US county case studies of COVID-19 transmission in Q4 2020

David Brackbill

Abstract

This paper addresses the disproportionate impact that COVID-19 can have on smaller counties by identifying which counties are experiencing the highest rates of infection and death per capita. Using data from John's Hopkins University and the Bureau of Labor Statistics, we create a profile of these small counties based on infection and employment statistics. This study finds that unemployment in the deadliest counties is a percentage point lower than the national average, suggesting that workforce participation may correlate with death rates in rural areas. We have provided ideas for expanding the scope of study to verify these claims.

Keywords: data analysis, rural, metropolitan, unemployment

1 Introduction

As the COVID-19 virus continues to rage its way through American communities, it's important to notice that some counties have proven remarkably resistant to the virus, while others have suffered great losses of life and economic prosperity. Why are certain counties resistant, while others fail to stem the tide?

Some of it may come down to inaccurate data and inconsistent case reporting. The case fatality rate (CFR) is the commonly used but flawed heuristic for the threat of disease. Computationally, it is the confirmed deaths from the disease / confirmed cases of the the disease * 100.[1] Because it relies on perfectly accurate reporting of cases and deaths is always going to be a heuristic and not a perfect measure of the actual infection fatality rate (which is what researchers discover about a disease in hindsight).

Further, the CFR is confounded by factors like age, level of healthcare, knowledge about the disease in the medical community and social protocols for disease prevention. These factors confound a purely biological view of the virus' deadliness, but can be used to model how deadly a virus is within a given society at a given time. We will be using CFR, among other measures of the COVID-19 virus' spread in our analysis on at-risk American counties.

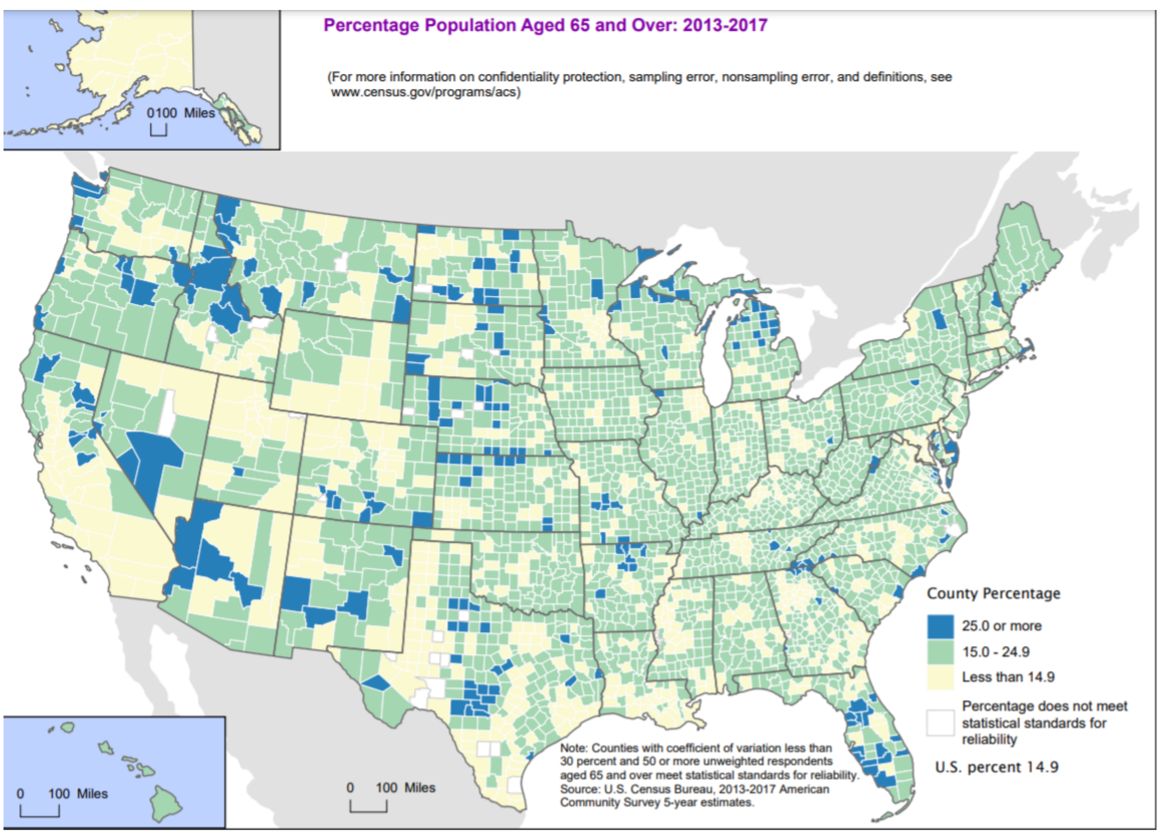
[1]https://ourworldindata.org/mortality-risk-covid

1.1 Summary of Studies

Population density is likely the leading factor in the transmission of the COVID-19 virus. According to Zhang et al., population density is the primary predictor of speed of community transmission in non-rural areas.[2] This study estimated that 43% of the variations in infection rate by county stemmed from population density. It follows from higher population densities in these areas (suburbs and cities) that infected and susceptible individuals have more in-person contacts per day, spreading the disease faster. In the SIR model, this corresponds to a higher beta parameter.

Interestingly, Zhang et al. found that population density was insignificant in comparisons of rural counties.[2] The study concluded that rural areas can have outsized infection rates for their population density and are also highly at-risk compared to metropolitan areas because of their higher rates of comorbidity, age and poverty. It is thus essential that analysis of at-risk counties take these risk factors into account.

FIGURE 1: Percentage of population aged 65+ per county, estimated from 2013-2017 by the American Community Survey.



[2]Zhang et al.

2 Method

Our methods involve graphical analysis of map data along with examining correlation between confirmed cases per capita and confirmed deaths per capita in the counties with the highest respective rates of death and infections.

FIGURE 2: 7-day smoothed average of confirmed cases per capita, taken on September 14.

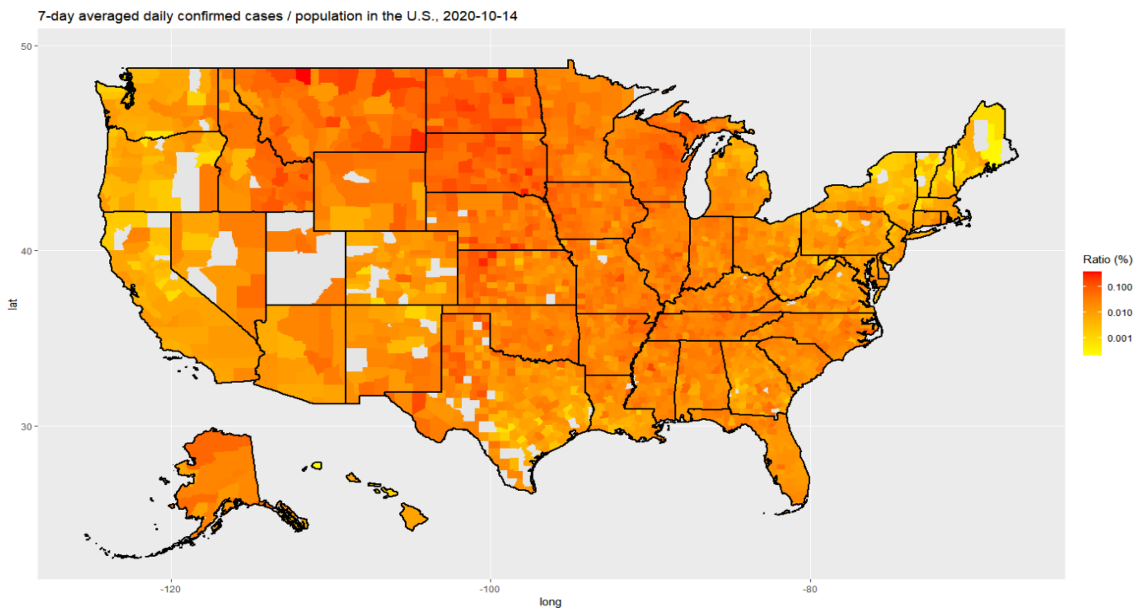


FIGURE 3: 7-day smoothed average of confirmed cases per capita, taken on October 14.

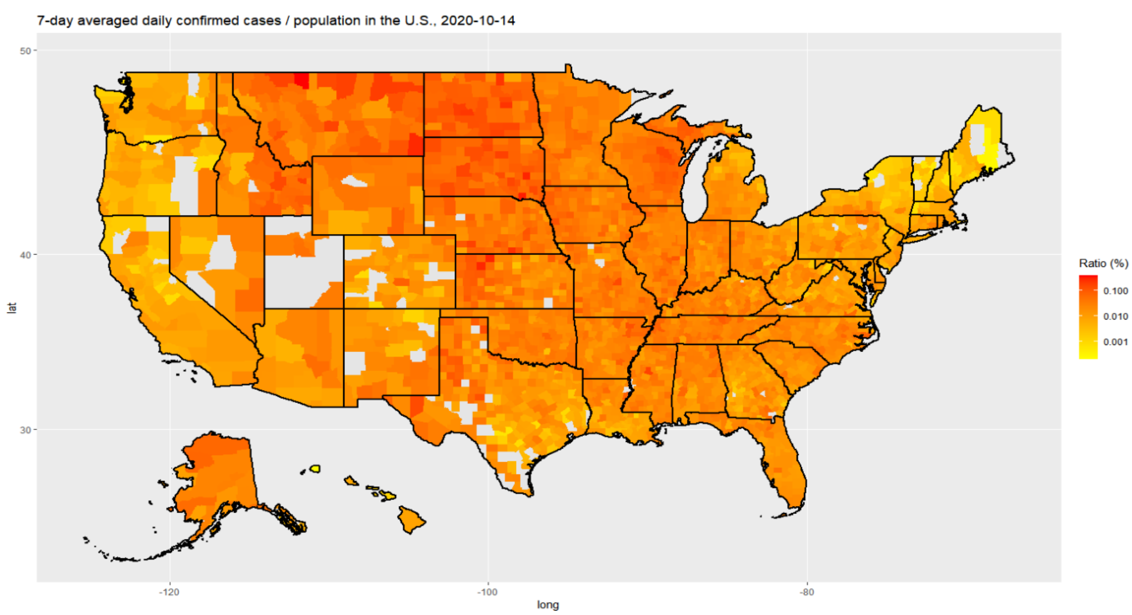


FIGURE 4: 7-day smoothed average of confirmed cases per capita, taken on November 14.

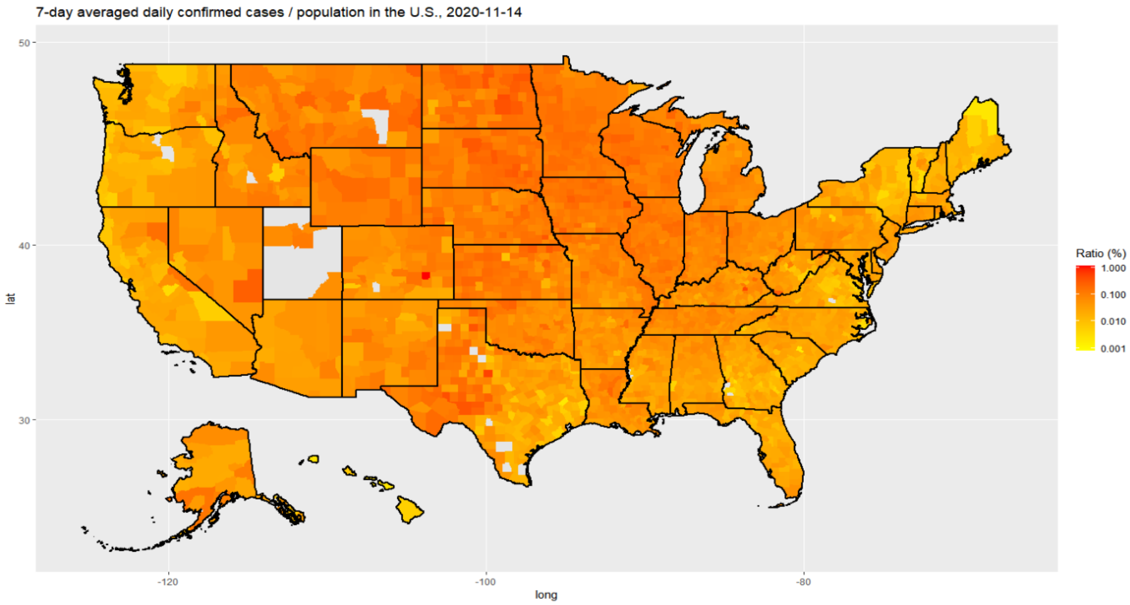
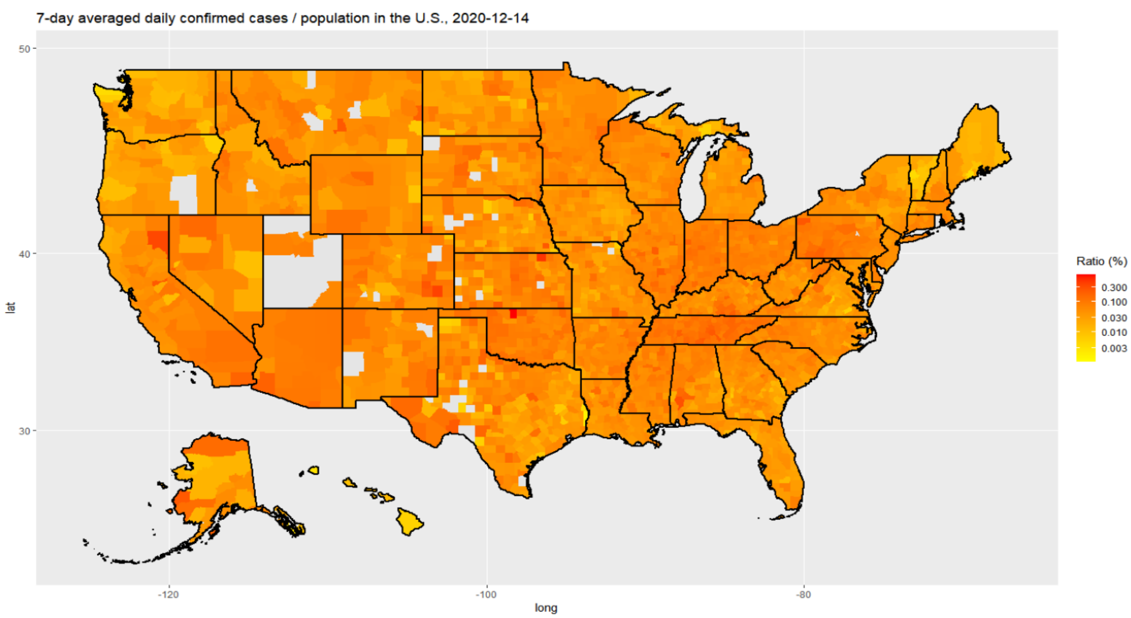


FIGURE 5: 7-day smoothed average of confirmed cases per capita, taken on December 14.



Using modified code and updated date ranges from the second data project, we have mapped 4 county snapshots (located above this text) of the 7-day average for the 14th of the month for the months of September, October, November and December. By visual comparison, we notice consistent sites of high infection and death rate.

We compare these areas of high infection and death rates to the proposed risk factors for infection and death within a county, namely unemployment rates and population of elderly. We reference the pre-existing map built from data from the American Community Survey

(Figure 1) that displays share of elderly population per county.

Using data analysis in R, we also have constructed maps of the unemployment rate from August, September and October. This data set comes from the Bureau of Labor and Statistics and includes information on rates of unemployment and total unemployment per month in 2020, all at the county level.

We merge this county unemployment data with our existing data frame of COVID-19 infections, deaths, populations, etc. at the county-level. We sort this data frame by high rates of unemployment, infection and death to create a subset of counties that we deem to be highly at-risk of economic recession in the coming years due to COVID-19.

2.1 Data sets

Here we provide a list of the data sets used and their file paths in the attached files:

- Bureau of Labor Statistics' unemployment by county, 2020:
- *Data/BLS_Unemployment _by _County _2020.xls*
- John Hopkins' Covid-19 infection data, from Data Report 2 code:
- */120c Report 2 Upload(Dec16).rmd*

3 Results

Perhaps because of the vastness of the American rural area, rural areas have high variance in reaction to the virus. Some rural areas fare better than their metropolitan counterparts, like H while others experience devastating amounts of losses. We wanted to know if smaller counties tended to be safer than larger counties. We can do this by comparing the mean cases and deaths per capita of counties with less than 30,000 individuals against counties with more than 30,000 individuals.

TABLE 1: Comparison of infection, deaths between 30,000+ and <30,000 population counties

Population of county	Avg cases per capita (%)	Average Deaths per capita (%)
30,000+	5.04%	0.08%
<30,000	6.02%	0.11%

Surprisingly, the rural areas actually fare worse in both measures of infection containment and death prevention. Considering the early stages of the outbreak swept through extremely populated counties in the New York and California Bay areas, this suggests that rural areas are positioned to outpace their metropolitan counterparts in deaths and cases.

By tallying up the cumulative COVID-19-caused deaths in each county on December 14, we can confirm what this table suggests: the worst-hit areas both in terms of infection and deaths aren't the metropolitan areas, but a scattered amount of rural counties.

FIGURE 6: Cumulative US county cases per capita on December 14.

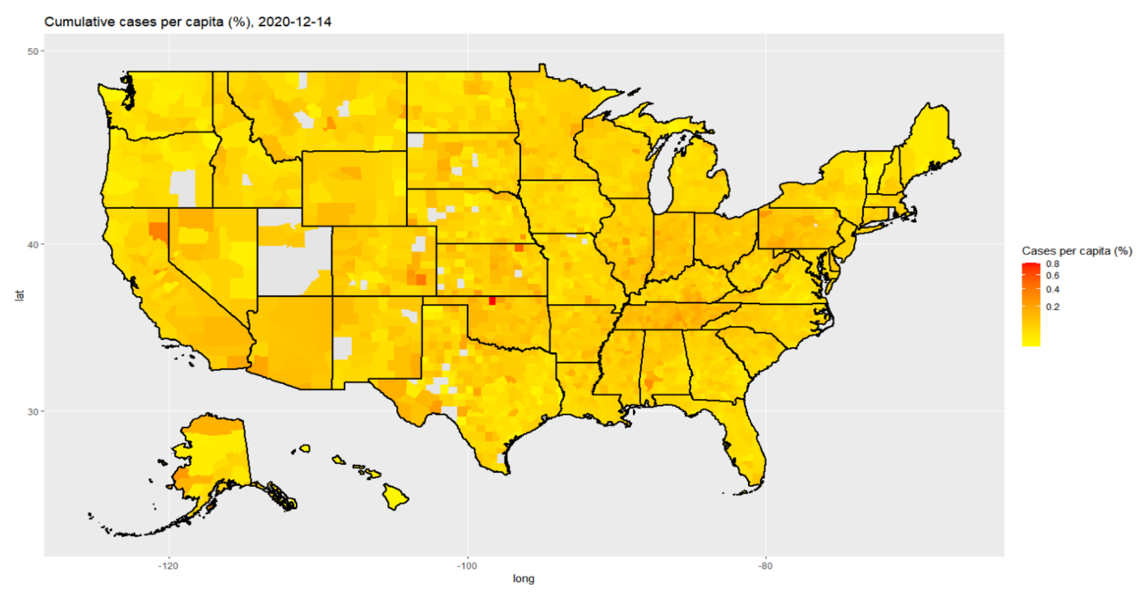
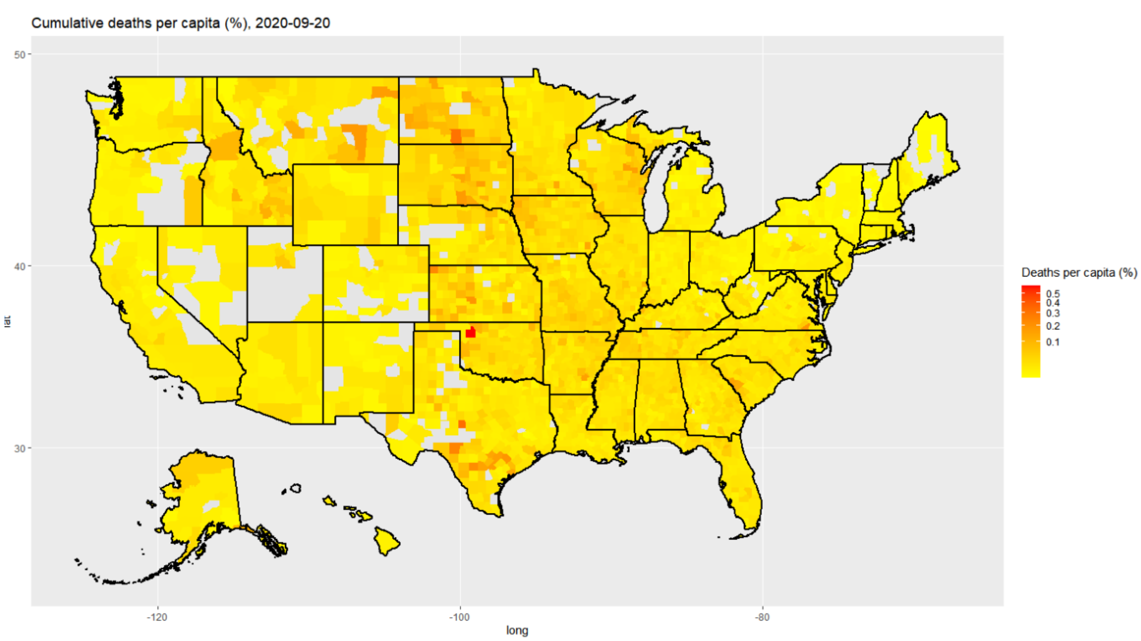


FIGURE 7: Cumulative US county deaths per capita on December 14.



Under-served areas in Oklahoma, Kansas and the Dakotas have experienced the worst of COVID-19's effects in the United States. Are they also at risk of economic recession?

TABLE 2: 10 highest cases per capita, county-level

County, State	Cumulative Cases	**Cases per capita (%)**	Cumulative Deaths	Deaths per capita (%)
Crowley, Colorado	1457	24.04%	10	0.16%
Norton, Kansas	1141	21.28%	24	0.45%
Bon Homme, South Dakota	1427	20.68%	21	0.30%
Lincoln, Arkansas	2680	20.58%	20	0.15%
Buffalo, South Dakota	401	20.44%	10	0.51%
Dewey, South Dakota	1170	19.86%	7	0.12%
Chattahoochee, Georgia	2144	19.66%	3	0.03%
Lake, Tennessee	1307	18.63%	11	0.16%
Trousdale, Tennessee	2023	17.93%	12	0.11%
Buena Vista, Iowa	3379	17.22%	21	0.11%

Interestingly, the deadliest counties per capita have experienced less unemployment since January. Compared to the national average of 5% employment, this group experiences 4.16$ unemployment, which would go against traditional models of increased disease susceptibility to highly unemployed areas.

However, lowering levels of workforce participation may skew the unemployment rate lower than it would normally be, based on total working-age population.[3]

4 Discussion

Because of their comprehension and size, these data sets are very promising for matched-pair causation analysis, where subsets of counties with similar demographics and densities

[3]Bureau of Labor Statistics.

TABLE 3: 10 highest deaths per capita, county-level

County, State	Cumulative Cases	Cases per capita (%)	Cumulative Deaths	**Deaths per capita (%)**
Gove, Kansas	315	11.95%	20	0.76%
Jerauld, South Dakota	252	12.52%	15	0.75%
Dickey, North Dakota	607	12.46%	32	0.66%
Foster, North Dakota	524	16.32%	19	0.59%
Gregory, South Dakota	463	11.06%	24	0.57%
Turner, South Dakota	895	10.68%	47	0.56%
Hancock, Georgia	554	6.55%	47	0.56%
Emporia, Virginia	372	6.96%	29	0.54%
Renville, North Dakota	268	11.52%	12	0.52%
Buffalo, South Dakota	401	20.44%	10	0.51%

but differing responses (controlled variables) can be created. Then, statistical analyses such as the t test or simple difference in sample means between two groups of responses can be used to determine what characteristics rebuff the virus.[4]

We chose a few case studies of counties to reduce the scope of this study. Further studies may want to examine the entirety of the county data sets to examine the total amounts of poverty stemming from the COVID-19 virus and government handling of the crisis. Economics-minded scholars may even be able to estimate the loss of GDP and human potential due to income restrictions, unemployment and death, rather than relying on existing data heuristics about the effect of unemployment on national GDP and mental health, as we have done in this study.

In this paper, we did not compare the SIR model's estimated beta parameters of the case-study counties. However, a motivated researcher could compare the estimated beta parameters of counties with similar population densities as measures of community adherence to social distancing.

[4]Reiter

TABLE 4: Unemployment vs cases in 10 most infectious counties per capita, 2020

County, State	**Cases per capita**	January Unemployment (%)	October Unemployment (%)	**9-Month Difference** (%)
Crowley, Colorado	0.24	4.8	5.2	0.4
Norton, Kansas	0.21	2	2.2	0.2
Bon Homme, South Dakota	0.21	3.8	3.3	-0.5
Lincoln, Arkansas	0.21	5.5	5.6	0.1
Buffalo, South Dakota	0.20	6.2	5.4	-0.8
Dewey, South Dakota	0.20	6.9	6.2	-0.7
Chattahoochee, Georgia	0.20	4.3	4.5	0.2
Lake, Tennessee	0.19	7.1	10.4	3.3
Trousdale, Tennessee	0.18	3.3	6.1	2.8
Buena Vista, Iowa	0.17	2.6	2.2	-0.4
AVG:	0.20	4.65	5.11	0.46

Additionally, we did not examine ICU and hospital bed shortages as a factor for vulnerability to the virus and predictor of death rates. However, this data is being curated on a county-level by the University of Minnesota COVID-19 Hospitalization Tracking Project. The implementation of this data into existing risk models could save lives by diverting aid, resources and National Guard assistance to areas with infection rates on the verge of wiping out access to medical care.

Further, average income per county data was available to this study, but due to time constraints was not added. High employment rates in the worst-affected counties may be spurred by low levels of savings from historically low income per capita. This can be easily checked and if correct would suggest that individuals worked even through pandemic conditions because they could not afford to lose their jobs.

Table 5: Unemployment vs deaths in 10 deadliest counties per capita, 2020

County, State	**Deaths per capita**	January Unemployment (%)	October Unemployment (%)	**9-Month Difference** (%)
Gove, Kansas	0.76%	2.4	2.1	-0.3
Jerauld, South Dakota	0.75%	2.3	2.2	-0.1
Dickey, North Dakota	0.66%	2	2.2	0.2
Foster, North Dakota	0.59%	3.3	2.6	-0.7
Gregory, South Dakota	0.57%	4.3	3	-1.3
Turner, South Dakota	0.56%	4.2	2.6	-1.6
Hancock, Georgia	0.56%	5.8	7.1	1.3
Emporia, Virginia	0.54%	5	10.6	5.6
Renville, North Dakota	0.52%	3.5	3.8	0.3
Buffalo, South Dakota	0.51%	6.2	5.4	-0.8
AVG:	0.60%	3.9	4.16	0.26

5 Conclusion

Although much concern about metropolitan areas is warranted, because of their high population density, small communities can experience high rates of transmission as well. These communities are at higher risk because of increased mean age and lower income per capita. These small counties should be examined and studied to understand how the virus spread so rapidly, because of the relative ease of contact tracing within a smaller network of people.

These counties may be protected from further outbreaks by increased government assistance, lowering the daily contacts per individual.

References

Bureau of Labor Statistics. (n.d.). Local area unemployment statistics home page. U.S. Bureau of Labor Statistics. https://www.bls.gov/lau/

Han, J., Meyer, B., & Sullivan, J. (2020). Income and Poverty in the COVID-19 Pandemic. doi:10.3386/w27729

John Hopkins Coronavirus Resource Center. (n.d.). United States cases by county. Johns Hopkins University Medicine. Retrieved Dec 14, 2020, from https://coronavirus.jhu.edu/us-map

Oxford. (n.d.). Mortality risk of COVID-19 - Statistics and research. Our World in Data. https://ourworldindata.org/mortality-risk-covid

Reiter, J. (2000). Using Statistics to Determine Causal Relationships. The American Mathematical Monthly, 107(1), 24. doi:10.2307/2589374
U.S. Census Bureau (2017). 2013-2017 American Community Survey. Retrieved from [https://www.census.gov/programs-surveys/acs/news/data-releases/2017/release.html].

University of Minnesota. (n.d.). COVID-19 hospitalization tracking project. Carlson School of Management. https://carlsonschool.umn.edu/mili-misrc-covid19-tracking-project