An Analysis of COVID-19 and Poverty in the United States

Abstract

The COVID-19 pandemic has left no corner of the United States untouched. In order to be better prepared for the reminder of the pandemic (as well as for any future pandemics) and ensure that healthcare resources are equitably allocated, epidemiologists and public health officials must work to further their understanding of who viral pandemics hit the hardest. Here, we report on descriptive and exploratory research on the impact of poverty rates and population densities of Illinois counties and counties nationwide on those counties’ COVID-19 cases and deaths per capita or per 1,000 residents (depending on the kind of analysis used). The study found the strongest correlations between poverty rates and 1) COVID-19 deaths per capita in rural counties nationwide, 2) COVID-19 deaths per capita in all counties nationwide, and 3) COVID-19 cases in non-rural counties nationwide. Overall, little to no correlation was found between poverty rates and COVID-19 cases or deaths in Illinois, though individual counties in the southeastern and central parts of the state displayed isolated correlations. By extrapolating the data to a national level, one can see that the heaviest correlations between poverty rates and COVID-19 cases and deaths occur in the American South and West, while the weakest correlations occur in the Midwest and Northeast. With Illinois serving as the focal point of our analysis, we have identified the areas of the state that would benefit most from more healthcare resources, and how the state’s COVID-19 response compares to that of other states — something both business professionals and policymakers alike can reflect on as they make decisions moving forward.
Introduction

The first confirmed COVID-19 case in the U.S. occurred in the middle of January 2020 and the virus soon spread to every state. The pace at which the disease spread prompted government-mandated shutdowns that prevented millions of members of the blue-collar workforce from working and sent their white-collar counterparts into work from home. At the April 2020 peak, approximately 20 million people were unemployed and 23 million filed for unemployment insurance. The economic impacts of COVID-19 have been profound, but further assessing the relationship between economic indicators and the spread of cases could expose disparities in the current public health infrastructure.

After collecting poverty rate and population density data from the U.S. Census Bureau and COVID-19 case and death data from The New York Times, we were able to use R—an open-source software for statistical computing—to synthesize the data into one, clean data set and analyze it through various methods of statistical (e.g., correlation and linear regression) and visual analysis.

This paper seeks to answer the question—How did poverty rates and population densities on a county by county basis both in the state of Illinois and, for context, nationwide influence COVID-19 cases and deaths in those counties? The study examines the relationships between these four variables (poverty rate, population density, COVID-19 cases, and COVID-19) deaths on a per-county basis throughout the country, with a specific focus on Illinois for more in-depth analysis.

Although poverty rate is the primary independent variable that we are using in our research, it is important to recognize that poverty looks and feels different depending on where it is happening. We are attempting to address this concern by grouping counties based on
population density. Counties will be labeled as either rural (≤ 500 residents per square kilometer) or urban (>500 residents/km²). In order to better create a level playing field, the analysis will only consider COVID-19 cases and deaths from April 1, 2020 to November 30, 2020—the timeframe before the first vaccine was distributed, but after mass distribution of testing equipment. Selecting this time period is intended to show the true impact COVID-19 had in each county.

Illinois was selected for analysis due to the state’s diversity of counties on a poverty rate and population density basis. The state hosts the third largest city in the U.S. by population (Chicago) as well as more scarcely populated cities in its western and southern regions. Similarly, there is a healthy blend of counties whose GDPs are primarily driven by a blue-collar workforce and those by a white-collar workforce.

Knowledge of how viruses like COVID-19 spread helps epidemiologists and public health officials to make decisions about the allocation of healthcare resources during pandemics like this one. The misallocation of these resources is a humanitarian issue, and one that must be avoided by researching the factors that influence the spread and deadliness of COVID-19. Through this study, we hope to generate a better understanding of who COVID-19 has hit the hardest and why.

**Literature Review**

This study will assess each county’s poverty rate based on 2019 figures to show figures before COVID-19’s impact. Research described by Han, Meyer, and Sullivan found that “income poverty fell shortly after the start of the COVID-19 pandemic in the United States” (Han et al. 3). The authors believed this trend could be attributed to government policy, including increased unemployment benefits and Economic Impact Payments (EIP). Per the aforementioned
article, the EIP provided $1,200 to individuals with less than $75,000 and to single parents with incomes below $112,500. Similarly, the CARES Act provided an additional $600/week to claimants. The authors noted that “these policies have the potential to significantly boost family incomes and lift many families, at least temporarily, out of poverty” (Han et al. 5). Given the temporary nature of the government intervention, this analysis will utilize pre-COVID-19 poverty rates to ensure a more accurate portrayal of each county.

Basic intuition surrounding news media coverage of COVID-19 and subsequent investigation dictates that poorer communities are the population segments hit hardest by the fallout of the virus. However, research conducted by Finch and Hernandez-Finch complicates this dynamic, at least in the first months of the pandemic. Their study sought to analyze the relationship between poverty and COVID-19 cases and deaths from January to April of 2020. In their research, the authors found that “[t]hrough the months of February and March, 2020 there were more confirmed cases of the virus in poorer counties, but by April 1, 2020 the relationship had shifted” such that the amount of COVID-19 cases was greater in more affluent counties (Finch, Hernandez-Finch 1). However, their analysis surrounding poverty and deaths found that “[r]esults for the number of deaths confirmed to be caused by Covid-19 demonstrated a pattern whereby the number of deaths was greater in areas of relatively greater poverty later in the pandemic” (Finch, Hernandez-Finch 1). Moreover, the effect was even further exacerbated in urban communities. Even though the relative amount of cases may have been increasing at a faster rate in more affluent communities during the early stages of the pandemic, poorer areas still faced a worse fallout. An important qualification mentioned in the author’s discussion of these findings notes that there was a distinct lack of testing available in poorer areas throughout their time window, possibly contributing to the observed change in COVID-19 case data between
affluent and disadvantaged counties. To best avoid a similar pitfall in our own study, we will be concentrating on a time horizon where testing was as uniformly available as possible: from April 1 to November 30.

General consensus maintains that increased global population density and the consequential increased biodiversity will lead to the next pandemic. The implications of this analysis will inform leaders in Illinois Public Health as to how to best prepare for this next eventual event. Current research from the World Bank indicates those who were poor heading into the pandemic are “bearing the brunt of the crisis” (World Bank Group 7). This is largely because their jobs are more easily impacted by recessions. For example, less-wealthy, blue-collar workers are less likely to be able to work from home relative to white-collar workers. This analysis will look to build on this finding by determining if already impoverished counties were more susceptible to COVID-19 cases and deaths.

Disparities in economic impact resulting from the COVID-19 pandemic also seem to be correlated to relative population density as well. In the research conducted by Mueller, Farrell, McConnell, and Burow, the study highlighted the worsened economic condition faced by those in rural counties compared to national averages. When surveyed for employment status during June and July of 2020, “a total of 13.4% of the labor force in our sample reported that they were temporarily unemployed in the month prior to the survey, which is more than double the national average (6.0%, reported by the Bureau of Labor Statistics)” (Mueller, Tom J., et al 3) Furthermore, the polled rural population suffered from worsened rates of income loss, mental distress, and economic confidence than the corresponding national averages. As a result, our analysis will denote each county as either rural, suburban, or urban to observe any similar findings.
A previous research article studying the elements that affect the spread of COVID-19 found that, along with air pollution, population density was one of the two “main variables driving [COVID-19] viral spread” (Aabed and Lashin 1185).

Poverty aside, the rural and urban populations are already divergent in terms of health. According to the Health Policy Institute, the rural population experiences higher rates of chronic conditions, limitations (e.g., difficulty walking, lifting, or standing), risky health behaviors (e.g., smoking and alcohol consumption), uninsurance, and unemployment. Additionally, the rural population experiences lower rates of testing for chronic conditions. Moreover, “the proportion of rural residents reporting fair to poor physical health is almost one and a half times the proportion of urban residents” (“Rural and Urban Health”). Given that “adults of any given age with certain underlying medical conditions [and] some people with certain disabilities” (“Underlying Health Conditions”) are more likely to get severe illness from COVID-19, one can imagine how the above-mentioned differences between rural and urban health could result in differences in COVID-19 cases and deaths between rural and urban places.

Economic resources are not the only ones that are unevenly distributed. A previous research article studying the “spatial accessibility of COVID-19 healthcare resources [in Illinois]” (Kang et al. 1) found that:

- “At the state level, ICU beds and ventilators are not equitably distributed, especially considering the number of COVID-19 confirmed cases” (Kang et al. 10)
- “[P]eople living in areas with low accessibility are more vulnerable to the external stresses, such as the COVID-19 spread, socioeconomic status, housing type and transportation, and household characteristics and disability” (Kang et al. 12)

This inequitable distribution of healthcare resources is a humanitarian issue, one that public
health officials may have been able to avoid if they had a better understanding of the factors that influence COVID-19 spread, which is what we aim to develop through our research.

Regarding methodology, there are a few notable studies that draw parallels between COVID-19 cases and poverty rates using publicly available datasets. In the aforementioned analysis conducted by Finch and Hernandez Finch, researchers obtained COVID-19 data from the New York Times national COVID-19 dataset and poverty data from the Poverty Solutions Initiative (PSI) database via the University of Michigan. Rather than use standard poverty rate metrics, however, the researchers opted to use something known as the “Index of Deep Disadvantage” or “The Poverty Index” that assigned a “relative poverty score” to each county based on variables such as income, health, education, and social mobility — income being the most heavily weighted variable in the index. Finch and Hernandez Finch then factored population metrics into their study, though they did not make population density one of their variables. Instead, they categorized counties as either “Urban” or “Not Urban” based on total gross population. Any county with a population over 1 million residents was deemed urban, while any county with a population under 1 million residents was deemed not urban.

One of the limitations of this approach, as referenced in Wong and Li’s study, is that the gross population does not account for how close residents live to each other. A county with over a million people could have lower population density than a county with 980,000 people, yet based on Finch and Hernandez Finch’s methodology, the less densely populated county would be considered Urban. It’s no secret that social distancing is key to slowing the spread of the virus, and, according to Wong and Li’s study, “how far people can be spatially separated is partly constrained by population density” (1). To account for this Wong and Li ran a “regression analysis using US Census Bureau Metrics from 2010 which were used to derive population
density information” (3). They then took a natural logarithm of daily confirmed cases, and
calculated a seven day moving average to “ensure that the frequency distribution would not be
severely influenced by outliers” (3). In doing this, they found that the impact of population
density on the spread of COVID-19 cases to be “modest and relatively stable over time,” (5)
though variations in testing frequency across communities remains to be another limitation. As
asserted in Diego F Cuadro et al’s study “Dynamics of the COVID-19 epidemic in urban and
rural areas in the United States,” people in rural areas “tend to have less access to testing and
general medical care than those in urban areas”.

Methods

Procedure and Data Collection

The foundation of this analysis is publicly available COVID-19 and demographic data at
a county level. We accessed the New York Times COVID-19 database through an R package on
Github called “covdata”. This provides access to large quantities of data including the Centers
for Disease Control and Prevention (CDC)’s provisional death count by week, state, and sex,
daily international COVID-19 cases and deaths, case and death count in the US by race, and the
New York Times’ excess mortality estimates. This analysis focused specifically on the daily
county-level data collected and provided by the New York Times. This dataset contains the
following six variables: date, county, state, fips (a unique code assigned to each county), cases,
and deaths. For each date, the dataset showed the cumulative amount of cases and deaths.
Because our analysis focused on the range from April 1, 2020 and November 30, 2020, we
subtracted each county’s April 1 figure from our November 30 figure. This provided us with the
COVID-19 case and death data that we would then use in our analysis.
The US Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) provides single-year income and poverty estimates for all US states and counties. From here, we gathered variables at a per county level including total poverty rate, under 18 poverty rate, median household income in dollars, population density, and total population. The poverty rate is based on the criteria laid out by the Census Bureau found here. Each of these variables were joined to their corresponding county in the other dataset described. The following variables were derived from existing variables: cases per capita, deaths per capita, cases per 1000 people, deaths per 1000 people, and county type. County type is either rural or non-rural and is based upon population density. A county with fewer than 500 people per square kilometer was labeled rural and those with more than 500 people per square kilometer were labeled non-rural. Cases per capita was calculated by dividing a county’s cases count by its population, and deaths per capita were calculated by dividing a county’s death count by its population. Per 1000 people figures were calculated by multiplying its per capita figure by 1000.

For our Illinois analysis, the larger dataset was filtered to only show Illinois counties and their corresponding attributes. Each county name was cleaned to ensure it matched the counties listed in R’s ggplot package for the sake of visualizing on a map. Cleaned data in both the Illinois dataset and the dataset of each county in the US made for a much more streamlined analysis process.

**Statistical Analysis**

Our goal from the onset was to collect data relevant to our research question and observe relationships between variables. This analysis utilized Pearson correlation to assess the relationship between variables because it is the optimal measure for linear relationships and does not seek to identify dependence between variables. Eliminating dependence was an important
aspect of the decision to use Pearson because this analysis was not striving to identify causality, considering the difficulty of confirming causality through publicly available data. Each Pearson test was run through R using RStudio.

Our analysis also benefited from ample data visualizations intended to make viewing the data and making interpretations simpler. Each county was plotted on a scatter plot with poverty rate on the x-axis and COVID-19 cases per 1000 on the y-axis. For the sake of making the graphs easier to read, we added an estimate of the conditional mean function using the LOESS method. This helps the reader identify trends in the data more easily.

Results

This paper explored how COVID-19 cases and deaths per capita correlate with poverty rates on a county by county basis. The study found that, in non-rural counties nationwide, poverty rates correlated to COVID-19 cases per capita at 0.2786 and to COVID-19 deaths per capita at -0.0896. The study also found that, in rural counties nationwide, poverty rates correlated to COVID-19 cases per capita at 0.1215 and to COVID-19 deaths per capita at 0.3127. In addition, poverty rates nationwide (regardless of whether counties were rural or non-rural) correlated with COVID-19 cases per capita at 0.1285 and with COVID-19 deaths per capita at 0.2999. Additionally, poverty rates in Illinois correlated with COVID-19 cases per capita at -0.0688 and with COVID-19 deaths per capita at 0.0567.

The strongest correlations that were found were between poverty rates and 1) COVID-19 deaths per capita in rural counties nationwide, 2) COVID-19 deaths per capita in all counties nationwide, and 3) COVID-19 cases in non-rural counties nationwide. Little to no correlation was found between poverty rates and COVID-19 cases per capita in Illinois. In addition, little to no correlation was found between poverty rates and COVID-19 deaths per capita in Illinois.
Table 1. Correlation of COVID-19 cases and deaths per capita with poverty rates

<table>
<thead>
<tr>
<th>Poverty rates</th>
<th>COVID-19 cases per capita</th>
<th>COVID-19 deaths per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-rural nationwide</td>
<td>0.2786</td>
<td>-0.0896</td>
</tr>
<tr>
<td>Rural nationwide</td>
<td>0.1215</td>
<td>0.3127</td>
</tr>
<tr>
<td>All nationwide</td>
<td>0.1285</td>
<td>0.2999</td>
</tr>
<tr>
<td>Illinois</td>
<td>-0.0688</td>
<td>0.0567</td>
</tr>
</tbody>
</table>

From Visualizations 1 and 2 in the Appendix, one can see that the correlation of COVID-19 cases with poverty rates and the correlation of COVID-19 deaths with poverty rates tended to be high in the same states (e.g., Arizona, Mississippi, and Rhode Island) and low in the same states (e.g., Nebraska, New Hampshire, and New York). Some states—such as Illinois, Maine, and Utah—actually flipped from having displayed a positive correlation of COVID-19 cases with poverty rates to having displayed a negative correlation of COVID-19 deaths with poverty rates, or vice versa.

From Visualizations 3 and 4, one can see that the correlation of COVID-19 cases with poverty rates and the correlation of COVID-19 deaths with poverty rates also tended to be high in the same areas (such as Southeastern Illinois and South Central Illinois) and low in the same areas (such as Southwest Illinois) in the state of Illinois. What Visualizations 3, 4, and 5 do not exhibit, however, is an easily identifiable correlation of COVID-19 cases and deaths to poverty rates on a county by county basis in Illinois.

Visualizations 6 and 7 are both scatter plots. Visualization 6 shows the relationship between COVID-19 cases per 1,000 residents and poverty rates by county nationwide. Visualization 7 shows the relationship between COVID-19 deaths per 1,000 residents and poverty rates by county nationwide. As one can see, both display a positive relationship between
their respective variables (the relationship between Visualization 6’s being the stronger of the two). COVID-19 cases per 1,000 residents ranged from 1.42 to 219.89, COVID-19 deaths per 1,000 residents from 0 to 7.63, and poverty rates from 2.7% to 47.7%.

Visualizations 8 and 9 are also both scatter plots. Visualization 8 shows the relationship between COVID-19 cases per 1,000 residents and poverty rates by county in Illinois. Visualization 9 shows the relationship between COVID-19 deaths per 1,000 residents and poverty rates by county in Illinois. As one can see, both display a U-shaped relationship between their respective variables (the relationship between Visualization 9’s being the more amplified of the two). COVID-19 cases per 1,000 residents ranged from 23.77 to 88.47, COVID-19 deaths per 1,000 residents from 0 to 2.34, and poverty rates from 4.0% to 25.4%.

Discussion and Interpretation

Our analysis makes use of a number correlation statistics and coefficients to both effectively and concisely compare poverty rates to COVID-19 cases and deaths based on population density. We utilize data coding to break down the population into segments and draw conclusions based on the results. As previously noted, poverty and COVID-19 deaths correlated most strongly in rural counties nationwide, while poverty and COVID-19 cases correlated most strongly in non-rural counties nationwide. Though this dichotomy between cases and deaths in rural and non rural counties may seem paradoxical at first glance, there are a few different possible interpretations. One is that cases are more visible in non rural counties because of a stronger societal impetus to get frequently tested for COVID-19. In major cities such as New York, Los Angeles, and Chicago, for example, many workplaces instituted mandatory COVID-19 testing requirements early in the pandemic to ensure offices and job sites would not become hot spots for the virus. Workplaces in rural counties, however, generally waited longer to
institute mandatory testing requirements, thereby increasing the likelihood of undetected cases. Beyond social reasons, people from rural counties generally have less access to testing sites and primary care facilities, making them less likely to seek impromptu medical care. Studies from Arcury et al. suggest that patients are less likely to travel to go see a doctor if they live far from one.

Where people impacted by urban poverty saw higher COVID-19 case counts per capita, people in rural areas saw higher COVID-19 death counts. One possible interpretation of this is the underlying demographic and psychographic factors that define rural poverty. Foremost, individuals affected by rural poverty are more likely to be older, unvaccinated, and managing a preexisting condition. Whether it be obesity, hypertension, chronic lung disease, or substance abuse disorder, there’s a greater concentration of people in rural counties afflicted by such ailments and each one puts individuals at a greater risk for death if they contract COVID-19. Similarly, age is a major factor. The older one is, the more susceptible they are to death upon contraction of COVID-19, and, according to 2020 US Census data, 17.8 percent of people in rural areas are aged 65 and older compared to just 13.8 percent in urban areas. Finally, studies via Monnat et al. indicate that 46 percent of individuals in rural counties were fully vaccinated compared to 60 percent in Urban counties as of September 2021. While each of the aforementioned interpretations aims to provide context and insight as to why there is a higher correlation between non rural counties and cases and rural counties and deaths, further research is needed for confirmation across the board.

Conclusions, Implications, and Recommendations

The most significant finding from our analysis is that, among all variables, poverty rates and COVID-19 deaths per capita are the most highly correlated. Further, this relationship is the
most prominent at a rural and national basis, suggesting that an individual’s poverty level most strongly affects their likelihood to die of the virus rather than just one’s probability of contracting the disease. Considering this reality, closer attention must be paid to those suffering from rural poverty in the national fight against COVID-19. When distributing economic aid to the most affected areas, the federal government should more seriously consider providing aid to rural areas at a higher concentration than it currently is. The largest vehicle for federal aid to America’s poverty stricken population has been the American Rescue Plan. Signed into law in March of 2021 by President Biden, the Treasury Department has distributed approximately $700 billion in aid over the past 8 months. The economic growth created by this plan has helped America’s most desperate populations regain their economic footing and has provided large amounts of support to the public. However, $300 billion of remaining aid has yet to be distributed, and this is where our findings can hopefully provide some helpful insight. By identifying the states in which poverty and COVID-19 cases and deaths are the most strongly correlated, future spending can be concentrated in these areas, especially to their rural communities. With the ever-growing threat of more severe strains of COVID-19, such as Delta and Omicron, the federal government needs to have a firm understanding of which already-impoverished areas are most severely affected by the virus so aid can be properly distributed to help these needy populations. By utilizing our findings, a better understanding can be formed of such areas, hopefully assisting in the continuing fight against the economic impacts COVID-19.

Moreover, the substantial difference between the correlations of poverty rates and rural and non-rural deaths is noteworthy. As highlighted in the literature review, there exist major differences in overall access to healthcare resources between rural and non-rural communities, causing rural populations to be generally unhealthier than their non-rural counterparts.
Healthcare resources may have been distributed to rural areas at a disproportionately low level due to their small populations and low population densities. In reality, more healthcare investment should have been allocated to rural areas because of the prevalence of pre-existing conditions and the generally lower level of access to quality care. If the health gap between these populations had been bridged properly by government resources, we should have observed the reverse of our findings; rural areas should have received disproportionately higher resources, causing deaths to occur in equal or lower concentrations than in non-rural areas. As a result, in order to best prepare for the spread of the Delta and Omicron variants, more substantial amounts of healthcare investment should be delivered to the worst affected rural areas of the nation.

The primary limitations of our research are twofold: the analysis does not identify a direct causal relationship between poverty rates and COVID-19 cases and deaths and the correlations of those same variables vary state by state to such an extent that some are negative and some are positive. In terms of the causal relationship between poverty and the COVID-19 variables, our commentary is able to offer some possible explanations like insufficient access to healthcare and economic resources, but considerably more qualitative and quantitative analysis needs to be conducted to solidify this relationship. Additionally, more investigation must be done into why the correlation between poverty rates and COVID-19 cases and deaths varies so significantly from state to state. It may perhaps be related to the abundance of rural poverty in each state, but more analysis must be conducted to finalize a conclusion. Our study provides solid groundwork for further exploration into both of these factors, hopefully furthering our understanding of COVID-19 and its effects on the overall wellbeing of the United States.
Appendix

Visualization 1. Correlation of COVID-19 cases with poverty rates on a state by state basis nationwide

Visualization 2. Correlation of COVID-19 deaths with poverty rates on a state by state basis nationwide
Visualization 3. COVID-19 cases per capita on a county by county basis in Illinois

Visualization 4. COVID-19 deaths per capita on a county by county basis in Illinois

Visualization 5. Poverty rates on a county by county basis in Illinois
Visualization 6. COVID-19 cases per 1,000 residents vs. poverty rates by county nationwide

Visualization 7. COVID-19 deaths per 1,000 residents vs. poverty rates by county nationwide
Visualization 8. COVID-19 cases per 1,000 residents vs. poverty rates by county in Illinois

Visualization 9. COVID-19 deaths per 1,000 residents vs. poverty rates by county in Illinois
“American Rescue Plan: Treasury's Progress And Impact After Six Months.”


