaware, Louisiana, and Washington, DC).

ArcMap, version 10.4.1 and ArcGIS Pro, version 2.5.1 were used for the analysis.

RESULTS

As of May 12, 2020, a total of 1,320,909 people had been diagnosed with COVID-19 in the U.S. Of the 2,844 counties included in the study (91.5%; 2,844 of 3,108), the highest case count was recorded in Trousdale County, Tennessee (16,525.22 per 100,000) and the lowest in Tehama County, California (1.54 per 100,000). Overall, COVID-19 case count was highest (at least 681 per 100,000) in New York, New Jersey, Massachusetts, Rhode Island, Connecticut, Delaware, Louisiana, and Washington, DC.

On the basis of the adjusted OLS regression, overall SVI and minority status and language were predictive of increased COVID-19 case counts. A percentile increase in overall SVI was associated with a 65% increase in COVID-19 case counts \( (e^\beta=1.65, p=0.03) \), and a percentile increase in minority status and language was associated with a 6.69-fold increase in case counts \( (e^\beta=6.69, p<0.001) \) (Table 1). Upon fitting the global model on the 15 individual variables that constitute the SVI, minority status \( (e^\beta=6.69, p<0.001) \) and lacking high school education \( (e^\beta=1.42, p=0.01) \) positively predicted case counts (Appendix Table 1, available online). Sensitivity analysis of 6 states (and District of Columbia) with the highest rate of testing affirmed the strong association between minority status and language and COVID-19 case counts \( (e^\beta=9.13, p<0.001) \). The results did not differ significantly when missing data from 264 counties were imputed using spatial prediction methods (Table 1).

The negative relationships between household composition and disability and case counts, and between housing and transportation and case counts were

Table 1. Multiple Linear Regression of Log-Transformed COVID-19 Case Counts (Per 100,000) on Social Vulnerability Factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1 ( (n=2,844) ) ( e^\beta )</th>
<th>Model 2 ( (N=3,108) ) ( e^\beta )</th>
<th>Model 3 ( (N=177) ) ( e^\beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall SVI</td>
<td>1.65 (0.03)</td>
<td>1.76 (0.02)</td>
<td>0.46 (0.71)</td>
</tr>
<tr>
<td>SES</td>
<td>0.61 (0.05)</td>
<td>0.57 (0.05)</td>
<td>0.63 (0.71)</td>
</tr>
<tr>
<td>Household composition and disability</td>
<td>0.72 (&lt;0.001)</td>
<td>0.72 (&lt;0.001)</td>
<td>0.87 (0.69)</td>
</tr>
<tr>
<td>Minority status and language</td>
<td>6.69 (&lt;0.001)</td>
<td>6.67 (&lt;0.001)</td>
<td>9.13 (&lt;0.001)</td>
</tr>
<tr>
<td>Housing and transportation</td>
<td>0.69 (&lt;0.001)</td>
<td>0.66 (&lt;0.001)</td>
<td>0.82 (0.53)</td>
</tr>
</tbody>
</table>

Model diagnostics

- Multiple \( R^2 \): 0.27, 0.28, 0.50
- Adjusted \( R^2 \): 0.27, 0.28, 0.46

- AICc: 8,362.40, 8,927.77, 487.70
- Joint F-statistic: 80.41 (<0.001), 92.33 (<0.001), 12.73 (<0.001)
- Joint Wald statistic: 1,589.06 (<0.001), 1,456.06 (<0.001), 4,743.15 (<0.001)
- Koenker (BP) statistic: 93.22 (<0.001), 89.83 (<0.001), 30.87 (<0.001)

Note: Boldface indicates statistical significance \( (p<0.05) \). All models adjusted for population size, population density, number of people tested, average daily sunlight, precipitation, air temperature, heat index, and fine PM (PM2.5).

- Original: 264 missing counties.
- Imputed for missing data.
- Select states and Washington, District of Columbia with high testing rates \( (\geq 40,000/\text{million}) \): New York, Massachusetts, Rhode Island, Louisiana, New Mexico, Utah, and District of Columbia.
- Outcome variable has been log-transformed; model coefficients have been exponentiated for ease of interpretation.

AICc: corrected Akaike Information Criterion; BP, Bruesch–Pagan; PM, particulate matter; SVI, Social Vulnerability Index.
counterintuitive. However, these associations were better described by local spatial variations between counties, as captured by the GWR model.

Although social vulnerability was associated with increased case counts at the global level (as demonstrated by the OLS), the direction and magnitude of this association varied among U.S. states. The authors explored this spatial nonstationarity with a GWR model (Figures 1–3). In the states of Washington and Oregon, minority status and language and household composition and disability were more predictive of COVID-19 case counts than housing and transportation. In the U.S. Gulf Coast states (especially Eastern Texas, Louisiana, Mississippi, Alabama, Florida), Southern Arkansas, and Western Tennessee, vulnerability to COVID-19 was better explained by housing and transportation than by minority status and language. Consistent with the OLS regression results, the GWR model confirmed a strong association with minority status and language than with other SVI themes (Figures 1–3). Counties (hot spots) identified as demonstrating both high SVI and high COVID-19 case counts are summarized in Appendix Table 1 (available online).

**DISCUSSION**

Disasters and emergencies, be they natural or anthropogenic, profoundly and inequitably affect the health of socially vulnerable populations. Such events expose inequities in areas including access to quality housing and education as well as issues of economic and environmental justice that create conditions that make it difficult to maintain health. After all types of disasters, minority communities have been documented in the literature to be faced with a disproportionate share of detrimental physical and mental health outcomes. These disparate health impacts may be due to an inability to fully understand warning advisories related to an impending hazard event or further caused by living in areas of high hazard vulnerability and poor access to health care. The principal finding from this study—increased COVID-19 cases in some of the most socially vulnerable counties—highlights the need for targeted interventions to mitigate the impact of COVID-19 on these communities.
vulnerable counties in the U.S.—underscores the importance of continuing to work to address inequities related to the social determinants of health.

At the global level, the finding that overall SVI and minority status and language were associated with increased COVID-19 case counts highlights the impact of social vulnerability on the pandemic. Notably, this study found that the relationship between social vulnerability and COVID-19 was not stationary but varied between U.S. counties. For example, the variables associated with increased case counts were not the same for southwest Georgia household composition and disability as they were for New York City minority status and language. These spatially variable findings contribute to previous studies that demonstrated the nonuniform distribution of social vulnerability to U.S. disasters. In the context of the COVID-19 pandemic, further research is needed to substantiate these findings and elucidate possible mechanisms for these associations.

In the U.S., 44 million adults are underinsured, including 48% of adults aged 19–64 years who face high copays and out-of-pocket costs. Nearly 2 million Americans lack running water in their homes, with Native Americans 19 times more likely and African Americans and Hispanics twice as likely to lack indoor plumbing than whites. This makes basic infection prevention practices, such as hand washing, less attainable for mitigating the spread of the disease. The lack of paid sick leave among low-wage earners, whose work often involves face-to-face interaction, also makes self-quarantining a less feasible option. In addition, because of low wages, many workers hold multiple jobs to support their families, which may further increase their risk of contracting or spreading COVID-19. For fear of retaliation and deportation, undocumented migrants are less likely to report COVID-19 symptoms or present themselves for medical screening, which could potentially heighten their risk of contracting or transmitting the disease.

The Families First Coronavirus Response Act enacted on March 18, 2020, addresses some disparities related to COVID-19, making provisions for paid sick leave to some employees as well as for providing free testing, nutrition assistance, and expanded unemployment benefits. However, as the U.S. nears the peak of the pandemic,
more could be done to mitigate the disproportionate burden of the pandemic among marginalized communities. For example, Massachusetts convened an emergency task force on coronavirus and equity to develop policy recommendations for communities especially vulnerable to COVID-19, and many of these recommendations could be applied to other states: prohibiting U.S. Immigration and Customs Enforcement from entering healthcare facilities and shelters and assuring residents of universal healthcare provision without question of immigration status; financially assisting community-based organizations to support immigrant populations; expanding unemployment benefits, paid time off, paid family leave, and parental leave to cover the vulnerable for as long as the pandemic lasts; providing safety training to public-facing employees (e.g., first responders); improving sanitation and access to testing and treatment for nursing home residents, the prison population, residents of detention centers, and the homeless; reinforcing nutrition assistance programs for adults and children; improving access to care for the disabled, especially those with disrupted support systems; improving access to quality internet connection for rural communities to support remote work; and financially covering unexpected costs that may be incurred by government agencies and health departments. In the face of limited resources, support can be prioritized to the more socially vulnerable counties, such as those identified in this study, because their populations may have a stronger social vulnerability to COVID-19.

**Limitations**

There are several limitations to this study. Social vulnerability was assessed using CDC’s SVI, a hierarchical index that has been shown to be associated with lower precision and weaker internal validity. The SVI’s precision was also reported to be sensitive to the weighting scheme chosen. Despite these inherent limitations, the authors used the SVI because it has demonstrated higher accuracy than other models and compares well with other indices of social vulnerability.
also publicly available and widely cited in the literature. Although the authors adjusted for demographic and environmental factors—population size, population density, number of people tested, average daily sunlight, precipitation, air temperature, heat index, and PM2.5—SVI variables explained only 38.9% ($R^2=0.389$) of the variability in COVID-19 case counts. Further studies are warranted to explore additional factors, which explain the burden of the pandemic. Given the potential heterogeneity of vulnerable regions within a county, census tract level might be more suitable for analysis. As the authors do not have publicly available data disaggregated on census tract level, this could be examined in future research. In addition, because disparities in exposure and susceptibility to pandemics may vary across time, it may be suitable to analyze the rate of increase in case counts at multiple time points for each county. The authors do not have longitudinal data to permit trend analysis, and this could be investigated in future studies. This study was further limited by COVID-19 case counts. In particular, potential substantial undocumented cases of COVID-19 infection in the U.S. may have limited the use of confirmed case counts as an outcome in the study. Finally, the exclusion of 8.5% of U.S. counties (264 of 3,108) owing to missing data is a limitation. However, the sensitivity analysis with imputed data yielded results similar to those of the original model (Table 1 and Appendix Figures 1–3, available online), thus suggesting that comparable findings would be obtained if data were not missing. Nonetheless, given the continuing spread of the disease, periodic updated spatial statistical analyses may be warranted.

CONCLUSIONS

Large-scale disasters, whether natural (e.g., COVID-19), traumatic (e.g., the World Trade Center attack or mass shootings), or environmental (e.g., the Deepwater Horizon spill), profoundly affect the health of socially vulnerable communities. In this study, minority status and language, household composition and disability, and housing and transportation were found to predict COVID-19 case counts in U.S. counties. Addressing the social determinants of health, such as housing, education, and environmental and economic justice, is essential to reducing inequities in the health impacts of disasters.

ACKNOWLEDGMENTS

All the authors made substantial contributions to this paper and take responsibility for it. IMK and JAH conceived of the presented idea; IMK collected and analyzed the data; IMK wrote the manuscript; JAH verified the analytic methods and reviewed and provided substantial edits to the manuscript; and both authors approved the final version of the manuscript.

No financial disclosures were reported by the authors of this paper.

SUPPLEMENTAL MATERIAL

Supplemental materials associated with this article can be found in the online version at https://doi.org/10.1016/j.amepre.2020.06.006.

REFERENCES

14. Collins S, Bhupal H, Doty M. Health insurance coverage eight years after the ACA: fewer uninsured Americans and shorter coverage gaps.


