State and County-Level Factors Associated with the Effectiveness of Stay-At-Home Orders Issued in the United States in Response to COVID-19

Forrest DL Barker,¹ Ali Mirzadadeh,² Neelam Sekhri Feachem³*

¹Institute of Global Health Sciences, University of California, San Francisco, San Francisco, California, United States of America

²Department of Epidemiology and Biostatistics, Institute for Global Health Sciences, University of California San Francisco, California, United States of America

³Institute of Global Health Sciences, University of California, San Francisco, San Francisco, California, United States of America

*Corresponding author

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Title

State and County-Level Factors Associated with the Effectiveness of Stay-At-Home Orders Issued in the United States in Response to COVID-19

Background

To slow the spread of COVID-19 and protect medical facilities from overflowing, Stay-At-Home Orders (SAHOs) were issued in the United States during the spring of 2020. These orders had variable levels of effectiveness and profound consequences that continue to manifest long after their termination. This study aimed to assess if state and county-level population characteristics could explain variability in SAHO effectiveness as measured by the effective reproductive number ($R_t$).

Methods

We calculated the $R_t$ for the 40 states which enacted SAHOs, and also for a sample of 289 counties that issued SAHOs in 2020, using EpiEstim R Package based on the states’ and counties’ daily case data. We determined SAHOs to be effective if, three weeks after their implementation, $R_t$ was equal to or less than one. Wilcoxon rank sum tests and logistic regression were used to determine if population characteristics (age, income, level of education, political orientation, percent of non-English speaking people, racial and ethnic compositions), and percentage of frontline workers, percent of eligible people vaccinated by July 2021, level of viral transmission, and other Non-Pharmaceutical-Interventions (NPI) enacted before the SAHO, were associated with effectiveness of SAHOs.
Results

SAHOs were effective in 20 (50%) states. No significant differences were found in the characteristics studied between states with effective and ineffective SAHOs. SAHOs were effective in 54% of counties. Counties with effective SAHOs had fewer days of NPIs before the SAHOs in comparison to counties with ineffective SAHOs (Median 24 vs. 34, p-value 0.005). All other characteristics considered showed non-significant differences. In multivariate analysis, days of NPIs before the SAHOs remained the only significant factor for effective SAHOs in studied counties.

Conclusion

Our analysis suggests that SAHO effectiveness may be influenced by the implementation of prior public health interventions but is not likely to be related to the other characteristics studied. These findings should be considered when assessing when and how to implement SAHOs in future epidemics to limit the spread of an infectious respiratory disease.
INTRODUCTION

In early December 2019, a cluster of idiopathic pneumonia cases was reported in Wuhan, China.¹ These are thought to be the first identified cases of a disease called COVID-19 caused by a novel coronavirus named Severe Acute Respiratory Coronavirus Syndrome-2 (SARS-CoV-2) or colloquially the COVID-19 virus.² By the end of January 2020, COVID-19 was classified as a public health emergency of international concern and cases had been detected in 21 countries spread across Asia, Europe, North America, and Oceania. A few weeks later, the World Health Organization (WHO) declared COVID-19 a global pandemic.³ As of April 1, 2023, the WHO had globally recorded 761,402,282 confirmed cases and 6,887,000 confirmed deaths, making COVID-19 one of the most severe pandemics in history.⁴ In the United States (U.S.) alone the COVID-19 virus has infected over 100 million individuals, more than one million of whom have died.⁵

The number of active COVID-19 cases has periodically spiked, causing hospitals to reach and surpass their capacity. This in turn has forced care and resources to be rationed.⁶–¹⁰ For example, medical facilities in the U.S. have, at times, needed to suspend non-emergent services to prioritize resources for overflowing emergency rooms and intensive care units (ICUs). At the same time medical supply chains have been disrupted by excessive demand and supply shortages.¹¹–¹³ This was particularly true at the start of the pandemic when treatments for COVID-19 were unrefined and vaccines were unavailable, leaving populations particularly vulnerable. The lack of pharmaceutical protection forced nations to implement Non-Pharmaceutical Interventions (NPI’s) to reduce transmission of COVID-19 and avert the collapse of their medical systems. In the U.S. this effort was termed “flattening the curve”.¹⁴ Without such measures, hospitals would overflow, and patients would be left untreated, drastically increasing all-cause mortality.
Throughout the pandemic many NPI’s have been utilized to reduce transmission with varying levels of success. Common NPIs have include mask wearing, quarantine (confinement of individuals who may be infected), isolation (confinement of individuals who are infected), restricting travel, prohibiting gatherings, requiring individuals to maintain distance between each other, closing schools, closing non-essential businesses, and full lockdowns or Stay-at-Home Orders (SAHOs). Implementation of these NPIs has come at the cost of reducing personal liberties and economic activity, increasing inequality, and a myriad of unintended public health side effects such as increased consumption of drugs and alcohol, higher rates of mental health disorders, and rising rates of child abuse and intimate partner violence.

Prior to COVID-19, the efficacy and consequences of NPIs were poorly researched and largely based on theoretical knowledge. However, throughout the pandemic, countries have implemented and relaxed various NPI’s in response to rising and falling infection rates. This has created a natural experiment which allows for real world assessment of the efficacy of interventions. As a result, an enormous amount of research on the subject has been conducted in the last few years. Assessments have utilized numerous methodologies including mathematical models, surveys, opinions, direct observations and measurements, and indirect observations via proxy data. In addition to the multitude of methodologies for assessing NPI effectiveness, there is substantial variability in the metrics used for quantification, making it challenging to compare analyses. In general, it has been found that effectiveness of NPIs varies between locations and that multiple interventions have stronger effects than any single intervention.

SAHOs, which inherently include other NPI’s, have been found to be very effective at reducing transmission of COVID-19. However, it is also estimated that SAHOs in the US resulted in up to $64 billion a week in lost revenue and brought a grim set of unintended consequences.
Often described as a blunt instrument or last resort, the far-reaching impacts of SAHOs warrant an in-depth investigation into the state and county-level factors that influence their efficacy. A number of investigations have shown that SAHO’s are effective at reducing population mobility and that SAHO’s are associated with a decrease in transmission and mortality at a state and county level. However, it has not been determined if certain population characteristics are associated with more or less effective SAHO’s. This paper investigates whether various state or county level factors could have influenced the efficacy of SAHOs issued in the U.S. during the spring of 2020. The threat of a future pandemic requiring lockdowns has drawn the attention of the world. Information presented here may help inform how future SAHOs can be coordinated in a manner that maximizes their impact while minimizing their duration and harm.

**MATERIALS AND METHODS**

We performed an independent cross-sectional analysis looking for differences in communities with effective and ineffective SAHOs during the spring of 2020. Effective SAHOs were defined by communities achieving a local $R_t$ (i.e., basic reproductive number at a point in time to measure transmissibility of the virus), of $\leq 1$ after three weeks of a SAHO. Ineffective SAHOs were defined by having an $R_t > 1$ three weeks after order implementation. We initially proposed quantifying SAHO effectiveness as a continuous variable defined by the change in $R_t$ from the first week of SAHO implementation to the third week of the SAHO, however, we found this definition favored communities with higher initial values of $R_t$ and did not accurately describe whether a SAHO accomplished the goal of reducing the number of cases in a community.

Factors we assessed for possibly influencing SAHO effectiveness included:

- Percentage of essential workers
- $R_t$ when order was enacted.
- Time between first case and SAHO enactment
We originally proposed to also study whether community mobility influenced SAHO effectiveness. However, this analysis was conducted and published by Wellenius et al. while we were conducting our investigation. As the influence of mobility on SAHOs was no longer a novel investigation we removed it from our list of variables, to prioritize others.

All 40 states that issued SAHOs were included in this sample. For county level analysis a sample of 429 counties was obtained from the American Community Survey (ACS). This sample is designed to be representative of the population of the US. The sample was filtered to remove counties without SAHOs during the study period, where there were too few cases of COVID-19 for accurate calculations of $R_t$, and where the first case was not reported until after the start of the SAHO. This filtering yielded a final sample of 289 counties spread across the U.S.

The reproductive number $R_t$ was calculated based on confirmed cases of COVID-19 reported in the Johns Hopkins University Coronavirus Resource Center Database. This dataset included cumulative daily anonymized COVID-19 cases aggregated by county. New daily cases were defined as the change in cumulative cases from one day to the next. Dates with negative total
daily cases were adjusted to zero cases for that day. For state level analysis, the daily case data for all counties in each state were summed. An attempt was made to remove cases known to have occurred in congregate living settings such as prisons and nursing homes. However, available data on these cases were insufficient to make a robust adjustment. R version 4.0.2. and R Studio were used for organization and manipulation of case data. EpiEstim R Package was then used to calculate $R_t$ for states and counties based on daily case data. A serial interval (SI) of 5.2 days with a standard deviation of 0.7 days, as reported in the meta-analysis by Alene et al, was used in calculating $R_t$. For each state and county, $R_t$ was measured at two time points, the week following the start of the SAHO and the week 21 to 27 days following the start of the SAHO. These time points are estimated to represent viral reproductive number during the week preceding the SAHO implementation and two weeks post implementation, accounting for the 6.5 day mean incubation period also reported by Alene et al. $R_t$ was calculated as an average over a seven-day time period to account for lower case reporting on weekends compared to weekdays. Locations where fewer than ten cases were reported during either of the seven-day periods used for calculating $R_t$ were excluded as so few cases did not allow for accurate calculation of $R_t$. Additionally, calculated $R_t$ values that were greater than 2 standard deviations above the mean were excluded as outliers. These outliers were thought to represent either an outbreak in a congregate setting, or a batch of previously unreported cases and thus not representative of community transmission.

Dates of public health orders including the start of SAHOs were retrieved from the CUSP database maintained by Boston University. For each location, we identified NPI's that preceded SAHO implementation. The number of days each NPI was in place prior to the SAHO were then summed to give a cumulative number of days with prior NPI orders. For example, if a location had one order that began ten days before the SAHO and another began five days before the SAHO, that location would be classified as having 15 days of prior NPI orders. The
orders included in this measurement were: declaring a state of emergency, school closures, business closures, and prevention of large gatherings.

The IPUMS database was used to access and organize a five-year sample (2015-2019) of the census bureau’s ACS community survey, which is designed to be 5% of the U.S. population. This sample contained over 15.9 million unique responses including data on individuals’ sex, income, age, race, ethnicity, occupational industry, ability to speak English, educational attainment, county and state of residence, and the weight each response should carry to represent the state population. Responses were filtered to exclude entries where county of residence was not reported. Data were then aggregated at both the state and county level and further filtered to only include locations with SAHOs and a sufficient number of reported cases. The final sample contained 5,787,898 responses across 289 counties, with responses from a mean of 2.8% of the population in each county (1.5% minimum and 5% maximum).

For each county and state/territory, the mean value was calculated for age and years of education. Years of education were coded from 0-11 where 0 represents no schooling, 1 represents schooling up to fourth grade, 2 represents schooling up to eighth grade, 3-10 each represent one additional grade or year of schooling above eighth grade, 11 represents five or more years of college. The median value was calculated for income as these data had outliers with high earnings. Income data were also altered by the ACS survey which reports incomes over $999,999 as $999,999. This is done to protect anonymity of respondents. The percentage of respondents who reported speaking little to no English was calculated and used to estimate percent of non-English speaking individuals. Similarly, the percentage of respondents who identified as Hispanic, and who identified as African American were calculated.

In the ACS survey, industry is logged as a 5-digit code based on the North American Industrial Classification System (NAICS). We identified 22 NAICS codes that contained most front-line
workers including: essential consumer business such as grocery stores, transportation workers, first responders, and medical practitioners. The codes included were 4470, 4971, 4972, 5070, 5391, 6080, 6180, 6170, 6390, 7690, 7970, 8090, 8170, 8180, 8191, 8192, 8270, 8290, 8370, 8380, and 8470. The percentage of respondents who reported working in one of these industries was then calculated as an estimate for the community’s percent of front-line workers.

We identified front line workers based on industry rather than occupation because industry has a larger influence on whether someone is a front-line worker. For example, a manager at a grocery store is a front-line worker while a manager of a software company is not. In this example the occupation of manager is not indicative of whether a job is front-line while industry, in this case, grocery store or software development, is.

The ACS survey lists response weights which indicate how many people in the population each response represents. These weights are based on state population data. Weighting was utilized for state level calculations but not for county level calculations. The unweighted calculations at the county level allowed us to capture intra-state variability better than the weighted sample.

Data on vaccination rates in each location was collected from the U.S. Centers for Disease Control and Prevention’s COVID-19 Data Tracker.5 Data on the percentage of people who voted for Trump in 2020 was obtained from MIT’s Election Lab Database.34

Utilizing R Version 4.0.2 and R Studio, we ran Wilcoxon rank sum tests with all the above-mentioned variables to compare their distribution in populations with successful or unsuccessful SAHOs. We utilized an alpha value of 0.05 as our rejection criteria. The Wilcoxon test was utilized for hypothesis testing because data on the variables did not have a normal distribution or contained outliers which could not be excluded as they represented valuable data points on unique communities. We then created logistic regression models controlling for each of the 12 factors assessed to determine if any factors were functioning as confounders or effect modifiers.
of other factors. We chose to classify variables as having a significant effect if their presence in
the logistic regression model reduced the p-value of other variables by more than 10%.

RESULTS

A summary of the sample of ACS data used for our analysis is shown in Table 1. Variables that
were not from the ACS survey, such as percent of eligible population vaccinated, were derived
from the whole population rather than a sample. Thus, the demographics of these population-

Table 1: ACS survey sample responses (n=5,787,898)

<table>
<thead>
<tr>
<th>Sex – no. (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2,832,597 (48.94)</td>
</tr>
<tr>
<td>Female</td>
<td>2,955,301 (51.06)</td>
</tr>
<tr>
<td>Race and ethnicity – no. (%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>4,234,838 (73.17)</td>
</tr>
<tr>
<td>Black</td>
<td>612,182 (10.58)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>33,890 (0.59)</td>
</tr>
<tr>
<td>Chinese</td>
<td>118,888 (2.05)</td>
</tr>
<tr>
<td>Japanese</td>
<td>21,174 (0.37)</td>
</tr>
<tr>
<td>Other or not reported</td>
<td>766,926 (13.25)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>983,002 (16.98)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>4,804,896 (83.02)</td>
</tr>
<tr>
<td>Age – median (IQR)</td>
<td>44 (27)</td>
</tr>
<tr>
<td>Educational attainment code – median (IQR)</td>
<td>7 (4)</td>
</tr>
<tr>
<td>Front-line workers – no (%)</td>
<td>1,120,470 (19.36)</td>
</tr>
<tr>
<td>Income – median (IQR)</td>
<td>36,711 (54,728)</td>
</tr>
<tr>
<td>English Speaking – no. (%)</td>
<td></td>
</tr>
<tr>
<td>Speaks no English/does not speak well</td>
<td>237,856 (4.11)</td>
</tr>
<tr>
<td>Speaks English well</td>
<td>555,0042 (95.89)</td>
</tr>
</tbody>
</table>

Figure 1 shows the locations of the counties which made up our sample for analysis at the county level.

Figure 1: Counties included in analysis n=289

For state level analysis, we first grouped states into effective or ineffective SAHOs. States that
achieved an R_t value of one or lower after three weeks of SAHO were said to have effective
orders while states that did not achieve an $R_t$ value of less than one during the same time period
were classified as having ineffective SAHOs. Based on these criteria it was found that, of the 40
states with SAHOs, 50% were successful. Table 2 shows the variation of characteristics we
assessed for states with effective and ineffective SAHOs.

Table 2: Difference in state level characteristics by SAHO effectiveness

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Effective (n=20)</th>
<th>Non-effective (n=20)</th>
<th>Wilcoxon rank sum test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time between first case and SAHO start (days)</td>
<td>18.5 9.5</td>
<td>23 7.5</td>
<td>254.5 0.144</td>
</tr>
<tr>
<td>$R_t$ at Start of SAHO</td>
<td>1.96 0.99</td>
<td>1.96 1.26</td>
<td>179 0.583</td>
</tr>
<tr>
<td>Cumulative days of NPI orders (days)</td>
<td>26 14.8</td>
<td>30 14.5</td>
<td>226.5 0.311</td>
</tr>
<tr>
<td>Front-line workers (%)</td>
<td>20.0 2.3</td>
<td>20.8 2.1</td>
<td>209 0.138</td>
</tr>
<tr>
<td>Educational attainment</td>
<td>7.70 0.31</td>
<td>7.74 0.43</td>
<td>164 0.950</td>
</tr>
<tr>
<td>Non-English speaking (%)</td>
<td>1.62 2.24</td>
<td>1.94 1.88</td>
<td>163 0.975</td>
</tr>
<tr>
<td>Income (2019 USD)</td>
<td>37,859 6,366</td>
<td>38,300 5,088</td>
<td>149 0.704</td>
</tr>
<tr>
<td>Age (years)</td>
<td>44 1</td>
<td>44 1.5</td>
<td>184.5 0.461</td>
</tr>
<tr>
<td>Eligible population vaccinated (%)</td>
<td>66.6 15.7</td>
<td>62.9 18.5</td>
<td>192.5 0.850</td>
</tr>
<tr>
<td>Voted for Trump in 2020 presidential election (%)</td>
<td>45.4 14.7</td>
<td>48.8 12.8</td>
<td>216 0.478</td>
</tr>
<tr>
<td>Identify as Hispanic (%)</td>
<td>8.35 2.70</td>
<td>7.24 7.34</td>
<td>143 0.573</td>
</tr>
<tr>
<td>Identify as African American (%)</td>
<td>11.9 12.7</td>
<td>10.4 10.9</td>
<td>164 0.950</td>
</tr>
</tbody>
</table>

As with states, our sample of counties was split into two groups, those with effective SAHOs
and those with ineffective SAHOs. Effectiveness was determined based on the same criteria as
states. Of the 289 counties in our sample, 158 (54%) were determined to be successful. Table 3
shows the variation in characteristics of counties with successful SAHOs and those with
unsuccessful SAHOs.

Table 3: Difference in county level characteristics by SAHO effectiveness

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Effective (n=185)</th>
<th>Non-effective (n=127)</th>
<th>Wilcoxon rank sum test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time between first case and SAHO start (days)</td>
<td>14 8.8</td>
<td>15 9.5</td>
<td>10678 0.642</td>
</tr>
</tbody>
</table>
The characteristics assessed were factors or proxies for factors we hypothesized could have influenced how well populations complied with SAHOs, and how effective SAHOs were at reducing transmission of COVID-19. Wilcoxon rank sum tests were utilized to test the null hypothesis that there was not a difference in the distribution of characteristics between communities with successful versus unsuccessful SAHOs. We utilized an alpha value of 0.05 as the rejection criteria.

Figures 2 through 13 show the distribution of data points at the state and county levels. Each graph shows how a characteristic varied among locations with successful SAHOs (green) and unsuccessful SAHOs (red). For clarity, data points on these graphs have been binned into 30 distinct groups evenly distributed along the Y axis.

- **Days between first case and SAHO start**

The first factor we considered was the number of days between the first case and the start of a SAHO. We suspected that as this time period grew, so would the number of undetected cases of COVID-19. Greater numbers of individuals with undetected cases could lead to increased household transmission once SAHOs were put in place, causing them to be less effective at reducing $R_t$. However, we did not find a difference in the distribution of this time period between...
communities (states or counties) with effective versus ineffective SAHOS (p=0.144 states; p=
0.642 counties).

**Figure 2:** Days between first case and SAHO Start. Non-successful SAHOS are shown on the left in red, successful SAHOS are shown on the right in green.

- **Rₜ value at start of SAHO**

We then considered the viral reproductive value at the start of SAHOS. Other studies have reported that NPIs, including SAHOS, can reduce Rₜ by finite amounts which would make it easier for SAHOS to be effective in communities with lower initial Rₜ values. Comparison of Rₜ distribution between groups yielded non-significant results (p=0.583 states; p=0.227 counties). Thus, we could not reject the null hypothesis that the distribution of initial viral reproduction values was the same for communities with effective versus ineffective SAHOS.

**Figure 3:** Rₜ calculated for the first week of SAHO. Non-successful SAHOS are shown on the left in red, successful SAHOS are shown on the right in green.

- **Cumulative days of NPI orders prior to SAHO start**

After looking at initial Rₜ values, we assessed if the cumulative number of days of alternate NPI orders, before SAHOS were implemented could have influenced effectiveness. Other studies have reported that multiple NPI orders were more effective than a single NPI order. Based on this we expected that a greater number of days of other NPI orders before a SAHO began would result in a SAHO being more likely to drop the Rₜ value to below 1. When we tested this hypothesis at the state level, we did not find a difference in the distribution of days of prior NPI orders between effective and ineffective SAHOS (p=0.311). However, we found a significant difference between these groups of SAHOS at the county level (p=0.005). Surprisingly this difference was not in our hypothesized direction. Rather we found that the median number of
days of prior NPI orders in counties with effective SAHOs was lower than the median in counties with ineffective SAHOs.

**Figure 4:** The cumulative number of days other NPI orders were in place before the start of a SAHO. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

- **Percent of front-line workers**

  The next characteristic we considered was the percent of front-line workers. The jobs of these workers were not affected by SAHOs, and they continued to have high levels of interaction and contact with other community members. Thus, we expected that the effectiveness of SAHOs would be reduced in communities with a greater proportion of front-line workers. Our analysis, however, did not find a difference in SAHO effectiveness for either states or counties based on the percent of front-line workers (p=0.138 states; p=0.905 counties).

**Figure 5:** The percent of ACS survey respondents who reported working in industries identified as front-line. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

- **Educational attainment**

  Educational attainment has commonly been associated with behavioral differences. For example, studies have found correlations between college education and political affiliations. We hypothesized that the level of education in a community could influence how that community responds to public health orders. We tested the null hypothesis that there was no difference in distribution of educational attainment between communities with successful and unsuccessful SAHOs and found that there was not a significant difference (p=0.950 states; p=0.252 counties).

**Figure 6:** Educational scores relating to the highest level of education attained by respondents of ACS survey. On this scale 0 represents no schooling and 11 represents 5+ years of college. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
• Non-English-speaking population.

Public health messaging in the US was largely carried out in English. This messaging spread awareness of SAHOs and gave explanations of their purpose, goals, and guidelines. It is possible that non-English speaking individuals were not exposed to this messaging, which could have altered their behavior while SAHOs were in effect. We assessed if the proportion of individuals in states or counties who spoke little to no English differed between effective and ineffective SAHOs. Our tests did not find a significant difference in English speaking abilities based on the success of SAHOs (p=0.975 states; p=0.515 counties).

Figure 7: The percent of ACS survey participants who reported they did not speak English or did not speak English well. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

• Income

The financial consequences of SAHOs issued in spring of 2020 were enormous. Compliance with SAHOs for many individuals meant loss of income and financial security while the businesses where they worked were closed or operating at reduced capacity. Additionally, many individuals and families faced large costs associated with working and schooling from home. To assess whether financial security could have influenced compliance with SAHOs thereby altering their effectiveness, we looked for a difference in the distribution of median incomes in states and counties with successful versus unsuccessful SAHOs. These tests yielded insignificant results, suggesting that the distribution of income is similar between the two groups at both the state and county level (p=0.704 states; p=0.136 counties).

Figure 8: Median income in thousands of US Dollars reported in ACS survey responses (adjusted for inflation to match 2019 USD value). Unsuccessful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
• **Age**

Age has a profound impact on individuals’ behavior including mobility and social practices. Studies have found that SAHO effectiveness is negatively correlated with mobility and number of social interactions. Given this finding we expected that the median age of a population could influence whether a SAHO could reduce the viral reproductive value to less than one after three weeks. Contrary to our expectation, analysis did not show any difference in the median age of populations based on SAHO successfulness ($p=0.461$ states; $p=0.104$ counties).

**Figure 9:** Median age of ACS survey respondents. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

• **Proportion of eligible population vaccinated.**

The COVID-19 pandemic has been plagued with mistrust of government officials and the medical community. This mistrust has led to a lack of compliance with public health orders and recommendations, such as SAHOs or receiving vaccinations. Because no quantification of mistrust in government and medical communities exists at the locational specificity needed for this analysis, we utilized the percentage of individuals who had received at least one dose of a COVID vaccine as a proxy for this mistrust. We expected that communities with unsuccessful SAHOs would have lower vaccination rates. To assess this, we tested the null hypothesis that the distribution of vaccination rates as of July 15, 2021, was the same in communities with effective and ineffective SAHOs. These tests yielded non-significant results, and the null hypothesis was not rejected ($p=0.850$ states; $p=0.080$ counties). This result can be extrapolated to suggest that mistrust in government and public health systems likely did not influence SAHO effectiveness.

**Figure 10:** The percent of eligible individuals vaccinated as of July 15, 2021. Unsuccessful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
• Proportion of population who voted for Trump in 2020 presidential elections.

Leading up to and during SAHOs in the spring of 2020 there were stark differences in how politicians discussed the threat of COVID-19, and the types of NPI’s they supported. This difference is commonly credited with affecting how populations responded to public health orders. Numerous media articles and scientific studies suggest that individuals who identified as Republican or supported Trump were less likely to adhere to public health orders such as mask mandates and SAHOs. We utilized the voting records from the 2020 presidential election to estimate the proportions of people who supported Trump or the Republican party. As with other variables, we then looked to see if the distribution of the percent of people who voted for Trump was different in communities based on SAHO effectiveness. Surprisingly, a significant difference was not found suggesting that support for Trump did not have a significant influence on SAHO effectiveness (p=0.478 states; p=0.895 counties).

Figure 11: Percent of population that voted for Trump in the 2020 general election. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

• Proportion of population who identify as Hispanic and African American.

The pandemic has highlighted and deepened the stark inequities that exist along racial and ethnic lines. Studies have shown that COVID-19 has disproportionately affected Black and Hispanic individuals. For example, both Black and Hispanic communities have seen a larger drop in life expectancy during the pandemic than white individuals. The larger impact of COVID-19 on these communities should not be attributed to racial or ethnic differences but rather the multitude of structural and systemic variables that disadvantage these individuals relative to white individuals. These structural and systemic variables, however, are too numerous and difficult to quantify for this analysis. For this reason, we utilized the percent of a
population identifying as African American and the percent of a population identifying as Hispanic as a proxy for the structural and systemic variables which have created a larger burden of COVID-19 in these communities. We tested the null hypothesis that the percentage of individuals identifying as African American or Hispanic in a community was equally distributed between communities with successful and unsuccessful SAHOs. These tests yielded non-significant results, suggesting that whether a SAHO was effective in a community was not associated with the proportion of African American individuals (p=0.573 states; p=0.493 counties) or Hispanic individuals (p=0.950 states; p=0.262 counties) in that community.

Figure 12: The percent of individuals who identified themselves as African American on the ACS survey. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 13: The percent of individuals who identified themselves as Hispanic on the ACS survey. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

After completing a bivariate analysis of these population characteristics, we conducted a multivariate logistic regression analysis to assess whether some of the variables were confounders or effect modifiers. In R Studio we controlled for each of the 12 variables and found that none of them had more than a 10% impact on any of the other variables. This suggests that there is no confounding or effect modification taking place amongst the variables considered.

The cumulative number of days of prior NPI orders remained the only significant result in our multivariate model.

DISCUSSION

The burden of COVID-19 has varied substantially across the US. Media reports, case studies, and scientific literature, have all commonly pointed to the characteristics of local populations as possible explanations for this variability.\textsuperscript{36–38} Our study aimed to determine which factors may have influenced the effectiveness of SAHOs issued in the U.S. in the spring of 2020. To answer
this question, we assessed whether the distribution of 12 characteristics was different in

communities with successful versus unsuccessful SAHOs. The characteristics considered were:

the time between the first case of COVID-19 and start of a SAHO, the value of $R_t$ at start of a

SAHO, the cumulative number of days of alternate NPIs prior to the SAHO, the proportion of

front-line workers in a community, the educational attainment of a community, the proportion of

individuals who lack fluency in English, the median income of a population, the mean age of a

population, the proportion of adults who had received at least one dose of a COVID-19 vaccine

by July 15th 2021, the percentage that voted for Trump in the 2020 general elections, the

proportion of a community that identifies as African American, and the proportion of a

community that identifies as Hispanic.

We expected our results to show that SAHOs were more effective in populations with

characteristics such as high median income compared to populations with contrasting

characteristics such as low median income. This would have been in line with other reports

linking population characteristics to variations in the burden of COVID-19. However, we found

that the only characteristic with a significant difference was the cumulative number of days of

alternate NPI orders prior to a SAHO; this showed that a fewer number of days of prior NPI

orders positively correlated with SAHO effectiveness. Interestingly, this result was found at the

county but not at the state level. When interpreting this result, it is important to recognize the

substantial variation in the data. The county data show that while the medians of the two groups

may be different, there are still many effective SAHOs which have a greater number of days of

prior NPI orders than ineffective SAHOs. Given the variability in the data it is possible our

finding is the result of an opportune sample and not representative of the population. If,

however, this finding is representative of the population, a possible explanation for the

difference could be fatigue. Counties that had endured a greater number of days of restrictive

public health orders prior to SAHO start may have already been stressed or fatigued to the point
where complying with SAHOs was not feasible. This reduction in compliance could have caused
a reduction in order effectiveness. Another, and perhaps more likely, explanation would be that
prior NPI’s minimized the impact of COVID-19 on communities. This could have caused
communities with more days of prior NPIs to view SAHOs as unnecessary or an overreaction. It
follows that such a perspective could lead populations to be less compliant with a SAHO
thereby reducing the order’s effectiveness. This explanation would be in line with other studies
that found combinations of less invasive NPIs were almost as effective at reducing spread of
COVID as SAHOs.23,39

Our findings should encourage those assessing the variable burden of COVID-19 to look
beyond the convenient explanations of demographics. Rather they should consider more
complex and detailed structural and systemic variables. Our assessment supports the idea that
greater in-depth research is necessary to explain the variation in the success of SAHOs. More
detailed assessment is particularly important given that polarization of the American public has
been increasing and negatively impacting public health.40 Inappropriately attributing health
outcomes, such as SAHO effectiveness, to factors that contrast communities could deepen the
division between them and build further inequity. This would result in undermining the efforts of
the public health community.

Limitations

While our findings highlight the importance of detailed assessment into structural and systemic
factors, the study also has significant limitations. This analysis is based on cases reported in the
Johns Hopkins University Coronavirus Resource Center Database. While this dataset is thought
to be one of the most complete records of COVID-19 cases in the US, it is still far from
representing the true number of COVID-19 infections. This is due to a multitude of factors
impacting our ability to identify cases throughout the pandemic. Early in the pandemic, a lack of
available tests left many people possibly infected with COVID-19 undiagnosed. This effect was compounded by restrictive CDC guidelines which outlined who should get tested for COVID-19. In early 2020 these guidelines restricted testing to individuals with multiple symptoms and travel history to various hot spots such as China, Italy, or Iran. Tests were also not equally available to all populations, creating testing deserts where fewer cases were detected. It is also well established that a large proportion of COVID-19 cases are asymptomatic and often undetected. One study found that in New York City, 82%-87% of cases were asymptomatic. Complicating this further, a higher proportion of asymptomatic cases occur in young and healthy individuals compared to older individuals and people with comorbidities. These factors and others have caused a substantial under-reporting of the number of COVID-19 cases in all datasets, including the one used in this analysis. Studies have attempted to map the proportion of unreported cases across the US. Such data could be used to adjust case numbers in the Johns Hopkins Database; however, this modification would lack the locational and temporal precision necessary for this analysis. The result of these sources of error in the case data is a measurement bias where cases are ubiquitously under-reported, however, the degree of under-reporting cannot be assumed to be uniform throughout time or between locations.

To further assess the quality of case data, we compared how reported cases in the Johns Hopkins dataset compared to reported cases in the CDC’s data set. Surprisingly we found substantial differences between the two. In some locations there was more than a 20% difference in cumulative cases reported. These differences are thought to stem from methods of data collection. Johns Hopkins actively collects case data from sources, while the CDC passively has public health departments send case data to them. This distinction led us to believe that the Johns Hopkins database is more accurate.
Variations in daily reported cases greatly influence our calculations of the viral reproductive number. The influence of errors in the number of reported cases is especially large when relatively few cases have been reported. Generally, the smaller communities in our sample reported fewer cases which caused them to be disproportionately affected by errors in case numbers. Unfortunately, it is not possible to determine how errors in case reporting influenced $R_t$ calculations without knowing how under-reporting varied temporally and geographically.

Another important limitation of our study is that calculations of $R_t$ are heavily dependent on the serial interval (SI) of the virus. SI is a measure of how long it takes one infected individual to pass the infection to another person. We used a serial interval of 5.2 days with a standard deviation of 0.7 days which was derived from a meta-analysis. However, other studies have found that SI varies substantially with location and is also influenced by various public health orders. For example, a serial interval in a congregate living environment such as a prison would be expected to be lower than the SI for a rural farming community. Similarly, an order prohibiting group gatherings would be expected to increase the SI as the order would reduce the opportunities for individuals to pass viral particles among each other. Knowing this we performed a sensitivity analysis of the SI and its standard deviation. Surprisingly we found that varying the SI by one day from 5.2 to 4.2 or 6.2 altered our values of $R_t$ by up to 50%. Varying the SI’s standard deviation by one day also impacted our values of $R_t$, in some cases by more than 20%. Notably, these variations also created substantial changes in which communities had effective and ineffective SAHO’s. Published literature on the COVID-19 virus’ reproductive value utilizes a wide range of SI’s which suggests that there may be lack of comparability between these studies, and that reported R values may not be accurate if the SI was not derived from cases geographically and temporally pertinent to the analysis. Numerous studies such as this one, have used a single SI value to derive R values for a range of times and locations. Our findings suggest that these methods are insufficient for generalized analyses, and such reports
should be considered with caution. We recommend that future research identifies a more robust method for either calculating $R_t$ or qualifying SAHO effectiveness in a manner that is not dependent on SI values. This is especially important given that policy makers have used $R$ values throughout the pandemic to help inform decisions with profound consequences, such as when to end a SAHO.

**CONCLUSION**

This analysis is a first look into which factors influence the effectiveness of SAHOs. We found that the cumulative number of days of prior NPI orders may have negatively influenced SAHO effectiveness; however, all other characteristics assessed likely had no impact. It is possible that variability in SAHO effectiveness is derived from systemic (for example access to public health messaging) or structural factors (for example, the financial burden created by a SAHO). It is also possible that the methods used for calculating $R_t$ were not sufficiently robust to accurately identify effective and ineffective SAHOs, causing associations to be missed.

If SAHOs become necessary again in the future, policy makers should recognize that the effectiveness of orders issued in spring of 2020 was variable, and that further research is needed to identify the factors responsible for that variability. Given the extreme impact of SAHOs on communities, determining factors that increase or decrease their effectiveness remains imperative. The threat of a future pandemic requiring lockdowns has drawn the attention of the world. Information presented here may help inform how future SAHOs can be implemented in a manner that maximizes their impact while minimizing their duration and harm.
ACKNOWLEDGEMENTS

Dr. Yea-Hung Chen has provided extremely valuable input and insight into how to accomplish the goals of this paper. This project required an enormous amount of data organization and manipulation. Dr. Chen’s guidance was essential to accomplishing this. Additionally, Dr. Chen’s encouragement and enthusiasm for determining factors that impacted performance of communities during the pandemic brought life to the project.

Dr. Brooke Jarvie has contributed to this project substantially by providing feedback, entertaining brainstorming discussions.

Dr. Kelly Sanders has also been a consistent supporter of this project. Her feedback on, and enthusiasm for this project has greatly improved it. Dr. Sanders’ willingness to invest time into this project has been tremendously appreciated and valued.
REFERENCES


5 CDC COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days (accessed March 1, 2023).


DOI:10.1126/science.abb9789.


Zaremba A, Kizys R, Aharon DY, Demir E. Infected Markets: Novel Coronavirus,


22 Pollard MS, Tucker JS, Green HD. Changes in Adult Alcohol Use and Consequences During the COVID-19 Pandemic in the US. *JAMA Netw open* 2020; **3**: e2022942.


26 The Double Pandemic Of Social Isolation And COVID-19: Cross-Sector Policy Must Address Both I Health Affairs.


28 Sikali K. The dangers of social distancing: How COVID-19 can reshape our social


Prepared for the Independent Panel for Pandemic Preparedness and Response. 2021,

Institute for Global Health Sciences, University of California San Francisco.


40 Allcott H, Boxell L, Conway JC, Gentzkow M, Thaler M, Yang DY. Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic. 2020; published online April 13. DOI:10.3386/W26946.


43 Subramanian R, He Q, Pascual M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc Natl Acad Sci* 2021; 118. DOI:10.1073/PNAS.2019716118.

44 Wu SL, Mertens AN, Crider YS, *et al.* Substantial underestimation of SARS-CoV-2...

Figure 1: Counties included in analysis n=289
Figure 2: Days between first case and SAHO Start. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 3: $R_t$ calculated for the first week of SAHO. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
Figure 4: The cumulative number of days other NPI orders were in place before the start of a SAHO. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 5: The percent of ACS survey respondents who reported working in industries identified as front-line. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
Figure 6: Educational scores relating to the highest level of education attained by respondents of ACS survey. On this scale 0 represents no schooling and 11 represents 5+ years of college. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

![Education Scores State vs County](image)

Figure 7: The percent of ACS survey participants who reported they did not speak English or did not speak English well. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

![Not Fluent in English State vs County](image)
Figure 8: Median income in thousands of US Dollars reported in ACS survey responses (adjusted for inflation to match 2019 USD value). Unsuccessful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 9: Median age of ACS survey respondents. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
Figure 10: The percent of eligible individuals vaccinated as of July 15, 2021. Unsuccessful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 11: Percent of population that voted for Trump in the 2020 general election. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.
Figure 12: The percent of individuals who identified themselves as African American on the ACS survey. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.

Figure 13: The percent of individuals who identified themselves as Hispanic on the ACS survey. Non-successful SAHOs are shown on the left in red, successful SAHOs are shown on the right in green.