ORIGINAL ARTICLE

The Association of Social Determinants of Health With COVID-19 Mortality in Rural and Urban Counties

Rajib Paul PhD[1] | **Ahmed Arif MBBS, PhD**[1] | **Kamana Pokhrel MS**[2] |
Subhanwita Ghosh MS[1]

[1] Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

[2] Health Analytics and Informatics, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

Correspondence
Rajib Paul, PhD, Department of Public Health Sciences, University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, USA.
Email: Rajib.Paul@uncc.edu

Abstract

Purpose: To identify the county-level effects of social determinants of health (SDoH) on COVID-19 (corona virus disease 2019) mortality rates by rural–urban residence and estimate county-level exceedance probabilities for detecting clusters.

Methods: The county-level data on COVID-19 death counts as of October 23, 2020, were obtained from the Johns Hopkins University. SDoH data were collected from the County Health Ranking and Roadmaps, the US Department of Agriculture, and the Bureau of Labor Statistics. Semiparametric negative binomial regressions with expected counts based on standardized mortality rates as offset variables were fitted using integrated Laplace approximation. Bayesian significance was assessed by 95% credible intervals (CrI) of risk ratios (RR). County-level mortality hotspots were identified by exceedance probabilities.

Findings: The COVID-19 mortality rates per 100,000 were 65.43 for the urban and 50.78 for the rural counties. Percent of Blacks, HIV, and diabetes rates were significantly associated with higher mortality in rural and urban counties, whereas the unemployment rate (adjusted RR = 1.479, CrI = 1.171, 1.867) and residential segregation (adjusted RR = 1.034, CrI = 1.019, 1.050) were associated with increased mortality in urban counties. Counties with a higher percentage of college or associate degrees had lower COVID-19 mortality rates.

Conclusions: SDoH plays an important role in explaining differential COVID-19 mortality rates and should be considered for resource allocations and policy decisions on operational needs for businesses and schools at county levels.

KEYWORDS
Bayesian inference, chronic diseases, cluster analysis, residential segregation, socioeconomic status

According to the World Health Organization's definition, social determinants of health (SDoH) are "the conditions in which people are born, grow, live, work and age."[1] The role social factors play in determining adverse health outcomes, such as cardiovascular diseases, cancer, chronic obstructive pulmonary disease (COPD), asthma, and flu, is now widely recognized.[2–5] Healthy People 2020[6] addressed 5 major components of the SDoH: (1) neighborhood and built environment,

(2) health and health care access, (3) social and community context, (4) education, and (5) economic stability. Since a complex mechanism impacts our health, it is imperative that a holistic intervention approach is necessary. Only improving health and health care access is not enough; focus should also be on the advancements of "education, child-care, housing, business, law, media, community planning, transportation, and agriculture."[6] According to the Unites States Census Bureau

report, about 19.3% of the population resides in rural areas.[7] Urban areas are rich in health care infrastructure; however, several barriers exist in accessing these facilities.

A large body of literature in social epidemiology identifies rurality itself as a cause of health disparities.[8] Many rural clinics struggle with health care workforce retention.[9] While access to quality health care has been the biggest priority for decades in rural health research, several studies have shown that population-based approaches which consider "rural culture health determinants" are effective in mitigating chronic health conditions prevalent in rural communities.[10] Rural residents exercise less, eat less nutritional food, and are more prone to obesity than urban residents.[10,11] The rural population is generally older, has lower high school graduation rates, and earns on average less than their urban counterparts. They are also less likely to seek preventative care,[12] placing them at a higher risk of COVID-19-related comorbidities. While the Affordable Care Act (ACA)[13] and telemedicine [14] helped in reducing the disparities to some extent, inequalities in health care and prevention still exist due to area deprivation in rural communities.[15]

Black populations living in rural areas in the southern United States face significant health disparities due to lack of social capital, fractiousness in churches, systematic racism, residential segregation, and poverty, and these lead to low utilization of preventive care and increased psychological stress.[10] Poverty, unemployment, mental stress, and social exclusion have been identified as primary causes of disparities in HIV rates among the Black populations in southeastern rural regions.[10]

While addressing SDoH in urban areas, one needs to address both spectrums of socioeconomic status, from slum areas, inner cities, middle-income communities, to higher-income neighborhoods.[16] Health disparities in urban areas are mostly caused by race, socioeconomic status,[17] and racial residential segregation.[18] Residential segregation is an important indicator of health disparities in the United States. It creates inequities in educational and employment opportunities, pathogenic residential conditions, neighborhood quality and social capital, crime rates, access to health care, and health behaviors, which are detrimental for keeping good health.[18] Residential segregation is also responsible for socioeconomic immobility.[19]

As COVID-19 spreads through the rural and urban communities in the United States, it is important to understand how economic and social factors related to SDoH impact the vulnerable populations so that the local, state, and federal resources, such as availability of test kits, personal protective equipment, financial needs for businesses and schools, and vaccines, can be appropriately allocated. While COVID-19 has been termed as a social equalizer[20] in terms of who gets infected, few studies have shown its differential effects on rural communities[21–24] and Black and Latino communities.[25–27] Studies have shown racial disparities,[27–32] including higher morbidity and mortality rates among Black communities.[20,33] Also, the COVID-19 infection and mortality rates have been rising at an alarming rate in rural communities.[34]

The objectives of this study were (1) to identify county-level SDoH variables that are associated with increased or decreased county-level mortality rates of COVID-19, and (2) estimate county-level exceedance probabilities of mortality to identify clusters of counties at an elevated risk of death from COVID-19.

MATERIALS AND METHODS

This research used crowdsourced (data that are submitted by a large group of people and organizations via the Internet, social media, and apps) COVID-19 data on confirmed positive cases and deaths from the data repository published by the Johns Hopkins University (JHU),[35] and publicly available data on SDoH from the County Health Rankings and Roadmaps (CHRR)[36] supported by the Robert Wood Johnson Foundation, and the US Bureau of Labor Statistics.[37] Specifically, we used CHRR's recently published 2020 estimates that were based on the 2014–2018 American Community Survey (ACS)[38] data.

We used the latest data (as of October 23, 2020) on the total number of confirmed positive cases and deaths from COVID-19 from the repository published by the Centers for Systems Science and Engineering (CSSE) at JHU on the GitHub website.[35] There are 3108 county and county-equivalents in the 48 contiguous United States. Historically, data from the five counties (New York (Manhattan), Kings (Brooklyn), Bronx, Richmond (Staten Island), and Queens) covering New York City were reported collectively as total infections and deaths from New York City. This brings down our number of counties to 3104. As of October 23, 2020, 3080 counties and county-equivalents reported at least one infection, and 2742 counties reported at least one death. Counties with no reported infections were treated as structural zeros in our model. County-level crude mortality rates were calculated as the number of deaths divided by the county population size.

Neighborhood and built environment

The data for variables related to neighborhood and built environment were retrieved from the County Health Rankings and Roadmaps[36] website. The *Percentage of Population under Severe Housing Cost Burden* measured the percent of households that spent greater than or equal to half (50%) of their income on housing from 2014 to 2018. The *Residential Segregation* variable was the 2020 estimates of the index of dissimilarity, showing the degree of segregation between the White and the Black county residents where they lived apart geographically from each other using data from the ACS 5-year estimates (2014–2018).[38] The residential segregation index ranges from 0 to 100; a higher value indicates more residential segregation with 0 indicating complete integration and 100 indicating complete segregation.

Economic stability

The *Unemployment Rates* were procured from the US Bureau of Labor Statistics that used the Current Employment Statistics (CES)[39] program, the Current Population Survey (CPS),[40] ACS,[38] and the state Unemployment and Insurance system as data sources. This variable

was the percentage of the annual average of seasonally unadjusted unemployment rates in 2019. The variable *Income Inequality Ratio* was procured from the County Health Rankings and Roadmaps that used the most recent ACS 5-year estimates as the data source. It was defined as the household income ratio at the 80th percentile to that of the 20th percentile from 2014 to 2018.

Education

County-level data on *Percent of Population with College or Associate Degree* were obtained from the US Department of Agriculture (USDA).[41] The USDA Economic Research Services obtained these estimates using the most recent 5-year (2014–2018) data from the ACS.

Social and community context

The data for variables of social and community context were abstracted from the County Health Rankings and Roadmaps website that used the Census Population Estimates[36] as the primary data source. The variable *Percent Black Population* was the percentage estimates of the non-Hispanic Black population in the year 2018. *Age 65 Years and Up* was the total percentage estimates of elderly aged 65 years and older in the year 2018. The *Female Percent* was the county-level estimates of female percentage in the year 2018.

Health and health care

We retrieved data on health and health care variables from the County Health Rankings and Roadmaps website. The Frequent Mental Distress was defined as the percentage of adults reporting poor mental health, such as depression, stress, and emotional problems, for 2 weeks or more per month in 2017. The *Percent Diabetes* variable was the unadjusted diabetes rates from the most recent comprehensive county-level 2016 data collected by the United States Diabetes Surveillance System (USDSS).[42]

The *HIV Rate Per 100,000 Population* reflected the total number of diagnosed human immuno deficiency virus (HIV) infection per 100,000 populations in the age group of 13 years and older in the year 2016. It was calculated by dividing the total number of HIV cases by the average number of people at risk of HIV. HIV rates were calculated based on the data from the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention.[43] The variable *Adult Smoker Percent* was the percent of current adult smokers in the year 2017 who smoked regularly or frequently with at least 100 cigarettes in their life. County Health Rankings and Roadmaps used data from the Behavioral Risk Factor Surveillance System (BRFSS)[44] to estimate county-level Adult Smoker Percent and Frequent Mental Distress Percent.

Rural/urban designation

A county-level binary variable designating rural and urban classification was created using RUCA (Rural-Urban Commuting Area) codes[45] based on the 2010 census data. Rural counties were micropolitan low commuting, core small towns, small towns with high and low commuting, and areas with the primary flow to tracts outside of urban areas or clusters. All metropolitan areas and high commuting micropolitans were classified as urban counties.

STATISTICAL ANALYSIS

Our data analyses comprised exploratory and inferential statistical techniques. In exploratory analyses, we first computed the correlation matrix to assess the multicollinearity among the SDoH characteristics. Next, we computed median and interquartile ranges for all SDoH variables, separately for rural and urban counties. Mann–Whitney two-sample test[46] was used to assess differences among these variables across rural and urban counties. As SDoH exhibit differential gradients for rural and urban counties, we used a binary measure on rurality as an effect modifier in our analysis.

We developed a spatial model using SDoH characteristics that would enable comparing mortality rates across counties and aids in identifying hotspot clusters. We constructed a semiparametric[35] negative binomial regression model on county centroids that enabled faster computation and addressed spatial nonstationarity across 3104 counties in the contiguous United States.

Denoting by y_i, the number of deaths from COVID-19 in the *i*th county, we assumed that y_i follows a Poisson distribution with mean $\lambda_i Z_i$. Further, a Gamma distribution with shape and rate parameters ψ was imposed on Z_i. When we integrate out Z_i, this Poisson-Gamma distribution resulted in a negative binomial distribution with overdispersion λ_i^2/ψ, where we modeled the mean parameters, λ_i, using a set of county-specific independent variables, $\mathbf{X}_i$, a set of basis functions $\mathbf{S}_i$ specified using the centroids of the *i*th county (see Equation 2 for further details), two sets of regression coefficients, β and θ, and county-specific random effects, v_i, as:

$$\log(\lambda_i) = log(E_i) + \beta \mathbf{X}_i + \theta \mathbf{S}_i + v_i, i = 1, \dots, N. \quad (1)$$

In Equation (1), E_i, the expected count was used as an offset term in the regression. While estimating mortality, we used E_i as the total population of the *i*th county times the overall mortality rate for all counties (SMR: standardized mortality rate). This modeling framework gives the flexibility to vary regression coefficients β based on the hypotheses under considerations. For example, one can use different sets of coefficients for rural and urban counties.

The terms in $\mathbf{S}_i$ were specified using bivariate basis functions using longitude and latitude $c_i = \{long_i, lat_i\}$ of the *i*th county centroid and a

set of k knot locations $\{u_1,\ldots,u_k\}$ defined on the entire domain of analysis. The number of knots, k, controls the smoothness of the regression surface. More knots lead to a less smooth regression surface. We used 50 knots and knot locations were selected using the *clara* algorithm.[47] Specifically, we used the following basis function to specify (i, j) term of the $\mathbf{S}_i$ vector:

$$S(i, j) = ||\mathbf{c_i} - -\mathbf{u_j}||^2 |log\, ||\mathbf{c_i} - -\mathbf{u_j}||, j = 1, \ldots, k, \qquad (2)$$

where $||\cdot||$ denotes the Euclidean distance between two vectors. The county-specific random effects terms, $\{v_i, i = 1,\ldots,N\}$, were assumed to follow independently and identically distributed zero-mean Gaussian distribution with variance σ^2.

In a traditional spatial model that is based on the first law of geography, one assumes that spatial dependence decays with distance. However, for COVID-19 data, we noticed that this assumption is not valid due to the mechanisms by which the virus spreads throughout the country. The main carrier of the virus is people and people travel throughout the country irrespective of geographic distances. Also, high-density areas, such as cities and large towns, are more susceptible to infection. In that regard, the virus transmission in New York City could be more similar to Los Angeles than Buffalo. Our semiparametric regression is not based on the first law of geography and provides flexibility for more general covariance structures. Our model was fitted to data in the Bayesian paradigm using integrated Laplace approximation (INLA). We used R/RStudio version 4.0.0.[48] and R-INLA[81] and Semipar[47] packages. An independent variable was considered (Bayesian) significant if the 95% credible interval (CrI) of the corresponding exponentiated regression coefficient, known as *Risk Ratio* (RR), did not include 1.

The exceedance probabilities of excess risks of mortality were calculated by:

$$\omega_i = exp\{\beta\mathbf{X}_i + \theta\mathbf{S}_i + v_i\}, i = 1, \ldots, N. \qquad (3)$$

In Equation (3), ω_i denotes the relative risk for the ith county. We computed the county-level probabilities of ω_i being greater than 1.5, which indicates 50% excess risk from the SMR for the ith county. Crude mortality rates and exceedance probabilities were mapped in ArcGIS Pro from ArcGIS Desktop version 10.8.[49]

We used a stepwise selection[50] procedure for subset selection to mitigate multicollinearity in the regression model. Additionally, we compared competitive models using Watnabe Akaike Information Criterion (WAIC);[50] lower values of WAIC are preferable. Two variables were dropped by our stepwise subset selection, housing cost burden, and percent of mental distress. The final adjusted regression model had the lowest WAIC of 18,422.48, about 11% lower than the WAIC that used all the independent variables from the unadjusted regressions.

RESULTS

As of October 23, 2020, the COVID-19 mortality rate was 65.43 per 100,000 in urban counties, and 50.78 per 100,000 in rural counties (P value < .0001). The housing cost burden and residential segregation indices were significantly higher in urban counties. The median unemployment rate was slightly but significantly higher in rural areas, whereas the median income inequality ratio was slightly but significantly elevated in urban counties. The physical and mental distress and diabetes rates were higher in rural counties, and HIV rates were higher in urban counties (Table 1).

Table 2 summarizes our findings on unadjusted and adjusted rate ratios (RR) from the multivariate semiparametric negative binomial regression. For every 5% increase in residential segregation among Whites and Blacks, the COVID-19 mortality rates increased by 3.4% (adjusted RR = 1.034, CrI = 1.019, 1.050) and for every 5% increase in unemployment rates, the mortality increased by 47.9% (adjusted RR = 1.479, CrI = 1.171, 1.867) in urban counties, when adjusted for other variables. Unemployment rates were significantly associated with mortality from COVID-19 (unadjusted RR = 1.868, CrI = 1.171, 2.127) in rural counties in the unadjusted analysis; however, this effect was not (Bayesian) significant in the adjusted model. Under the adjusted regression, for every 5% increase in income disparity in rural counties, the mortality increased by 0.03% (adjusted RR = 1.003, CrI = 1.000, 1.006). Counties with a high percent of the population with college or associate degrees had lower mortality rates, whereas counties with a higher percentage of Blacks and the female population (Bayesian) significantly increased the mortality rates in rural and urban counties. The rural counties with a higher percentage of the population 65 years and older had slightly reduced mortality rates. Additionally, we found that diabetes (adjusted RR for rural counties = 1.094, CrI = 1.038, 1.153; adjusted RR for urban counties = 1.121, CrI = 1.031, 1.219) and HIV (adjusted RR for rural counties = 1.003 CrI = 1.001, 1.006; adjusted RR for urban counties = 1.001, CrI = 1.000, 1.002) per 100,000 substantially increased the COVID-19 mortality in rural and urban counties. For every 5% increase in frequent mental distress, mortality rate increased by 86.2% (RR = 1.862, CrI = 1.704, 2.035) in rural counties and 96.2% (RR = 1.962, CrI = 1.783, 2.158) in urban counties, in the unadjusted analyses. However, this effect was removed in the adjusted model by our stepwise subset selection procedure.

Figure 1A shows the county-level total infection per 100,000 population, and Figure 1B shows the crude mortality rates from COVID-19 per 100,000 as of October 23, 2020. Figure 2A shows the estimated mortality rates from COVID-19 per 100,000 using the fitted semiparametric negative binomial regression. These estimated rates are a denoised version of the crude mortality rates. Exceedance probabilities of the county-wise mortality rates to be greater by 50% from its SMRs based on our semiparametric negative binomial regression were mapped in Figure 2B. From the maps of exceedance probabilities, we detected clusters of high mortality rates in northeastern US counties,

TABLE 1 Descriptive statistics of the study sample

Variables	Rural (n = 2107)	Urban (n = 997)	P value
Crude mortality rate per 100,000	50.78	65.43	<.0001[a]
Neighborhood and Built Environment			
Percent severe housing cost burden	10.226 (3.914)	12.253 (4.573)	<.0001
Residential segregation index	33.540 (46.630)	45.270 (22.243)	<.0001
Economic Stability			
Percent unemployed	3.800 (1.700)	3.500 (1.200)	<.0001
Income inequality ratio	4.393 (0.867)	4.420 (0.844)	.02882
Education			
Percent some college or associate degree	30.750 (7.300)	30.100 (6.050)	<.0001
Social and Community Context			
Percent Black population	1.279 (5.062)	7.019 (15.067)	<.0001
Percent above 65 years	19.870 (4.958)	16.582 (4.786)	<.0001
Percent female population	50.110 (1.660)	50.741 (1.230)	<.0001
Health and Health Care			
Percent frequent physical distress	12.139 (3.526)	11.698 (2.752)	<.0001
Percent frequent mental distress	13.163 (2.956)	12.697 (2.349)	<.0001
Percent diabetes	12.100 (5.600)	11.100 (4.500)	<.0001
HIV rate per 100,000	61 (128)	150 (173)	<.0001
Percent of adult smokers	17.257 (4.973)	16.715 (4.485)	<.0001

[a]Proportions are compared using Chi-Square Test. For the remaining variables, medians and interquartile ranges are displayed and compared using Mann–Whitney two-sample test.

TABLE 2 Unadjusted and adjusted risk ratios for mortality rates using negative binomial regressions

	Unadjusted		Adjusted	
Variables	**Rural**	**Urban**	**Rural**	**Urban**
Neighborhood and Built Environment				
Percent severe housing cost burden	1.317 (1.247, 1.392)	1.291 (1.229, 1.355)	–	–
Residential segregation index	1.020 (1.010, 1.029)	1.019 (1.010, 1.028)	1.003 (0.995, 1.010)	1.034 (1.019, 1.050)
Economic Stability				
Percent unemployed	1.868 (1.644, 2.217)	2.146 (1.838, 2.513)	1.077 (0.903, 1.284)	1.479 (1.171, 1.867)
Income inequality ratio	1.023 (1.021, 1.025)	1.024 (1.021, 1.026)	1.003 (1, 1.006)	1.000 (0.995, 1.004)
Education				
Percent some college or associate degree	0.759 (0.733, 0.786)	0.759 (0.732, 0.787)	0.938 (0.895, 0.982)	0.826 (0.779, 0.876)
Social and Community Context				
Percent Black population	1.151 (1.135, 1.166)	1.125 (1.107, 1.144)	1.051 (1.029, 1.074)	1.041 (1.014, 1.068)
Percent 65 years and above	0.878 (0.845, 0.913)	0.886 (0.847, 0.927)	0.929 (0.884, 0.976)	1.023 (0.953, 1.097)
Percent female population	1.154 (1.078, 1.154)	1.158 (1.107, 1.144)	1.211 (1.116, 1.314)	1.278 (1.162, 1.406)
Health and Health Care				
Percent frequent mental distress	1.862 (1.704, 2.035)	1.962 (1.783, 2.158)	–	–
Percent diabetes	1.302 (1.247, 1.359)	1.360 (1.291, 1.434)	1.094 (1.038, 1.153)	1.121 (1.031, 1.219)
HIV rate per 100,000 populations	1.012 (1.011, 1.014)	1.008 (1.006, 1.009)	1.003 (1.002, 1.005)	1.001 (1, 1.002)
Percent adult smoker	1.328 (1.264, 1.395)	1.376 (1.301, 1.456)	1.039 (0.955, 1.130)	0.954 (0.861, 1.057)

In the table, posterior medians of risk ratios and within parentheses its 95% credible Intervals are exhibited under five units change in the independent variable. The independent variables in the final adjusted models were selected using correlation analyses, stepwise subset selection, and Watnabe Akaike Information Criterion (WAIC). The adjusted model with all the variables had WAIC = 20,711.8 and the adjusted model without percent severe housing cost burden and percent frequent mental distress had WAIC = 18,422.48. Lower WAIC is better.
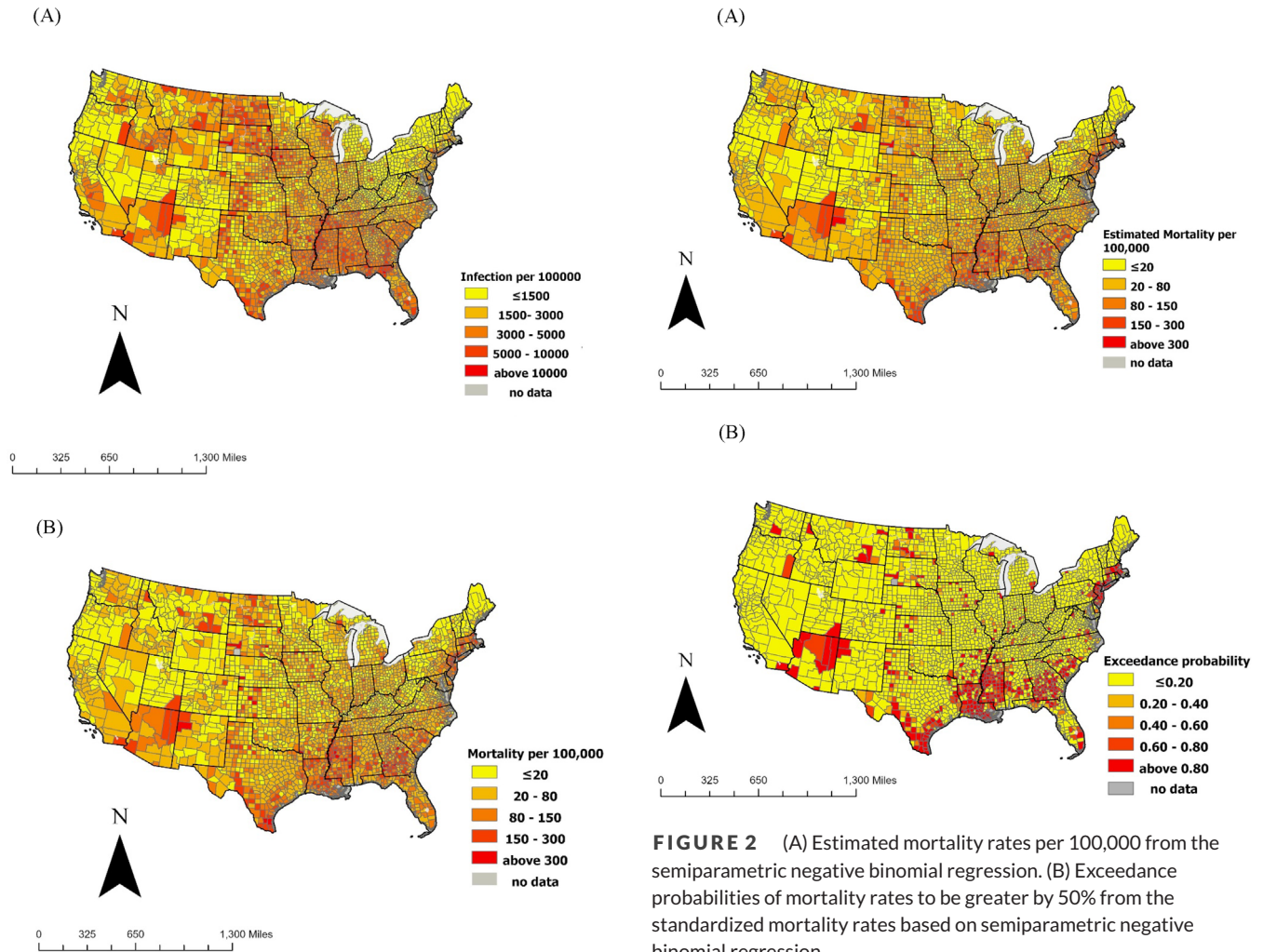
(A)

Infection per 100000
- ≤1500
- 1500- 3000
- 3000 - 5000
- 5000 - 10000
- above 10000
- no data

N

0   325   650   1,300 Miles

(B)

Mortality per 100,000
- ≤20
- 20 - 80
- 80 - 150
- 150 - 300
- above 300
- no data

N

0   325   650   1,300 Miles

FIGURE 1 (A) County-level total COVID-19 infection rates per 100,000 population. (B) County-level crude mortality rates per 100,000

(A)

Estimated Mortality per 100,000
- ≤20
- 20 - 80
- 80 - 150
- 150 - 300
- above 300
- no data

N

0   325   650   1,300 Miles

(B)

Exceedance probability
- ≤0.20
- 0.20 - 0.40
- 0.40 - 0.60
- 0.60 - 0.80
- above 0.80
- no data

N

0   325   650   1,300 Miles

FIGURE 2 (A) Estimated mortality rates per 100,000 from the semiparametric negative binomial regression. (B) Exceedance probabilities of mortality rates to be greater by 50% from the standardized mortality rates based on semiparametric negative binomial regression

east Michigan, and southern counties of Texas, Louisiana, Mississippi, Alabama, Georgia, and Florida. Additionally, several rural counties in Arizona, such as Mohave, Coconino, Apache, and Navajo, and two counties of New Mexico, San Juan, and McKinley, formed a cluster of high mortality rates.

DISCUSSION

We found substantial associations of several SDoH factors with COVID-19 mortality. Recall that county-level diabetes and HIV rates were significantly associated with COVID-19 mortality rates. Unemployment rates were significantly associated with higher mortality rates in urban counties only. While the percent of college and associate degrees substantially lowered mortality rates from COVID-19 in both urban and rural counties, residential segregation significantly increased mortality only in urban counties. Counties with a high percentage of blacks and females had a higher COVID-19 mortality rate.

Education is an important socioeconomic variable and influences one's well-being throughout life.[51] Previous studies found a strong association between higher education and good health.[52] Unemployment rates are highest among those with less than high school education.[53] Lower education and unemployment increase the risk of poor health, social isolation, and chronic diseases,[54–56] including mental distress and diabetes.[57] Adults with lower education often work in jobs that do not provide the opportunity to work from home, hence exposing them to the risk of COVID-19 infection.[58,59] In the current study, the rising unemployment rate was positively associated with COVID-19 mortality in urban counties, whereas higher education rate was negatively associated with COVID-19 mortality in both urban and rural counties. In prior studies, researchers reported from individual-level data analysis that poorly controlled diabetes was associated with the elevated risks of COVID-19 mortality.[60–62] Consistently, we found county-level elevated rates of diabetes associated with increased risk of COVID-19 mortality in both urban and rural counties.

We found a significant association of HIV rates with COVID-19 mortality in rural and urban counties. There are limited data on HIV as a comorbid condition for COVID-19. Some clinical reports[63] suggest that coinfection with COVID-19 may compromise the immune

system of patients with HIV. Moreover, advanced age and the presence of underlying comorbid conditions place people with HIV at an elevated risk of COVID-19 mortality.[64] Further research is needed to identify factors that may place HIV patients at higher risk of COVID-19 mortality.

Consistent with earlier reports,[20,26,27,33] we found significantly higher mortality rates from COVID-19 infection in counties with a higher percentage of the Black population. In the present study, the association of percent Black population with the elevated COVID-19 mortality was similar in both urban and rural counties. While 13% of the US population is Black, about 25% of the essential workforce is made up of Black adults,[65] placing them at higher exposure to COVID-19. Further, Blacks have higher rates of chronic diseases, like diabetes,[66] COPD,[67] and severe heart conditions,[68] increasing the risk of COVID-19 mortality.[69]

Residential segregation, an important indicator of structural racism, negatively impacts the health of the Black population,[70] in general. Residential segregation creates socioeconomic immobility and, predisposes vulnerable populations to social, physical, and environmental risks imperative to their well-being.[18] We found that the residential segregation index was significantly higher in urban counties compared to the rural counterparts and was associated with the elevated COVID-19 mortality rates in rural and urban counties in the unadjusted analysis only. When adjusted for other factors in the morrow, the association remained significant for urban counties only. The study by Yang et al.[71] found that high residential segregation among Blacks and Whites increased health disparities by 25%. Residential segregation also explains geographic differences in health.[72] Racially segregated areas have higher poverty and crime rates and less infrastructure and development.[71] Hao et al.[73] found that poor quality of life in the cancer survivors was also noticed in residents living in the highly segregated black counties. Del Brutto et al.[74] found that in rural communities of Ecuador, for each unit increase in social risk measured by the Gijon scale, the infection rates of COVID-19 increased by 15%. Additionally, they found a strong association between the COVID-19 seropositivity rate and lack of housing and basic home facilities. Our findings suggest that residential segregation between the Black and the White population should be considered for mitigating COVID-19 severity and mortality risks.

LIMITATIONS

Our study has certain limitations. The associations that we found from our county-level analyses of data are susceptible to ecological fallacy and should not be used for causal inference. For example, a negative association with percent of the population above 65 years with mortality should not be interpreted as lower death rates among that age group. Age is an individual-level nonmodifiable risk factor for death, and our analyses did not address any individual-level risk factors. At an aggregate level, a negative association simply tells us that county-level mortality rates decreased with percent of the population above

65 years. However, our findings provide a roadmap for further individual-level clinical research that is needed for better insights into the factors related to COVID-19 mortality. The current study, like others,[21,25,26] relied on the JHU data repository as the basis for COVID-19 research. While the JHU repository provides the most comprehensive data, it is still susceptible to certain errors and underreporting. Readers should consider these shortcomings when interpreting the findings of the research that used such data.

CONCLUSION

This study found that county-level SDoH variables were associated with the COVID-19 mortality in the United States. Whether it is a natural disaster (hurricane and forest fire) or a pandemic, identifying vulnerable populations is important for mitigating severity.[75,76] SDoH aids in identifying high need communities.

One's health outcomes and its associations with risk factors and clinical conditions are very complex relations.[77] Identifying the subgroups at higher risk of COVID-19 complications and mortality requires the consideration of behavioral factors (such as smoking, alcohol drinking, physical activity, diet, etc.), genetic compositions and family history, economic factors like income and access to health care, demographic factors (such as age, gender, race, etc.), and of course, environmental exposures.[78,79] A multilevel analysis[78,79] via hierarchical models that incorporates not only individual-level data, but also the population-based SDoH variables informing one's neighborhood and exposure is needed for comprehensive analysis that would give a pathway for planning targeted interventions in high-risk communities/clusters for mitigating COVID-19 infection and mortality. Such targeted interventions could include providing test kits, personal protective equipment, ventilators, financial needs for businesses and schools, and vaccines, when available. Information on factors related to SDoH,[80] such as one's "neighborhood and built environment, economic stability, education, social and community context, and health and health care" is often unavailable at the individual level, and aggregated data collected at the county or neighborhood level[28] help generate hypotheses that can be tested using analytical epidemiological studies in high-risk communities.

ORCID

Rajib Paul PhD 🔟 https://orcid.org/0000-0002-5120-7525
Ahmed Arif MBBS, PhD 🔟 https://orcid.org/0000-0002-0237-9937
Subhanwita Ghosh MS 🔟 https://orcid.org/0000-0002-0510-2917

REFERENCES

1. Social Determinants of Health. Available at: https://www.who.int/social_determinants/sdh_definition/en/. 2020. Accessed July 14, 2020.
2. Cockerham WC, Hamby BW, Oates GR. The social determinants of chronic disease. *Am J Prev Med.* 2017;52(1S1):S5-S12.
3. Marmot M. Social determinants of health inequalities. *Lancet.* 2005;365(9464):1099-1104.

4. Marmot M, Wilkinson R, Eds. *Social Determinants of Health*. Oxford University Press; 2005.

5. Adler NE, Glymour MM, Fielding J. Addressing social determinants of health and health inequalities. *JAMA*. 2016;316(16):1641-1642.

6. Social Determinants of Health. Available at: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health. 2020. Accessed November 4, 2020.

7. Staff AC. One in Five Americans Live in Rural Areas. Available at: https://www.census.gov/library/stories/2017/08/rural-america.html. 2017. Accessed January 21, 2021.

8. Lutfiyya MN, McCullough JE, Haller IV, Waring SC, Bianco JA, Lipsky MS. Rurality as a root or fundamental social determinant of health. *Dis Mon*. 2012;58(11):620-628.

9. Malatzky C, Cosgrave C, Gillespie J. The utility of conceptualisations of place and belonging in workforce retention: a proposal for future rural health research. *Health Place*. 2020;62:102279.

10. Hartley D. Rural health disparities, population health, and rural culture. *Am J Public Health*. 2004;94(10):1675-1678.

11. Downey LH. Rural populations and health: determinants, disparities, and solutions. *Prev Chronic Dis*. 2013;10:E104.

12. Taylor MM. *Rural Health Disparities: Public Health, Policy, and Planning Approaches*. Cham, Switzerland: Springer; 2019.

13. Haakenstad A, Hawkins SS, Pace LE, Cohen J. Rural-urban disparities in colonoscopies after the elimination of patient cost-sharing by the Affordable Care Act. *Prev Med*. 2019;129:105877.

14. Khairat S, Haithcoat T, Liu S, et al. Advancing health equity and access using telemedicine: a geospatial assessment. *J Am Med Inform Assoc*. 2019;26(8-9):796-805.

15. Kurani SS, McCoy RG, Lampman MA, et al. Association of neighborhood measures of social determinants of health with breast, cervical, and colorectal cancer screening rates in the US Midwest. *JAMA Netw Open*. 2020;3(3):e200618.

16. Ompad DC, Galea S, Caiaffa WT, Vlahov D. Social determinants of the health of urban populations: methodologic considerations. *J Urban Health*. 2007;84(3 Suppl):i42-i53.

17. Alicea-Alvarez N, Reeves K, Lucas M, et al. Urban communities: preparing future healthcare providers for "neighborhood-engaged care" through a community engagement course intervention. *J Urban Health*. 2016;93(4):732-743.

18. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Rep*. 2001;116(5):404-416.

19. Nieuwenhuis J, Tammaru T, Van Ham M, Hedman L, Manley D. Does segregation reduce socio-spatial mobility? Evidence from four European countries with different inequality and segregation contexts. *Urban Stud*. 2020;57(1):176-197.

20. Abrams EM, Szefler SJ. COVID-19 and the impact of social determinants of health. *Lancet Respir Med*. 2020;8(7):659-661.

21. Paul R, Arif AA, Adeyemi O, Ghosh S, Han D. Progression of COVID-19 from urban to rural areas in the United States: a spatiotemporal analysis of prevalence rates. *J Rural Health*. 2020;36(4):591-601.

22. Peters DJ. Community susceptibility and resiliency to COVID-19 across the rural-urban continuum in the United States. *J Rural Health*. 2020;36(3):446-456.

23. Henning-Smith C, Tuttle M, Kozhimannil KB. Unequal distribution of COVID-19 risk among rural residents by race and ethnicity. *J Rural Health*. 2021;37(1):224-226.

24. Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of May 2020. *J Rural Health*. 2020;36(3):433-445.

25. Guha A, Bonsu J, Dey A, Addison D. Community and socioeconomic factors associated with COVID-19 in the United States: zip code level cross sectional analysis. 2020. http://doi.org/10.1101/2020.04.19.20071944.

26. Millett GA, Jones AT, Benkeser D, et al. Assessing differential impacts of COVID-19 on black communities. *Ann Epidemiol*. 2020;47:37-44.

27. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891-1892.

28. Laurencin CT, McClinton A. The COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. *J Racial Ethn Health Disparities*. 2020;7(3):398-402.

29. McLaren J. Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census Data. NBER Working Papers 27407. National Bureau of Economic Research, Inc.; 2020

30. Chowkwanyun M, Reed Jr AL. Racial health disparities and Covid-19—caution and context. *N Engl J Med*. 2020;383:201-203.

31. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *Lancet*. 2020;395(10232):1243-1244.

32. Raifman MA, Raifman JR. Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. *Am J Prev Med*. 2020;59(1):137-139.

33. DiMaggio C, Klein M, Berry C, Frangos S. Blacks/African Americans are 5 times more likely to develop COVID-19: spatial modeling of New York city ZIP code-level testing results. *Ann Epidemiol*. 2020;51:7-13.

34. Marema T. Rural America's Daily Rate of New Infections Climbs 150% in Last Month. The Daily Yonder. July 15, 2020. Available at: https://dailyyonder.com/rural-americas-daily-rate-of-new-infections-climbs-150-in-last-month/2020/07/15/. Accessed July 13, 2020.

35. CSSEGISandData/COVID-19. Available at: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports. 2020. Accessed July 13, 2020.

36. Ranking Data & Documentation. Available at: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation. 2020. Accessed July 14, 2020.

37. U.S. Bureau of Labor Statistis. Available at: https://www.bls.gov/. Accessed June 27, 2020.

38. American Community Survey 5-Year Data (2009–2018). Available at: https://www.census.gov/data/developers/data-sets/acs-5year.html. 2019. Accessed July 7, 2020.

39. State and Metro Area Employment, Hours, & Earnings. Available at: https://www.bls.gov/sae/home.htm. 2020. Accessed July 14, 2020.

40. Current Population Survey (CPS). Available at: https://www.census.gov/programs-surveys/cps.html. Accessed July 14, 2020.

41. Economic Research Service: United States Department of Agriculture. County-level Data Sets. 2020. Available at: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/. Accessed July 14, 2020.

42. U.S. Diabetes Surveillance System. Available at: https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html. Accessed July 14, 2020.

43. NCHHSTP AtlasPlus. Available at: https://www.cdc.gov/nchhstp/atlas/index.htm. 2019. Accessed June 30, 2020.

44. Behavioral Risk Factor Surveillance System. Available at: https://www.cdc.gov/brfss/data_documentation/index.htm. 2019. Accessed July 14, 2020.

45. Economic Research Service: United States Department of Agriculture. Rural-Urban Commuting Area Codes. Available at: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx. 2019. Accessed July 11, 2020.

46. Hollander M, Wolfe DA, Chicken E. *Nonparametric Statistical Methods*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.

47. Wand M, Coull B, French J, et al. SemiPar 1.0. R Package. 2013. Available at: http://cranr-projectorg. Accessed November 7, 2020.

48. Team RC. R: A Language and Environment for Statistical Computing (Version 3.0. 2)[Computer software]. Vienna, Austria: R Foundation for Statistical Computing; 2019.

49. Institute CESR. ArcGIS Desktop: Release [computer program]. Version 10.8. Redlands, CA: ESRI; 2019.

50. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med*. 2016;4(7):136.

51. McGillc N. Social determinants of health: education attainment linked to health throughout life span. *Am J Public Health*. 2016;106(10):1719-1719.

52. Peters E, Baker DP, Dieckmann NF, Leon J, Collins J. Explaining the effect of education on health: a field study in Ghana. *Psychol Sci*. 2010;21(10):1369-1376.

53. Chen A. More Education: Lower Unemployment, Higher Earnings. Available at: https://www.bls.gov/careeroutlook/2017/data-on-display/more-education.htm?view_full. 2017. Accessed July 14, 2020.

54. Krug G, Eberl A. What explains the negative effect of unemployment on health? An analysis accounting for reverse causality. *Res Soc Strat Mob*. 2018;55:25-39.

55. Norström F, Waenerlund AK, Lindholm L, Nygren R, Sahlén KG, Brydsten A. Does unemployment contribute to poorer health-related quality of life among Swedish adults?. *BMC Public Health*. 2019;19(1):457.

56. Herbig B, Dragano N, Angerer P. Health in the long-term unemployed. *Dtsch Arztebl Int*. 2013;110(23-24):413-419.

57. Janssen EM, Longo DR, Bardsley JK, Bridges JF. Education and patient preferences for treating type 2 diabetes: a stratified discrete-choice experiment. *Patient Prefer Adherence*. 2017;11:1729-1736.

58. African American Employment. Available at: https://blackdemographics.com/economics/employment/. 2020. Accessed May 7, 2020.

59. Kantamneni N. The impact of the COVID-19 pandemic on marginalized populations in the United States: a research agenda. *J Voc Behav*. 2020;119:103439.

60. Peric S, Stulnig TM. Diabetes and COVID-19: disease-management-people. *Wien Klin Wochenschr*. 2020;132(13-14):356-361.

61. Wu ZH, Tang Y, Cheng Q. Diabetes increases the mortality of patients with COVID-19: a meta-analysis. *Acta Diabetol*. 2020:1-6. https://doi.org/10.1007/s00592-020-01546-0

62. Singh AK, Singh R. Does poor glucose control increase the severity and mortality in patients with diabetes and COVID-19? *Diabetes Metab Syndr*. 2020;14(5):725-727.

63. Wang M, Luo L, Bu H, Xia H. One case of coronavirus disease 2019 (COVID-19) in a patient co-infected by HIV with a low CD4$^+$ T-cell count. *Int J Infect Dis*. 2020;96:148-150.

64. Center for Disease Control and Prevention. What to Know About HIV and COVID-19. Available at: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/hiv.html. 2020. Accessed July 19, 2020.

65. Labor Force Characteristics by Race and Ethnicity, 2018. Available at: https://www.bls.gov/opub/reports/race-and-ethnicity/2018/home.htm. Accessed November 7, 2020.

66. Heidemann DL, Joseph NA, Kuchipudi A, Perkins DW, Drake S. Racial and economic disparities in diabetes in a large primary care patient population. *Ethn Dis*. 2016;26(1):85-90.

67. Ejike CO, Dransfield MT, Hansel NN, et al. Chronic obstructive pulmonary disease in America's Black population. *Am J Respir Crit Care Med*. 2019;200(4):423-430.

68. Van Dyke M, Greer S, Odom E, et al. Heart disease death rates among Blacks and Whites aged ≥35 years — United States, 1968–2015. *MMWR Surveill Summ*. 2018;67(SS-5):1-11.

69. Centers for Disease Control and Prevention. COVID-19: People with Certain Medical Conditions. Available at: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions. 2020. Accessed July 22, 2020.

70. Lutfi K, Trepka MJ, Fennie KP, Ibanez G, Gladwin H. Racial residential segregation and risky sexual behavior among non-Hispanic Blacks, 2006–2010. *Soc Sci Med*. 2015;140:95-103.

71. Yang TC, Zhao Y, Song Q. Residential segregation and racial disparities in self-rated health: how do dimensions of residential segregation matter? *Soc Sci Res*. 2017;61:29-42.

72. White K, Haas JS, Williams DR. Elucidating the role of place in health care disparities: the example of racial/ethnic residential segregation. *Health Serv Res*. 2012;47(3 Pt 2):1278-1299.

73. Hao Y, Landrine H, Smith T, Kaw C, Corral I, Stein K. Residential segregation and disparities in health-related quality of life among Black and White cancer survivors. *Health Psychol*. 2011;30(2):137-144.

74. Del Brutto OH, Mera RM, Recalde BY, Costa AF. Social determinants of health and risk of SARS-CoV-2 infection in community-dwelling older adults living in a rural Latin American setting. *J Community Health*. 2020:1-6. https://doi.org/10.1007/s10900-020-00887-9

75. Miller IF, Becker AD, Grenfell BT. Disease and healthcare burden of COVID-19 in the United States. *Nat Med*. 2020;26:1212-1217.

76. Wallace E. Vulnerable Populations in the Path of Hurricane Florence. Available at: https://www.policymap.com/2018/09/vulnerable-populations-in-the-path-of-hurricane-florence/. 2018. Accessed November 6, 2020.

77. Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health*. 2000;21:171-192.

78. Leyland AH, Goldstein H. *Multilevel Modelling of Health Statistics*. Hoboken, NJ: John Wiley & Sons; 2001.

79. Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Stat Med*. 2002;21(21):3291-3315.

80. Social Determinants of Health. Available at: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health. Accessed July 14, 2020.

81. Lindgren Finn, Rue Håvard. Bayesian Spatial Modelling withR-INLA. *Journal of Statistical Software*. 2015;63(19). http://doi.org/10.18637/jss.v063.i19.