Disparities in COVID-19 Mortality Outcomes Associated with the Community-Level Risk Factors

Vivek Kantamani

Background

As of November 2021, the coronavirus disease 2019 (COVID-19) pandemic has resulted in 46,180,190 reported cases and 747,970 deaths in the United States.[1] Data on COVID-19 from the Centers for Disease Control and Prevention suggests that the disease has a disproportionate morbidity and mortality burden associated with racial, ethnic, geographic, and socioeconomic disparities, including social and structural determinants of health – economic and educational disadvantages, health care access and quality, housing access, social network stability, etc.[2]

In particular, while underlying medical comorbidities such as older age, diabetes, and obesity have been identified as resulting in more severe COVID-19 outcomes[3], socioeconomically disadvantaged and minority populations have experienced a greater incidence of COVID-19 related mortality.[4] For instance, estimates of the COVID-19 pandemic's mortality burden (which measures years of life lost from premature mortality while accounting for quality-of-life differences) found that Black and Hispanic populations experienced nearly twice the mortality burden of any other race or ethnic group.[5]

Motivation and Research Question

Emerging data on the disproportionate COVID-19 burden stemming from racial, ethnic, and socioeconomic disparities across various social and structural determinants of health is concerning. While ongoing research has presented numerous medical risk factors, comorbidities, and social risk factors associated with COVID-19, the lack of state and local municipality COVID-19 morbidity and mortality data indexed by pertinent sociodemographic characteristics make it difficult to assess the impact of community-wide social risk factors on COVID-19 morbidity and mortality risk. Thus, research has focused on examining the association between racial, ethnic, socioeconomic, and medical risk factor disparities in COVID-19 morbidity and mortality outcomes at the individual and racial/ethnic group level.

However, racial and ethnic minority groups in the United States are more likely to live in areas characterized by social and structural disadvantages – high unemployment and poverty rates, inadequate housing, limited access to public transportation, etc.[6] Data on these social risk factors and sociodemographic vulnerabilities is typically available in local municipality census data associated with particular census tracts (subdivisions of counties for which the census collects statistical data), not state and local municipality COVID-19 morbidity and mortality data.[7]

The purpose of this study is to extend current research on the association between social risk factors and sociodemographic vulnerabilities with COVID-19 morbidity and mortality data by examining these risk factors and vulnerabilities at the community level, in order to better inform COVID-19 local policy and community response to the COVID-19 pandemic.

Data Collection and Limitations

In response to the COVID-19 pandemic, the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) has released provisional counts of COVID-19 related deaths from the National Vital Statistics System (NVSS). The released COVID-19 morbidity and mortality surveillance datasets index COVID-19 mortality data (the primary response variable) by several explanatory variables – various time periods (weekly, monthly, etc.), geographic regions (counties, local jurisdictions, health service areas, hospital referral regions, etc.), sociodemographic factors (age, sex, race, educational attainment, etc.), individual health factors (health conditions, comorbidities), and community factors (social vulnerability index, access to health services, estimates of excess deaths etc.) in order to facilitate analysis into COVID-19 mortality. The datasets are representative of a retrospective observational study, as the NVSS tabulates provisional counts of COVID-19 related deaths from within the 50 states and the District of Columbia as they are coded and documented on a rolling basis.

First, we must acknowledge the limitations of our proposed datasets. According to the data documentation, while the surveillance datasets are updated regularly, the process of receiving, coding, and tabulating COVID-19 related deaths may take up to several weeks. Therefore, data may be incomplete and not include all deaths that occurred in each time period, especially for more recent time periods. Thus, a limitation of our data is a potential underrepresentation of COVID-19 mortality in each time period.

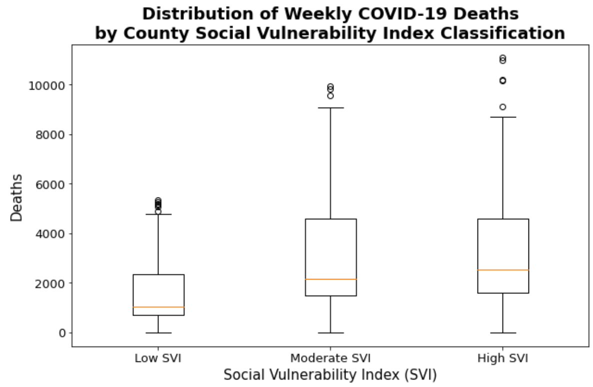
Furthermore, the data documentation notes that deaths attributed to COVID-19 on death certificates either list COVID-19 as either the underlying cause of death, or as a contributing cause of death. Inaccuracy or lack of consistency of reporting according to similar standards across the country may cause the number of COVID-19 related deaths to be underestimated if COVID-19 contributed to the cause of death but was not reported on the death certificate.

Despite these limitations, the datasets sample from the entire population of the United States using reliable sources of mortality statistics – thus, the results of our analysis can reliably be extended to the entire population of the United States.

Therefore, we will analyze the NVSS datasets to examine community risk factors (i.e., social risk factors and sociodemographic vulnerabilities) with respect to COVID-19 morbidity and mortality outcomes to better inform local policy regarding community response to the COVID-19 pandemic.

Data Source: https://www.cdc.gov/nchs/covid19/covid-19-mortality-data-files.htm

COVID-19 Incidence by County Social Vulnerability Index:



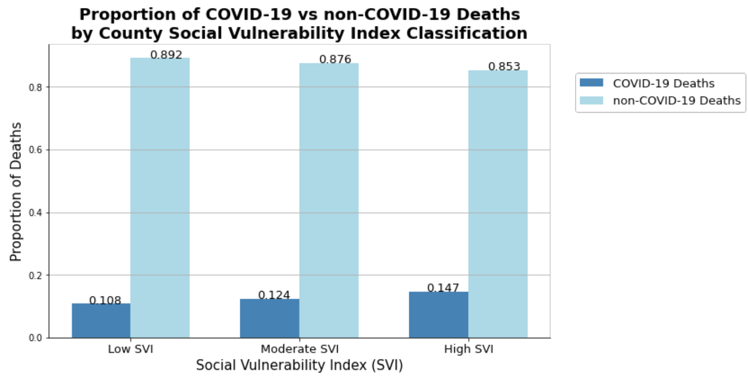


Figure 1: Grouped Boxplot of COVID-19 Deaths by County Social Vulnerability Index Classification
Figure 2: Grouped Bar Chart of the Proportion of COVID-19 Deaths by County Social Vulnerability Index Classification

The comparative boxplots (*Figure 1*) compare COVID-19 mortality for counties ranked by their social vulnerability index (SVI) classification. The SVI rating is a measure used by the CDC to evaluate a community's resilience and capacity to respond to emergency situations, using indices like poverty, access to transportation, housing density, etc. The boxplots demonstrate significant right skew across SVI levels with several (large) outliers. The median number of COVID-19 deaths increases with SVI rating, and as SVI rating increases, the spread of data (measured by the IQR), maximum and minimum bars, and extreme nature of outliers increases.

This figure suggests that a county's SVI rating and the capacity of its communities to respond to disaster is directly associated with COVID-19 mortality outcomes. This provides us with the ability to target resources more pragmatically to vulnerable communities, namely those with a low SVI rating.

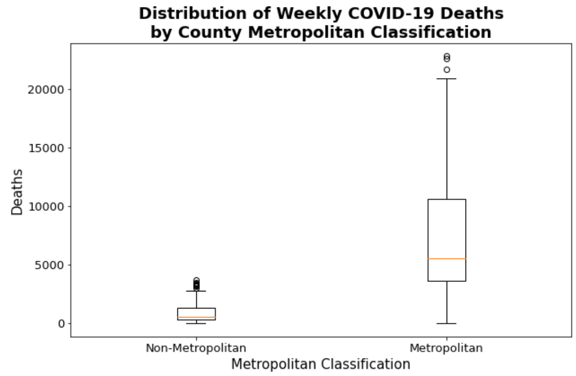
The grouped bar chart (*Figure 2*) demonstrates that while non-COVID-19 deaths do not systematically increase with SVI score, COVID-19 deaths do. In other words, while social vulnerabilities may certainly contribute to mortality outcomes in general situations, the SVI score is more indicative of the number of deaths a community will face during a pandemic, such as the COVID-19 pandemic.

We will compare the mean number of COVID-19 deaths for counties with a high SVI rating with counties that have a low SVI rating. The average number of weekly COVID-19 deaths for high SVI counties is 3399.93 deaths (95% CI: [2852.23, 3947.64]). For low SVI counties, the average number of weekly COVID-19 deaths is 1705.55 deaths (95% CI: [1399.35, 2011.76]).

The mean number of COVID-19 deaths for counties with high SVI ratings exceeded the mean number of COVID-19 deaths for counties with low SVI ratings by 1694.38 deaths (95% CI = [1066.89, 2321.87]). These data provide strong evidence (given a two-sample t-test p-value of $p<0.001$) that the mean number of COVID-19 deaths was greater for counties classified as having a high SVI rating compared to those with a low SVI rating.

Racial and ethnic minority groups in the United States are more likely to live in areas characterized by social and structural disadvantages, namely inadequate housing facilities alongside a high population density. These social risk factors and sociodemographic vulnerabilities are metrics used to determine a county's SVI status – thus a lower SVI rating is indicative of a county's vulnerability to a disproportionate number of deaths due to the COVID-19 pandemic.

COVID-19 Mortality Incidence by County Metropolitan Classification:



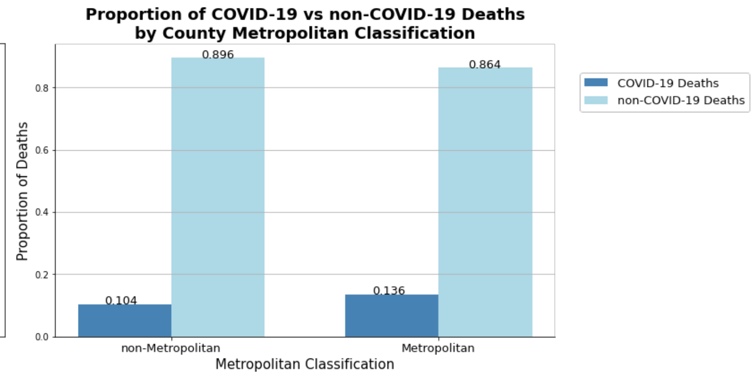


Figure 3: Grouped Boxplot of Weekly COVID-19 Deaths by County Metropolitan Classification
Figure 4: Grouped Bar Chart of the Proportion of COVID-19 Deaths by County Metropolitan Classification

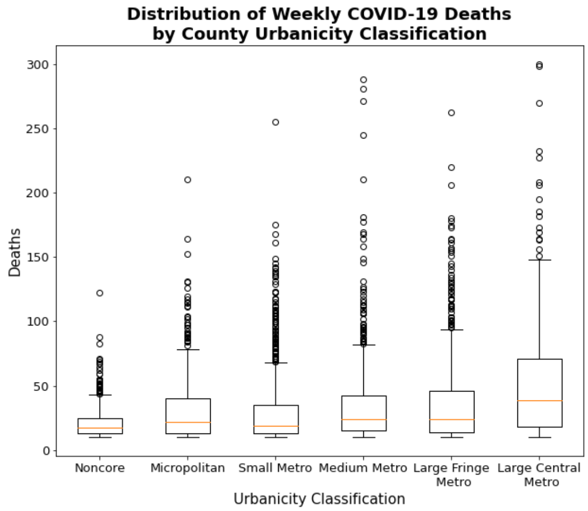
When assessing COVID-19 mortality incidence by a county's metropolitan classification (*Figure 3*), significant right skew is demonstrated alongside an increasing spread as population density increases (a social vulnerability that is tied directly to a county's metropolitan status). Metropolitan counties have a substantially greater median number of deaths than do non-metropolitan counties. The grouped bar chart (*Figure 4*) indicates that as metropolitan classification indicates larger metropolitan areas with greater population density, the proportion of COVID-19 deaths increases.

We will compare the mean number of COVID-19 deaths by a county's metropolitan classification as a metropolitan or non-metropolitan county. The average number of weekly COVID-19 deaths for metropolitan counties is 7385.82 deaths (95% CI: [6204.58, 8567.06]). For non-metropolitan counties, the average number of weekly COVID-19 deaths is 986.98 deaths (95% CI: [785.19,1188.76]).

The mean number of COVID-19 deaths for metropolitan areas exceeded the mean for non-metropolitan areas by 6398.84 deaths (95% CI = [5200.49,7597.19]). These data provide strong evidence (given a two-sample t-test p-value of $p<0.001$) that the mean number of COVID-19 deaths was greater for counties classified as metropolitan areas compared to non-metropolitan areas.

Racial and ethnic minority groups in the United States are more likely to live in densely population metropolitan areas characterized by social and structural disadvantages, namely high unemployment and poverty rates, inadequate housing, limited access to public transportation, etc. A county's classification as a metropolitan area is therefore indicative of a county's vulnerability to a larger number of COVID-19 deaths.

COVID-19 Mortality Incidence by Urbanicity Classification:



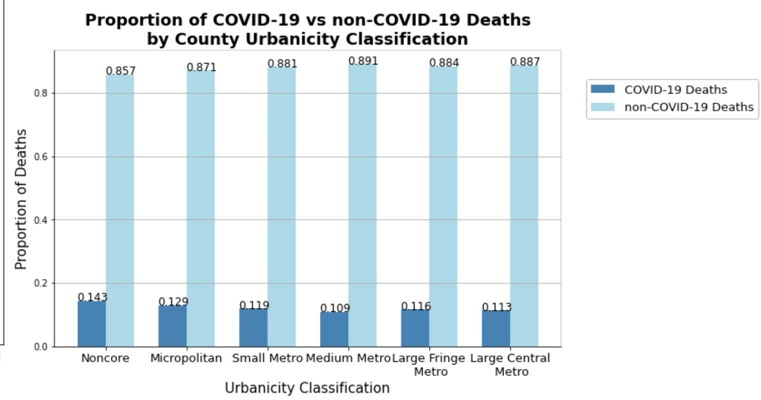


Figure 5: Grouped Boxplot of Weekly COVID-19 Deaths by County Urbanicity Classification
Figure 6: Grouped Bar Chart of the Proportion of COVID-19 Deaths by County Urbanicity Classification

A county's metropolitan classification is indicative of vulnerability to a disproportionate number of COVID-19 deaths. Subdivided by a county's urbanicity classification (*Figure 5*), a more focused granularity than metropolitan classification alone, COVID-19 mortality demonstrates right skew with increasing variability as population density increases (a social vulnerability that is tied directly to a county's metropolitan status). In particular, the right skew is so pronounced that all urbanicity classifications have a considerable number of positive outliers – indicating pronounced variability in large numbers of COVID-19 deaths by county. Generally, larger and more metropolitan counties have an increasing median number of deaths.

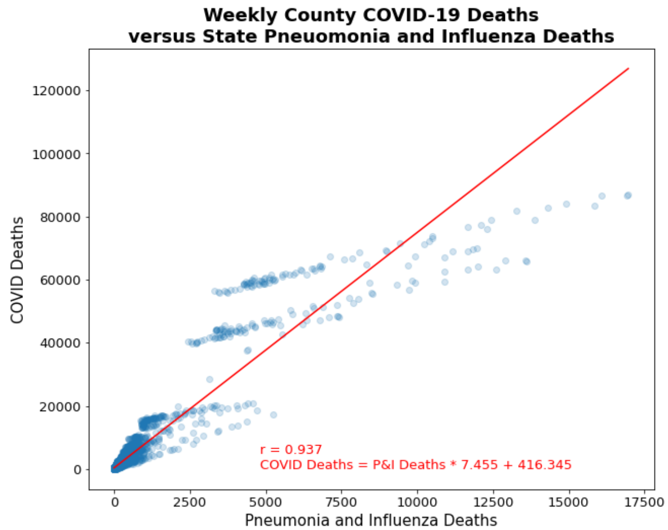
The grouped bar chart (*Figure 6*) indicates that as urbanicity classification indicates larger urban areas with greater population density, the proportion of COVID-19 deaths generally increase. Of course, for this granularity of urbanicity classification, the trend is not as clear. For instance, medium metropolitan areas have a greater spread in outliers than other groups for instance, though its boxplot appears relatively compact in terms of spread and symmetry. As a result, medium metropolitan areas demonstrate a reduced proportion of COVID-19 deaths even compared to some smaller urban areas (e.g., non-core or micropolitan areas).

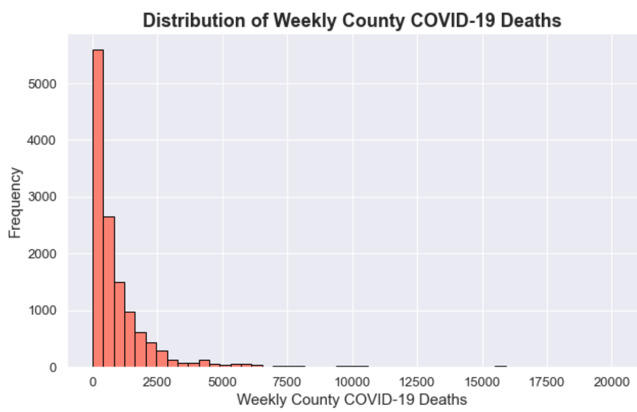
We will compare the mean number of COVID-19 deaths for counties with a large fringe metropolitan urban classification with counties that have a medium metropolitan urban classification. The average number of weekly COVID-19 deaths for large fringe metropolitan urban areas is 36.10 deaths (95% CI: [33.98,38.22]). For medium metropolitan urban areas, the average number of weekly COVID-19 deaths is 34.11 deaths (95% CI: [32.07,36.15]).

The mean number of weekly COVID-19 deaths for large fringe metropolitan urban areas exceeded the medium metropolitan urban area mean by 1.99 deaths (95% CI = [-0.96, 4.93]). These data fail to provide evidence (given a two-sample t-test p-value of $p=0.186$) that the mean number of COVID-19 deaths was greater for counties classified as large fringe metropolitan urban areas compared to medium metropolitan urban areas.

Previously, we found that a county's classification as a larger metropolitan area is indicative of a county's vulnerability to a larger number of COVID-19 deaths. At the increased granularity provided by a county's urbanicity classification however, large fringe metropolitan urban areas do not have a significantly larger mean number of COVID-19 deaths than medium metropolitan urban areas. This indicates that the social and structural disadvantages faced by those in large fringe metropolitan areas are not as indicative of increased COVID-19 deaths. Understanding how these social and structural outcomes in these areas differ from other metropolitan areas will elucidate the structural issues most pertinent to the COVID-19 pandemic.

Estimating of the Disease Burden of COVID-19:





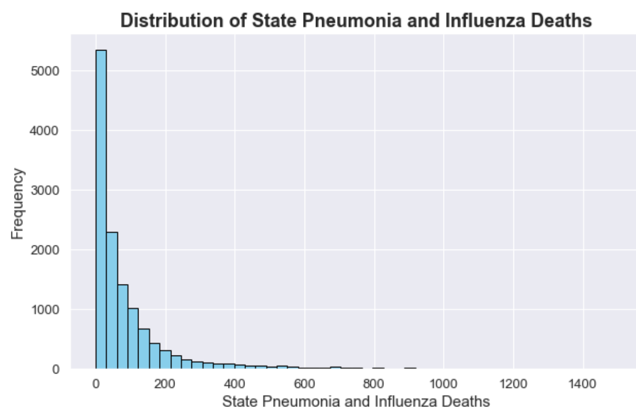


Figure 7: Scatterplot of Weekly County COVID-19 Deaths versus State Pneumonia and Influenza Deaths
Figure 8: Histogram of Weekly County COVID-19 Deaths
Figure 9: Histogram of State Pneumonia and Influenza Deaths

One issue of concern is access to healthcare amidst the COVID-19 pandemic. Annual influenza epidemics occur seasonally and present a substantial disease burden on the American healthcare system due to influenza's effect on the respiratory system and the potential for pneumonia comorbidities among patients. Examining the relationship between COVID-19 deaths and seasonal influenza and pneumonia deaths will help to elucidate the disease burden of COVID-19.

The scatterplot (*Figure 7*) demonstrates a strong positive relationship between COVID-19 deaths and pneumonia and influenza deaths by county. While there are several points that will exert leverage in a regression analysis, this graph can be used to accurately exemplify the relationship between COVID-19 mortality with pneumonia and influenza deaths. Thus, this graph can be used to describe the relationship between COVID-19 and access to health care resources, as COVID-19 mortality is coterminous with deaths associated with preventable diseases.

The histogram of weekly county COVID-19 deaths (*Figure 8*) and the histogram of state pneumonia and influenza deaths (*Figure 9*) indicate the distributions of the response and predictor respectively. While the distributions are non-normal with significant right skew, their symmetry facilitates the validity of the regression analysis.

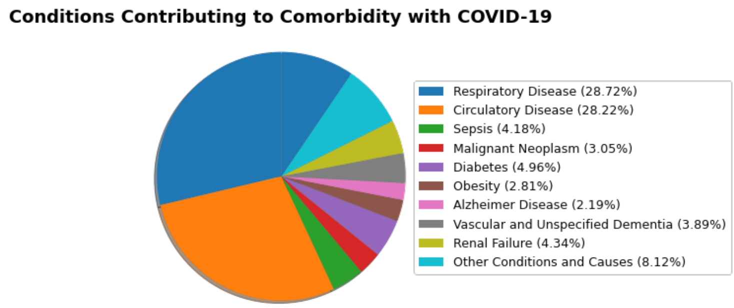
The correlation coefficient between weekly county COVID-19 deaths (the response) and state pneumonia and influenza deaths (the predictor) is 0.937 (given a p-value of $p<0.001$), indicating a strong positive linear relationship between the predictor and response. The data provide strong evidence (given an F-test p-value of $p<0.001$) of a statistically significant regression line. The regression analysis produces a regression equation of

$$Y = 7.455(X) + 416.345$$
$$Y = Weekly\ County\ COVID-19\ Deaths$$
$$X = State\ Pneumonia\ and\ Influenza\ Deaths$$

with a statistically significant slope parameter (given a p-value of $p<0.001$) of 7.455 (95% CI = [7.408, 7.502]).

The regression slope parameter of 7.455 indicates that for every additional death in a given state due to seasonal influenza and pneumonia, we would expect the mean number of weekly county COVID-19 deaths in the state to increase by 7.455. As a measure of disease burden, the regression analysis indicates that (on average), COVID-19 presents 7.455 times the disease burden of seasonal influenza and pneumonia deaths for a given state, which are responsible one of the greatest seasonal disease burdens on the US healthcare system. This is a strong indicator of a pertinent negative externality of the COVID-19 pandemic - its disproportionate impact on access to healthcare resources.

COVID-19 Contributing Comorbidities



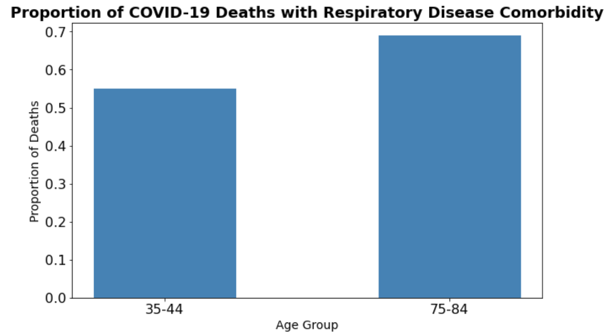


Figure 10: Pie Chart of Medical Conditions Contributing to Comorbidity with COVID-19
Figure 11: Bar Chart of the Proportion of COVID-19 Deaths Associated with a Respiratory Disease Comorbidity

As indicated by the pie chart (*Figure 10*) respiratory and circulatory diseases present the greatest comorbidity risk in conjunction with COVID-19. Many of the conditions listed in the data like sepsis, malignant neoplasm, and renal failure represent coterminal risks with contracting COVID-19.

As indicated in the bar chart (*Figure 11*), the proportion of individuals in the 35-44 year old age group who died from COVID-19 while suffering from a respiratory disease comorbidity was 54.96% (95% CI: [54.64, 55.27]). For individuals in the 75-84 year old age group, the proportion who died from COVID-19 while suffering from a respiratory disease comorbidity was 68.94% (95% CI: [68.82, 69.07]).

It is known that older individuals face a greater risk from COVID-19. As indicated above, respiratory disease presents the greatest comorbidity risk with COVID-19 among all other disease types. It is worth examining whether death from COVID-19 given a respiratory disease comorbidity is independent of age.

We will compare the proportion of comorbidity outcomes for individuals who died from COVID-19 who have a respiratory disease with those who do not between the age groups 35-44 years and 75-85 years. According to CDC standard, these age groups are used to determine baseline death rates for middle-aged and elderly individuals respectively.

Of the 96,411 individuals in the 35-44 year old age group, 54.96% who died from COVID-19 had a respiratory disease comorbidity. However, among the 529,129 individuals in the 75-84 year old age group, 68.94% who died from COVID-19 had a respiratory disease comorbidity. The data provides strong evidence (given a chi-squared test of independence p-value of $p<0.001$) that the probability of dying from COVID-19 given a respiratory disease comorbidity was greater for those in the 75-84 year old age group, as the proportion difference was -13.98% (95% CI = [-14.32, -13.65]). This provides strong evidence that older age groups are more vulnerable to death from COVID-19 given a respiratory disease comorbidity.

Conclusion

Our statistical analyses provided evidence of community-level risk factors (i.e., social risk factors and sociodemographic vulnerabilities) that led to increased COVID-19 mortality risk. We suggest that a county's social vulnerability index (SVI) rating (used by the CDC to evaluate a community's capacity to respond to emergency situations) is particularly indicative of vulnerability to COVID-19 mortality outcomes. Furthermore, while larger and more densely populated metropolitan areas are generally more susceptible to COVID-19 mortality outcomes, when examined at the granularity of urbanicity classification, large fringe metropolitan urban areas do not have a significantly larger mean number of COVID-19 deaths than medium metropolitan urban areas. This indicates that the variation in social and structural disadvantages faced by those in metropolitan areas of various sizes are not consistent, and that further research is needed to identify the structural issues most pertinent to the COVID-19 pandemic.

Furthermore, the statistical relationship between weekly county COVID-19 deaths and associated state seasonal influenza and pneumonia deaths provide an estimate of the disease burden of the COVID-19 pandemic – approximately 7.455 times the disease burden of seasonal influenza and pneumonia deaths for a given state. Because seasonal influenza and pneumonia deaths represent one of the greatest recurrent disease burdens in the US healthcare system, this provides a tangible estimate of the COVID-19 pandemic's impact on healthcare access. Finally, we suggest that there is strong evidence that older age groups are more vulnerable to death from COVID-19 given a respiratory disease comorbidity.

Since this was an observational study, a causal conclusion - that community-level risk factors cause changes in the mean number of COVID-19 deaths - does not follow from the analysis of the difference in COVID-19 mortality among various groups. However, the study does support the notion that community-level sociodemographic vulnerabilities are evident in COVID-19 mortality data, and that communities with greater vulnerabilities (characterized by population density, structural disadvantages like limited access to healthcare, greater incidence of known COVID-19 comorbidities, etc.) should be targeted with local policy and community response to mitigate the COVID-19 pandemic at the county level.

References

1. Centers for Disease Control and Prevention. (2021, November 5). *Covid Data Tracker Weekly Review*. Centers for Disease Control and Prevention. Retrieved November 5, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html.

2. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891-1892. doi:10.1001/jama.2020.6548

3. Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 United States, March 1-30, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):458-464.

4. Karmakar M, Lantz PM, Tipirneni R. Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US. *JAMA Netw Open.* 2021;4(1):e2036462. doi:10.1001/jamanetworkopen.2020.36462

5. Reif J, Heun-Johnson H, Tysinger B, Lakdawalla D. Measuring the COVID-19 Mortality Burden in the United States: A Microsimulation Study. Ann Intern Med. 2021 Sep 21:M21-2239. doi: 10.7326/M21-2239. Epub ahead of print. PMID: 34543588; PMCID: PMC8462514.

6. Ulmer, J. T., Harris, C. T., & Steffensmeier, D. (2012). Racial and Ethnic Disparities in Structural Disadvantage and Crime: White, Black, and Hispanic Comparisons. *Social science quarterly*, *93*(3), 799–819. https://doi.org/10.1111/j.1540-6237.2012.00868.x

7. Centers for Disease Control and Prevention. (2021, April 28). *CDC/ATSDR's social vulnerability index (SVI)*. Centers for Disease Control and Prevention. Retrieved November 6, 2021, from https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.