



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major Article

Visualizing and assessing US county-level COVID19 vulnerability

Gina Cahill MPH^a, Carleigh Kutac MPH^b, Nicholas L. Rider DO^{a,*}^a Baylor College of Medicine and Texas Children's Hospital, Section of Immunology Allergy and Retrovirology, Houston, TX^b Baylor Saint Luke's Medical Center, Houston, TX

Key Words:

Novel Coronavirus
Population-health
Data visualization
SARS-CoV-2

Background: Like most of the world, the United States' public health and economy are impacted by the COVID19 pandemic. However, discrete pandemic effects may not be fully realized on the macro-scale. With this perspective, our goal is to visualize spread of the pandemic and measure county-level features which may portend vulnerability.

Methods: We accessed the New York Times GitHub repository COVID19 data and 2018 United States Census data for all United States Counties. The disparate datasets were merged and filtered to allow for visualization and assessments about case fatality rate (CFR%) and associated demographic, ethnic and economic features.

Results: Our results suggest that county-level COVID19 fatality rates are related to advanced population age ($P < .001$) and less diversity as evidenced by higher proportion of Caucasians in High CFR% counties ($P < .001$). Also, lower CFR% counties had a greater proportion of the population reporting having 2 or more races ($P < .001$). We noted no significant differences between High and Low CFR% counties with respect to mean income or poverty rate.

Conclusions: Unique COVID19 impacts are realized at the county level. Use of public datasets, data science skills and information visualization can yield helpful insights to drive understanding about community-level vulnerability.

© 2020 Association for Professionals in Infection Control and Epidemiology, Inc. Published by Elsevier Inc. All rights reserved.

In December 2019, an unknown form of pneumonia was identified in Wuhan Province China which ultimately heralded emergence of the COVID19 pandemic. Since that time, spread of SARS-CoV-2 resulted in over 14 million individuals infected and over 600,000 deaths worldwide.^{1,2} At present, COVID19 has spread to every continent on Earth except for Antarctica.^{2,3} In the United States (US), the first confirmed COVID19 case was reported by a team in Snohomish County Washington.⁴ In the ensuing 7 months, the US reported over 3.8 million infections and over 140,000 deaths.² Given the totality of health and economic devastation levied by COVID19, tremendous energy has been invested towards understanding features which could predict individual and community impact.^{5–10}

Understanding community vulnerability with COVID19 is important and may enable pre-emptive and ongoing interventions to stem

public health and economic crises.^{11–13} However the explosion of data about COVID19 has led to much debate and skepticism around best measures for analysis and assessment of impact.¹⁴ Measuring severity, which is the disease's ability to cause death, is key to developing appropriate public health responses for reducing the spread of infection. The 2 most commonly used severity metrics are infection fatality ratio and case fatality rate (CFR). In order to accurately measure infection fatality ratio, the total number of disease related infections and deaths must be known. This typically involves serological testing of a random population sample to estimate the true number of infections. However, in the US widespread serological testing of the population was not accomplished at the time of our analysis, making it an inappropriate measure for COVID in this study. CFRs on the other hand can be calculated by taking the number of confirmed COVID deaths and dividing it by the number of confirmed cases. Despite these limitations, CFRs can be easily calculated with accessible aggregate cases and death counts making them a practical measure of severity early-on in an epidemic. Once widespread disease reporting is in place, CFR% can reasonably be used.^{15–18} With widespread US testing and reporting of COVID19 cases, we used assessment of CFR% across US counties coupled with US Census data to

* Address correspondence to Nicholas L. Rider, DO, Baylor College of Medicine and Texas Children's Hospital, Section of Immunology Allergy and Retrovirology, Houston, TX.

E-mail address: nlrider@bcm.edu (N.L. Rider).

Conflicts of interest: GC and CK have nothing to disclose. NLR received consulting fees for scientific advisory activities with Takeda Pharmaceuticals, Horizon Therapeutics and CSL Behring. He also receives royalties from Wolters Kluwer for topic contribution to UpToDate.

visualize and evaluate regional differences in COVID impact. Our analysis was focused on 4 months of pandemic activity (March to June 2020) for US Counties. Here we show pandemic spread via CFR%

visualized across US counties stratified by both CFR% and case burden. Additionally, we show relevant Census-mined data attributes associated with high and low CFR% counties.

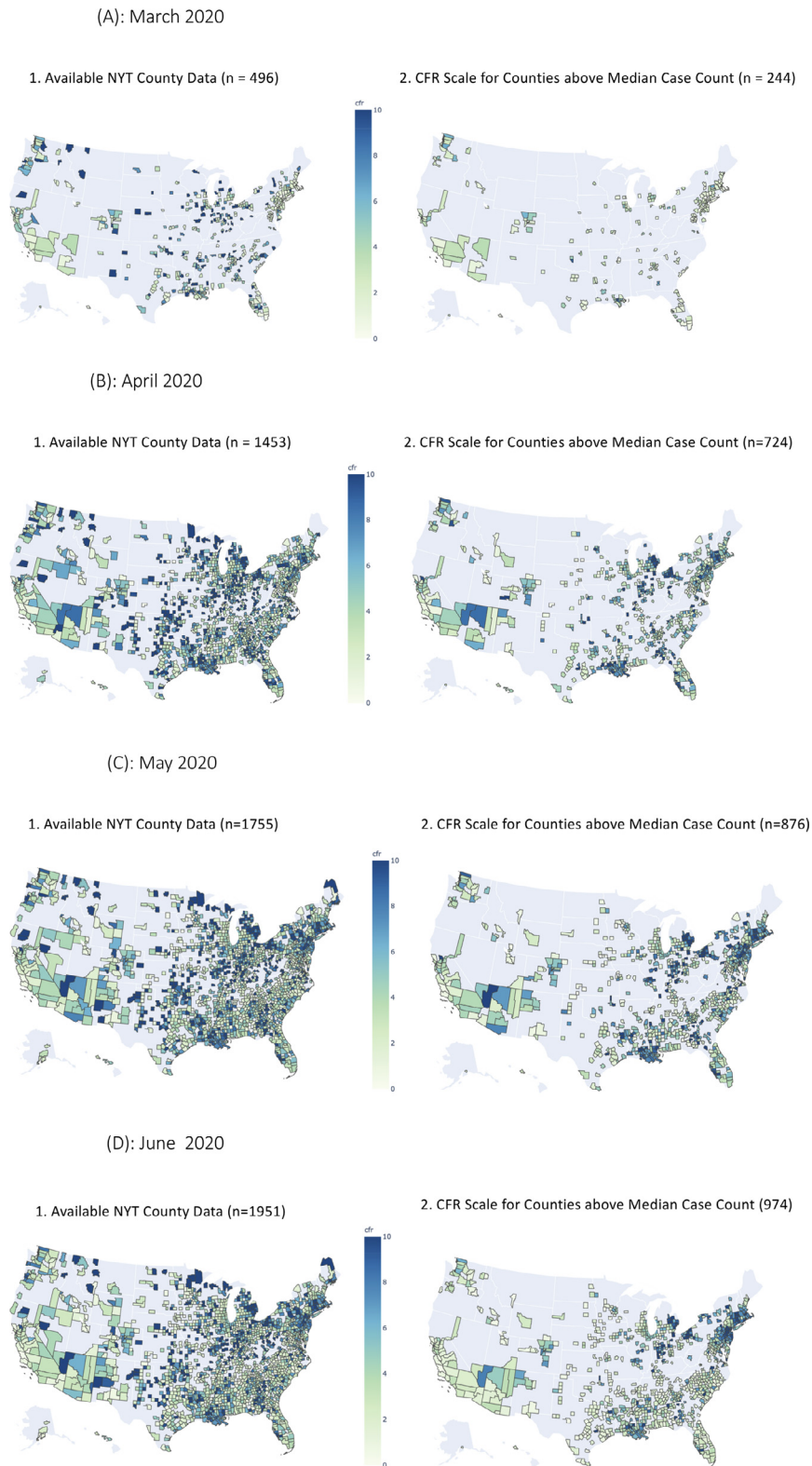


Fig 1. (A-D) Each panel shows US Counties visualized by relative scale CFR% for the specified month. Left side panels display all available counties within the CFR% range (0.1-99.9) and right-side panels show CFR% for counties with case counts about the National median. Horizontal bar-scale shows relative CFR% as plotted from 0 - 10 relative. County numbers shown above each map. CFR, case fatality rate.

METHODS

Dataset extraction & creation

We extracted US county data about the COVID19 pandemic from the New York Times Github repository (<https://github.com/nytimes/covid-19-data>) and American Community Survey 5-year Estimates Detailed Tables (2013-2018) census data about county-level demographics, economic factors and ethnicities (<https://www.census.gov>). Data was extracted on July 6, 2020. Using JupyterLab through the Anaconda distribution (v. 2020.02) Python version 3.8.3 with Pandas (version 1.0.5) we merged and filtered the final dataset for analysis. Our code is available here (<https://github.com/nlrider/COVID-Public-Health-Data>). Unique counties with a CFR% between 0.1 and 100 and a Federal Information Processing Standard (fips) code were analyzed for the last date of each month studied (March to June 2020).

Calculating CFR% & county descriptive statistics

CFR% was calculated by dividing the total number of COVID-19 suspected or confirmed deaths by the total number of confirmed COVID cases. The cumulative numbers of deaths and cases were included per month. We excluded counties with a CFR of 0 or 100 that had active case counts as these extremes in CFR were suspicious for under or over reporting outcomes.

Data visualization & inferential statistics

From the list of unique counties with a CFR% between 0.1 and 100, we assessed median case count per month per county. We then visualized both the total number of counties available at end of each month and also counties above the median by relative scale CFR%

(Fig 1A-D). County level data visualization was done with plotly (v. 4.8.2) in JupyterLab for available counties by fips codes at monthly time intervals reported on the last date of each month.

Descriptive statistics for each monthly time capsule was performed in JupyterLab using Pandas (Supplemental Tables 1-4).

Taking all counties with a CFR% between 0.1 and 100 (ie, excluding CFR=0 and CFR=100), high CFR counties as of June 30, 2020 were defined by a county with a CFR% at or above the upper quartile (>6.25%). Low CFR counties as of June 30, 2020 were defined by a county with a CFR% falling below the upper quartile (6.25% or lower). Subsequent significance testing for county data as of June 30, 2020 was performed with STATA version 12.1 Welch's t-tests were used to determine whether or not there was a significant difference in the mean values of variables of interest (NYT COVID and US Census Features) between high and low CFR counties, as this method of testing accounts for unequal variances between the 2 groups (Table 2) Rank of US states by number of high or low CFR counties as of June 30, 2020 were plotted and sorted using Tableau (v.2020.2; Figs. 2 and 3).

RESULTS

Data extraction from the New York Times COVID19 GitHub repository (<https://github.com/nytimes/covid-19-data>) revealed 3,060 unique fips reflecting the same number of counties and associated COVID data. Filtering data to include counties with a CFR% between 0.1%-100% reduced this number of unique counties to 1996. This number of counties was the total available for our 4-month analysis, however pandemic data available at each monthly time interval varied as shown in Table 1. For each month interval assessed, an increase in the number of counties available for analysis is noted in alignment with the expansion of the pandemic. Similarly, the median per county case count increased and was paralleled by deaths across the

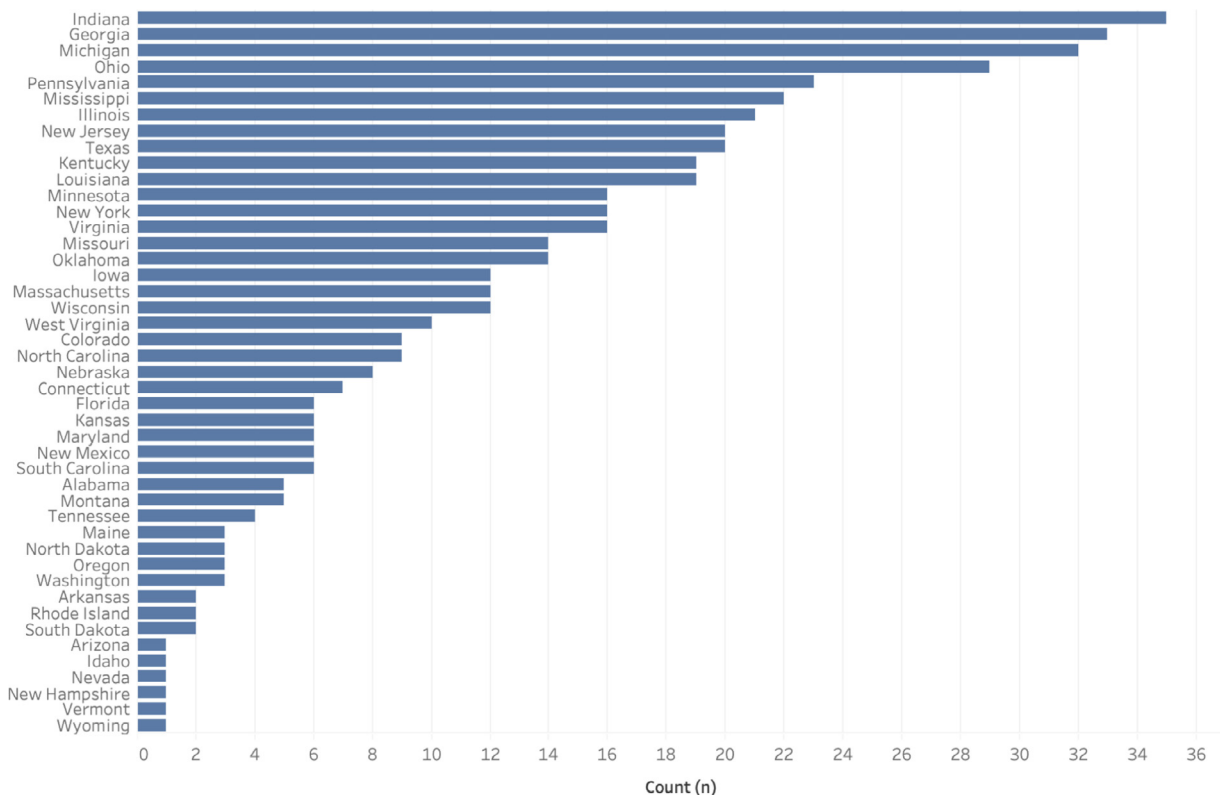


Fig 2. Horizontal bar chart depiction of states and corresponding county counts within the highest US National quartile (ie, CFR% > 6.25). States sorted in descending order. CFR, case fatality rate.

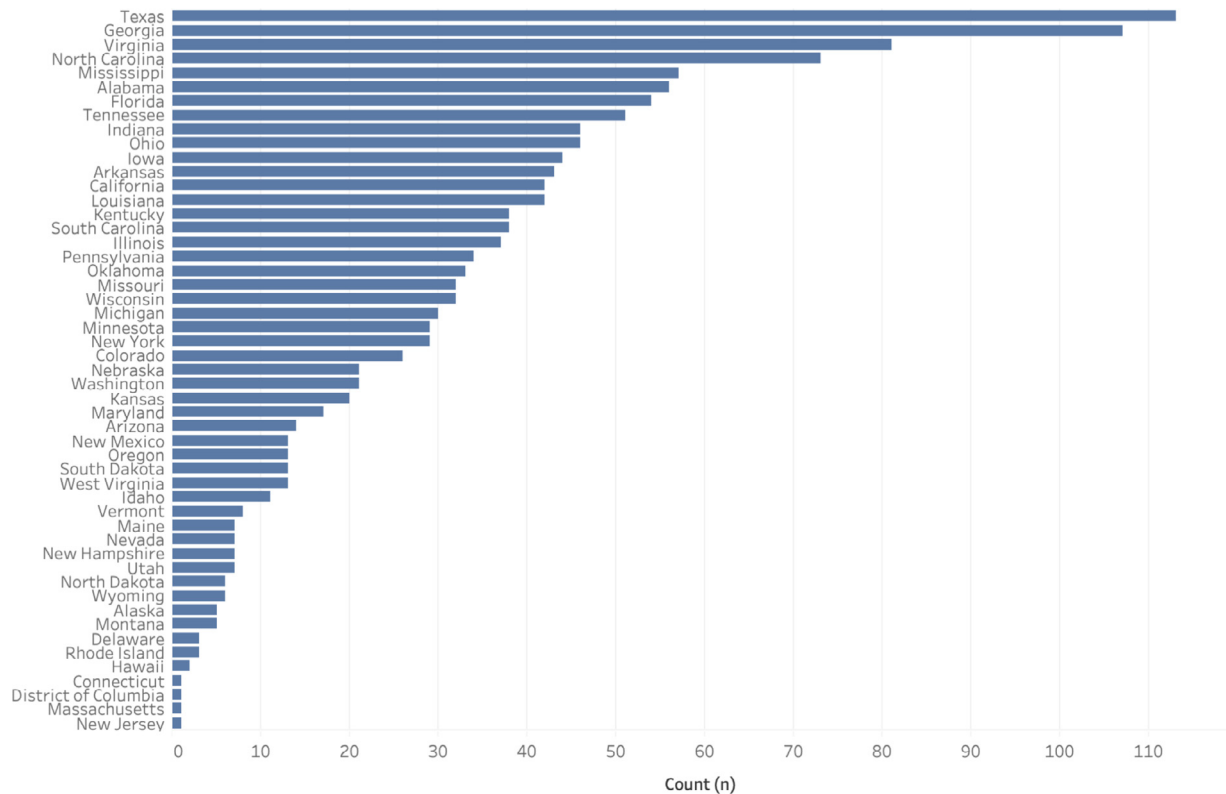


Fig 3. Horizontal bar chart depiction of states and corresponding county counts within the lower 3 US National quartiles (ie, CFR% \leq 6.25). States sorted in descending order. CFR, case fatality rate.

Table 1.
Aggregate county-level COVID19 statistics by month studied

	March	April	May	June
County count (n)	496	1453	1755	1,951
Mean per county case count (n)	339	722	1007	1,337
Median per county case Count(n)	55	88	144	228
Mean per county death count (n)	8	43	59	65
Mean county CFR (%)	7.1	6.3	5.3	4.6
Counties > median case Count/month (n)	244	724	876	974

4 months assessed (244-974 cases/county). However, slope of the increase in case counts and deaths reported was not identical as noted by the gradual trend down in CFR% from end of March to end of June (7.1%-4.6%) as seen in Table 1.

Progression of the pandemic by CFR% at the last date of each month is shown in Figure 1 with the associated high-case counties plotted according to CFR%. The left-side map in each panel shows total cases for counties that met our criteria by CFR%; whereas, the right-side map shows CFR% for counties with case counts above the US median. The CFR% scale is arbitrary and only shows relative differences.

State level county assessments are shown in Figures 2 and 3. Figure 2 shows the county counts for each state sorted in descending order within the high CFR% category (CFR% > 6.25; upper quartile). Figure 3 shows state-level county counts of those within the bottom 3 quartiles (CFR% \leq 6.25). Figure 4 shows all available state data without lower or upper boundaries for case count as of July 6, 2020.

Descriptive aggregate COVID19 statistics per month studied are shown in Table 1. Table 2 shows significance testing across socioeconomic, demographic and ethnic county features in comparing top quartile counties with those in the bottom 3 quartiles. From this

analysis we found that significant, normalized differences between High CFR% and Low CFR% counties included more of the following in high CFR% counties: % of population over 16 years of age, % of population above 62 years of age, % of population over 25 years of age with a high school degree, % of population reporting as Caucasian. We also noted significantly higher numbers of the following in Low CFR% counties: total and % of population reporting as Asian, total and % of population reporting as Pacific Island Native and/or Hawaiian as well as total and % of counties reporting 2 or more races.

DISCUSSION

Spread of COVID19 across the US has been rapid and hot spots such as New York City, Seattle Washington, New Orleans Louisiana and Los Angeles California emerged early and can be seen in our mapped data (Fig 1A1 and 2).^{19,20} We note persistence of “hot spot” High CFR% counties within the Northeast, the Gulf Coast region, the Upper Midwest and Desert Southwest, Southern California and the Pacific Northwest (Fig 1: progression from A to D). Mean CFR values for the 4 months analyzed gradually decreased over time (7.1 in March to 4.6 in June). According to the WHO, in the beginning of outbreaks surveillance tends to be focused on the more symptomatic patients who seek care while individuals with mild, unsymptomatic forms of the illness go undetected, which can lead to the overestimation of the CFR. As testing and surveillance increase, the CFR generally decreases, which corresponds with our findings.¹⁸ We do see such a dynamic occur via the 4-month time course (Table 1). Also, the reduction of mean county CFR% shown in Table 1 from March to June reflects a sharp rise in mean case count over the timespan.

Interestingly, our visualizations show that High CFR% US counties are not isolated to urban regions (Fig 1A-D, Fig 3). Rather, wide

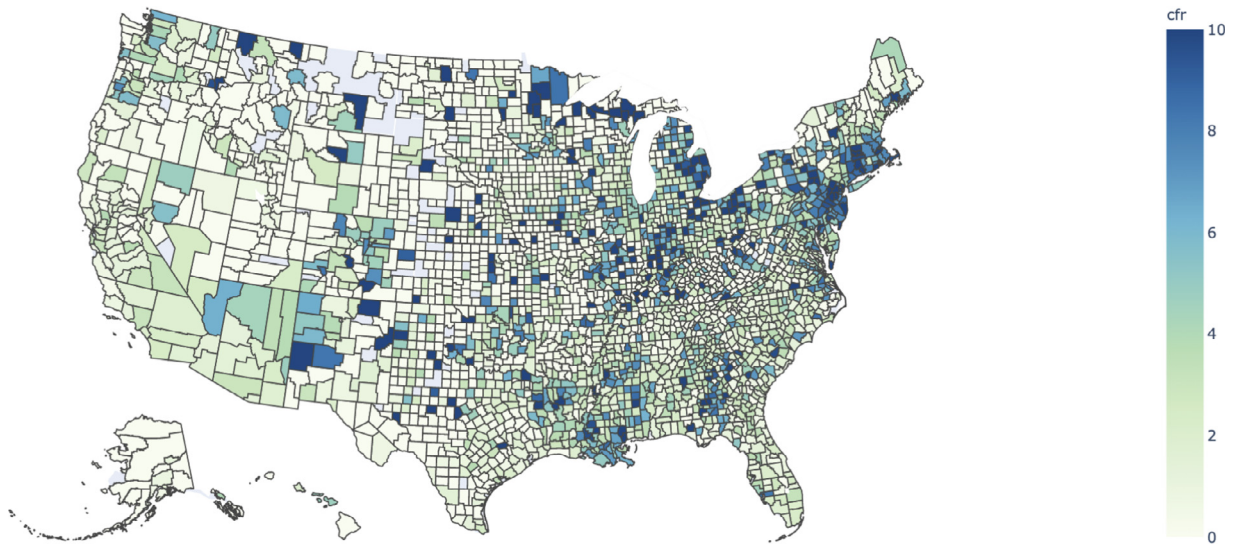


Fig 4. Plot of all available US County CFR% data (n = 3,060) without case count filter. Color range in relative scale 0-10 as of July 6, 2020. CFR, case fatality rate.

Table 2.
Comparison of high and low CFR counties socio-demographic features

County Feature	CFR Class	Mean (STD)	P value
Total population	Low	159,128 (438,503)	.007
	High	117,591 (226,235)	
Population below poverty line	Low	22,539 (66,667)	.003
	High	14,570 (29,438)	
% Population below poverty line	Low	15.9 (0.17)	.222
	High	15.6 (0.30)	
Total population >16 years of age	Low	129,470 (354,037)	.01
	High	97,621 (186,687)	
% Population >16 years of age	Low	82.8 (5.9)	.001
	High	83.9 (5.7)	
Unemployment rate	Low	6.1 (2.5)	.52
	High	5.9 (2.6)	
Total population >62 years of age	Low	29,365 (73,066)	.1
	High	24,800 (45,203)	
% Population > 62 years of age	Low	21.4 (4.9)	<.001
	High	23.5 (4.5)	
Median income (USD)	Low	53,006 (14,964)	.97
	High	52,982 (15,809)	
Population >25 years of age and HS graduate	Low	28,585 (66,645)	.07
	High	24,008 (40,751)	
% Population >25 years of age and HS graduate	Low	23.3 (6.2)	<.001
	High	25.2 (5.7)	
Population with college degree	Low	21,505 (65,331)	.06
	High	16,850 (38,256)	
% Population with college degree	Low	10.2 (4.1)	.95
	High	10.2 (4.2)	
Total Caucasian population	Low	116,080 (271,635)	.03
	High	93,703 (16,829)	
% Population Caucasian	Low	81.5 (17.6)	<.001
	High	84.8 (17.9)	
Total black population	Low	21,520 (71,286)	.01
	High	14,496 (45,623)	
% Population black	Low	13.7 (17.6)	.1
	High	12.1 (18.2)	
Total American Indian population	Low	1,322 (4,888)	<.001
	High	707 (2,637)	
% Population American Indian	Low	1.4 (5.9)	.8
	High	1.4 (5.2)	
Total Asian population	Low	9,430 (56,395)	.01
	High	5,204 (18,778)	
% Population Asian	Low	1.9 (3.2)	.01
	High	1.5 (2.5)	
Total Hawaiian native population	Low	352 (2,915)	<.001
	High	47 (113)	

(continued)

Table 2. (Continued)

County Feature	CFR Class	Mean (STD)	P value
% Hawaiian native	Low	0.1 (0.4)	<.001
	High	0.04 (0.09)	
Population with 2 or more races	Low	5,563 (18,319)	<.001
	High	3,203 (6,689)	
% Population with 2 or more races	Low	2.6 (1.9)	<.001
	High	2.3 (1.5)	

CFR, case fatality rate.

P value calculated with Welch's t test for unequal variances.

distribution of high CFR% counties suggests that unique demographic, geographic or other features may affect county-level vulnerability. Our CFR% mapping, not surprisingly aligns with other group's work related to community risk.^{21,22} Additionally, an individual state may have a significant number of both High and Low CFR% Counties as shown in Figures 2 and 3. If we focus on the top 10 states with most High CFR% counties, we see that they represent the Midwest, South East, South, and Northeast regions. Top 10 states for Low CFR% county counts also overlap with the High CFR% group in that Texas, Georgia, Mississippi, Indiana and Ohio are present in both groups (Figs. 2 and 3). This underscores the notion that community vulnerability may be intrinsic and likely independent of climate, local governmental policy or even population density. It is also important to note that a priori county-level risk and actual CFR% are not uniformly aligned.

Our assessment of county-level features which may contribute to higher or lower CFR% included Census information taken from demographic, economic and ethnicity tables. From these, our data suggest that a more diverse and younger population is better able to weather COVID19 (Table 2). This census-based analysis on an unbalanced set of counties with high CFR% (highest quartile) with a CFR% >6.25 to that of the bottom 3 quartiles, shows that predominantly Caucasian counties (% Caucasian) were found in greater numbers within the High CFR% group. Conversely, counties with proportionally more Asians and Pacific Islander and/or Hawaiian Natives were found in the Low CFR% group. We also note that counties reporting as having a higher proportion of the population representing 2 or more races, were more likely to be in the Low CFR% county group. However, while looking at the severity of COVID from a county lens is helpful for getting an overall view of the COVID epidemic in the US it is not as effective looking at the association between race and/or racial diversity and COVID burden. These questions are better addressed at the local or census tract level. Importantly, no differences were observed between High and Low CFR% groups in their levels of unemployment, median income or proportion of population below the poverty line. Also, education level did not seem to differ in that college degree holders were in similar numbers across both groups. While there were more high school diplomates over age 25 years in the High CFR%, this may relate to risk associated with age.

The COVID19 pandemic remains dynamic without clear evidence for which communities or individuals may fare worse from the outset. Factors such as population health features, number of long-term care facilities, prisons or other workplaces known to drive outbreaks are important and were not addressed here.^{23–25} Also, analysis of a comprehensive public health dataset is important for teasing out community-level risks. However, no single or amalgamated resource retrospectively analyzed will capture all needed features for complete risk assessment. Lastly, some of our results are in contrast to previously reported findings which show important regional ethnic risk factors for outcomes in COVID19.²⁶ Such differences may relate to intrinsic reporting bias with US Census data as noted previously.²⁷

CONCLUSION

In summary, we present an analysis of county-level CFR% via merged COVID19 and US Census data. Our findings suggest that younger and more diverse counties fared better over the 4 months studied for the US COVID19 experience. We also find that community risk level must be assessed at granularity level below that of the state or region as scrutiny of specific population-level features may yield greater insights for enabling population resiliency. Implementation of policies and systems which foster prospective quality assessments and enable rigorous analysis are needed for the US Healthcare System and will likely need to be implemented at that level of detail.

Acknowledgments

We wish to thank the team and investigators at the New York Times for making COVID19 project and associated GitHub repository data publicly available here - <https://github.com/nytimes/covid-19-data>

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.ajic.2020.12.009>.

References

- Zhu F, Zhang M, Gao M, et al. Effects of respiratory rehabilitation on patients with novel coronavirus (COVID-19) pneumonia in the rehabilitation phase: protocol for a systematic review and meta-analysis. *BMJ Open*. 2020;10:e039771.
- J. H. University. *Coronavirus Resource Center*. 2020. Available at: <https://coronavirus.jhu.edu/map.html>. Accessed July 20, 2020.
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265–269.
- Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382:929–936.
- Arneson D, Elliott M, Mosenia A, et al. CovidCounties is an interactive real time track of the COVID-19 pandemic at the level of US counties. *Scientific Data*. 2020;7:1–10.
- Khose S, Moore JX, Wang HE. Epidemiology of the 2020 Pandemic of COVID-19 in the state of texas: the first month of community spread. *J Community Health*. 2020; 45:696–701.
- Mollalo A, Rivera KM, Vahedi B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *Int J Environ Res Public Health*. 2020;17:4204–4217.
- Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ*. 2020;728:138884.
- Fox TA, Troy-Barnes E, Kirkwood AA, et al. Clinical outcomes and risk factors for severe COVID-19 infection in patients with haematological disorders receiving chemo- or immunotherapy. *Br J Haematol*. 2020:194–206.
- Verhagen MD, Brazel DM, Dowd JB, Kashnitsky I, Mills MC. Forecasting spatial, socioeconomic and demographic variation in COVID-19 health care demand in England and Wales. *BMC Med*. 2020;18:203.
- Danielli S, Patria R, Donnelly P, Ashrafian H, Darzi A. Economic interventions to ameliorate the impact of COVID-19 on the economy and health: an international comparison. *J Public Health (Oxf)*. 2020:1–5.
- Sheervalilou R, Shirvaliloo M, Dadashzadeh N, et al. COVID-19 under spotlight: a close look at the origin, transmission, diagnosis, and treatment of the 2019-nCoV disease. *J Cell Physiol*. 2020:8873–8924.

13. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. *Appl Geogr*. 2020;118:102202.
14. Hoseinpour Dehkordi A, Alizadeh M, Derakhshan P, Babazadeh P, Jahandideh A. Understanding epidemic data and statistics: a case study of COVID-19. *J Med Virol*. 2020;92:868–882.
15. Undela K, Gudi SK. Assumptions for disparities in case-fatality rates of coronavirus disease (COVID-19) across the globe. *Eur Rev Med Pharmacol Sci*. 2020;24:5180–5182.
16. Abdollahi E, Champredon D, Langley JM, Galvani AP, Moghadas SM. Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. *CMAJ*. 2020;192:E666–E670.
17. Jen TH, Chien TW, Yeh YT, Lin JJ, Kuo SC, Chou W. Geographic risk assessment of COVID-19 transmission using recent data: an observational study. *Medicine (Baltimore)*. 2020;99:e20774.
18. World Health Organization. *Estimating Mortality From COVID-19*. 2020. Available at: <https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19>. Accessed August 4, 2020.
19. Fauver JR, Petrone² Mary E, Hodcroft EB, et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell*. 2020;181:990–996. e5.
20. Patel A, Jernigan DB, V. C. D. C. R. T. nCo. Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak - United States, December 31, 2019–February 4, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:140–146.
21. Harvard Global Health Initiative. *Epidemics Explained: COVID19 Risk Levels-United States*. 2020. Available at: <https://globalepidemics.org/key-metrics-for-covid-suppression/>. Accessed July 21, 2020.
22. Georgia Tech University. *COVID-19 Event Risk Assessment Planning Tool*. 2020. Available at: <https://covid19risk.biosci.gatech.edu>. Accessed July 21, 2020.
23. McMichael TM, Clark S, Pogosjans S, et al. COVID-19 in a long-term care facility - King County, Washington, February 27–March 9, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:339–342.
24. Zheng Z, Peng F, Xu B, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect*. 2020:e16–e25.
25. Reinhart E, Chen D. Incarceration and its disseminations: COVID-19 pandemic lessons from Chicago's cook county jail. *Health Aff (Millwood)*. 2020;39:1412–1418.
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. US Census Nonresponse Bias Analysis, "2018 National survey of children's health: nonresponse bias analysis," US Census Bureau. Available at: <https://www.census.gov/programs-surveys/nsch/technical-documentation/nonresponse-bias-analysis.html>. Accessed December 28, 2020.

Receive AJIC Table of Contents Via E-Mail

Get a first glance at the latest issue with a Table of Contents e-Alert.

Sign up through our website www.ajicjournal.org

Go to the **FEATURES** section on the home page, click on **Register for Email Alerts** and follow the instructions.

Table of Contents Email Alerts are sent out when each new **AJIC** issue is posted to www.ajicjournal.org