

THE ROLE OF STATE-WIDE STAY-AT-HOME POLICIES ON CONFIRMED COVID-19 CASES IN THE UNITED STATES: A DETERMINISTIC SIR MODEL

David Chen¹, Seungmin Lee² and Jason Sang³

¹University of Western Ontario, London, Canada

²University of Toronto, Toronto, Canada

³University of Waterloo, Waterloo, Canada

ABSTRACT

In January 2020, the first confirmed case of the novel severe acute respiratory syndrome coronavirus 2 emerged in the United States of America. By March 2020, the USA had declared a national emergency and implemented stay-at-home policies subject to the individual initiative of health authorities of each state. However, ambiguity in the literature exists about the extent to which temporal variation of stay-at-home implementation contributes to an effective stay-at-home order. To examine the role of the implementation of stay-at-home policy at the county level on outbreak progression, we compiled the case count data and dates of policy commencement for 1720 counties from the US Counties: Socio-Health Data database. Measures of central tendency and rate of change identified correlation between the change of confirmed case counts compared to time, quantified by comparing four successive time points of 5 days to the initial date of each county's stay-at-home implementation. We then used a deterministic county-level SIR epidemiological model to predict post stay-at-home case counts based on pre-stay-at-home parameters and compared the model to actual post-stay-at-home case counts to identify the degree of error Mean Squared Error (MSE). Our analyses demonstrated the high error between time since stay-at-home implementation and change in actual case counts compared to predicted case counts, which suggests an interaction between policy and COVID-19 transmission. Our findings shine light on the confounding variables of stay-at-home policy at the county level and the promising outlook of stay-at-home policy in the USA.

KEYWORDS

Stay-at-home, SIR model, deterministic.

1. INTRODUCTION

1.1. Covid-19: An Emergent Epidemic

As the coronavirus epidemic (COVID-19) expands globally, reports of clinical outcomes and contributing risk factors to the rate of confirmed case growth are emerging. Previous SARS-2003 and MERS-2012 epidemics have set the genetic precedent for the coronavirus family. Within the scope of COVID-19, the hotspot for the initial outbreak was first reported at a live animal market on December 12, 2019 in the city of Wuhan in Hubei Province and was declared a global pandemic on March 11, 2020 by the World Health Organization.

The United States has the highest number of cases at over 1.7 million cases and over 100,000 deaths because of the ongoing epidemic as of May 30, 2020 [1]. During the timeframe since the first confirmed COVID-19 case was reported on January 21, 2020 by the Centre for Disease

Control, transmission of COVID-19 gained traction at an exponential rate [1]. Federal travel bans from travellers from China in the last 14 days were instated on January 31, 2020. Due to the high rate of change of epidemiological metrics such as estimated transmission rates and confirmed case count per 1 million people compared to similar developed nations, the US government has been criticized for its inaction at the early stages of the national pandemic.

1.2. The United States Stay-at-home Order

On March 13, 2020, President Donald Trump declared a national emergency that enforced changes to measures in public policy and conduct to limit the spread of COVID-19. These measures included closures of non-essential services, discouraging gatherings of more than 50 people, and promoting work to be done remotely. However, there exists an uncertain outlook of COVID-19 and the efforts of the healthcare system to sustain infection control measures at the national level. The public health system of the United States is subject to the individual initiative of its 50 member states, each with their own health authorities with fragmented jurisdiction based on geopolitical boundaries and variable healthcare resources. Due to this incongruent response, there exists variation in when stay-at-home procedures were implemented, to what extent they were reinforced, and the associated epidemiological consequences on a case by case basis at the county level. As seen in Figure 1, the general response to initiate medium and high severity stay-at-home directives were gradual; some states also did not initiate any form of directive, such as the Midwest states of Nebraska, South Dakota, and North Dakota. The spatial and temporal variation in response contributes to an incongruent nation-wide response that can lead to confusion and disruption between states and their public health policies needed to support an effective stay-at-home [2].

1.3. Current Literature on Stay-at-home Order

Several studies have been conducted evaluating the effectiveness of stay-at-home measures on controlling the spread of COVID-19. Several studies have quantified the effect of stay-at-home measures in Wuhan, the origin of the COVID-19 outbreak and the city that accounts for most of the China's confirmed cases and deaths [3]. One study, focusing specifically on Wuhan, looked at the effects of imposing the stay-at-home and flight restrictions on the number of COVID-19 cases. The researchers found a significant increase in doubling time, the time it takes for the number of confirmed case counts to double, from two days to four days before and after imposing stay-at-home orders while also noting a significant decrease in the correlation coefficient between domestic air travel and COVID19 transmission after imposing the flight restrictions [4]. A second study evaluated the correlation of population emigration out of Wuhan before the implementation of stay-at-home measures and found a correlation coefficient of 0.943 between the number of provincial cases and emigration from Wuhan [3]. Recent research has been directed towards quantifying the effectiveness of stay-at-home measures in other areas of the world. Researchers in the USA predict that stay-at-home measures decreased the number of weekly cases by 30.2% after week one, 40.0% after week two and 48.6% after week three compared to countries that did not impose these measures [5]. The existing literature supports the proposal that the effect of stay-at-home orders can be quantified to evaluate its role in reducing rate of increase of confirmed case counts.

1.4. Proposal of Investigation

This paper intends to investigate the role of state-wide stay-at-home policies on attenuating rates of change of confirmed COVID-19 case counts. Initial forecasts at the city-level, such as in the case of New York City, have been proposed to be effective in reducing the growth rate of novel

confirmed cases. However, estimates indicate that almost half of transmissions may originate from asymptomatic carriers [6].

These estimates can suggest that stay-at-home policies may not be enough to attenuate transmission rates. This poses the need to investigate the effectiveness of stay-at-home policies at the county-level to critically understand if and to what extent these policies are effective using robust correlational analyses and models. We aim to fill the gap in literature surrounding the role of stay-at-home order in the United States through statistical analyses and models translated from similar quantitative approaches used by existing epidemiological literature. Thereby, our investigation develops a robust framework by which the United States can identify key strengths and weaknesses of their stay-at-home policies at the county level to modify their policy response according to the most recent data.



Figure 1. Visualization of implementation of state-wide stay-at-home directives in the United States of America based on severity (low: white, medium: orange, high: red) during the week of March 23 to March 30 for each state (n=50).

2. RESEARCH METHODS

2.1. Dataset Sources

The data was sourced from the Kaiser Family Foundation, the New York Times, and the US Census. They were joined thanks to a US Counties COVID 19 Socio-Health Data Notebook by John Davis [7]. The dataset contains a large set of geographically linked data about the population of the United States, including statistics such as suicide rate, number of mental health providers, and percent single household rate. The smallest geographical division is county. The data was collected from applicable counties with at least one confirmed case of COVID-19 starting from January 1, 2020 to April 30, 2020 inclusive. We hereby examine and use only a few features for the purpose of this investigation. We use the county dimension of the dataset to investigate the effect of stay at home order. We join this county data with their state name to account for repetitive counties. In addition, the available data for COVID-19 cases was different for each county and created some unique challenges for inference that we address in the investigation. Other important features to consider include the date of stay at home order was implemented and the number of new infected cases by day.

2.2. Statistics in the Search Range

We investigated the difference in rates of change between day zero (day of stay-at-home implementation) and 5, 10, 15 and 20 days after. Comparisons between the rate of infections became necessary to evaluate the efficacy of stay-at-home orders. We looked at the percent

change in cases before and after stay at home orders and quantified them in terms of Daily Percent Change (DPC) of cases. We hereby represent the DPC time series using Equation 1. Where the DPC of a given day is computed by taking the new cases on the day of and dividing by the sum of new cases of today and yesterday.

$$DPC(n) = \frac{f(n)}{f(n) + f(n - 1)}$$

Equation 1: Daily Percent Change

Due to the exponential changes in cases, we expected the average DPC to decrease after the stay at home order. Overtime, the DPC is expected to decrease slowly rather than sharply due to the long incubation period of COVID19. Therefore, we wanted to evaluate the data at multi-step intervals after the date of stay-at-home (Day 0) as seen in Equation 2.

$$\text{Search Range} = \text{Day } 0 + 5, 10, 15, 20 \text{ days}$$

Equation 2: Daily Percent Change

This was done to compare the average DPC at two time points within these intervals: setting day zero as the date of stay-at-home implementation, we statistically compared DPC after 5, 10, 15 and 20 days to produce a comparison between the two time points. We propose a set of four comparative statistics across the Search Range as part of our methodology, outlined in Table 1. We first evaluate summary statistics of the impact of stay at home order, by identifying the percentage of counties that witnessed a decrease of average DPC. Next, we examine the average of counties' absolute percent change of DPC to evaluate of empirical effectiveness of the stay-at-home-order. Different counties might have had drastically different absolute change rates of infection. Therefore, we can further explore the empirical change by comparing the percent change of DPC with relation to itself before the stay-at-home order. Lastly, data availability due to the difference in tracking start date can create biases in our analysis as our Search Range expands. As such, we examine and offer the percentage of counties that was dropped out of calculation because of unavailable data; a more in-depth discussion about the biases will be accounted for in the results and discussion section.

2.3. SIR Model Structure

Predictive mathematical models for epidemics are crucial in comprehending the trajectory of an epidemic and plan for effective policy strategies, particularly through a regional perspective. The SIR model is one of the most common epidemiological models used to describe human to human transmission through three mutually exclusive and contiguous states of infection: susceptible (S), infected (I), and recovered (R). Peirlinck et al., among other researchers, have similarly modelled COVID-19 transmission by estimating clinical severity and forecasts for death rates using the SIR model [8]. The SIR model was chosen for its simplicity, minimal number of model parameters and its proven ability to outline the epidemiological factors associated with COVID-19 at the regional scale. The SIR model is particularly well suited in distinguishing important features and developing insights into public health policy. The parsimonious assumptions used to set up the SIR model are used to increase generalizability and avoid overfitting on limited and incomplete datasets.

The SIR model is used to describe human to human transmission through three mutually exclusive and contiguous states of infection: susceptible (S), infected (I), and recovered (R).

Peirlink et al., among other researchers, have similarly modelled COVID-19 transmission by estimating clinical severity and forecasts for death rates using the SIR model [8].

We employed the SIR model to estimate the expected number of COVID-19 case counts for each county in the dataset across a time series as an extension of our investigation on rate of change in confirmed case counts following stay-at-home implementation. The model uses data from before the date of stay-at-home implementation to forecast future case counts. We then compared estimates from the SIR model used to forecast the number of confirmed case counts in a deterministic scenario where stay-at-home had not been implemented against the actual data collected at the county-level to detect the degree of deviance between the two trends.

Correlation between the SIR model that estimates case counts and the actual trend of case counts across a time series can suggest that stay-at-home policies did not play a role in the decrease of the rate of confirmed case count growth to some extent. Limited to no correlation between the SIR model that estimates case counts and the actual trend of case counts across a time series can suggest that stay-at-home policies could play a role in the decrease of the rate of confirmed case count growth to some extent. Further investigation is needed to determine causality and the relative temporal order of stay-at-home policy implementation and rate of change of confirmed case counts at the county level.

We hypothesize that our models will observe a correlation with actual case count data before the stay-at-home order but divert as the stay-at-home order is in effect. This is used to further evaluate the correlation and potential role that stay-at-home plays in decreasing rate of increase of confirmed case counts, and to what power of statistical significance. Instead of evaluating historical data as proposed in previous studies, our novel approach involves comparing modelled forecasts based on pre-stay-at-home data with the trend of actual data. The SIR model measured the dependent variable of three counts of people categorized based on infection state as a function of the independent variable of time as seen in Equation Set 3. We then derived the rate of change for susceptible, recovered, and infected individuals as a function of time as seen in Equation 4, 5, and 6, respectively.

Certain assumptions were made regarding fitting the SIR model within the context of COVID-19.

1. The effect of immigration and emigration on population count was ignored due to the significant decrease in immigration and emigration during the COVID-19 timeframe in the USA.
2. The only way individuals can leave the susceptible group is if they become infected
3. The rate of change of susceptibility over time depends on the number of people susceptible, the number of people infected and the degree of contact between infected and susceptible populations.
4. Homogeneous mixing of the population and random contact between infected and susceptible populations.
5. Each infected individual has a fixed number (b) of contacts each day that transmit the disease to susceptible individuals
6. Each infected individual has a fixed fraction (k) that will recover each day

$$\begin{aligned}S &= S(t) \\I &= I(t) \\R &= RS(t)\end{aligned}$$

Equation 3: SIR Model derived from # of susceptible people in a population $S(t)$, the # of infectious individuals in a population $I(t)$, and # of people who recovered in a population $R(t)$

$$\frac{ds}{dt} = -b \times s(t) \times i(t)$$

Equation 4: Rate of change of the susceptible subpopulation as a function of time

$$\frac{dr}{dt} = -k \times i(t)$$

Equation 5: Rate of change of the recovered subpopulation as a function of time

$$\frac{di}{dt} = [-b \times s(t)] - [k \times i(t)]$$

Equation 6: Rate of change of the infected subpopulation as a function of time

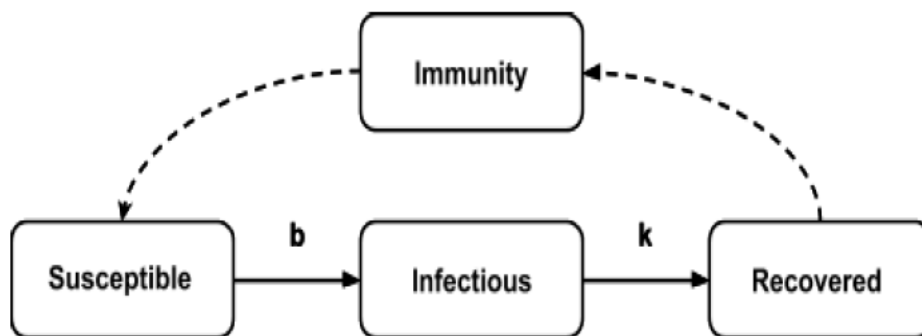


Figure 2. The SIR Model is an epidemiological model is characterized by the number of susceptible people (S), number of people infected (I), and the number of people who have recovered (R) and computes the theoretical number of people infected by a contagious illness in a closed population over time. The number of contacts per day sufficient for the spread of the disease is represented by b . The fixed fraction of the infected group that recovers during any given day is represented by k .

Table 1. Sample statistics of daily percent change of counties in Search Range time points since the date of stay-at-home order

| Statistic | Use Case | Type Representation |
|--|---|---------------------|
| Proportion of counties that reported a decrease in the rate of confirmed case counts | Determine if there was a significant proportion of counties within the sample that reflected a decrease before and after the date of stay-at-home | Percentage |
| Absolute average difference of confirmed case counts in the Search Range | Determine if the reported change in confirmed case counts before and after the date of stay-at-home is empirically significant | Percentage |
| Percentage difference of Daily Percent Change the Search Range | Determine if the Daily Percent Change of confirmed case counts before and after the date of stay-at-home is empirically significant | Percentage |
| Proportion of 1720 total counties that reported data in the Search Range | Determine the proportion of counties that contribute data to the Daily Percent Change statistics at each Search Range interval | Percentage |
| Equation for Each Statistics (Top to Bottom Order) | | |

| |
|--|
| $\frac{\text{Total percent change of all counties after stay at home} - \text{Total percent change of all counties before stay at home}}{\# \text{ of counties}}$ |
| $\frac{\text{Total percent change of all counties after stay at home} - \text{Total percent change of all counties before stay at home}}{\# \text{ of counties}}$ |
| $\text{Percent difference of daily avg. for county X} = \frac{\text{Pct. change before stay at home} - \text{Pct. change after stay at home}}{\text{Pct. change before stay at home}}$ $\frac{\text{Sum(Pct. diff. of daily avg. for all counties)}}{\text{Total Counties}}$ |
| $\frac{\# \text{ Counties that contain data X days before and after the stay at home order}}{\# \text{ of counties}}$ |

The SIR model is fit using maximum likelihood estimation with a Poisson likelihood based on the rates of change of susceptible, recovered and infected subpopulations as a function of time as outlined in Equation 4, 5, and 6 respectively. . The model sensitivity to parameterization highlights the inherent variability between epidemiological projections of outbreak and the need for thorough, population-wide testing for the most accurate parameterization. Using the SIR model, if the rate of change of the infected subpopulation as seen in Equation 6 decreases during an outbreak progression, namely through population-wide intervention measures such as quarantine and social distancing, the rate of increase in new infections within the total population will decrease. However, we caution that if the rate of change of the infected subpopulation is not

zero, an outbreak can reemerge and produce distinct changes in the trajectory of new confirmed cases compared to the original SIR model. Thereby, the use of the SIR model is to model different scenarios under known parameters to illustrate the effectiveness of quarantine public health measures and stress the potential ramifications if these measures are eased too soon.

2.4. SIR Model Parameters

The SIR virology model is parameterized with the mean contact rate (MCR, also represented as k) and the mean recovery rate (MRR, also represented as b), as visualized in Figure 2. The selection of the MCR was applied differently for each individual city to determine the best fit. Specifically, the MCR fitted each city’s case count from the point when community transmission began to the day a stay at home order was in effect. This produced a case count curve for each of the counties with an estimate for future cases. Assuming the effectiveness of stay-at-home policies in decreasing transmission rates, the model was expected to diverge from the actual case count as the incubation period passes.

Model fit was determined using MSE scores, a measure of the goodness of fit of a regression model. The MSE score of the SIR Model before the stay-at-home policies were implemented would show a high correlation and thus a low error between data and function as the SIR model was fitted to fit pre-policy implementation case counts. However, once the stay-at-home policies were implemented, we predict that the MSE score would diverge from the data as the rate of increase in total confirmed cases begins to decrease. By measuring the divergence of MSE, we will gain a better understanding of the infection curve in relation to the SIR model and how stay at home order has changed that dynamic.

3. RESULTS

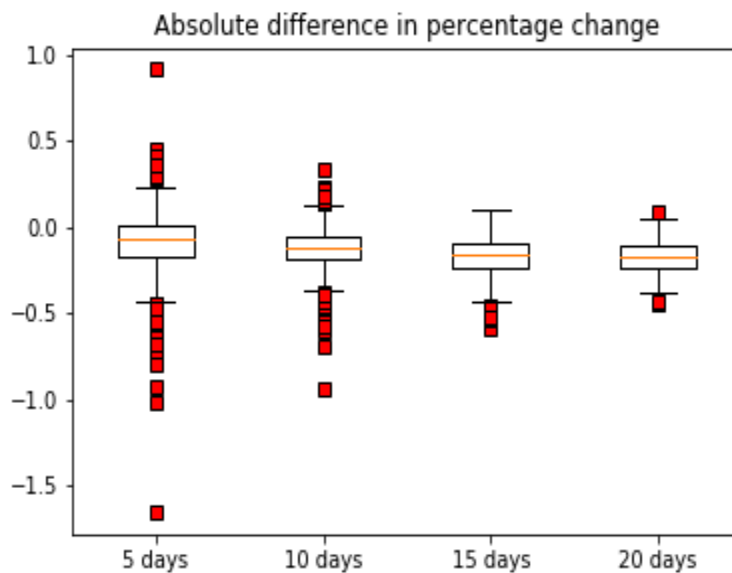


Figure 3. Time since stay-at-home order (Day 0) is negatively correlated with the absolute average difference of confirmed case counts across all counties with valid data in the Search Range. Difference between Day 0 and Day 5 to 20 had absolute average differences of -0.0917%, -0.129%, -0.171% and -0.176% respectively.

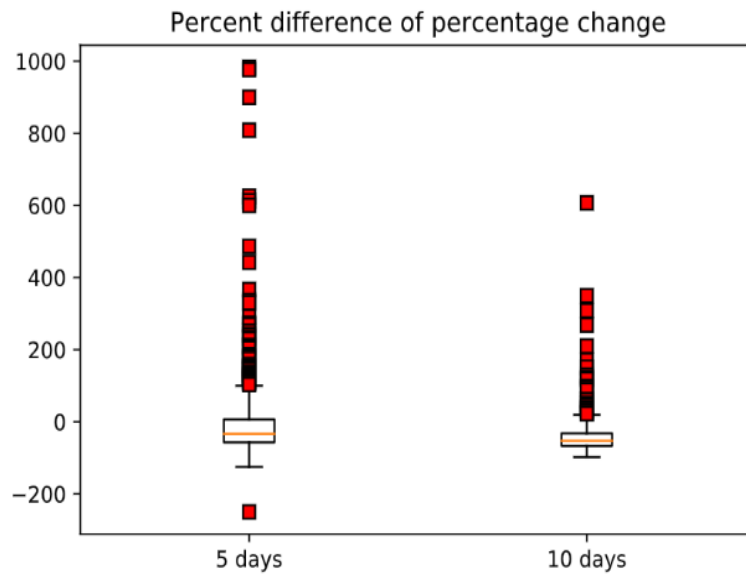


Figure 4. Time since stay-at-home order (Day 0) is negatively correlated with the percentage difference of Daily percentage change. The average percent difference of percentage change for 5 days and 10 days were +3.0% and -34.9%.

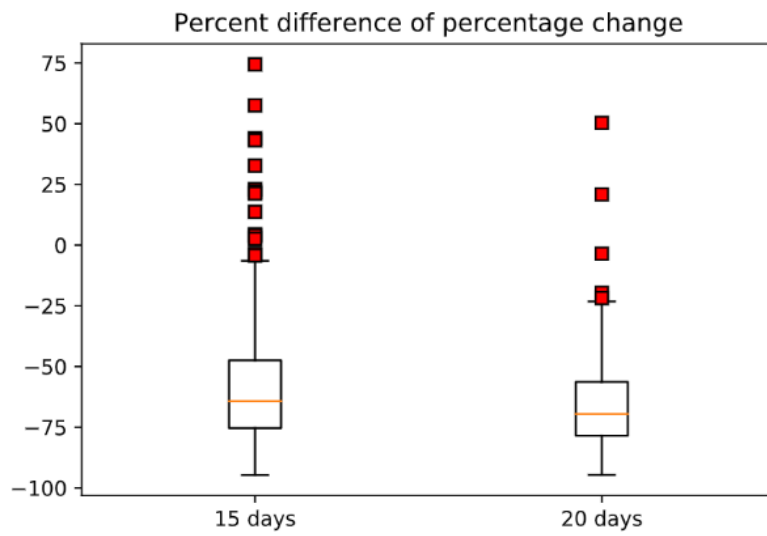


Figure 5. Time since stay-at-home order (Day 0) is negatively correlated with the percentage difference of Daily percentage change. The average percent difference of percentage change for 5 days and 10 days were +34.9% and -57.2%.

Table 2. Logarithmically transformed average MSE score for county case studies before and after stay-at-home order

| Statistic | Search Range (Day 0 compared to X Days after Time Point) | | | |
|--|--|---------|---------|---------|
| | 5 day | 10 day | 15 day | 20 day |
| Proportion of counties that reported a decrease in the rate of confirmed case counts | 75.1% | 87.9% | 93.4% | 97.4% |
| Absolute average difference of confirmed case counts | -0.0917% | -0.129% | -0.171% | -0.176% |
| Percentage difference of Daily Percent Change | +3.0% | -34.9% | -53.6% | -57.2% |
| Proportion of 1720 total counties that reported data in the Search Range | 35.8% | 23.5% | 12.0% | 5.5% |

Table 3. Logarithmically transformed average MSE score for county case studies before and after stay-at-home order

| | Before stay-at-home | After stay-at-home |
|----------------|---------------------|--------------------|
| Mean MSE Score | 2.095 | 9.179 |

Table 4. Logarithmically transformed MSE score for county case studies before and after stay-at-home order

| County | County Type | Before stay-at-home | After stay-at-home |
|---------------|--------------------|---------------------|--------------------|
| Los Angeles | Large Metropolitan | 1.633 | 13.065 |
| Cobb County | Medium City | 3.577 | 11.105 |
| Sonoma County | Medium Township | 1.083 | 3.639 |

Table 2 displays statistics for the sampled counties that compares the effect of the stay-at-home policy at two time points: the first time point at the day of policy implementation in a given county (Day 0) and the second time point at a specified number of days after the policy’s implementation.

As previously indicated, we tracked four key statistics from the dataset. Overall, with a larger Search Range, we observed a great certainty in the decrease of the DPC, evident from both absolute average difference over time as well as relative percent change. However, given the lack of centralized reporting over a number of counties as well as the variation in availability of testing in certain regions of the United States, the proportion of counties which contains valid data for the entirety of the Search Range, up to 20 days since the start of the stay-at-home order, also decreases which can present skew within smaller subsets of data. We will discuss more in depth on the confounding variables as well as bias this may introduce in the discussion section.

Table 2 displays the sample statistics comparing Day Zero of stay-at-home order implementation and increasing time points since then, as originally described in Table 1. We observed that the proportion of counties that reported a decrease in the rate of confirmed case counts compared to Day 0 increased as time increased, suggesting a positive correlation. The absolute average difference of confirmed case counts decreased over time, which suggests a negative correlation.

A similar negative correlation is noted for Percent difference of Daily Percent Change. It is important to note the positive percentage difference of Daily Percent Change comparing between Day 5 and Day 0. Finally, we observed that the proportion of the 1720 total counties that participated in this study decreased as we increased our Search Range, due to limitations within the dataset that led to underreporting and lack of complete data. We excluded these data points as part of our methodology to only include counties with valid data within the Search Range.

Table 3 shows the logarithmically transformed average MSE score for county case studies before and after the stay-at-home order. We see that there is significant error compared to the error before the stay-at-home order that indicates a poor fit between the SIR model trained on pre stay-at-home case count data and modelled under the scenario that the stay-at-home order did not exist compared to the actual case count data on average. This can suggest a deviance from expected trends following stay-at-home order.

Table 4 shows the logarithmically transformed MSE score for county case studies before and after the stay-at-home order. The 3 counties, Los Angeles, Cobb County and Sonoma County, were chosen to be representative case studies whose results can be generalized to a limited degree to other similar sized and featured cities.

4. DISCUSSION

4.1. A Retrospective Analysis on the Impact of Stay-at-home Policy

The study's purpose was to investigate the role of stay-at-home policy through comparison of the projected models of case counts in the scenario that stay-at-home did not exist compared to actual case counts over a time series at the county-level. The results comparing the statistics based on Daily Percent Change from Table 2 are consistent with the literature and reveal how stay-at-home policies correlate with a decrease in Daily Percent Change on average across all counties with valid data within the Search Range. There was a positive correlation observed between the size of the interval within the Search Range since day zero of policy implementation and the proportion of counties that reported a decrease in the rate of increase confirmed case counts.

Surprisingly, the majority of counties reported a decrease in confirmed case count within 5 days of policy implementation and continued to reflect a decrease in subsequent intervals, which we interpreted as an indication of the sweeping implementation of stay-at-home policy and its correlated effect. The positive correlation parallels historical and epidemiological case reports of commensurate stay-at-home policies during epidemics. In particular, this compares to the precedent scenario within the coronavirus family, the response to the 2003 pandemic of severe acute respiratory syndrome, which shares similar border control, stay-at-home and population level surveillance policies at scale and subsequent decrease in confirmed case counts following implementation. An important aspect to note is that this positive correlation does not indicate temporal order of which variable came first, or the causation of stay-at-home policy implementation leading to the observed decrease of confirmed case counts across all counties on average.

There was a negative correlation observed from Table 2 between the size of the Search Range interval since day zero of policy implementation and the average absolute decrease of rate of change of confirmed cases across all counties with valid data within the Search Range. This correlation is consistent throughout the intervals of the Search Range, which we interpreted as an increase in the longevity and robustness of stay-at-home policy implementation since Day Zero. This supports the conclusion of a study by Jiang et al. that concluded that longer periods of stay-at-home could be more effective in preventing viral transmission, in response to new findings

that median incubation periods of all infected were 1.8 days longer than previously estimated [9]. Thereby, we would expect based on literature that as the time since Day Zero of policy implementation increases, the decrease in Percent Daily Change at increasing time points compared to Day Zero increases, and this is reflected in the negative correlation observed.

To further illustrate the difference in rate of change, we observed a negative correlation from Table 2 between the size of the Search Range interval and the average relative decrease of rate of change of confirmed cases across all counties with valid data within the Search Range. We noted that the average relative decrease of rate of change of confirmed cases at post-10, 15, and 20 days reflected the negative correlation observed. However, we noted a low increase of 3.0% in average relative decrease of rate of change of confirmed cases at 5 days after Day Zero.

This anomaly at Day 5 leads to two mutually inclusive conclusions: the presence of outliers in the county data at Day 5 that artificially skew the averaged mean of the statistic to be higher than expected and the implementation of stay-at-home policies did not have a significant effect that translates to confirmed case counts at Day 5. We believe both the incubation period and the presence of outliers outside of the interquartile range of the sample data may skew the data at Day 5 compared to subsequent intervals and be represented as more positive than expected.

As seen in Figure 4, the boxplot at Day 5 illustrates a significant skew of positive value outliers that increase the overall mean of the statistic, since averaged values tend to be sensitive to statistical outliers. This supports the conclusion that the presence of outliers had an effect on the statistics on Day 5.

We also believe that there was no significant effect at Day 5 observed due to the median incubation period that persists beyond Day 5, which reduces the reported case counts compared to subsequent intervals. This observation was determined by the near-zero average relative decrease of rate of change of confirmed cases at Day 5, which indicates that relative to 5 days before Day Zero, there was no empirically significant difference in rate of change. This in turn leads us to believe that at Day 5, there is no empirically significant relationship between implementation of stay-at-home policies and relative decrease in rate of change of confirmed cases. Based on literature that purports a median incubation period of 5.1 days [10], we would expect that 50% of the infected population on average would show symptoms by Day 5, and even less would have already been diagnosed on the day that they show symptoms. Thereby, it is unlikely that the implementation of the stay-at-home policy would have an observable correlation with confirmed case counts by Day 5.

Further investigation is needed to identify if and to what extent the two mutually inclusive conclusions contribute to a systematic decrease in the relative rate of change of confirmed case counts at Day 5.

It is important to note the limitation that as the Search Range increased at 5 day steps, starting at 5 days to 20 days since Day Zero of stay-at-home policy implementation, the number of counties with valid data within the Search Range decreased as part of a negative correlation as observed in Table 2. This is because the Search Range extended beyond the existing data for certain counties, which were then excluded based on the scope of the Search Range as part of the methodology. The decreasing subset of data as the Search Range increased was likely due to the inconsistent nature of reporting of confirmed cases at the county-level to the centralized national dataset. With respect to the interval at the 20 days post-Day Zero of policy implementation, 5.5% of the original 1720 counties had data, which can then lead to outliers having a greater skew on averaged data within a smaller subset of sample data.

4.2. SIR Model Case Studies

The SIR Model is proposed to explain the projected infection dynamics by predicting the number of case counts based on pre-stay-at-home data in the scenario where stay-at-home did not exist compared to the actual number of confirmed case counts across a time series for each county. Underestimation and overestimation can be particularly impactful in the early stages of an epidemic due to lack of reliable information. After the SIR model parameters have been estimated based on the clinical case count data, we evaluated the effect of the implementation of the stay-at-home policy by comparing the error between trends of actual number of case counts and forecasted number of case counts from the SIR model.

We used the first MSE value for each county's time series data to investigate how well the SIR model correlated with the actual case counts over time before the stay-at-home, and the second MSE value to investigate how well the SIR model correlated with the actual case counts over time before the stay-at-home. The mean MSE score transformed logarithmically of all counties before stay-at-home was 2.095, which suggests relatively lower error indicative of good fit between the SIR model and the actual case count data, that is consistent with existing literature on the accuracy of SIR modelling with trained data and relevant clinical input parameters. The mean MSE score transformed logarithmically of all counties after stay-at-home was 9.179, which suggests relatively higher error indicative of poor fit between the SIR model and the actual case count data. This indicates that the significant deviance between the case counts in the scenario predicted by the SIR model where stay-at-home did not exist compared to the actual case count data post-stay at-home likely occurred due to a factor that strongly correlates with rapid change in transmission rates. The literature establishes that stay-at-home is purported to be an effective public health measure to decrease rate of transmission, and this is reflected in the deviance between the SIR model and the actual case counts. It is understood that the model is limited due to the sensitivity of the input parameters initiated in the methodology in recreating accurate models of epidemic transmission, such as the rate of testing and consistency of reporting at the county level, that can contribute to variation in input parameters and measures of central tendency at the national level.

As an extension of our investigation, we observed three categories of trends between case count forecasts from the SIR model compared to the actual case counts after the implementation of stay-at-home policy.

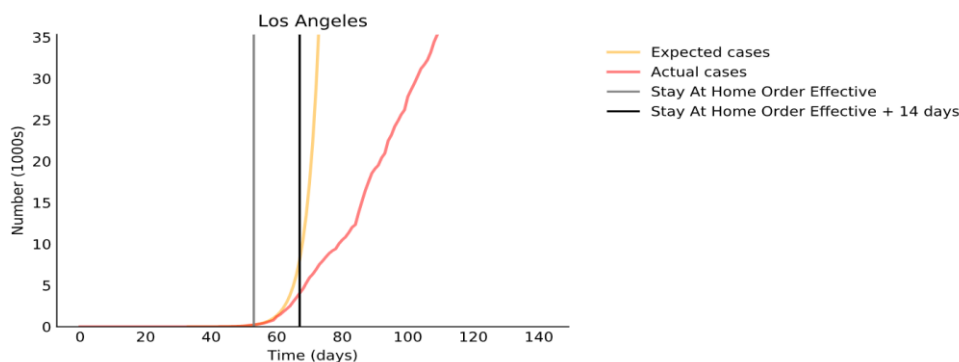


Figure 6. SIR Model (yellow) prediction of expected case counts in the scenario that stay-at-home orders were not implemented compared to actual case counts of Los Angeles county, California. The high deviation (MSE = 13.065) between expected and actual case counts post stay-at-home implementation suggests that there may be a significant factor that influenced the number of actual case counts in Los Angeles county.

Figure 6 outlines the forecast of case counts based on the SIR model and the actual case counts after the implementation of stay-at-home for the city of Los Angeles, California. Here, we observed a strong correlation ($MSE = 1.633$) that represents a very good fit by the SIR model on the actual case count data pre-stay-at-home. It is important to note that after stay-at-home, a strong deviation ($MSE = 13.065$) between the forecast of case counts from the SIR model in the scenario in which stay-at-home was not implemented contrasted with the actual number of case counts reported after stay-at-home was implemented in the real world. Los Angeles case count data shows high error between the date since the start of the stay-at-home and the total confirmed case counts relative to the expected number of case counts. This deviation is representative of similar trends of other large metropolitan cities such as San Diego, New York City, and Houston defined by over one million population and regarded as economic hubs across the United States. We believe that these consistent empirical similarities in trends based on forecasted and actual data suggests that for large cities, there is a high error between the date since the start of the stay-at-home and the rate of increase of transmission rates that translates to confirmed case counts. Thereby, large cities will likely experience a decrease in rate of change of confirmed case counts following implementation of stay-at-home policies compared to the scenario where these policies were not implemented. We noted that for larger cities, the reporting of confirmed cases were more likely to be consistent due to established healthcare systems and the lesser effect of bias from underreporting or overreporting of subsets of case counts on the total number of confirmed case counts.

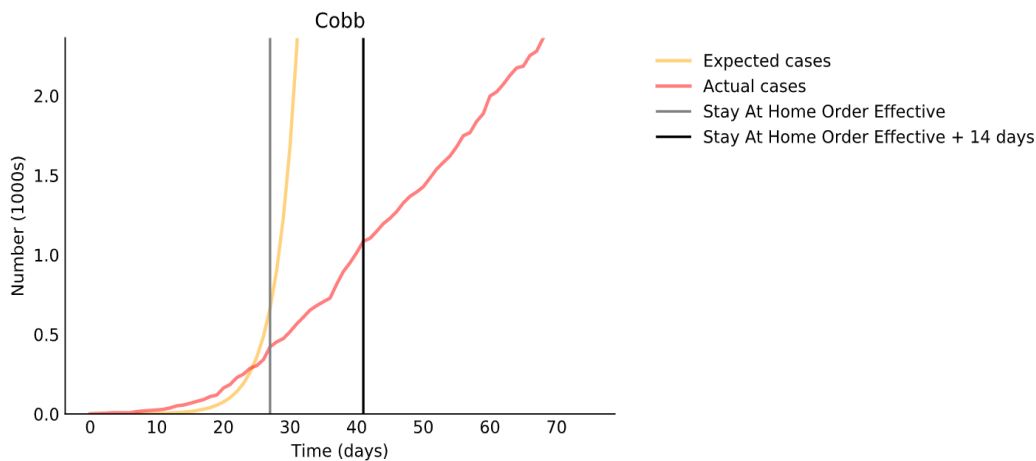


Figure 7. SIR Model (yellow) prediction of expected case counts in the scenario that stay-at-home orders were not implemented compared to actual case counts (red) of Cobb county, Georgia. The high deviation ($MSE = 11.105$) between expected and actual case counts post stay-at-home implementation suggests that there may be a significant factor that influenced the number of actual case counts in Cobb county.

Figure 7 outlines the forecast of case counts based on the SIR model and the actual case counts after the implementation of stay-at-home for Cobb county in Georgia. Here, we observed a medium correlation ($MSE = 3.577$) that represents a good fit by the SIR model on the actual case count data pre-stay-at-home. It is important to note that after stay-at-home, we observed a strong deviation ($MSE = 11.105$) between the forecast of case counts from the SIR model in the scenario in which stay-at-home was not implemented contrasted with the actual number of case counts reported after stay-at-home was implemented in the real world. We observe that Cobb county shows a high error between the date since the start of the stay-at-home and the total confirmed case counts relative to the expected number of case counts. Cobb county is home to a suburban population of 760,141 with a large, in-city university campus and access to major transportation hubs and airports in nearby counties [11]. Cobb county has a less population count

and more variable population density than comparatively homogeneous and dense metropolitan areas such as Los Angeles. From this, we noted that the reporting of confirmed cases were less likely to be consistent due to less centralized healthcare systems and the greater effect of bias from underreporting or overreporting of subsets of case counts on the total number of confirmed case counts. However, the contrast between forecasted and actual confirmed case counts deviates similarly to counties with larger populations that are more robust to skew from underreporting or overreporting such as Los Angeles. From this, we believe that Cobb county likely experienced a decrease in rate of change of confirmed case counts following implementation of stay-at-home policy compared to the scenario where these policies were not implemented. Further investigation is warranted to investigate to what extent population size plays a role in reinforcing the robustness of the SIR predictive modelling to underreporting and overreporting.

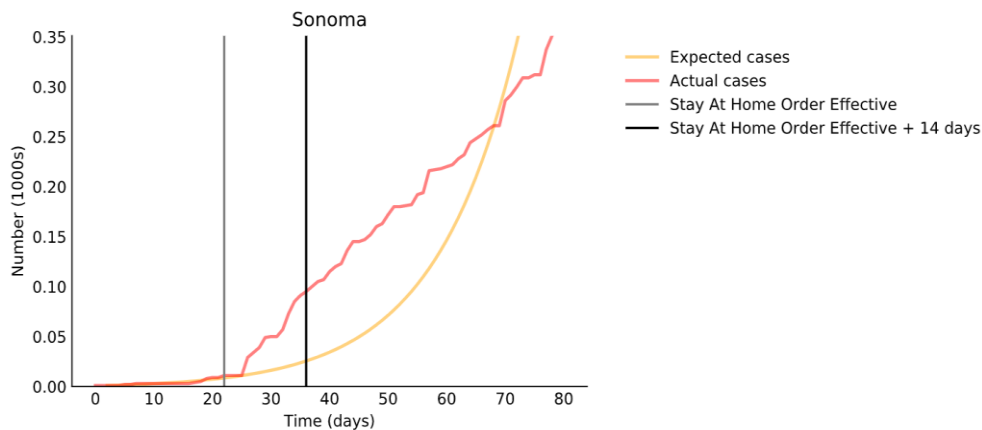


Figure 8. SIR Model (yellow) prediction of expected case counts in the scenario that stay-at-home orders were not implemented compared to actual case counts (red) of Sonoma county, California. The poor fit of the SIR model to post stay-at-home actual case count data (MSE = 3.639) invites further investigation of the anomaly of actual reported case counts and alternative models for edge cases such as Sonoma county.

Figure 8 outlines the prediction of case counts based on the SIR model compared to the actual case counts after the implementation of stay-at-home for Sonoma county in California. Here, we observe a surprising trend where the expected number of cases predicted by the SIR model in the scenario where stay-at-home did not exist was predicted to have a lower number of case counts from Day 25 to Day 68 compared to the actual confirmed case counts. We believe that this suggests that following the implementation of stay-at-home policy in Sonoma county, the stay-at-home policy may not have played a role with significant empirical observations of correlation between the date since the start of the stay-at-home and the number of confirmed case counts. This conclusion is supported by the poor fit of the SIR Model (MSE = 3.639) to the actual case count data post-stay-at-home.

The low MSE value for post-stay-at-home compares to the high MSE values of Los Angeles and Cobb County at around 13.065 and 11.105 respectively, which serves as an anomaly that we investigated further. Following Day 68, however, the expected number of cases forecasted by the SIR model exceeds the actual number of cases that indicates that there is an external variable influencing the confirmed case count, which we suggest can be due to stay-at-home implementation.

Sonoma county is home to a rural population of 499,942 and hosts an agricultural and tourism industry, with accessible pathways to major transportation hubs, airports, and railroads [12]. Sonoma county has a significantly lower population count and more variable population density

than comparatively homogeneous and dense metropolitan areas such as Los Angeles. Like Cobb county, we noted that the reporting of confirmed cases were less likely to be consistent due to less centralized healthcare systems and the greater effect of bias from underreporting or overreporting of subsets of case counts on the total number of confirmed case counts. The SIR model reported a strong fit to the pre-stay-at-home actual case count data (MSE = 1.083) but a poor fit to post stay-at-home actual case count data (MSE = 3.639), which suggests the possibility of overfitting of the SIR model on the pre-stay-at-home data. We propose that the anomalous nature of Sonoma county when comparing the SIR model forecast with the actual case number counts is indicative of overfitting of the SIR model based on under-reported case counts at the earliest stages of county-specific transmission before the date of stay-at-home implementation. One possible reason for overfitting is a sudden change in rate of increase of actual case counts post-stay-at-home that the SIR model could not account for since it was fitted based on pre-stay-at-home case count data, which can be seen on Day 25 of Figure X. Further investigation is warranted to investigate anomalous specific county-level stay-at-home policies and the extent to which they influence transmission rates.

4.3. SIR Model Use Cases and Limitations

The SIR model used in this study assumes that all individuals have the same probability to contract the disease and that there is homogenous mixing of the population. These assumptions fail to consider several factors relevant to COVID-19 transmission. This model considers individuals in self-stay-at-home before lockdown measures were officially announced as equally susceptible to those who did not self-stay-at-home. It also does not take into consideration variation in susceptibility in the population, for example, due to age comorbidity. The model also assumes that infection confers immunity; however, research to validate this claim is ongoing. Despite these limitations however, the SIR model is one of the most used models for epidemics and is useful in evaluating the difference in the rate of change in COVID-19 cases before and after lockdown measures were implemented. In fact, several variations of the SIR model exist and the one specifically used for this study, a time dependent SIR model, have been used by several other past studies to model COVID-19 transmission [13] [14]. This model is not only more adaptive than a static SIR model approach but also more predictive than many other traditional estimation methods.

4.4. Variation in Stay-at-home Compliance

This study focused on the effect of implementing a lockdown but not on the specific lockdown policies themselves. The extent of the lockdown policy, how strictly they were enforced and how this information was conveyed to residents, differed among the counties. These interdisciplinary factors likely play a role in determining the role of stay-at-home and warrants further investigation through correlational analyses.

One factor that may contribute to variation in stay-at-home compliance during COVID-19 is the knowledge about the emerging outbreak of disease and communication of practical protocols. In the case of 5 Australian schools closed during the H1N1 flu pandemic, the unclear protocol led the public to create their own protocols based on their own accepted interpretations of the symptoms of the disease, degree of contact with infected individuals and overall transmission risk [15]. On the other hand, hyper-strict protocols can be difficult to understand. In this second case, health professionals stay-at-homed in Senegal during the Ebola epidemic did not consistently adhere to stay-at-home measures since they believed that the protocols were over precautionary [16]. The effectiveness of communicating stay-at-home policies will introduce random variation on if and when a statistically significant change is observed in confirmed case counts.

A second factor that may have contributed to this variation in stay-at-home compliance is politics. Specifically in the context of the USA, the COVID-19 situation has created a large political division between Republicans and Democrats. By April 22nd, only eight states (Arkansas, Iowa, Nebraska, North and South Dakota, Oklahoma, Utah, and Wyoming) had not implemented stay-at-home orders; interestingly, these are all typically Republican-leaning states [17]. As a general trend, Republican-led states have not only been slower, and less likely, to impose restrictions but tended to downplay the situation as well. Several Republican leaders have even encouraged citizens to continue to visit bars and restaurants amid the pandemic. In addition, recent polls suggest that Democrats in general express more concern regarding COVID-19 and have changed their personal behaviours to a greater extent than Republicans [18]. As such, politics may have produced variation in stay-at-home compliance and could have affected the effectiveness of stay-at-home policies.

4.5. Variation in Testing

The number of tests administered was also not considered in this study. Because the number of cases per day is dependent on the number of tests administered, the perceived effectiveness or ineffectiveness of implementing stay-at-home measures could be due to a significant difference in the number of administered tests before and after lockdown measures. Since March 14th, the number of tests administered in the USA has been steadily increasing, reaching almost 4 million tests on April 18th [19]. However, the number of testing kits available and the number of tests administered vary drastically between counties. Not only this but several states have been reported as not distinguishing between the overall number of tests performed and the total number of individuals tested. In addition, there have also been issues with the inclusion of serology or antibody tests, which are not diagnostic tests, in their total number of tests [20]. This highlights a significant limitation to not only to this study but to the available COVID-19 case data in the USA as a whole.

4.6. Epidemiological Considerations of Herd Immunity

Herd immunity is defined as the protection from infection conferred to susceptible individuals when a proportion of immune individuals exist within the population [21]. This stems from the sum of the effects of individual immunity scaled to a population and leads to two primary scenarios: the first scenario where a pathogen can freely transmit through susceptible hosts of a naive population and the second scenario where a proportion of the population has conferred immunity from previous infection or vaccines that prevents further pathogenic transmission. Kissler et al. state that unless a previous immunizing infection that confers the same degree of immunity to COVID-19 had taken place, there is otherwise little evidence to suggest that herd immunity is a key factor in reducing transmission rates in May 2020 within the scope of the United States during the implementation of the stay-at-home policy [22].

Herd immunity is expected to take up to 18 months in the absence of a vaccine, and given the timeframe since the first confirmed case of COVID-19 in the United States on January 22, 2020, it is unlikely that the propagation of immunity was expedited. Furthermore, a study by Bao et. al found that rhesus macaques were immune to reinfection one month after the first viral infection, which can suggest a degree of short-term immunity [23]. Despite promising findings in the early stages of immunological research, it remains to be determined to what extent herd immunity currently scales to the size of the American population and the potential effect on reducing confirmed case counts in comparison to stay-at-home measures. Further mass serological testing is required to estimate how close the population of the United States is to reaching the minimum herd immunity threshold.

5. CONCLUSION

This study evaluated the role of stay-at-home measures in decreasing the transmission rate of COVID-19. Our findings suggest that while stay-at-home measures were effective overall, they are more likely to be effective in larger, higher density counties. This is due to the nature of counties with smaller populations and population distributions.

An SIR model was first proposed for each county to model COVID-19 transmission before stay-at-home measures were implemented. Based on this model, a Percent Daily Change was determined by comparing the actual number of cases with the predicted number of cases 5, 10, 15, and 20 days after these measures were implemented. Overall, most counties (75.1%) saw a decrease in the number of reported cases within 5 days of implementation with almost all (97.4%) counties showing a decrease by 20 days. Interestingly, at 5 days after stay-at-home measures were implemented, we observed a slight increase of 3.0% in the percentage difference in the Percent Daily Change, but this can be explained due the median incubation period of COVID-19 being 5.1 days.

We then focused our attention to three key counties (Los Angeles, Cobb County, Sonoma County) to represent a large metropolitan city, a medium-sized city, and medium-sized Township. While both Los Angeles and Cobb County showed a strong fit with the SIR model before stay-at-home was implemented (MSE = 1.633, 3.577 respectively) and a weak fit after (MSE = 13.065, 11.105 respectively), Sonoma County produced unexpected results. Unlike the other two counties, stay-at-home measures produced higher numbers of cases than predicted using the SIR model. We suggest that a lower population count, more variable population density, less centralized healthcare systems leading to inconsistent reporting as well as greater effects of bias from underreporting or overreporting as potential reasons for these observations. As such, based on these results, we propose that stay-at-home home measures are most effective when implemented in larger cities. We also note a promising outlook for the role of quarantine on attenuating the increase in actual case counts on average across sampled counties.

Further investigation is still needed to better understand why and to what extent these anomalies are affecting the effectiveness of stay-at-home measures at the county-level. Since this was a correlational study, we propose potential for further investigation on the temporal order of the stay-at-home implementation and decrease in rate of case count change and experimental validation for causal factors. Subtle interplays between causation and association across the variable nature of counties across the United States should be noted. In fact, improvements in different factors such as testing rates, resource allocation and knowledge dissemination can decrease efficiency of transmission to variable extents, which individually should be analysed to provide a holistic outlook of outbreak progression. This will vastly improve the efficacy of stay-at-home policies at the county level, as policies can be uniquely devised to fit the needs of that county.

Our study contributes to existing literature of public health policy by determining that the emergent nature of COVID-19 transmission and stay-at-home policy can be quantified to better prepare the public health response at scale across the United States. On a larger scale, the results of this study can be applied globally as well as towards future outbreaks. Several additional factors, such as variations in testing and compliance rates, that may affect the effectiveness of quarantine policies were also suggested and a further investigation on these topics can provide more localized, targeted means of slowing down the transmission rates of infectious diseases.

ACKNOWLEDGEMENTS

We would like to thank Mr. Wu for his assistance in the designing the research methodology of this study.

SUPPLEMENTARY

The model used in this research article and relevant graphs can be found at <https://github.com/JZSang/david-datascience-paper>.

REFERENCES

- [1] The New York Times. Coronavirus in the u.s.: Latest map and case count. Coronavirus in the US: Latest Map and Case Count, Mar 2020.
- [2] A.P.A. Katz, M.P.H. Rebecca PhD, Andrea M.P.H. Vaught, Adrienne M.S. Formentos, and Jordan B.A. Capizola. Raising the yellow flag. *Journal of Public Health Management and Practice*: July/August, 24(Issue 4):380–384.
- [3] Zeliang Chen, Qi Zhang, Yi Lu, Xi Zhang, Wenjun Zhang, Cheng Guo, Conghui Liao, Qianlin Li, Xiaohu Han, Jiahai Lu, and et al. Distribution of the 2019-ncov epidemic and correlation with population emigration from wuhan, china. *Chinese Medical Journal*, Dec 2020.
- [4] Hien Lau, Veria Khosrawipour, Piotr Kocbach, Agata Mikolajczyk, Justyna Schubert, Jacek Bania, and Tanja Khosrawipour. The positive impact of lockdown in wuhan on containing the covid-19 outbreak in china. *Journal of Travel Medicine*, 27(3), 2020.
- [5] James H Fowler, Seth J Hill, Nick Obradovich, and Remy Levin. The effect of stay-at-home orders on covid-19 cases and fatalities in the united states. *MedRxiv*, Apr 2020.
- [6] Covid-19: Study estimates rate of 'silent transmission'.
- [7] John Davis. Us counties: Covid19 weather socio/health data, May 2020.
- [8] Mathias Peirlinck, Kevin Linka, Francisco Sahli Costabal, and Ellen Kuhl. Outbreak dynamics of covid-19 in china and the united states. *Biomechanics and Modeling in Mechanobiology*, Apr 2020.
- [9] Xue Jiang, Yawei Niu, Xiong Li, Lin Li, Wenxiang Cai, Yucan Chen, Bo Liao, and Edwin Wang. Is a 14-day quarantine period optimal for effectively controlling coronavirus disease 2019 (covid-19)? 2020.
- [10] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9):577–582, May 2020.
- [11] U.s. census bureau quickfacts: Cobb county, georgia.
- [12] Acs demographic and housing estimates, Nov 2019.
- [13] Hyeontae Jo, Hwijae Son, Se Young Jung, and Hyung Ju Hwang. Analysis of covid-19 spread in south korea using the sir model with time-dependent parameters and deep learning. 2020.
- [14] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of populationwide interventions in italy. *Nature Medicine*, 2020.
- [15] Annette Braunack-Mayer, Rebecca Tooher, Joanne E Collins, Jackie M Street, and Helen Marshall. Understanding the school community's response to school closures during the h1n1 2009 influenza pandemic. *BMC Public Health*, 13(1), 2013.
- [16] Alice Desclaux, Dioumel Badji, Albert Gautier Ndione, and Khoudia Sow. Accepted monitoring or endured quarantine? ebola contacts perceptions in senegal. *Social Science Medicine*, 178:38–45, 2017.
- [17] These eight states haven't issued stay-at-home orders to fight the coronavirus outbreak, Apr 2020.
- [18] Ronald Brownstein. Red and blue america aren't experiencing the same pandemic, Mar 2020.
- [19] Andy Kiersz. The us has now tested more than 4 million people for the coronavirus, but the number of tests being done per day is plateauing, Apr 2020.
- [20] All state comparison of testing efforts.
- [21] Haley E. Randolph and Luis B. Barreiro. Herd immunity: Understanding covid-19. *Immunity*,

52(5):737–741, 2020.

- [22] Stephen M. Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.
- [23] Linlin Bao, Wei Deng, Hong Gao, Chong Xiao, Jiayi Liu, Jing Xue, Qi Lv, Jiangning Liu, Pin Yu, Yanfeng Xu, and et al. Lack of reinfection in rhesus macaques infected with sars-cov-2. 2020.

AUTHORS

David Chen: Interdisciplinary Medical Sciences student. Affiliated with the Schulich School of Medicine and Dentistry at the University of Western Ontario.

Seungmin Lee: Pharmacology and Toxicology student. Affiliated with the University of Toronto.

Jason Sang: Affiliated with the University of Waterloo, Computer Science department.

Summers Wu: Affiliated with the University of Western Ontario, Software Engineering department.