

Neighborhood conditions and the initial outbreak of COVID-19: the case of Louisiana

Eric Joseph van Holm¹, Christopher K. Wyczalkowski², Prentiss A. Dantzer²

¹Department of Political Science, University of New Orleans, New Orleans, LA 70148, USA

²Urban Studies Institute, Georgia State University, Atlanta, GA 30302, USA

Address correspondence to Eric Joseph van Holm, Email: evanholm@uno.edu

ABSTRACT

The early outbreak of coronavirus disease-2019 (COVID)-19 became associated with various ‘hot spots’ in the USA, particularly in large cities. However, despite the widespread nature of the outbreak, much of what is known about the virus’ impact and clusters is understood either for individuals, or at the state level. This paper assesses the predictors of outbreaks at the neighborhood level. Using data from the Louisiana Department of Health, we use spatial regression models to analyze the case count through 3 May 2020 and its relationship to individual and geographic neighborhood characteristics at the census tract level. We find a particularly strong and large correlation between race and COVID-19 cases, robust to model specification and spatial autocorrelation. In addition, neighborhoods with lower rates of poverty and those with fewer residents over 70 have fewer cases. Policy makers should adjust testing strategies to better service the hardest hit populations, particularly minorities and the elderly. In addition, the results are greater evidence of the impact of systemic issues on health, which require a long-term strategy for redress.

Introduction

The early outbreak of coronavirus disease-2019 (COVID-19) is associated with various ‘hot spots’ in the USA, particularly in large cities. However, cases were not long contained within just large metropolitan areas and by early May, two-thirds of counties across the USA reported at least one confirmed case from COVID-19. Despite the widespread nature of the outbreak, much of what is known about the virus’s impact and clusters is understood either for individuals, or at much larger scales, such as the county or state level. However, state and even county level aggregations mask heterogeneity inherent over such large areas, while individual characteristics can overlook structural effects.

Neighborhoods have long been found to affect an individual’s health. For instance, low neighborhood scores on walkability suggest reduced activity, and are associated with increased body mass index and lower physical health.^{1–3} In addition, residential segregation is associated with worse health outcomes for minorities across a spectrum of outcomes.^{4–6} The neighborhood effects dictating the differential spread of diseases begs considerable attention as health professions and policymakers respond to COVID-19.

In this paper, we analyze case counts at the census tract level and evaluate the relationship of COVID-19 cases with

pre-existing neighborhood conditions in Louisiana. In the following section, we briefly review the methodology and data used in the study, before discussing the results. Robust to several model specifications and accounting for spatial autocorrelation, we find a strong positive relationship between the percent of minority population and cases of COVID-19, holding socioeconomic and structural factors constant.

Study design

We treat this initial spread of the virus as one time period, and examine the relationship between these initial COVID-19 cases with pre-existing cross-sectional neighborhood socioeconomic conditions in the State of Louisiana, USA. In late April reported new COVID-19 cases in Louisiana slowed, and on May 15 Louisiana Governor John Edwards began Phase 1 of reopening,⁷ making this an appropriate period for study of the initial outbreak.

Eric Joseph van Holm, Assistant Professor

Christopher Wyczalkowski, Postdoctoral Researcher

Prentiss A. Dantzer, Assistant Professor

COVID-19 Cases at Tract Level in Louisiana, through May3, 2020

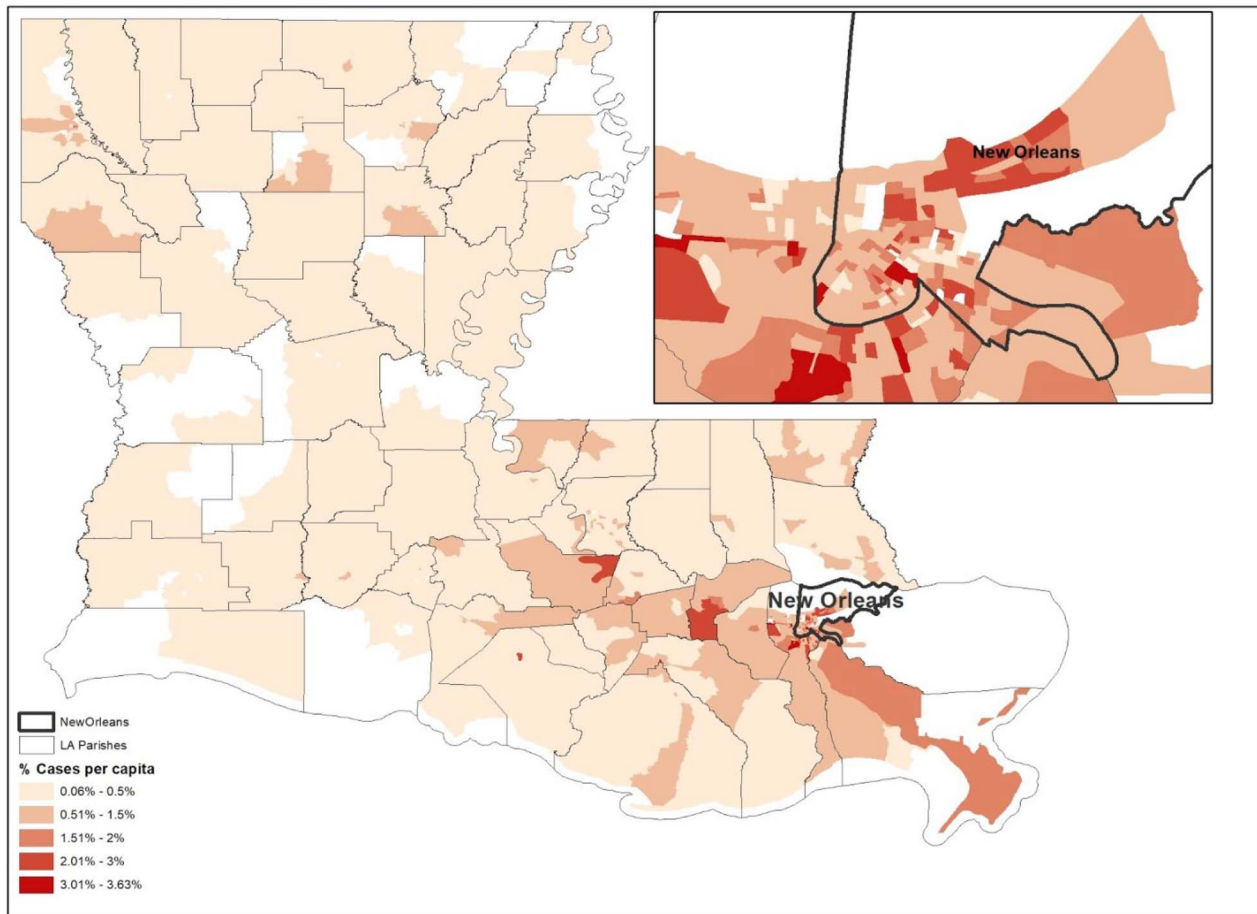


Fig. 1 Covid-19 cases at tract level in Louisiana, through 3 May 2020.

Data

To understand how neighborhood characteristics impact the size of the COVID-19 outbreak, we use data from the Louisiana Department of Health (LDH) and the US Census. Our primary dependent variable is the number of confirmed cases at the census tract level in the state of Louisiana, obtained online from the LDH. The number of COVID-19 cases should be expected to be influenced by the number of residents in a census tract, similarly to most other counts within a population, so we use the per capita figure (per 1000 residents) as our dependent variable. LDH publishes data each week, we used the May 3rd edition to capture the outbreak of COVID-19 prior to reopening.

Cases per capita are mapped as shown in [Figure 1](#). The highest concentrations of cases per capita are in the more urbanized areas in and near New Orleans.

The LDH data present several limitations, including missing data for low population tracts and aggregation

of tract with fewer than five cases into a 1–5 category. We defaulted to using the five for all these census tracts, but adjusting the figure imputed for each digit between 1 and 5 had no meaningful effect on the regression results.

Variables related to neighborhood conditions are gathered from the 2018 5-Year American Community Survey at the census tract level. We model five aspects of neighborhoods for their ability to explain the size of COVID-19 outbreaks: Density (population density and % of individuals who commute to work by any form of public transit), travel (percentage of the population employed in tourism and the percentage that commute to a different county for work), socioeconomic status (percentage of college graduates, percentage in poverty and the percentage that do not have health insurance), race (percentage Black, percentage Latinx, percentage Asian and percentage other minority), and demographics (percentage male and percentage over 70).

Table 1 Summary statistics

Variable	Mean	SD	Min	Max	Description
Cases per capita (logged)	6.13	6.34	0.58	36.26	Number of cases per 1000 residents (log transformed in model)
Population density (logged)	2567	3230	0.81	31450	Population per square mile of land (log transformed in model)
% Taking public Transit	0.02	0.05	0	0.39	% of census tract who use any form of public transit as primary means of transportation to work
% Working outside county	0.29	0.18	0	0.86	% of census tract that work in a different county
% Tourism employment	0.12	0.07	0	0.47	% of census tract employed in arts, entertainment, and recreation, and accommodation and food services
% College graduate	0.23	0.17	0	1	% of census tract who graduated college or higher
% In Poverty	0.22	0.13	0	0.84	% of census tract in poverty
% No Insurance	0.11	0.05	0	0.38	% of census tract without health insurance
% Black	0.37	0.3	0	1	% of census tract black
% Latinx	0.05	0.06	0	0.49	% of census tract Latinx
% Asian	0.02	0.03	0	0.44	% of census tract Asian
% Other minority	0.02	0.03	0	0.25	% of census tract other non-White
Diversity index	0.39	0.17	0	0.72	How far the census tract is from a balanced shared of different racial groups
% Over 70	0.1	0.04	0	0.5	% of census tract over the age of 70
Male	0.49	0.06	0.28	1	% of census tract that is male

Descriptive statistics for the dependent and independent variable along with a description of their measures are in Table 1.

Method

Ordinary least squares regression (OLS) has been used to evaluate neighborhood scale socioeconomic relationships, but suffers from potential spatial autocorrelation.^{8,9} The first law of geography suggests that everything is related to everything else, but things near to each other are more related. Therefore, the spatial structure of a place can also have an effect on any modeled neighborhood scale socioeconomic relationship.⁸ This applies to both measured and unmeasured characteristics of a neighborhood. Spatial models account for spatial endogeneity by identifying spatial relationships in a weight matrix, which is used to address omitted variables and spatial error.^{10,11} Omitted variables, such as can be expected when modeling using a limited set of explanatory variables, can exacerbate the bias arising from spatial dependence¹²; A more detailed discussion is provided by Anselin,⁸ Anselin *et al.*¹¹ and Conway *et al.*⁹ among others. We address spatial endogeneity with spatial lag and spatial error models.

We first fit the standard OLS model, ignoring potential spatial dependencies.

$$Y = X\beta + \varepsilon,$$

where Y is the dependent variable (natural log of cases) and β is the coefficient on the vector of control variables. If location is a contributing driver of the effect, the resulting problem for an OLS model is correlation of the residuals in the error term, violating the uncorrelated error assumption, and leading to potentially biased and inconsistent results.

The spatial lag model attempts to correct for spatial dependence by introducing a term for the effect of the spatially lagged Y on Y ; i.e. the outcome variable in location i is influenced by adjoining location j .

$$Y = \rho WY + X\beta + \varepsilon,$$

where Y is the dependent variable, ρ is the lag coefficient, W is the spatial weight matrix, β is the coefficient for a vector of neighborhood characteristics, and ε is the error term.

A spatial error model assumes that there is correlation in the error term, and uses the spatial weight matrix in the error term to model error from potential misspecifications.^{9,11}

$$Y = X\beta + \lambda W\varepsilon + \nu,$$

where W is again the weight matrix, λ is the coefficient, ε is the residual error matrix, and ν is the independent error transformed to a distribution with a mean of zero. We present findings for all three models, the OLS, Spatial lag and spatial

Table 2 Combined regression results

	<i>Dependent variable</i>		
	Cases per 1000 residents		
	<i>OLS</i> (1)	<i>Spatial error</i> (2)	<i>Spatial lag</i> (3)
Population density (logged) (0.012) (0.012) (0.009)	0.101***	0.015	0.034***
% Taking public transit (0.460) (0.399)	4.063*** (0.324)	0.352	0.969***
% Working outside county	1.325*** (0.100)	0.117 (0.101)	0.454*** (0.071)
% Tourism employment	0.540* (0.315)	0.066 (0.216)	0.172 (0.216)
% College graduate	0.173 (0.138)	−0.303** (0.135)	−0.030 (0.095)
% In poverty	−1.096*** (0.194)	−0.501*** (0.149)	−0.532*** (0.135)
% No insurance	−0.976** (0.412)	0.122 (0.280)	0.041 (0.284)
% Black	1.417*** (0.090)	0.965*** (0.078)	0.762*** (0.067)
% Latinx	3.340*** (0.367)	0.679** (0.294)	1.148*** (0.258)
% Asian	2.202*** (0.521)	0.445 (0.476)	0.457 (0.361)
% Other minority	2.025*** (0.691)	0.247 (0.519)	0.771 (0.475)
Diversity Index	−0.228* (0.118)	−0.065 (0.092)	−0.022 (0.082)
% Over 70	3.290*** (0.437)	2.135*** (0.316)	2.193*** (0.301)
Male	0.244 (0.377)	0.377 (0.236)	0.261 (0.259)
Constant	−0.393* (0.231)	0.893*** (0.186)	−0.472*** (0.159)
Observations	1,063	1,063	1,063
R2	0.544	0.781	0.792
Akaike info criterion	1594.04	960.663	995.104

* $P < 0.1$; ** $P < 0.05$; *** $P < 0.01$.

error, to compare results and place the greatest confidence in those variables with consistent results across the different specifications.

Results

The OLS model finds that most of the variables behave as predicted. Both variables for measuring density,

population density and public transit, are significant and positive, indicating that people being in closer proximity is associated with higher case counts of COVID-19. Although tourism employment is weakly significant, the share of individuals that work in different county is associated with a higher number of COVID-19 cases per capita.

The share of individuals in poverty and the percent of the population that is uninsured is associated with fewer cases, contrary to expectations. In addition, the share of college

Table 3 Lagrange multiplier diagnostics for spatial dependence

Model	Statistic
Spatial lag	757.19***
Spatial error	549.15***

*** $P < 0.01$.

graduates makes an insignificant difference, although it too is not in the predicted direction.

There is a consistent and strong relationship between race and COVID-19. Across all four minority groups, there is a significant and positive relationship with the size of the per capita outbreak. Diversity within a neighborhood however makes only a weak difference.

Finally, the size of the elderly population within a neighborhood, specifically the share of residents that are over 70, has a large and positive effect, while the share that is male is insignificant.

It should be noted, OLS results can suffer from spatial autocorrelation as outlined above; a test of spatial autocorrelation Morans I finds significant clustering. Spatial error model separates out and weights the component of the error term that corresponds to the spatial autocorrelation. In the absence of spatial autocorrelation, the spatial error model would yield results identical to OLS.

However, there are small changes in the results between the first two columns in Table 3. The direction of variables had no significant variables change, although the magnitude of the differences do shift between the two models. In addition, population density, public transit ridership, working outside of the county, the share of uninsured and two racial variables lose significance.

Finally, we test a spatial lag model. The Lagrange multiplier statistic tests the hypothesis that an outcome variable in location i is influenced by the value of the outcome variable in location j , i.e., if spatial lag is present in the data.⁸ Although Conway *et al.*⁹ suggest the ad hoc rule that the model with the higher Lagrange Multiplier test statistic is the more appropriate model, which would suggest utilizing the spatial lag model. However, the test statistics are not very different in Table 3 (757.19 versus 549.15 for the lag and error models, respectively).

The spatial lag model finds fewer differences from the OLS results. Here, there is a positive and significant relationship between population density and the use of public transit. Unlike the spatial error mode, the percentage of residents working in a different county is significant and positive. However, the share of poverty continues to have a significant

and negative relationship with the number of confirmed COVID-19 cases. Finally, we find a consistently significant relationship between the percentage of Blacks and Latinx in a neighborhood and the number of COVID-19 cases per capita.

Conclusion

In this paper we tested the impact of proximity, travel, socioeconomic status, and race on the level of outbreak from COVID-19 at the neighborhood level. Race, poverty, and age have the most consistent effect on the size of COVID-19 outbreaks. In particular, the share of African Americans or Latinx in a neighborhood is a significant and large positive predictor for the size of the outbreak per capita. These results indicate these communities have been hardest hit by the novel coronavirus, which supports most contemporary reporting on the problem and thus add to the mounting evidence of how the virus has disproportionately impacted minorities and their communities.

However, with a neighborhood level analysis we are unable to identify the exact causes of the findings and several mechanisms may be at play. There are several possible explanations for why minorities, and African Americans in particular, would be more likely to live in areas with higher counts. We have been able to account for their potential greater reliance on public transit and living in higher housing density, higher rates of poverty, and lack of access to health insurance, all of which may explain the higher counts of COVID-19. Thus, these factors do not fully explain the differences. It is possible that such communities have not followed government guidelines to prevent the spread of the virus, but van Holm *et al.*,¹³ showed nearly zero significant differences between the races with regard to trips taken, social distancing, hand washing or other associated behaviors.

A virus is not motivated by racial animus. The heightened impacts on minorities is thus evidence of the way that American society shapes the lives and environments of different groups, putting minority communities at greater risk. To fully understand the causes of the COVID-19 outbreak in minority communities, researchers will have to grapple with the cumulative impact of economic, social and spatial inequities throughout society.

The relationship between socioeconomic status and COVID-19 is somewhat contradictory in the model. The percentage of college graduates and poverty percentage, despite being negatively correlated, both predict a negative relationship with the number of COVID-19 cases in a census tract (although education is only significant in the spatial error model). That leads to the question of why higher

socioeconomic status would both predict more and fewer cases? That finding likely reflects that the two measures are capturing different aspects of socioeconomic status. Individuals with a high socioeconomic status were likely able to protect themselves and more strictly social distance, reducing their hazard with the virus. For those with low socioeconomic status, such distancing would be difficult, but they may be less likely to get tested, thus reducing the case count.

These results are collectively indicative of the complexity of socioeconomic status and the COVID-19 outbreak, and the challenges of using confirmed cases as a measure. A more aggressive testing procedure will raise the number of cases in a community, while those areas where only the very ill are tested will appear to have fewer cases. Future research may be able to better disentangle the relationship between socioeconomic status and COVID-19 using results from antibody tests, or future data once testing is more regularly and consistently available.

There are lessons for cities, as well. Cities have seen a rise in inner-city residents and a return to the cities over the last two decades, and the forces causing this effect are likely to continue. New York City public transit has received significant attention for the potential (and hotly disputed) role in the outbreak in New York and surrounding states. But in the case of Louisiana we see similar trends with more dense neighborhoods with higher public transit ridership having larger per-capita outbreaks. Governments should reverse the funding shortfalls and promote increased capacity and systems to achieve safe use of public transportation.

The results here show that living in a suburban community does not inoculate one from COVID-19. The variable measuring the share of individuals that work in a different county was significant and positive, indicating that longer commutes help to spread the virus. However, that pattern follows a specific trend, as urban areas have far fewer individuals commuting out of the county for work. As the variable is positive, it is likely that individuals working in the city, which are denser and have higher populations, pick up the virus and carry it out to outlying areas.

Conflict of interest

There are no conflicts of interest to report.

References

- 1 Frank LD, Sallis JF, Conway TL *et al.* Many pathways from land use to health: associations between Neighborhood walkability and active transportation, body mass index, and air quality. *J Am Plann Assoc* 2006;**72**(1):75–87.
- 2 Speck J. *Walkable City: How Downtown Can Save America, One Step at a Time*. New York, NY: Macmillan, 2013.
- 3 Villanueva K, Knuiman M, Nathan A *et al.* The impact of neighborhood walkability on walking: does it differ across adult life stage and does neighborhood buffer size matter? *Health Place* 2014;**25**:43–6.
- 4 Acevedo-Garcia D. Zip code-level risk factors for tuberculosis: Neighborhood environment and residential segregation in New Jersey, 1985–1992. *Am J Public Health* 2001;**91**(5):734–41.
- 5 Greer S, Casper M, Kramer M *et al.* Racial residential segregation and stroke mortality in Atlanta. *Ethn Dis* 2011;**21**(4):437–43.
- 6 White K, Borrell LN. Racial/ethnic residential segregation: framing the context of health risk and health disparities. *Health Place* 2011;**17**(2):438–48.
- 7 Edwards, J. B. (2020). *Louisiana will Move to Phase One Statewide on May 15*. <https://gov.louisiana.gov/index.cfm/newsroom/detail/2488> (19 August 2020, date last accessed).
- 8 Anselin L. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geogr Analysis* 1988;**20**(1):1–17.
- 9 Conway D, Li CQ, Wolch J *et al.* A spatial autocorrelation approach for examining the effects of urban Greenspace on residential property values. *J Real Estate Financ Econ* 2010;**41**(2):150–69.
- 10 Anselin L. *Spatial Econometrics: Methods and Models*. Berlin/Heidelberg, Germany: Springer Science & Business Media, 2013.
- 11 Anselin L, Bera AK, Florax R *et al.* Simple diagnostic tests for spatial dependence. *Regional Sci Urban Econ* 1996;**26**(1):77–104.
- 12 Pace, R. K., & LeSage, J. P. (2010). Omitted Variable Biases of OLS and Spatial Lag Models. In A. Páez, J. Gallo, R. N. Buliung, & S. Dall’erba (Eds.), *Progress in Spatial Analysis: Methods and Applications* (pp. 17–28). Springer.
- 13 van Holm, E. J., Monaghan, J., Shahar, D., *et al.* (2020). *The Impact of Political Ideology on Concern and Behavior During COVID-19*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3573224 (19 August 2020, date last accessed).