ORIGINAL ARTICLE

Progression of COVID-19 From Urban to Rural Areas in the United States: A Spatiotemporal Analysis of Prevalence Rates

Rajib Paul, PhD ;[1] Ahmed A. Arif, MBBS, PhD ;[1] Oluwaseun Adeyemi, MBChB, MPH;[1] Subhanwita Ghosh, MS;[1] & Dan Han, PhD[2]

1 Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina
2 Department of Mathematics, University of Louisville, Louisville, Kentucky

Abstract

Purpose: There are growing signs that the COVID-19 virus has started to spread to rural areas and can impact the rural health care system that is already stretched and lacks resources. To aid in the legislative decision process and proper channelizing of resources, we estimated and compared the county-level change in prevalence rates of COVID-19 by rural-urban status over 3 weeks. Additionally, we identified hotspots based on estimated prevalence rates.

Methods: We used crowdsourced data on COVID-19 and linked them to county-level demographics, smoking rates, and chronic diseases. We fitted a Bayesian hierarchical spatiotemporal model using the Markov Chain Monte Carlo algorithm in R-studio. We mapped the estimated prevalence rates using ArcGIS 10.8, and identified hotspots using Gettis-Ord local statistics.

Findings: In the rural counties, the mean prevalence of COVID-19 increased from 3.6 per 100,000 population to 43.6 per 100,000 within 3 weeks from April 3 to April 22, 2020. In the urban counties, the median prevalence of COVID-19 increased from 10.1 per 100,000 population to 107.6 per 100,000 within the same period. The COVID-19 adjusted prevalence rates in rural counties were substantially elevated in counties with higher black populations, smoking rates, and obesity rates. Counties with high rates of people aged 25-49 years had increased COVID-19 prevalence rates.

Conclusions: Our findings show a rapid spread of COVID-19 across urban and rural areas in 21 days. Studies based on quality data are needed to explain further the role of social determinants of health on COVID-19 prevalence.

Key words Bayesian influence, disease hotspots, pandemic, geographic disparity, respiratory disease.

COVID-19 is a highly contagious novel coronavirus that has affected more than 7 million people worldwide, resulting in more than 418,000 deaths as of June 11, 2020.[1] In the United States, more than 113,000 people have died due to COVID-19, as of June 11, 2020.[2] The exact mechanism by which COVID-19 spreads from person to person is still under investigation. However, the virus is thought to spread mainly through respiratory droplets and environmental surfaces. It can cause severe lower respiratory illnesses like pneumonia, resulting in death. The large percentage of hospitalizations and deaths due to COVID-19 are among older individuals, aged 65 and above. However, younger people are getting infected by the virus at higher rates.[3] In the absence of vaccine availability, measures such as safe physical distancing and banning gathering are primary preventative measures in reducing the spread of infection and flattening the epidemiological curve.[4]

About 63% of the US counties are classified as rural. However, only 15% of the US population live in rural areas.[5] The rural population is mostly white, poor, older, has higher smoking rates, high blood pressure, and high

rates of obesity as compared to their urban counterparts.[5] The mortality rates from heart disease, cancer, respiratory diseases, and stroke are higher among people living in the rural areas as compared to those living in the urban areas.[6,7]

The first reported case of COVID-19 in a rural county was on February 20 in Humboldt County, Northern California.[8] There are growing signs that the COVID-19 virus has started to spread to rural areas.[9] Most rural areas lack public health infrastructure, and the current health care system, which is already stretched and lacks resources, may not be ready to deal with the sudden influx of patients.[10,11]

We used crowdsourced data, and spatiotemporal Bayesian models[12,13] to (1) estimate and compare the county-wise change in prevalence rates of COVID-19 by rural-urban status, (2) identify hotspots based on estimated prevalence rates, (3) find the association of demographic, smoking, and chronic diseases with COVID-19 prevalence rates and how they vary by rural/urban designation of counties, and (4) identify counties showing a significant increase or decrease of the percentage change in prevalence rates over 14 days. To the best of our knowledge, our research is the first attempt in estimating prevalence at the county level using Bayesian models that takes into account spatiotemporal autocorrelations.[14]

Methods

Study Design

This space-time study used a panel design, with the US counties and county-equivalents (hereafter referred to as "counties") as the spatial units of analysis. We restricted our analysis to the US contiguous states. Each county was identified using the 2018 5-digit Federal Information Processing Standards (FIPS) codes.[15] The county-level cumulative counts of COVID-19 infection and deaths were obtained from the publicly available data repository of the Johns Hopkins University (JHU).[16]

Selection Criteria

The cumulative confirmed COVID-19 county-level data between March 15 and April 22, 2020 were extracted as a time-series format. The daily county-level COVID-19 prevalence rates were computed as the difference between the reported cumulative count of the day of interest and the cumulative counts of the preceding 21 days (assuming on average a 3-week recovery period[17]) divided by the estimated county population. The daily deaths counts of the preceding day were removed from the numerator. Thus, in our inferential analysis, we used

prevalence rates over 3 weeks: April 3, 2020 to April 22, 2020. Data cleaning was achieved by assessing the range of daily prevalence counts. An a priori decision that daily incidence counts will be zero or higher was made, and dates with data entry inconsistency were corrected by selecting the counts of the preceding days.

County Characteristics

We used county-level demographics, smoking rates, and rates of chronic diseases (including diabetes and obesity) as independent variables. County demographics include county population size, age, and racial distributions. This information was extracted from the American Community Survey's (ACS) 2018 estimates.[18] County-level diabetes and obesity rates were obtained from the County Health Rankings website.[19]

Variables Definition

The outcome variable was daily county-level COVID-19 prevalence rates over $T = 20$ (April 3, 2020 to April 22, 2020) days across n = 3,108 counties in the continental United States. The daily prevalence rate of COVID-19 was measured as the number of active cases per 100,000 population. RUCA codes were used to designate urban and rural counties.[20] Urban areas were classified as all metropolitan areas as well as high commuting micropolitan. Rural counties included 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. Other county-level independent variables included the percent of residents aged 25-49 years, percent of the black population, adult smoking rates, diabetes rates, and obesity rates.

Data Analysis

Our data analyses consisted of descriptive and inferential statistical techniques. In descriptive analyses, we computed summary statistics for county-level demographics, smoking rates, and health conditions using chi-square and Mann-Whitney 2-sample test for rural and urban counties.

Our inferential analysis was based on a Bayesian Spatiotemporal Model (BSTM)[12,13,21] that used low-rank spatial autocorrelation techniques.[22] Recall that our outcome measure was daily county-level prevalence rates per 100,000 from April 3 through April 22, 2020. Since prevalence data were semicontinuous due to occurrences of zero, we needed a statistical distribution that incorporates point mass at zero for the counties with no COVID-19 case and a continuous distribution on the counties where we

had nonzero prevalence. Specifically, we used a skewed Tobit model for $Y(\mathbf{s}_i, t)$, the prevalence for ith county on tth day and modeled as

$$Y(\mathbf{s}_i, t) = \max(0, W(\mathbf{s}_i, t))^{\lambda}, \qquad (1)$$

where λ is a power transformation parameter that takes into account the skewness in the data and $W(\mathbf{s}_i, t)$ is a latent Gaussian space-time distribution that is modeled using a set of p independent variables $\mathbf{x}_p(\mathbf{s}_i, t)$ as

$$W(\mathbf{s}_i, t) \equiv \beta \mathbf{x}_p(\mathbf{s}_i, t) + \mathbf{A}(\mathbf{s}_i)\psi(t) + \nu(\mathbf{s}_i, t). \qquad (2)$$

The first term on the left-hand side of Equation (2) characterizes prevalence in terms of independent variables via regression coefficients β. The flexibility of this modeling approach allows one to make these regression coefficients dynamic and varying over counties. The second term $\psi(t)$ is a spatial-temporal autoregressive Gaussian process of dimension $m << n$ defined over m knot locations on the entire United States:

$$\psi(t) = \rho\psi(t-1) + \varepsilon(t). \qquad (3)$$

The parameter ρ captures how the process evolves with time; a positive value indicates an increase over time. The term $\varepsilon(t)$ is a zero-mean spatially correlated Gaussian process with covariance matrix $\boldsymbol{\Sigma}$ that is characterized by exponential covariance function using geodesic distances between county centroids (see Chapters 2 and 3 of Ref. 14). The vector $\mathbf{A}(\mathbf{s}_i)$ in Equation (2) maps the original process on m knot locations. Finally, the last term $\nu(\mathbf{s}_i, t)$ in Equation (2) is a white-noise zero-mean process with a constant variance that denoises the data (see Ref. 12).

We fitted our model using Markov Chain Monte Carlo algorithms (see Ref. 23) using R/RStudio version 3.6.3,[24,25] and spate, spTimer packages.[21,26,27] The effects of county-level independent variables were assessed using 95% credible intervals (CrI). Finally, we computed percent changes in daily prevalence from the fitted prevalence curves and evaluated whether it significantly increased or decreased using a linear trend equation of time and t-statistics. We plotted our estimated prevalence rates from model fitting using ArcMap 10.8.[28] Additionally, we conducted hotspot analysis using Getis-Ord local statistics.[29] Hotspots are defined as high values of prevalence rates concentrated in spatial clusters. We used 90%, 95%, and 99% cut-off values for assessing the significance of hotspots.

Results

Recall that we used data from March 15 through April 22, 2020, for calculating prevalence rates assuming a 3-week recovery period. This gave us 3 weeks of data on prevalence from April 3 to April 22, 2020, for inferential analysis and model fitting. The overall mean prevalence rate of COVID-19 as of April 3, 2020 was 5.7 per 100,000 population. The value increased to 23.6 per 100,000 on April 22, 2020—a 400% increase. The mean prevalence in urban counties increased from 10.1 per 100,000 on April 3, 2020 to 107.6 per 100,000 population on April 22, 2020. In the rural counties, the mean prevalence of COVID-19 increased from 3.6 per 100,000 population to 43.6 per 100,000 (Figure 1).

Table 1 summarizes the variables used in terms of proportions or medians and interquartile ranges (IQR). As of March 15, 2020, 79% urban and 3% rural counties had confirmed COVID-19 cases. These percentages increased to 98% urban and 84% rural counties within 5 weeks. Based on median estimates, in urban counties, about 37.8% of the population was between 25 and 49 years as compared to 35% in rural counties ($P < .0001$). Most of the rural population were white, obese, and smokers as compared to urban counterparts ($P < .0001$). There was also a statistically significant difference in diabetes rates between urban and rural counties (Table 1).

The overall prevalence rate of COVID-19 infection increased by 3.19 (95% CrI: 3.05, 3.32) per 100,000 population for 1% increase in population aged 25-49 years (Table 2). Figure 2 shows plots of estimated prevalence rates for all rural and urban counties in the United States. For ease of comparison, the square root of rates was plotted.

Figure 2(a) represents square root of prevalence curves for 2,107 rural counties, and Figure 2(b) represents square root of prevalence curves for 1,001 urban counties. The median square root prevalence rate for urban counties over the 20-day study period increased at a steeper rate than the median square root prevalence rate for rural counties (black lines). The red line in Figure 2(a) denotes the square root prevalence for Plaquemines Parish, Louisiana. The red line in Figure 2(b) denotes the square root prevalence for New York City, and the green line indicates the same for New Orleans, Louisiana. The prevalence curve for New York City was increasing, while the curves for Plaquemines Parish and New Orleans were quadratic.

Table 3 displays the adjusted prevalence rate ratio and changes in prevalence rates for urban and rural counties. The county-level COVID-19 prevalence rate ratio was 0.78 times (95% CrI = 0.77, 0.80) lower in rural counties as compared to urban counties, adjusted for covariates. The population aged 25-49 years had substantially higher prevalence in rural counties. Similarly, the COVID-19 adjusted prevalence rates were substantially elevated in counties with higher black populations; the prevalence increased by 0.57 per 100,000 for each percent increase in the black population (95% CrI: 0.51, 0.63). The association was more influential in rural counties. The

Figure 1 Median Prevalence Trend of COVID-19 Infection From the Observed Data Before Denoising. The triangles represent urban median prevalence rates and the circles represent rural median prevalence rates.
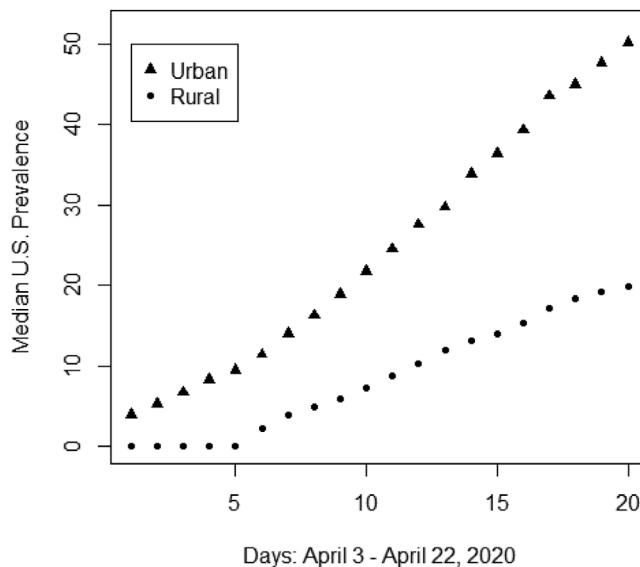

Table 1 Summary Statistics of the Study Sample

Variable	Urban	Rural	*P* value
Proportion of counties with confirmed cases as of 3/15[a]	0.79	0.03	<.0001
Proportion of counties with confirmed cases as of 4/22[a]	0.98	0.84	<.0001
Percent of population between 25 and 49 years	37.79 (3.44)	34.85 (4.10)	<.0001
Percent of African American population	7.41 (15.55)	1.44 (4.99)	<.0001
Diabetes mellitus rate	11 (4)	12 (4)	<.0001
Percent of adult smokers	17.00 (4.68)	17.51 (5.29)	<.0001
Obesity rate in percentage	31.40 (6.10)	32.70 (5.40)	<.0001

[a]Proportions are used as summary statistics and compared using chi-square tests. For the remaining variables, medians and interquartile ranges, IQR, are displayed and compared using Mann-Whitney 2-sample test.

Table 2 Age-Adjusted Rate Changes in COVID-19 Infection in Rural and Urban Counties in the United States

County variables	Overall Rate change (95% CrI)	Rural Rate change (95% CrI)	Urban Rate change (95% CrI)
Age 25-49 years[a]	3.19 (3.05, 3.32)	2.36 (2.22, 2.51)	2.80 (2.66, 2.93)

[a]Only the percent of population aged 25-49 years was used in the spatiotemporal model.

county-level smoking and obesity rates were positively associated with COVID-19 infection. However, the prevalence rates were negatively associated with county-level diabetes prevalence. Each percent increase in adult smokers in rural counties increased the prevalence rate by 0.46 per 100,000 population, and in urban counties, this increment was 0.51 per 100,000 population, when adjusted for covariates. Obesity rates were associated with increased

prevalence in urban counties only, whereas diabetes had a negative association with COVID-19 prevalence.

Figure 3 shows the estimated COVID-19 prevalence rates for 5 selected days during the 20-day study period. In the beginning, on April 3, 2020, the prevalence rates were spatially smooth. By April 22, the COVID-19 infection had spread to most northeast and southern states, and several hotspots were noted in large metropolitan as

Figure 2 Estimated (Denoised) Prevalence Rates From Fitted Spatiotemporal Model for (a) Rural and (b) Urban Counties. The black lines indicate median prevalence rates. Gray lines represent prevalence curves for 2,107 rural and 1,001 urban counties. Square root of rates are plotted for better comparison. The red line in plot (a) denotes the prevalence for Plaquemines Parish, Louisiana. The red line in plot (b) denotes the prevalence for New York City and the green line indicates the prevalence plot for New Orleans, Louisiana.
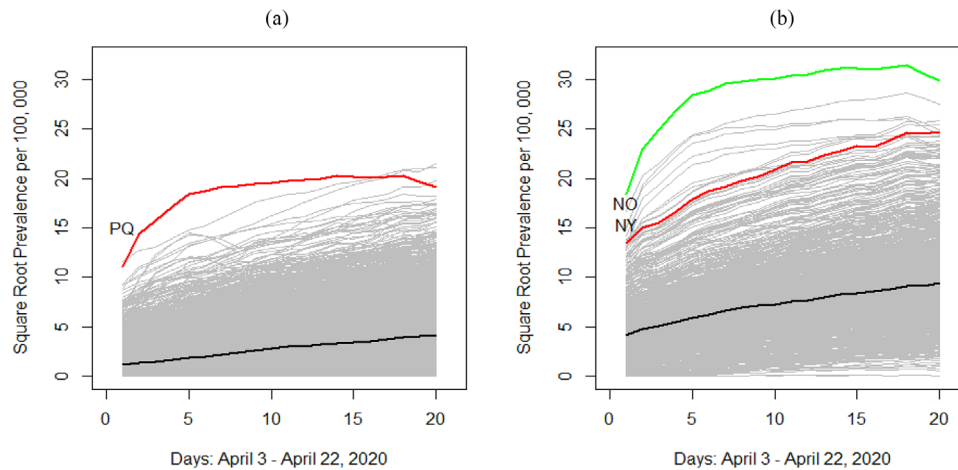

Table 3 Spatiotemporal Analysis of COVID-19 Infection in Rural and Urban Counties in the United States

County variables	Overall Adjusted rate change (95% CrI)	Rural Adjusted rate change (95% CrI)	Urban Adjusted rate change (95% CrI)
Rural	0.78 (0.77, 0.80)	–	–
Percent of age 25-49 years	1.82 (1.66, 1.97)	2.39 (2.14, 2.64)	0.39 (0.13, 0.66)
Percent of black	0.57 (0.51, 0.63)	0.67 (0.59, 0.75)	0.51 (0.41, 0.61)
Percent of smokers	0.59 (0.40, 0.79)	0.46 (0.24, 0.67)	0.51 (0.23, 0.79)
Percent with diabetes	−2.65 (−2.98, −2.31)	−1.94 (−2.47, −1.42)	−4.13 (−4.63, −3.63)
Percent with obesity	0.24 (0.08, 0.39)	0.16 (−0.09, 0.41)	0.34 (0.03, 0.65)

Note: The adjusted model was controlled for all variables included in the table simultaneously. Rural variable is binary and if its interval includes 1, we concluded evidence of no influence. For the remaining continuous independent variables, the inclusion of 0 indicates evidence of no influence.

well as small rural counties such as Apache, Navajo, and Coconino in Arizona (Figure 4).

Lastly, we mapped counties showing a significant increasing, decreasing, or stable pattern of daily percentage change of prevalence rates over 14 days (Figure 5). There were 580 counties, mostly in the southern and southeastern states, that showed significantly decreasing percent change. However, a cluster of 5 counties in Nevada (Churchill, Elko, Eureka, Lander, and Perishing), 2 counties in Arizona (Gila and Yavapai), and 1 county in Kansas (Wallace) showed significant (at 5% level of significance) increasing percentage change over a 14-day period. For most of the counties, the percentage change in the COVID-19 prevalence was stable as of April 22, 2020.

Discussion

This study demonstrates the spatiotemporal association of demographic, smoking, and chronic diseases with COVID-19 prevalence at a granular level in rural and urban counties. Urban counties, on average, had a substantially higher prevalence of COVID-19. The increasing county-level population of blacks, those aged 25-49 years, smokers, and obese were associated with increased rural COVID-19 prevalence rates.

COVID-19 infection spread rapidly from March 15 to April 22, affecting 98% of urban counties and 84% of rural counties in the United States. Earlier studies have reported substantially higher prevalence rates in urban counties as compared to rural counties.[30,31] Our results are in line with findings from other authors, but additionally identified hotspots of COVID-19 infection in rural counties as well.

In this study, the prevalence of COVID-19 was higher among blacks in both urban and rural counties. Earlier studies have reported the increased prevalence of COVID-19 infections among black and minority populations.[3,32-35] The higher infection rates among blacks is likely indicative of disparities in access to health care, health inequities,

Figure 3 Estimated COVID-19 (Denoised) Prevalence per 100,000 Population From the Fitted Spatiotemporal Model: April 3 to April 22, 2020.
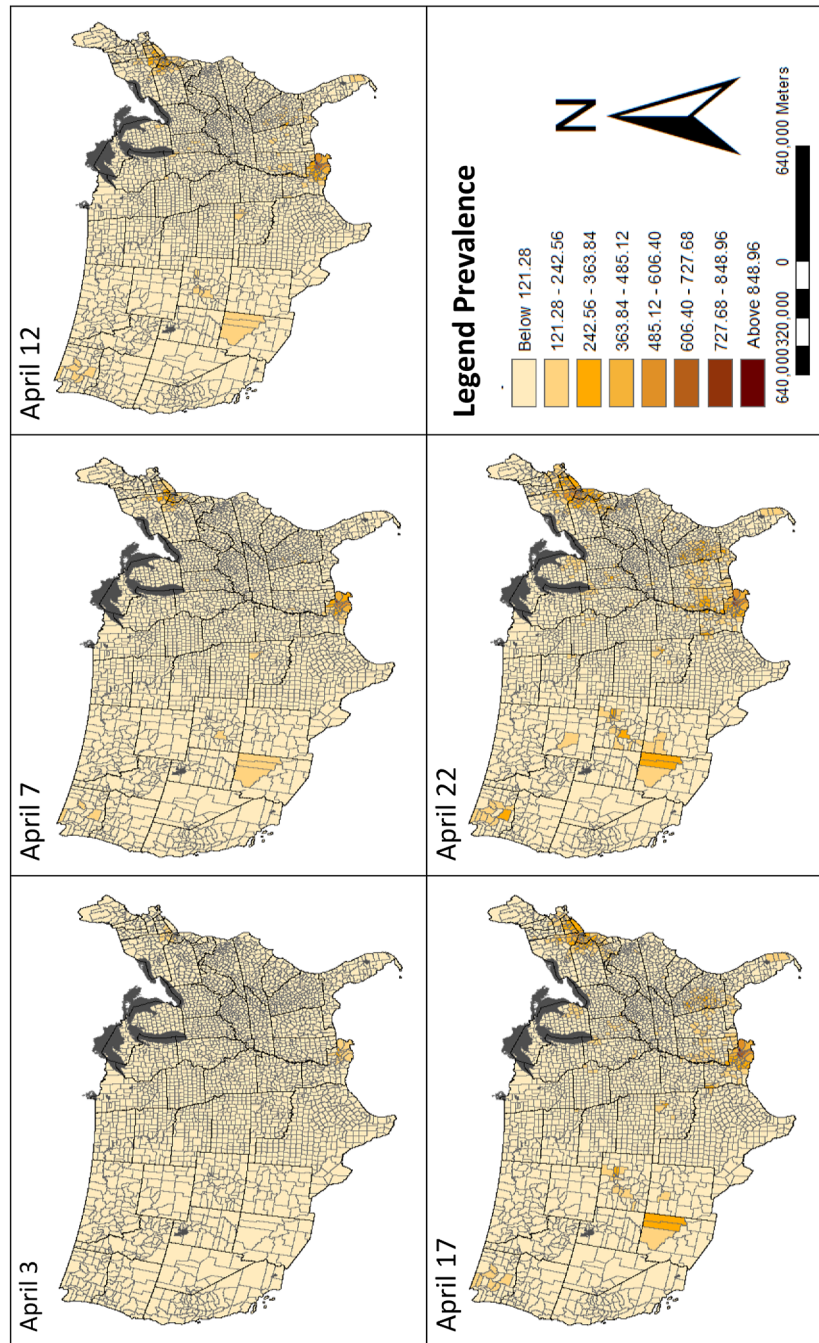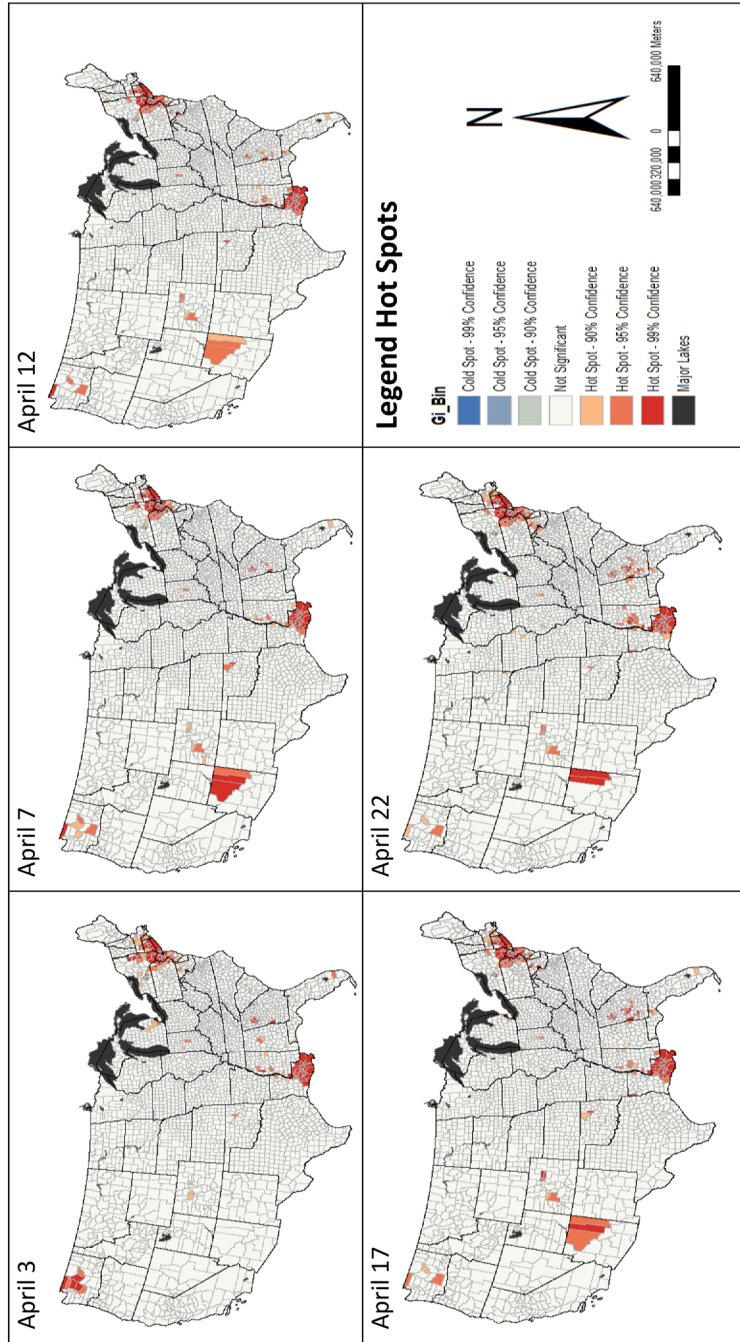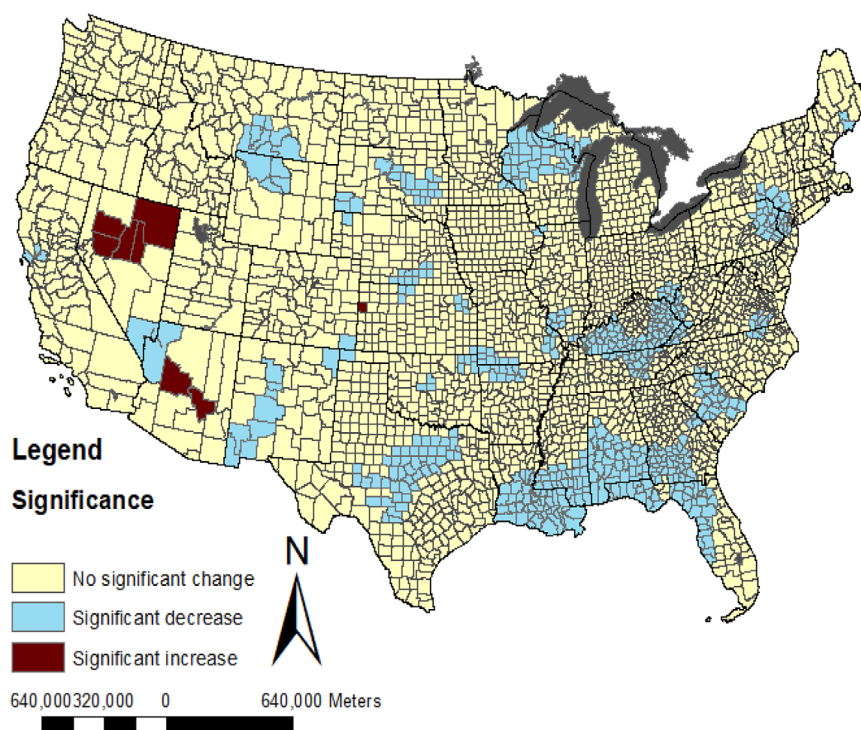
Figure 4 Hotspots of COVID-19 Estimated (Denoised) Prevalence: April 3 to April 22, 2020.

Figure 5 Significant Increase or Decrease of Percentage Change in Prevalence Over a 14-Day Period.

and underlying preexisting health conditions. Blacks are also more likely to work in "essential" jobs where the infection risk is higher.[36]

In this study, adults aged 25-49 years had a substantially higher prevalence of COVID-19 in rural counties ($t_{statistics}$: 18.9) as compared to urban ($t_{statistic}$: 2.8), adjusted for covariates. The Centers for Disease Control and Prevention analyzed data from February 12 to March 16, 2020, and reported that of 4,226 cases, 29% were adults ages 22-44.[3] While the mortality rates from COVID-19 infection are higher in older adults (ages 65 and older), the infection rates are higher among younger and middle-aged adults.[2] Our results show that the prevalence of COVID-19 rate change was 6 times more among young to middle-aged adults in rural counties as compared to urban counties. We are not aware of any other reports that have examined the data by urban-rural status.

Smoking is a major risk factor for cardiorespiratory diseases, including COPD. In the United States, the prevalence of smoking is about 14%; the prevalence among adults aged 25-44 is 16.5%.[37] Recently, some studies presented the "nicotine" hypothesis that nicotine in smoking is protective against the COVID-19 infection and hospitalization.[38] Our results were different. Smoking was associated with an increased prevalence of COVID-19 in both urban and rural counties.

In this study, we found a positive association of obesity with COVID-19 infection and a negative association with diabetes. About 40% of the US population is obese[39] whereas the prevalence of diabetes is about 11%.[40] Obesity increases the risk of outpatient visits from respiratory infections and hospitalization due to influenza virus.[41,42] Similarly, diabetes increases the risk of lung infections, hospitalization, and death.[43] However, there are conflicting reports about diabetes being an independent risk factor for infection-related mortality.[44,45] Recently, several studies have reported a higher prevalence of COVID-19 among patients with diabetes.[46-49] We observed a negative relationship between COVID-19 prevalence rates and diabetes that persisted despite adjusting for covariates. The negative association of diabetes in our study is likely due to the lower prevalence of diabetes in young and middle-aged adults (25-49 years) compared to older adults (65 years and above).[50]

The interpretation of counties with a significant decrease in percentage change of COVID-19 prevalence rates requires some caution. A considerable percentage change decrease does not mean that those counties with such results are ready for phased openings. A sustained decline in prevalence rates (Figure 3) supported by evidence of a significant percentage decrease (Figure 5) should inform the decision on phased openings.

These maps are, no doubt, powerful tools in aiding such decisions.

A relatively large volume of COVID-19-focused research has been dedicated to predicting when the epidemics will peak.[51-53] Noteworthy was the Institute for Health Metrics and Evaluation (IHME)[54] prediction model that provided state-level estimates for the next 4 months using a nonlinear mixed-effects model with an incorporated parametrized Gaussian structure for cumulative error rates. Unlike the IHME, our focus was a short-term county-level analysis on a daily scale. With COVID-19 infection pattern subject to dynamics of human interaction, our short-term approach is appropriate to capture the rapidly evolving county-level and county-specific situations.

We used a space-time Bayesian hierarchical model (BHM) approach using the reduced rank predictive process models.[12,55,56] These models are apt for semi-continuous data to which COVID-19 incidence counts belong.[21,26,27] Also, the models successfully addressed the spatial and temporal autocorrelations that arose from the spread of the disease. Our modeling approach denoised the crowdsourced data that have considerable reporting errors, making our estimated prevalence more reliable than the raw data.

This study has its limitations. It is an ecological study, and causal relationships cannot be established. It is important to note that our analyses are based on confirmed cases of COVID-19, a measure that is strongly dependent on the testing rate by county.[30,57] Also, coverage error is a concern as the reported confirmed cases of COVID-19 at state and county levels might be grossly underreported. We denoised the crowdsourced data, but data reporting and processing errors cannot be completely eliminated. Other than diabetes and obesity, data on other chronic diseases at the county level are unavailable with substantial coverage.

Our study is strengthened by the county-level prevalence and hotspots analysis that can guide legislation and policy relating to COVID-19 emergency preparedness, rural health infrastructure, and county-specific economy-reopening decisions. Also, various health departments can use our estimated prevalence for channelizing resources. The fast computation for large databases that uses the reduced rank predictive process models used in this study makes sequentially updating our estimates achievable as additional data emerge. This efficient modeling technique will produce real-time results. Our flexible modeling approach will enable the testing of several other hypotheses and control variables that we did not measure in this study.

Conclusion

With crowdsourced data requiring data cleaning, validation, and smoothing, we applied the appropriate level of rigor in cleaning the data and validation. Our space-time model played an essential role in the data smoothing, which filtered out the noise for more accurate inference. Our findings showed how COVID-19 spread from urban to rural areas in 21 days. With a limited facility of ICU beds and ventilators, it would be challenging for the rural health care system to cope with the influx. Our findings show geographic disparities in COVID-19 prevalence and how smoking, race, obesity, and age explain, to some extent, that disparity. In the future, as additional quality data on social distancing measures become available, we will be able to assess how such measures impact change in prevalence rates.

References

1. World Health Organization. *Coronavirus disease 2019 (COVID-19) situation report—97.* 2020; https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200426-sitrep-97-covid-19.pdf? Accessed May 7, 2020.

2. Centers for Disease Control and Prevention. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12-March 16, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(12):343-346.

3. Centers for Disease Control and Prevention. Cases of Coronavirus Disease (COVID-19) in the U.S. 2020; https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Accessed April 26, 2020.

4. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MU, … Wang B. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science.* 2020;368(6491):638-642.

5. Centers for Disease Control and Prevention. About Rural Health. 2017; https://www.cdc.gov/ruralhealth/about.html. Accessed April 3, 2020, 2020.

6. Atkins GT, Kim T, Munson J. Residence in rural areas of the United States and lung cancer mortality. Disease incidence, treatment disparities, and stage-specific survival. *Ann Am Thorac Soc.* 2017;14(3):403-411.

7. Moy E, Garcia MC, Bastian B, et al. Leading causes of death in nonmetropolitan and metropolitan areas—United States, 1999–2014. *MMWR Surveill Summ.* 2017;66(1):1-8.

8. New-York-Times. Coronavirus was slow to spread to Rural America. Not anymore; https://www.nytimes.com/interactive/2020/04/08/us/coronavirus-rural-america-cases.html. Accessed April 10, 2020.

9. Healy J, Tavernise S, Gebeloff R, Cai W. Coronavirus was slow to spread to rural America. Not anymore. *The New York Times*. April 8, 2020.

10. American Academy of Pediatric News. CDC offers COVID-19 guidance to rural communities. 2020; https://www.aappublications.org/news/2020/03/23/coronavirus032320. Accessed April 4, 2020.

11. Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Mapping the burden of COVID-19 in the United States. *medRxiv*. 2020. https://doi.org/10.1101/2020.04.05.20054700.

12. Sahu SK, Bakar KS. Hierarchical Bayesian auto-regressive models for large space time data with applications to ozone concentration modelling by Sujit Kumar Sahu and Khandoker Shuvo Bakar: rejoinder. *Appl Stochast Models Bus Indus*. 2012;28(5):418-419.

13. Cressie N, Wikle CK. *Statistics for Spatiotemporal Data*. New York: John Wiley & Sons; 2015.

14. Cressie N. *Statistics for Spatial Data*. New York: John Wiley & Sons; 1991.

15. United States Census Bureau. 2018 FIPS Codes. 2019; https://www.census.gov/geographies/reference-files/2018/demo/popest/2018-fips.html. Accessed April 11, 2020.

16. Center for Systems Science and Engineering. Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. 2020; https://systems.jhu.edu/research/public-health/ncov/. Accessed May 8, 2020.

17. WHO Director-General's opening remarks at the media briefing on COVID-19. 24 February 2020; https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19%201424-february-2020. Accessed May 7, 2020.

18. United States Census Bureau. 2018 American Community Survey Summary File Data. 2018; https://www.census.gov/programs-surveys/acs. Accessed May 7, 2020.

19. University of Wisconsin Population Health Institute. 2020 County Health Rankings State Reports. *County Health Ranking and Road Maps*. 2020; https://www.countyhealthrankings.org/. Accessed April 11, 2020.

20. Economic Research Services. Rural-Urban Commuting Area Codes. 2019; https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes. Accessed April 11, 2020.

21. Sigrist F, Künsch HR, Stahel WA. Stochastic partial differential equation based modelling of large spacetime data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2015;77(1):3-33.

22. Paul R, Jelsema CM, Lau KW. A flexible class of reduced rank spatial models for large non-gaussian dataset, peer reviewed book chapter on "Current Trends in Bayesian Methodology with Applications". edited jointly by Satyanshu K. Upadhyay, Umesh Singh, Dipak K. Dey, and Appaia Loganathan, (pp. 477-502), published by Chapman and Hall/CRC, New York; 2015.

23. Robert C, Casella G. *Monte Carlo Statistical Methods*. Monte Carlo statistical methods. Springer Science & Business Media; 2013. New York.

24. *RStudio: Integrated Development for R*. [computer program]. Boston, MA: RStudio, Inc.; 2019.

25. *R: A language and environment for statistical computing* [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2019.

26. Bakar KS, Sahu SK. spTimer: spatiotemporal Bayesian modelling using R. *J Stat Softw*. 2015;63(15):1-32.

27. Sigrist F, Künsch HR, Stahel WA. A dynamic nonstationary spatiotemporal model for short term prediction of precipitation. *Ann Appl Stat*. 2012;6(4):1452-1477.

28. *ArcGIS Desktop: Release* [computer program]. Version 10.8. Redlands, CA: Environmental Systems Research Institute; 2020.

29. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis*. edited by Anselin, Luc, Rey, Sergio J. (pp. 127-145). Springer, Berlin, Heidelberg. 2010.

30. CDC COVID-19 Response Team. Geographic differences in COVID-19 cases, deaths, and incidence—United States, February 12-April 7, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):465-471.

31. Fehr R, Kates J, Cox C, Michaud J. COVID-19 in rural America—is there cause for concern? 2020. https://www.kff.org/other/issue-brief/covid-19-in-rural-america-is-there-cause-for-concern/. Accessed May 7, 2020.

32. Zahnd WE. The COVID-19 pandemic illuminates persistent and emerging disparities among rural black populations. *J Rural Health*. Epub ahead of print, May 3, 2020. https://doi.org/10.1111/jrh.12460

33. Sood L, Sood V. Being African American and rural: a double jeopardy from Covid-19. *J Rural Health*. Epub ahead of print June 8, 2020. https://doi.org/10.1111/jrh.12459.

34. Rios E, Rangarajan S. COVID-19 has infected and killed black people at alarming rates. This data proves it. *Mother Jones*. April 17, 2020.

35. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *The Washington Post*. April 8, 2020.

36. Black Demographics. African American employment. *The African American Population* 2020; https://blackdemographics.com/economics/employment/. Accessed May 7, 2020.

37. Center for Disease Control and Prevention. Current cigarette smoking among adults in the United States. *Smoking & Tobacco Use*. 2019; https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm. Accessed May 7, 2020.

38. Farsalinos K, Niaura R, Le Houezec J, et al. Editorial: nicotine and SARS-CoV-2: COVID-19 may be a disease of

the nicotinic cholinergic system. *Toxicol Rep*. 2020;7:658-663.

39. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017–2018. *NCHS Data Brief*. 2017 Oct;(288):1-8.

40. Centers for Disease Control and Prevention. Estimates of diabetes and its burden in the United States. *National Diabetes Statistics Report*. 2020. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf. Accessed May 7, 2020.

41. Moser J-AS, Galindo-Fraga A, Ortiz-Hernández AA, et al. Underweight, overweight, and obesity as independent risk factors for hospitalization in adults and children from influenza and other respiratory viruses. *Influenza Other Respir Viruses*. 2019;13(1):3-9.

42. Campitelli MA, Rosella LC, Kwong JC. The association between obesity and outpatient visits for acute respiratory infections in Ontario, Canada. *Int J Obes (Lond)*. 2014;38(1):113-119.

43. Klekotka RB, Mizgała E, Król W. The etiology of lower respiratory tract infections in people with diabetes. *Pneumonol i Alerg Polska*. 2015;83(5):401-408.

44. Stegenga ME, Vincent J-L, Vail GM, et al. Diabetes does not alter mortality or hemostatic and inflammatory responses in patients with severe sepsis. *Crit Care Med*. 2010;38(2):539-545.

45. Guo W, Li M, Dong Y, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diab/Metab Res Rev*. 2020;e3319. https://doi.org/10.1002/dmrr.3319

46. Madsbad S. COVID-19 infection in people with diabetes. *Touch Endocrinology*. 2020. https://www.touchendocrinology.com/insight/covid-19-infection-in-people-with-diabetes/. Accessed June 10, 2020.

47. Bloomgarden ZT. Diabetes and COVID-19. *J Diab*. 2020;12(4):347-348.

48. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: knowledge in progress. *Diab Res Clin Pract*. 2020;162:108142.

49. Muniyappa R, Gubbi S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. *Am J Physiol Endocrinol Metab*. 2020;318(5):E736-e741.

50. Selvin E, Parrinello CM. Age-related differences in glycaemic control in diabetes. *Diabetologia*. 2013;56(12):2549-2551.

51. Nesteruk I. Estimations of the coronavirus epidemic dynamics in South Korea with the use of SIR model. *Preprint, ResearchGate*. 2020. https://doi.org/10.13140/RG.2.2.15489.40807

52. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv* 2020. https://doi.org/10.1101/2020.02.16.20023465

53. Centers for Disease Control and Prevention. COVID-19 Forecasts. *Coronavirus Disease 2019 (COVID-19)*2020; https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html. Accessed May 7, 2020.

54. COVID IHME, Murray CJ. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*. 2020.

55. Ver Hoef JM, Cressie N, Barry RP. Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT). *J Comput Graph Stat*. 2004;13(2):265-282.

56. Banerjee S, Gelfand AE, Finley AO, Sang H. Gaussian predictive process models for large spatial data sets. *J R Stat Soc: Ser B (Stat Meth)*. 2008;70(4):825-848.

57. Omori R, Mizumoto K, Chowell G. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *Int J Infect Dis*. 2020;94:116-118.