The Social Vulnerability Index and the COVID-19 Pandemic

Rebecca Klein

Introduction

There is no question that the COVID-19 pandemic has been devastating. There have been more than 1 million deaths due to the virus in just the United States [1], but some people were more at risk than others. There are obvious health reasons, like being immunocompromised, but there are also social vulnerabilities that made it so some communities were disproportionately impacted by the pandemic. The government had a responsibility to help these more vulnerable communities, which could be identified using the Social Vulnerability Index (SVI). The CDC claims that public health officials did use the SVI to "identify and support areas across the country that are at increased risk of becoming exposed to [COVID]." [2] It is important to analyze how well the SVI was used and how much of an impact the actions taken based on it had. This will allow public health officials to apply more strategically what they have learned during the last pandemic to future emergencies.

Background/Related Work

The Social Vulnerability Index (more formally known as the CDC/ATSDR Social Vulnerability Index) is a measurement of the relative social vulnerability on the county or state level. It was created to "help public health officials and emergency response planners identify and map the communities that will most likely need support before, during, and after a hazardous event." [3] Social vulnerability is well documented to be associated with contracting COVID-19 [4][5][6][7]. The pandemic has had an even more disproportional effect on communities of color [2][8]. Since social vulnerability is proven to be an important factor in emergencies, a 2021 paper included a pilot survey that was sent to emergency managers in the US to see how many used, or even knew of, a "community vulnerability assessment tool" like SVI. [9] The small sample size of the survey prevents generalization, but the results imply that many did not use or know about such tools, and those who did were less likely to be located in low-income areas.

Thus, though the Social Vulnerability Index is a powerful tool, it is underutilized in general, especially where it may be needed most. This leads to my research question: how well was SVI used during the COVID-19 pandemic? More specifically, I had two hypotheses. The first is that there are a lower vaccination rates and higher COVID cases in vulnerable compared to non-vulnerable counties in Indiana according to the Social Vulnerability Index. My second is that there were more mask mandates in Indiana in vulnerable compared to non-vulnerable counties based on SVI.

Methodology

For my analysis, I focused on the state of Indiana. This was chosen based on the county I was assigned in Part 1 – Marion, Indiana. I used 4 datasets for this project. The first was the timeline of mask mandates by county published by the CDC. The second, a cumulative count of COVID cases nationally by county sourced through Kaggle from Johns Hopkins University. The third is a timeseries of vaccination data by county in Indiana also through the CDC. And the fourth is of course the 2020 Social Vulnerability Index, which has data by state and county.

For the number of cases, I normalized by dividing by the overall county population so I could directly compare counties of different sizes. I also normalized by the 12+ and 18+ populations due to the age requirements for vaccination earlier in the pandemic. These population counts were available through the CDC vaccination dataset. Beyond just looking at the cumulative count of COVID cases, I also created a variable for the rate of cases using numpy's gradient.

The vaccination dataset contains several possible variables to use including booster shots, but I decided to focus on the percentage of the population that has completed the initial vaccination series (2 shots except for the Johnson & Johnson vaccine that only needed 1). I chose this because I was interested in what was done earlier in the pandemic. Similar to the normalized cases count, I including vaccination rates for the overall population, 12+ and 18+, as well as 65+ populations.

The social vulnerability index calculates an overall vulnerability and by 4 themes: Socioeconomic Status, Household Characteristics, Racial & Ethnic Minority Status, and Housing Type & Transportation. There are a total of 15 tracts associated with the 4 themes and they come from the U.S. Census (see Figure 1). I ran my analysis using the 4 themes above and the overall vulnerability, thus studying 5 variables. The reason to do so was to gain additional granularity as some themes may have been more impacted by the pandemic than others. For each of these themes I created a new variable indicating if a county is in the top 10% of vulnerability for that category. I chose to define vulnerable counties as in the top 10% because I wanted it to be large enough to have a good sample size, and small enough to hopefully only include significantly vulnerable counties.

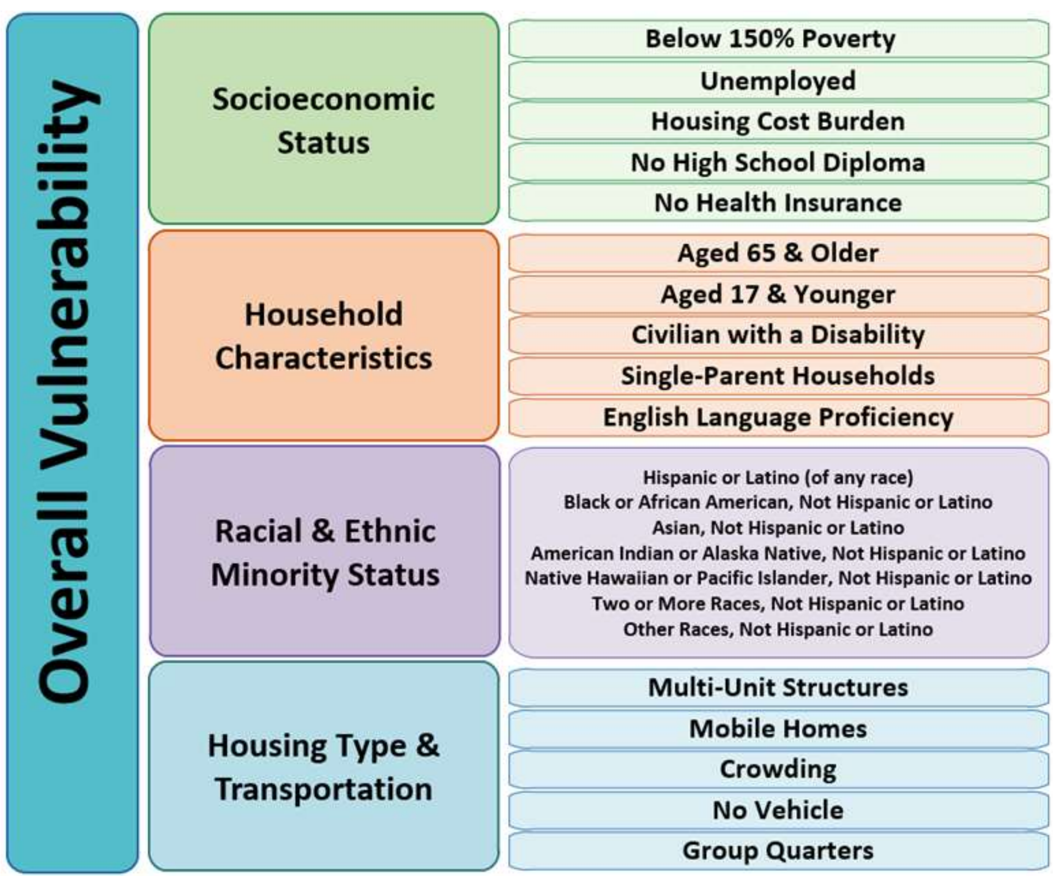

Figure 1: Social Vulnerability Index themes and tracts

In order to compare the cases and vaccination between vulnerable and non- or less-vulnerable counties, I ran a set of t-tests. I chose 6 dates that are 3 months apart (1/1/2021, 4/1/2021, 7/1/2021, 10/1/2021, 1/1/2022, 4/1/2022) for all of my cumulative variables like vaccination. Because these variables are cumulative, I only needed a snapshot in time to get a full picture around that date. For cases rate, I took the mean over the whole 3-month period (1/21-3/21, 4/21-6/21, 7/21-9/21, 10/21-12/21, 1/22-3/22) since the rate varies from day to day. It was important to look at a range of points in time in order to see if there were differences earlier and later in the pandemic.

Findings

Counties with higher social vulnerability indices had higher average rates COVID cases, as were counties with categories of lower socioeconomic status and racial and ethnic minority status. Racial and ethnic minorities by county in general had higher average vaccination rates.

Table 1 shows dates (or month ranges) that had a statistically significant difference in average COVID cases/rate or vaccination rates as determined by p-value <0.05. All of the significant results for cases rate were even more significant with a p-value of less than 0.01. The dates in red have a mean that is worse for vulnerable counties. The dates in **bold and green** have a mean that is better for vulnerable counties. As a note, the cases count normalized by both the 12+ and 18+ populations had the same results and very similar p-values. The same is true for the percentage of 12+ and 18+ population with a complete vaccination series.

	Overall	**Socioeconomic Status**	**Household Characteristics**	**Racial & Ethnic Minority Status**	**Housing Type & Transportation**
Cases Rate (p-values less than 0.01)	1/21-3/21, 4/21-6/21, 7/21-9/21, 10/21-12/21, 1/22-3/22	1/21-3/21, 4/21-6/21, 7/21-9/21, 10/21-12/21, 1/22-3/22	**1/21-3/21, 4/21-6/21, 7/21-9/21, 10/21-12/21, 1/22-3/22**	1/21-3/21, 4/21-6/21, 7/21-9/21, 10/21-12/21, 1/22-3/22	NS
Cases Count Normalized by Total Population	NS	NS	NS	7/1/2021	NS
Cases Count Normalized by 12+ and 18+ Populations	NS	NS	NS	1/1/2021, 4/1/2021, 7/1/2021	NS
Percentage of Population with Complete Vaccination Series	NS	NS	NS	4/1/2021, **1/1/2022, 4/1/2022**	NS
Percentage of 12+ and 18+ Population with Complete Vaccination Series	NS	NS	NS	4/1/2021, **7/1/2021, 10/1/2021, 1/1/2022, 4/1/2022**	NS
Percentage of 65+ Population with Complete Vaccination Series	NS	7/1/21, 10/1/21	NS	NS	NS

Table 1: Dates with a significant p-value for the difference in means between vulnerable and non-vulnerable counties. NS is non-significant for all dates studied.

For all 5 tests run for cases rate (3-month ranges) those with greater vulnerability by household characteristic had a lower average COVID case rate. In addition, vaccination was significantly better for counties with high racial and ethnic minority status for 2-4 out of 6 dates studied. Most differences, however, were worse for vulnerable counties. While the rate of cases was better by household characteristics, it was worse overall and by racial and ethnic minority status for all months. Vaccination was also worse in April of 2021 for racial and ethnic minorities, implying that any successful efforts made in this area were done later in the pandemic.

Though I also explored differences in mask mandates, all counties in Indiana had the same mask mandates for the same time periods.

Discussion/Implications

This analysis shows a clear pattern of the greatest impact to vulnerable populations in the themes of socioeconomic status and racial and ethnic minority status. The case rate variable is highly significant, and all month ranges analyzed had a higher average COVID rate for those vulnerable populations. This confirms the literature pointing to communities of color being at higher risk.

The best ways to reduce the rate of COVID cases is through masking and vaccination. At least in Indiana, mask mandates were similar across the state and thus no conclusions can be made in the area of mask mandates or policies as they relate to vulnerable counties. It is possible that stronger mask mandates could have impacted COVID rates across the state, but this analysis cannot answer that question. In terms of vaccination, the data illustrates an initial lower average vaccination rate in vulnerable populations by race and ethnicity that then flips to higher vaccination rates in those communities. It is possible that the SVI was used to lead campaigns that positively impacted vaccination rates in those communities. This is a hypothesis based on the improvements in vaccination seen over time in those vulnerable groups.

There were a number of areas where results were non-significant. Any non-significant results may mean that vulnerable communities were no worse off in those areas – possibly due to intervention – though more study would be needed to better understand why. Perhaps there were also interventions in terms of vaccination for populations with low socioeconomic status, which brought up the mean vaccination rate.

Further research should include a national sample evaluating states overall and possibly counties within individual states, as well as using all 15 tracts, 4 themes, and overall SVI. This would hopefully also include data about what interventions were attempted and in what locations. In addition, research could focus more narrowly by using thick data such as interviews with people where interventions were attempted. This could answer the question of whether people in vulnerable populations felt like the intervention was effective, as well as give advice on what could be done in future hazardous events. Interventions for situations like this will be effective based on local needs, and with so many confounding factors only the people they are impacting will have the full answers thus focusing on the human aspect of study design.

Limitations

This study used several retrospective data sources which are limited in number of variables available and potential by veracity of the data. A prospective randomized study including an intervention and non-intervention group of similar vulnerabilities would be a stronger methodology, but not a cost-effective study approach. Without knowing where interventions were attempted only inferences about the reason for results can be made. In addition, the number of t-tests run could result in significant p-values by chance. I did not apply a Bonferroni correction, thus potentially impacting some results.

Conclusion

The coronavirus disproportionately affected vulnerable populations, especially communities of color. Public officials have a responsibility to help those communities, but their effectiveness, at least in Indiana, is of concern. While there may have been some effective interventions in terms of vaccination for racial and ethnic minorities, the overall case rate was still significantly worse. The Social Vulnerability Index is a great tool to help focus efforts on vulnerable communities, but just knowing where these communities are is not enough. The SVI can also be a great tool to for data scientists to understand and evaluate interventions made in order to positively impact these vulnerable populations. There has been a large push recently to make data driven decisions, but just because its data driven doesn't mean it can't also be human driven.

References

[1] https://covid.cdc.gov/covid-data-tracker

[2] https://www.atsdr.cdc.gov/placeandhealth/project_snapshots/svitool_covid.html

[3] https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2020.html

[4] https://www.sciencedirect.com/science/article/pii/S0749379720302592

[5] https://www.medrxiv.org/content/10.1101/2020.04.10.20060962v2

[6] https://www.medrxiv.org/content/10.1101/2020.08.03.20166983v1

[7] https://link.springer.com/article/10.1007/s11606-020-05882-3

[8] https://onlinelibrary.wiley.com/doi/10.1111/puar.13264

[9] https://www.sciencedirect.com/science/article/abs/pii/S2212420921004258

Data Sources

Mask mandates: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

Covid cases: https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university

Vaccination: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh

SVI: https://www.atsdr.cdc.gov/placeandhealth/svi/interactive_map.html