Does social distancing work?
The effect of social distancing on COVID19

Observational Study

Kayla Morris¹, Anna Romero¹
Emily Berglund²


Abstract

COVID19 has changed the world as we know it. One way the world is responding to this pandemic is social distancing. Social distancing is used to slow the spread of the disease, as the SARS-CoV-2 virus is primarily spread by droplets. The idea is that if people aren't close enough to one another for droplet transmission to occur, the spread of the disease will be slowed. While most that understand the concept support social distancing, those that do not understand strongly oppose it. This research aims at answering if social distancing is effective. The total confirmed cases of COVID19 were tracked daily for nine states. The rate adjusted increase in cases from the start of each state's stay at home order until April 29th were used in a statistical analysis. The results of the analysis support the hypothesis that social distancing is an effective way to reduce the spread of COVID19.

Keywords: COVID19, SARS-CoV-2, Social distancing, Flatten the curve, Coronavirus

Introduction

‘Coronavirus,’ ‘SARS-CoV-2,’ ‘COVID19,’ and ‘social distancing’ are posted everywhere, but what do they mean? SARS-CoV-2 is the virus responsible for causing the coronavirus pandemic, with symptoms categorized as the “COVID19” disease. SARS-CoV-2 is highly transmissible, and scientists as well as physicians do not have a great understanding of the virus or disease, as it is novel. What is understood is the disease is very contagious and cases are exponentially growing. With no vaccine, treatment, or cure, it is imperative to follow social distancing. According to the CDC, "limiting face-to-face contact with others is the best way to reduce the spread of coronavirus disease 2019 (COVID-19)." Social distancing requires remaining at least 6 feet away from others unless people live together already, as well as not gathering in groups, and staying away from crowded spaces. Social distancing is important as COVID19 primarily spreads between people in close contact via droplets from the mouth and nose.¹ Social distancing is the most important factor that can be controlled.² Important factors in the spread of COVID19 are "how inherently contagious the virus is, how susceptible people are to infection, the number of contacts between people, and the duration of those contacts."² The first two factors cannot be managed, but the last two can be. Slowing the spread of the virus is important to keep from overwhelming healthcare systems. This is known as ‘flattening the curve,’ in which the main goal is reducing how fast the virus moves through populations, reducing how
many patients will seek intensive medical care at once. At the time of this research, the best way to flatten the curve is via social distancing. Italy and China were hit hard by COVID19, resulting in strict lockdowns to slow the spread. The American government left social distancing guidelines up to each state’s governor, and most governors eventually signed a stay at home order into law. This research aims to answer the question: How effective is social distancing and does it work with regards to COVID19?

Methods

Population
The population for this study was the United States of America. A random sample of nine states were chosen in order to mitigate bias. The sample consisted of Arkansas, California, Hawaii, Idaho, New York, Montana, Oregon, Virginia, and Washington. States were chosen from three different categories that indicated varying levels of infection at the beginning of the study, in order to account for different levels of infection at the start of the data collection, as well as varying state populations. This further strengthens the efficacy of the study.

Procedure
Information was collected via the WorldOMeter website which tracks the number of cases of COVID19 in every state and territory that makes up the United States, as well as a worldwide count. Every day, from the 22nd of March to the 29th of April, information was gathered from the website for the nine states. The value from the “Total Cases” column in the table was recorded. The information was then recalculated to represent the amount of cases per 100,000 people, illustrating the increase of infection over time. Stay at home information was gathered from The Institute for Health Metrics and Evaluation (IHME). Using the rate adjusted value for each state when the stay at home order was placed, states were grouped into early and late response groups. A state that had less than 7 cases per 100,000 people when the stay at home order was enacted, were placed in the early group. Any state with more than 7 cases per 100,000, or never placing a stay at home order, were grouped in the late response.

Statistics
Rate adjusted values per 100,000 people in each state were used for the statistical analysis. To calculate the change in rate, the final value collected on April 29th was subtracted from the rate adjusted value from the day each state implemented social distancing. The change in rate value for each state was added to the corresponding early and late group in a spreadsheet. Using the coding software RStudio, an ANOVA test was conducted.

In RStudio, the dataset was uploaded, and codes were run to determine supplementary data such as a five-number summary of the dataset and boxplots to illustrate the data. Lastly, an ANOVA test was run using the summary(aov) code. The code used in this experiment is attached as an extra material.

ANOVA was used for the statistical analysis portion of this study due to the classification of the data. The separation of data between states that responded earlier and later to social distancing, is a categorical distinction. There were five states that were classified as early responses and four as late responses. The rate adjusted value for the number of cases per 10,000 people is numerical data. Furthermore, both variables are independent of one another. ANOVA is used when comparing these two types of data to determine a correlational relationship between the two.

Results
States were placed into an early or late response group for statistical analysis. The categorization was based on the rate adjusted confirmed cases on the date of the stay at home order.
Table 1: Comparison of change in rate of COVID19 cases per 100,000 people from the stay at home (SAH) order date listed to April 29th. The rate adjusted cases, on the date the SAH was ordered per state, is listed below. States are grouped into early (<7 cases per 100,000) and late (>7 cases per 100,000, or no SAH order).

<table>
<thead>
<tr>
<th>States</th>
<th>Date SAH ordered</th>
<th>RA cases</th>
<th>State Response</th>
<th>Change in Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>3/19/20</td>
<td>4.557</td>
<td>Early</td>
<td>113.26</td>
</tr>
<tr>
<td>Idaho</td>
<td>3/25/20</td>
<td>5.092</td>
<td>Early</td>
<td>104.14</td>
</tr>
<tr>
<td>Montana</td>
<td>3/26/20</td>
<td>6.683</td>
<td>Early</td>
<td>34.924</td>
</tr>
<tr>
<td>Hawaii</td>
<td>3/25/20</td>
<td>6.709</td>
<td>Early</td>
<td>36.303</td>
</tr>
<tr>
<td>Oregon</td>
<td>3/23/20</td>
<td>4.557</td>
<td>Early</td>
<td>52.354</td>
</tr>
<tr>
<td>New York</td>
<td>3/22/20</td>
<td>81.17</td>
<td>Late</td>
<td>1487.1</td>
</tr>
<tr>
<td>Washington</td>
<td>3/23/20</td>
<td>29.17</td>
<td>Late</td>
<td>152.61</td>
</tr>
<tr>
<td>Arkansas</td>
<td>N/A</td>
<td>N/A</td>
<td>Late</td>
<td>103.95</td>
</tr>
<tr>
<td>Virginia</td>
<td>3/30/20</td>
<td>14.65</td>
<td>Late</td>
<td>160.64</td>
</tr>
</tbody>
</table>

The results of the ANOVA between early and late response groups showed that there was not a statistically significant difference ($p=0.212$) between states that enacted stay at home orders earlier or later. The variance was high, with New York as an outlier, suggesting it skewed the data. This is represented by the black dot in Figure 1.

Figure 1: Relationship between the two categories, early and late responses to COVID19 (x-axis), and the difference of the rate adjusted value from the start of the SAH order until April 29th (y-axis). The dataset used for this graph consisted of all nine states, with the black dot indicating New York as a probable outlier. (five number summaries can be found in Table 2)

The boxplot on the left indicates a much lower mean, median, and quartiles, which look significant just by observation. A second ANOVA was then run, removing New York. The results of this showed a statistically significant difference between states that responded early and states that responded late ($p=0.0342$). Figure 2 represents this relationship.

Figure 2: Relationship between the two categories, early and late responses to COVID19 (x-axis), and the difference of the rate adjusted value from the start of the SAH order until April 29th (y-axis). The dataset used for this graph consisted of eight states, excluding data for New York. (five number summaries can be found in Table 2)

The boxplots from Figure 2 illustrate a greater difference between the two response categories than Figure 1. The whiskers on the boxplot minimally overlap, and the majority of the data in the left box plot seems significantly lower than the one on the right. This is consistent with $p$-value found. A comparison of the summary statistics of the two analyses with and without New York showed the skewness of the outlier. The inclusion of New York greatly skewed the mean, third quartile, and maximum.

Table 2: Comparison of summary statistics comparing the early group, the late group including New York, and the late group excluding New York.

<table>
<thead>
<tr>
<th>Group</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>34.92</td>
<td>36.3</td>
<td>52.35</td>
<td>68.2</td>
<td>104.14</td>
<td>113.26</td>
</tr>
<tr>
<td>Late W/ NY</td>
<td>103.9</td>
<td>140.4</td>
<td>156.6</td>
<td>476.1</td>
<td>491.3</td>
<td>1487.1</td>
</tr>
<tr>
<td>Late W/O NY</td>
<td>103.9</td>
<td>128.3</td>
<td>152.6</td>
<td>139.1</td>
<td>156.6</td>
<td>160.6</td>
</tr>
</tbody>
</table>
**Discussion**

The first ANOVA test conducted included all nine states and resulted in a p-value greater than 0.05. This suggests there was no significant difference in COVID19 case increase between states enacting social distancing earlier versus later. This was unexpected, as the summary statistics and box plots clearly depict a large difference in values between the two categories. For example, the mean value for the rate adjusted cases in the early category is 68.2 and the mean value for the late category is 476.1. The mean of the latter category is more than six times greater than the former. Similar results were found for all summary statistic values listed in Table 2. Additionally, Figure 1 clearly shows the values in the early category do not have much overlap in values, seemingly indicating that there is a statistical difference between the categories. Based on the aforementioned values and graphs, an insignificant p-value was not expected. After further investigation, it was determined that New York’s data—which has a much larger rate adjusted value than the rest of the states—could be skewing the data. Therefore, it was excluded in the next analysis. The second ANOVA test was run with only eight states: Arkansas, California, Hawaii, Idaho, Montana, Oregon, Virginia, and Washington. Excluding New York changed the values for the late category, as shown in Table 2. After running ANOVA excluding New York, a significant difference was found. The differences in the mean between the two categories was not as large, but more representative of the sample size, whereas including New York was not. The summary statistics for this revised set of data in the late category remained much greater than the early category. Additionally, the p-value indicates that smaller rate adjusted values for the early category is statistically different. Due to these, it can be inferred that New York was acting as an outlier—skewing the data. The revised dataset and the second ANOVA test can be used to determine an accurate conclusion as opposed to the original including New York State.

**Conclusion**

The available data and research strongly indicates that social distancing is the best way to reduce spread of COVID19.¹ ² The analysis of early and late states excluding New York, supports the hypothesis that social distancing is effective. The data collected and analyzed further indicates that there is a correlation between the state’s response, whether early or late, and the case increase since health precautions were taken. In this case, a later response resulted in a greater case increase per 10,000 people, in comparison to an early response. Social distancing is the most important measure that can be used to slow the spread of disease and flatten the curve. Continued research is needed, following COVID19 cases in all 50 states through the end of the pandemic to have an increased understanding on how effective social distancing is with regards to COVID19.

**Acknowledgements**

We would like to thank the Carroll College Biology department, as well as Professor Emily Berglund for her ongoing support in this research and our individual career goals.

**References**


