JOURNAL OF
EMERGING INVESTIGATORS

Article

Relating socioeconomic position (SEP) and vaccination with Covid-19 rates in select populations

Sweekrit Bhatnagar[1], Yue Cao[1] , Diya Hasteer[1] , Joshua Li[1] , Larry N. McMahan[1]
[1] Aspiring Scholars Directed Research Program, Freemont, California

SUMMARY

In the COVID-19 pandemic, the world saw countless communities negatively impacted, and low-income communities were the most affected by the disease (29). In this study, we aimed to investigate and identify the relationship between key socioeconomic factors, COVID-19 infection rate, and vaccination, with hopes to apply this research to support low-income communities that are at a higher risk of infection. We hypothesized that there is a negative correlation between COVID-19 infection rate and positive economic factors such as per-capita GDP, and that there is a positive correlation between vaccination rates and similar factors. In our model for both infection rate and vaccination, we observed that although a negative relationship for some factors did exist, not all economic values had an observable correlation between the two . These findings could be used to further identify socioeconomic factors with the highest impact on disease infection rate, and therefore redirect public and private funding to solve underlying issues that allow the spread of disease.

INTRODUCTION

The COVID-19 pandemic demonstrated how people in lower socioeconomic positions (SEPs) experience disease more dramatically than those that are in higher socioeconomic standings. Many people in lower SEPs must make sacrifices and risk their and their families' health in order to maintain a level of economic security (1). Many epidemiological studies have confirmed the negative relationship between SEP and disease prevalence (even finding occasional contradictory results); however, few have investigated socioeconomic factors past the SEP Index (2-6).

A multi-disease study in Vietnam provided background results on the relation between SEP and diseases (2). The study demonstrated that population density has a positive correlation with oral-borne disease. Additionally, it associated high percentages of illiteracy with diarrhea, shigellosis, dengue fever, malaria, and rabies. Meanwhile, a study on household SEP with rates of individual infectious diseases risk in Kenya showed an increase in SEP with an increase in the overall health of the population (3). The study found that individuals in households with the lowest SEP were at the greatest risk of infection from Plasmodium falciparum, hookworm, and Entamoeba histolytica/dispar, as well as coinfection from each pathogen.

However, data from both studies demonstrated a negative correlation between SEP and infectious disease. The Vietnam study found that population density has a negative correlation with deadly infectious diseases such as measles and mumps (2), while the Kenya study found that infection with Mycobacterium tuberculosis was most likely in households with the highest SEP (3). This demonstrated a discrepancy in the values that are being presented which can create uncertainty in the relationship between socioeconomic position and disease prevalence.

Studies in the United Kingdom and the United States of America have also related COVID-19 to socioeconomic status through data collection and surveys. In the UK, a study analyzed the association between a lifestyle score, socioeconomic status, and COVID-19 outcomes within the UK Biobank cohort (a large prospective cohort of 502,536 participants aged 37–73 years) (4). They found that in fully adjusted models, there was evidence of nonlinear association of socioeconomic status with COVID-19 mortality, but a lack of association with severe COVID-19 (4). An analysis performed by another UK-based research group pointed to a multitude of factors that increase the exposure of people of low SEP to disease—overcrowded housing, less work-from-home opportunities, unstable income, and limited access to healthcare are overarching reasons behind this clear divide (5). A study in the United States, analyzing how COVID-19 disproportionately impacts minorities and individuals of lower SES showed how compared to non-Hispanic white individuals black, Asian/Pacific Islander, and Hispanic people had a significantly higher likelihood of being unable to work, and those with the lowest income group (≤$25,000) had the most serious impact from COVID-19 (6).

We noticed that although studies like that in Vietnam (2) and the United States (6) demonstrate a negative association with SEP and disease prevalence, discrepancies arise in studies such as that in Kenya (3), where a positive association was found, and that in the UK, where there was a lack of association between SEP and severe COVID-19 (4).

Our study related SEP with rates of COVID-19 through an agent-based modeling to represent the near-ideal spread of disease (accounting for no public health restrictions or other external factors) and the iteration through data to determine the correlation between the two factors. In addition, we also analyzed the relationship between socioeconomic factors and vaccination rates to account for this third factor. In doing so, our study surpassed analyzing just SEP, but also accounts for 11 individual socioeconomic variables in addition to providing a population analysis through agent-based modeling.

We hypothesized that relating the COVID-19 infection rate to economic factors would reveal a negative relationship between the two, because an increase in socioeconomic position increases access to hygiene resources and less congested living. Meanwhile, we hypothesized relating vaccination rates to economic factors would show a positive relationship between the two, as an increase in socioeconomic position increases transportation access to vaccination centers and the concentration of vaccination centers. Our results revealed that socioeconomic variables such as gross domestic product (GDP) and net taxable assessed value caused a difference in simulated and actual

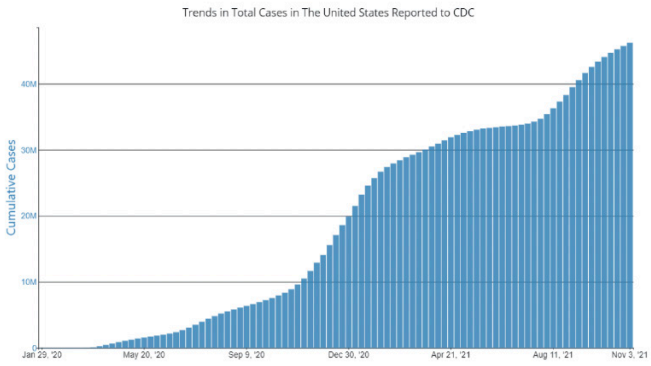

Figure 1: **Trends in the total COVID-19 cases in the United States reported to the CDC.** Cumulative total cases of COVID-19 from January 23, 2020, to November 3, 2021 as reported to the

disease; however, not all socioeconomic variables (SEVs) demonstrate significant relationships between vaccinations and disease spread.

RESULTS

Part 1: Agent-Based Modeling

We first wanted to gain an initial understanding of the spread of disease and how socioeconomic variables (among others) affect disease. As the modeling assumed that all people interacted as normal (no social distancing or quarantining) and people recovered randomly based on probability of recovery per day, our hypothesis was that simulated disease will follow a much more dramatic epicurve than that of the actual pandemic (i.e., it will peak within only a few days), and recovery would occur much earlier. We also hypothesized that the more initial infected cases there are, the more final cases there will be, due to the fact that there are more potential cases that can be infected by disease.

In each iteration, or step in a virtual 2D grid (agents move to the top, bottom, left, or right box), of the code, diseased and healthy computer agents interact in a torus map. By adjusting the amount of initial infected agents with values of recovery and disease, we simulated COVID- 19 in a probabilistic scenario. Each simulation only included the spread of disease and a constant recovery rate. Other factors, such as climate, vaccination, and political position, would be present in the real world—by not accounting for these factors, we created a baseline for analysis.

The primary value for comparison was the cumulative infection rate. We took note of a few key trends in the CDC data and our modeling data **(Figures 1 and 2)** (7). The first was that the curves shown by the agent-based modeling were extremely similar to those of point source epidemics (8), yet we know that due to its person-to-person nature, COVID-19 is a propagated epidemic (with elements of common source) (9, 10). The second was that the simulated pandemic ended much faster than the real-world one did, with most trials ending before the 100- day mark. This is in opposition to the real-world pandemic, which had a rising rate of infection (an average slope of 71,115 people per day) (7).

From these observations, we accepted the hypothesis that the simulated epidemic follows a much more dramatic curve than real-world COVID-19 and that recovery occurs quicker. However, we rejected the hypothesis that the more infected cases there are, the more final cases there are and the longer the epidemic lasts—this is because all simulations ended quickly (before 100 days passed), all with 0 final infected cases.

By making the key observations comparing real world and simulated disease, we were able to deduce that a significant factor in the real world (hypothesized by us to be SEVs) causes such a difference.

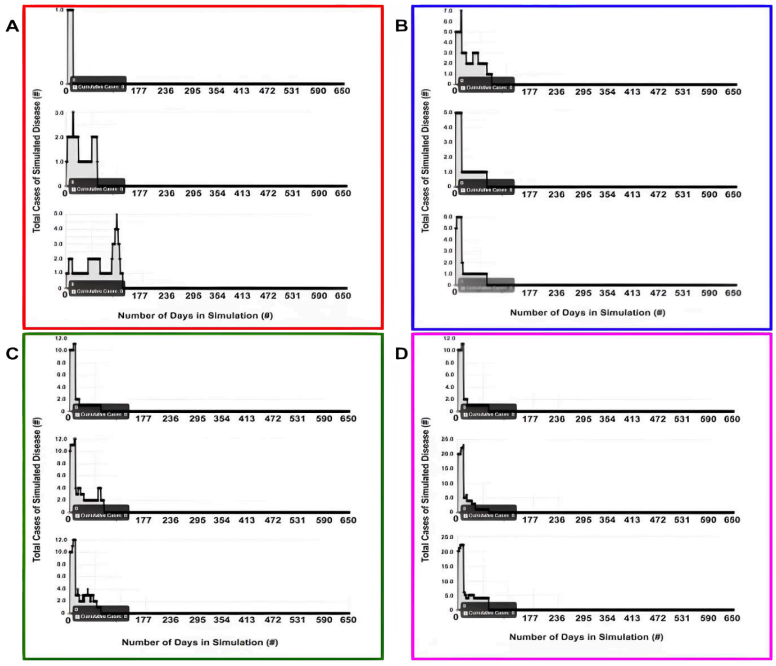

Figure 2: Gini coefficients for three trials of COVID-19 agent-based modeling. The Gini coefficient over each step of the code, reaching 650 steps. Values were graphed by conducting a Mesa library agent-based modeling to represent disease spread. A) 1 index case, B) 5 index

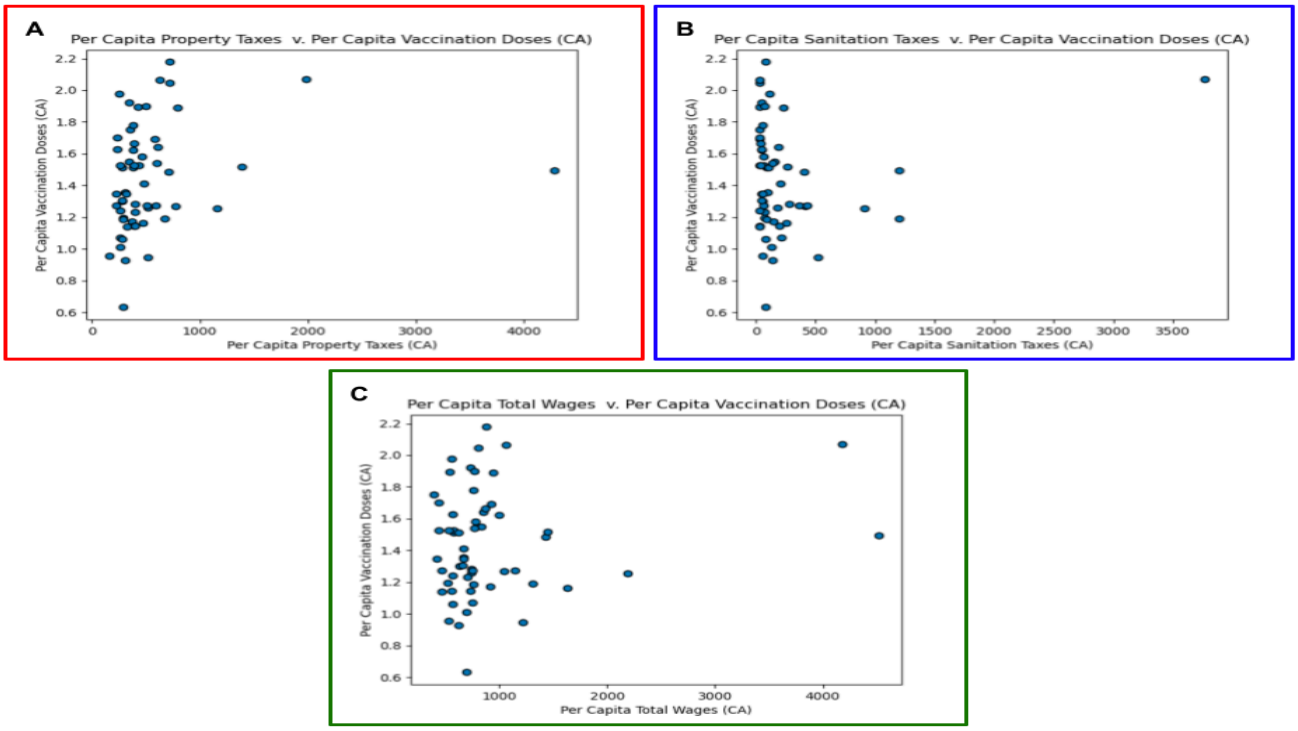

Figure 3. Insignificant socioeconomic factors versus vaccination rates in California. Vaccine rates were compared to A) average property taxes (r = 0.197, p = 0.138), B) sanitation investment (r = 0.134, p = 0.316), and C) total wages (r = 0.154, p = 0.246). Data was

Part 2: Infection Rate-Socioeconomics Analysis

We next evaluated the extent to which SEVs affected COVID-19 spread. This experiment was an analysis on the severity of COVID-19 and economic factors, in order to determine the relationship between SEP and disease prevalence. We ran a Pearson statistical analysis through sci.py on data derived from state (California) and national datasets for three periods of the pandemic (2/1/2020 to 10/1/2020, 10/2/2020 to 5/4/2021, and 5/5/2021 to 11/3/2021), determining the proportionality factor of each region in relation to the socioeconomic factors of the area (11, 7). Here, we hypothesized that there is a negative correlation between infection rates and socioeconomic variables.

For each period, as output, we generated a table with three columns: the socioeconomic factors, r-values ($1 \leq r \leq -1$), and p-values. The r-value determined the correlation between disease spread (with positive numbers indicating a positive correlation and negative numbers indicating a negative correlation) and the SEV, and a $p \leq 0.05$ indicated that said correlation is statistically significant.

We observed correlations between various SEVs and COVID-19 infection rate. For the first period, GDP was the only significant factor (r = -0.2541, p = 0.04). For the second period, significant SEVs included public assistance

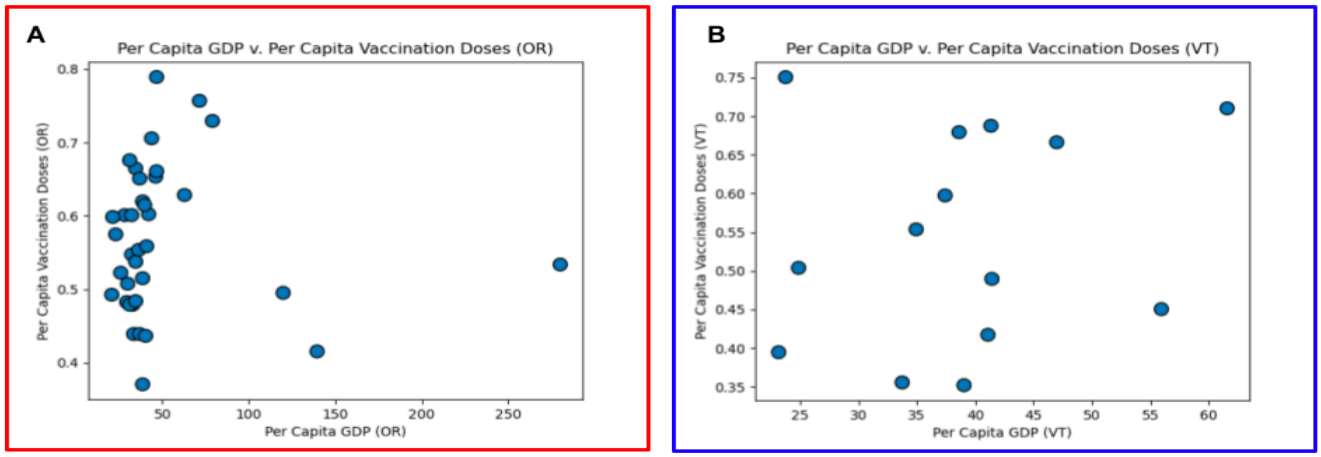

Figure 4. GDP versus vaccination rates in Oregon and Vermont. Per-capita rates of vaccination compared to GDP rates of **A)** Oregon with the median population of the US (r = -0.048, p = 0.781) and **B)** Vermont with the lowest population of the US (r = 0.183, p = 0.531). Data was analyzed with Pearson statistic functions.

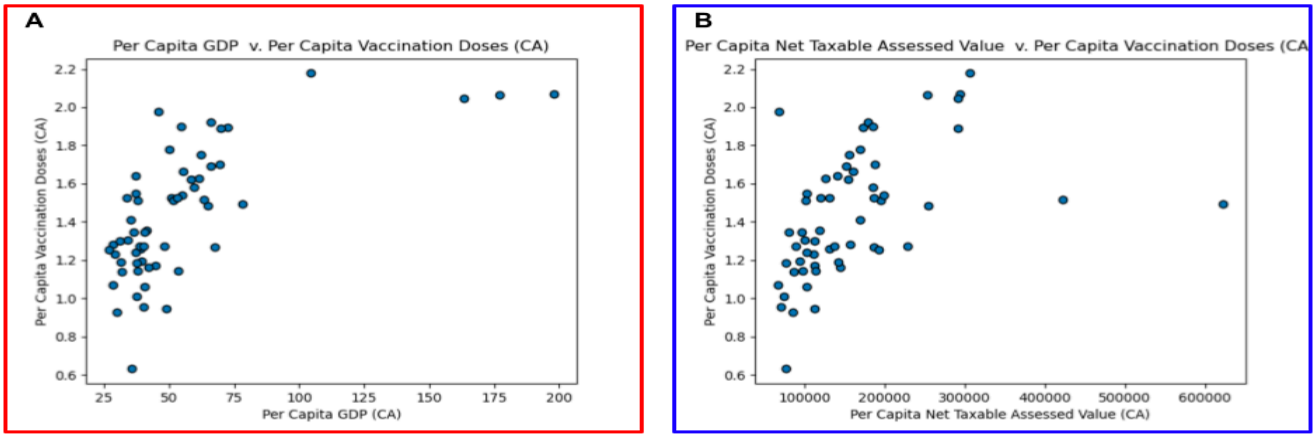

Figure 5. Significant socioeconomic factors versus vaccination rates. Per-capita rates of vaccination compared to **A)** GDP (r = 0.664, p = 1.302E-08) and **B)** average net taxable assessed value (r = 0.472, p = 0.00018). These images demonstrate a significant linear relationship; however, the Pearson statistic simply proves that there is one, even if the relationship is nonlinear. Therefore, the Pearson statistic may not

(r = -0.2743, p = 0.04) and population (r = 0.3396, p = 0.01). Finally, in the third period, general government spending (r = -0.2804, p = 0.04), health spending (r = -0.2762, p = 0.04), and migration (r = 0.4024, p = 0.00) were significant factors.

Using the results above, we found that the SEVs above did have a negative correlation with disease prevalence, supporting our hypothesis.

Part 3: Vaccination Ratio to Socioeconomic Data

This experiment was an analysis on the rates of vaccination and economic factors, in order to determine the relationship between socioeconomic variables and vaccination rates. We also ran a Pearson statistical analysis through sci.py on data derived from state (California) and national datasets; however, this analysis fitted a larger scope, analyzing and comparing data from California, Oregon, and Vermont (11, 12). Unlike the infection rate-socioeconomic analysis, the vaccination-socioeconomic analysis used data from the start of the vaccination process in the three states, as to get a holistic view of the correlation. Here, we hypothesized that there would be a positive relationship between SEVs and vaccination rates.

We observed that the relationship between sanitation, property tax, and total wages to rates of vaccination was insignificant and revealed a lack of correlation between these values (Figure 3). This was also supported by the r-values for the relationship between vaccination rates and these economic parameters: 0.197, 0.134, and 0.155 for property taxes, sanitation, and total wages, respectively.

We also noticed that other than California, GDP correlation between vaccination was not high. In Oregon, the analysis was shown to be insignificant (p = 0.781) (Figure 4); meanwhile in Vermont, the value was once again insignificant (p = 0.531) (Figure 4). However, California demonstrated significant relationships in GDP (r = 0.664, p < 0.001) and net taxable assessed value (r = 0.473, p = 0.00017) with vaccination rates (Figure 5). Therefore, we can accept our hypothesis for California's GDP and net taxable assessed value correlation with vaccination.

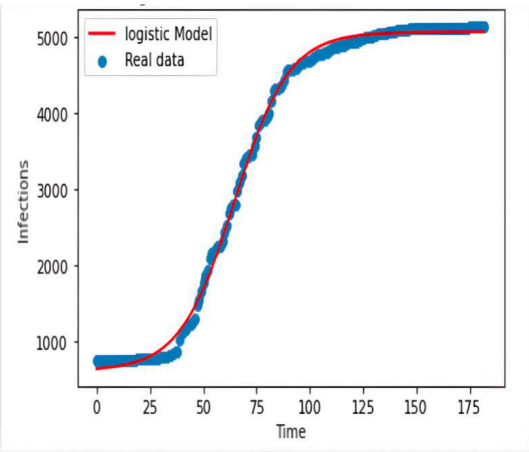

Figure 6. Logistic model representing COVID-19 infections in Lassen, California during the first period. Infections over time from 2/1/2020 to 10/1/2020, overlapping logistic model and real data. Period data over three spikes of COVID-19 were used to create a model for real data (blue), and then Pearson statistic functions in sci.py were used to determine a logistic model (red) for it. The unit

DISCUSSION

The agent-based modeling provided us with a few key observations regarding simulated and real-world disease (13). The first was that the simulated epidemic follows a pattern similar to a point source epidemic while the COVID-19 pandemic follows a propagated spread (9, 10). The second was that recovery for simulated disease occurs much faster, as our simulated diseases reached constant values within 100 days, while the real-world pandemic still has a rising case rate after 650 days (7). Finally, we understood that the initial case number does not cause a significant change in the graphs, and the final infection rate was the same for all trials. This led us to conclude that a factor in the real world causes such a difference. We hypothesized, and in our study proved to some extent, that these factors are socioeconomic variables. The agent-based modeling allowed for a baseline to be set on the extent of the impact of outside factors on disease spread.

However, this was not a perfect analysis, due to the fact that the agent-based modeling did not isolate SEVs solely. This is because of the inherent complexity of the real world—

to isolate SEVs would mean to account for vaccination, population density, climate change, etc.

For a definite conclusion to be drawn on the impact on SEVs solely through an agent-based modeling, we would need to simulate the status quo more effectively. Another caveat is that no matter how elaborate, models cannot simulate human behavior. For example, we cannot account for differences in infectious doses and therefore unique infection rates per person in our model yet.

In our infection rate-SEV analysis, we compared socioeconomic variables with a variety of economic factors in California's 55 counties. Our general data revealed that there is a slight negative correlation between SEVs and the spread of COVID-19. We speculate that this may be due to a poorer area having more crowded conditions and less hygienic living, increasing the spread of disease. We found that GDP, public assistance, population, government spending, health spending, single family housing units, and migration are all factors which had significant correlation with COVID-19 spread.

One key observation that we made is that the three different periods of the pandemic that we used had different factors that are significant, with little overlap. This implied that more data was required in order to obtain a more definite SEV that is correlated to COVID-19 spread. We saw this in our attempts to model k (proportionality factor) as a function of independent socioeconomic variables, which had been unsuccessful due to a lack of data points—each county provided only one data point for one period. It is also worth noting that although population was used as an SEV, population density which accounts for area is ultimately more accurate. Furthermore, for future research, to make our analysis more accurate, we want to use a nonlinear least squares regression model to show the spread of disease as a function of time, vaccination status, and socioeconomic factors, in order to get a more holistic analysis.

We would also like to find ways to determine if a nonlinear multivariable model is statistically significant, to try to solve the gradient explosion problem, which is when large error gradients accumulate and therefore data analysis is inaccurate. Furthermore, we would like to look into the susceptible, infected, and recovered (SER) model in epidemiology, which can help to provide us with more varied and accurate conclusions.

In our vaccination-SEV analysis, we took data from California, Oregon, and Vermont, to gain a holistic view on how relationships between socioeconomic variables and vaccination act in different populations. Those three states were specifically chosen, as California has the highest population, Oregon has the median population, and Vermont has the lowest population in the United States (14-16).

Our data revealed that the only area where GDP-vaccination correlation is significant is in California, most likely to numerous counties in the state (and therefore more abundant data), in addition to political positions in the state influencing rates of vaccination. In addition to that, net taxable assessed value showed significant correlation to vaccination. We speculated that this correlation is present because people in better situations (living in areas with higher GDP and more money to be taxed) do not have to make the choice between health and money.

It is worth noting, however, that raising the per capita GDP overall raises a variety of other socioeconomic factors as well (17). As a result, we could not assume that simply raising the GDP is a cure-all to problems related to vaccination—in reality, the issue is much more complex, and raising GDP is the broadest conclusion to it. More research needs to be done on which factors affected by the GDP causes a movement towards vaccination. In addition, Pearson's statistic requires independent data points; since we already know that GDP is related to a variety of economic factors, this study is limited in the conclusions that it can draw regarding other economic factors. Although it is likely that many of them do not have an impact on vaccination, definite conclusions cannot be drawn from them due to the nature of Pearson's statistics.

Furthermore, we wish to conduct regression and p-value analysis on more states, and control for external factors such as political position. This is because the usage of three states, despite being the maximum, median, and lowest population, is not totally representative of the COVID-19 epidemiology and socioeconomic factors of all US states. Differences in race, gender, and even political position can be significant factors when considering vaccine-SEV correlation, which we did not account for in this study. We also want to extend our reach to other countries and use a python library for spatial autocorrelation, or the presence of spatial variation in a variable compared to another value. Future research in this area could more effectively direct vaccination resources to populations who need it, while at the same time aiding these populations to become self-sustaining vaccination providers through helping their economy.

Overall, future research in this area would isolate variables that impact low-income communities the most, providing them with aid in specific areas instead of distributing resources ineffectively. Rather than trying to raise per-capita GDP for example, with this knowledge policymakers could try to focus on other factors that help mitigate the spread of disease. Also, since our data only includes counties of California, Vermont, and Oregon, it is not representative of the epidemiology of the whole United States.

This study, in its analysis of individual SEVs in addition to its usage of an agent-based modeling to serve as a basis of analysis, demonstrated different economic factors related to both infection rate and disease. By further studying and iterating upon the results obtained, we can gain a set of socioeconomic factors key to the livelihoods of people during pandemics, an use those to influence future policymaking regarding disease.

MATERIALS AND METHODS

All analysis performed in this section was done with the Pearson statistic function from the sci.py library, and all code is available at github.com/Mouse05/relatingsepwithcovid19prevalence, with the part 2 code being available at github.com/YYCCaa/relatingsepwithcovid19prevalence. The Pearson statistic measures the linear relationship between two sets of data (18). We note that due to the linear relationship outlined, the Pearson statistic can provide us with proof of a positive/negative relationship, but not the complexity of it. Our p-value threshold was chosen to be the standard quantity of 0.05 (19).

Part 1: Agent-Based Modeling

We analyzed three periods of data in Part 2 of our study, with the final period ending on 11/3/2021. Therefore, we can take the CDC's graph from the start of the pandemic (01/23/2020)
to 11/3/2021, which is a total of 650 days, as a basis for comparison (Figure 1). From here, w can build the metrics for our agent-based modeling (13).

The agent-based modeling consisted of a map of 100 agents in a never-ending torus map, represented by a 30 by 30-unit square. Each of the agents had the property of having or not having the disease, which was transferred between them randomly when they were in the same square. In each "step" of the code, agents randomly moved into an adjacent square, with 29.59% chance of obtaining the disease if a diseased agent is also present. This was found with the ratio of the reproductive number of COVID-19 (3.38:1), giving 1/3.38, or 28.58% (20). The reasoning behind this is that there is a 1/3.38 chance that the uninfected agent is one of the
3.38 the infected agents cause disease in. In addition to that, in each step, diseased agents get cured after 14, 42, or 73 steps of the code. This was done by taking a case-by-case analysis of recovery. In most cases, recovery rate is 14 days (21). However, in severe cases (shown to be
0.091% of patients (7)), recovery rates are up to six weeks (42 days) (21, 22). Finally, in long COVID patients, recovery rates are on average 73 days, with 14.2% of victims having this (23, 24). The program plotted the total number of cases.

We iterated through 4 sets of simulations, with 3 trials in each, changing the amount of index cases to 1, 5, 10, and 20 for 650 days. This allowed us to consider the spread of the cases in an environment with vaccination and socioeconomic factors, creating an effective benchmark to compare to real-world infection rate.

Part 1: Agent-Based Modeling

We analyzed three different spikes of COVID-19 in California, depending on the infection rate throughout the pandemic. To obtain this data, we used the California government open dataset about the state; the first spreadsheet had data regarding COVID-19 cases, deaths, number of years, and population size from February 1, 2020, to mid-October 2021 (7). We also modeled the data present in these spikes in a logistic growth function, to determine the proportionality
factor from the graph; the proportionality factor represents the speed of COVID-19 spread in the population.

A logistical growth model was used to model the infection rate of COVID-19 in specific regions, and samples demonstrated an almost perfect logistical growth of infections over time (Figure 3). However, there were several spikes of COVID-19 cases, and we had to divide the entire time span (from February 2020 to November 2021) into three periods. The first period was from 2/1/2020 to 10/1/2020; the second period was from 10/2/2020 to 5/4/2021; the third
period was from 5/5/2021 to 11/3/2021. This made our time versus COVID-19 analysis clear and provides accurate data on different periods.

As aforementioned, we modeled the infection rate of COVID-19 as a logistical growth
mode (Figure 6). The process is listed below.

$$\frac{dP}{dt} = kP(1 - \frac{p}{L}) \quad \text{[Eqn 1]}$$

We knew that the differential equation of the logistical growth model is represented as shown:
Here, *t* was the time since the start of a new wave of COVID-19 in a specific region, *P* was the number of infected people in a particular wave of COVID-19 in a specific region (separate from cumulative infection rate in a specific region), *k* was the

$$\frac{dP}{dt},$$

proportionality factor, and *L* was the carrying capacity—the maximum number of people a particular wave of COVID-19 can infect in a specific region, i.e., the number of susceptible populations before any infection.
was the instantaneous rate of change of the infected population, which might have been a good metric of severity,

$$P = \frac{L}{1+ce^{-kt}} \quad \text{[Eqn 2]}$$

but it contained an independent variable and was not normalized with respect to the susceptible population.
The general solution was where c shifted the graph horizontally and had no impact on severity.

However, some regions had a larger population, and thus the infection rate curve had a much steeper slope (infected people per day), although the pandemic situation might not have been perceived as severe as the regions with much

$$pL = \frac{L}{1+ce^{-kt}} \Rightarrow p = \frac{1}{1+ce^{-kt}}$$

lower population. Therefore, it was important to normalize the metric we use to quantify the speed of infection.

$$t = \frac{ln(\frac{1-p}{p}) - ln(c)}{-k} \quad \text{[Eqn 3]}$$

From (2) we observed that if we solve for t when the infection rate was a percentage of the carrying capacity,
and the solution did not depend on *L*. Solving this equation,

$$t = \frac{ln(\frac{1-p_1}{p_1}) - ln(c)}{-k} - \frac{ln(\frac{1-p_2}{p_2}) - ln(c)}{-k} = \frac{ln(\frac{1-p_1}{p_1}) - ln(\frac{1-p_2}{p_2})}{-k} \quad \text{[Eqn 4]}$$

we got:
Representing *p* as a percentage of carrying capacity, to find the time from *p1* to *p2*, we obtained the following equation

Since $1 > p_1 > p_2$, $1\text{-}p_1 < 1\text{-}p_2$, $ln(\frac{1-p_1}{p_1}) - ln(\frac{1-p_2}{p_2}) < 0$,

from (3):

Therefore, *c* should not have been considered in this analysis. Only *k*, the proportionality factor, mattered.

$$\frac{dP}{dt}$$

if we set *p2* close to 0 and *p1* close to 1, we could get the time it takes for the specific wave of virus to infect the maximum possible number of people. This *s* was proportional to 1/*k*, so

the speed was proportional to *k*. Also, note from the equation (1) that ceteris paribus,
was proportional to *k*. Therefore, *k* was a good metric to measure how fast COVID-19 spreads, and we called it "the proportionality factor".

We used *k* (the proportionality factor) to measure the infection rate of COVID-19. We then found the correlation between *k* and socioeconomic factors. In this case, each county for each period represented a data point, and each Pearson statistical analysis was performed on all California counties during a particular period/wave. These data points were considered as separate and independent, as we assumed a closed system for each county when we modeled the spread of COVID-19. The r-squared-value of the model was quite high (mostly 99% or higher according to data in period 1q.csv, period 2q.csv, period 3q.csv), which proved the closed-system nature of the spread.

For our socioeconomic data, we extracted the following factors from California's aforementioned database, using the second dataset which contained GDP, net taxable assessed value, property taxes, roads, facilities, transportation, and total wages from 2018 (11). We replaced the value for GDP with a more recent and therefore representative value for the COVID-19 pandemic, as GDP represents the net worth of each county in the state. Then, we divided said data by the population count for each county to determine the per capita value for each economic factor.

From there, we used SciPy's Pearson statistic function to find the correlation between the economic data and the infection rate of COVID-19. We first created a list of the data from the socioeconomic factors and infection rate. These lists allowed us to find the correlation between each value using SciPy's Pearson statistic function, giving us the r and p-values for our code. We used data from all 55 counties to understand the relation between socioeconomic data and the infection rate of disease for all 3 spikes of disease.

Part 3: Vaccination-Socioeconomics Analysis

We used the California government open portal to extract data specific to California's net taxable assessed value, total property taxes, transportation, and total wages (11). However, we also utilized the Bureau of Economic Data to get a more recent value for the GDP of California, allowing us to more accurately represent the relationship between socioeconomic data and vaccination (25). In addition to this, in this analysis, we also compared vaccination to GDP in Oregon and Vermont, drawing the data from the Bureau of Economic Data (25). We then divided the data by the total population count, which is from World Population Review's data on state-divided county population.

We derived our data on vaccination from the CDC's and California's vaccination database (12, 26). This provided us with vaccination data divided by state and by county, allowing us to perform an accurate correlation.

From there, we used SciPy's Pearson statistic function again, allowing us to derive the rand p-values for each type of economic data in California in addition to GDP and vaccination in Oregon and Vermont, in which the r-value represents the correlation between disease severity and a particular socioeconomic factor, and the p-value represents how confident we are that the result is true (anything below 0.05 is accepted). In addition to that, we used the Matplotlib library to plot the data on each of California's socioeconomic factors in comparison to vaccination rates, creating a scatter plot we can conclude from.

Received: March 5, 2022
Accepted: December 6, 2022
Published: July 25, 2023

REFERENCES

1. "US: Address Impact of COVID-19 on Poor." Human Rights Watch, 19 Mar. 2020, www.hrw.org/news/2020/03/19/us-address-impact-covid-19-poor.
2. Phung, Dung, et al. "The effects of socioecological factors on variation of communicable diseases: A multiple-disease study at the national scale of Vietnam." PLOS One, 1 Mar. 2018, doi.org/10.1371/journal.pone.0193246.
3. De Glanville, W. A., et al. "Household socio-economic position and individual infectious disease risk in rural Kenya." Scientific Reports, 27 Feb. 2019. doi:10.17638/datacat.liverpool.ac.uk/447.
4. Foster, Hamish M.E. et al. "The Association Between a Lifestyle Score, Socioeconomic Status, and COVID-19 Outcomes Within the UK Biobank Cohort." BMC Infectious Diseases, vol. 22, no. 273, 30 Mar. 2022. doi:doi.org/10.1186/s12879-022-07132-9.
5. Patel, J.A. et al. "Poverty, Inequality, and COVID-19: The Forgotten Variable." Public Health, vol. 183, no. 110-111, 14 May 2020, doi:10.1016/j.puhe.2020.05.006.
6. Geranios, Karina et al. "Impact of COVID-19 and Socioeconomic Status on Delayed Care and Unemployment." Health Equity, vol. 6(1), no. 91-97, 2 Feb. 2022, doi:10.1089/heq.2021.0115.
7. "COVID-19 Probable Cases." California Open Data Portal, data.ca.gov/dataset/covid-19-probable-cases/resource/ad565d59-27d2-4638-884d-e8c2d8bdea0c. Accessed 3 November 2022.
8. "Using an Epi Curve to Determine Mode of Spread." Centers for Disease Control and Prevention, www.cdc.gov/training/QuickLearns/epimode/4.html. Accessed 22 November 2022.
9. Chen, Ting, et al. "Epidemic Characteristics of the COVID-19 Outbreak in Tianjin, a Well- Developed City in China.", AM J Infect Control, vol. 48(9), no. 1068-1073, 12 Jun. 2020, doi:10.1016/j.ajic.2020.06.006.
10. Kolozsvári, László R., et al. "Predicting the Epidemic Curve of the Coronavirus (SARSCoV-2) Disease (COVID-19) Using Artificial Intelligence." MedRxiv, 27 Jan. 2021, doi.org/10.1101/2020.04.17.20069666.
11. California Economic Datapile. California State Association of Counties, 16 Dec. 2019, www.counties.org/post/datapile.
12. "COVID-19 Vaccinations in the United States, County." Centers for Disease Control and Prevention, data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States- County/8xkx-amqh/data. Accessed 11 January 2022.
13. Lifferth, Will. "Modeling Inequality with Mesa." Replit, replit.com/@WillLifferth/ modelinginequality- with-mesa. Accessed 22 November 2022.
14. "Census Data for Oregon." Population Research Center, Portland State University,
15. www.pdx.edu/population-research/census-data-oregon.

Accessed 11 January 2022.
16. "Population of Counties in California (2021)." World Population Review, circa 2021, worldpopulationreview.com/us-counties/states/ca.
17. "Population of Counties in Vermont (2021)." World Population Review, circa 2021, worldpopulationreview.com/us-counties/states/vt.
18. Bondarenko, Peter. "Gross Domestic Product." Encyclopedia Britannica, 28 Feb. 2017, www.britannica.com/topic/gross-domestic-product. Accessed 25 November 2022.
19. "Scipy.stats.pearsonr." SciPy Documentation, The SciPy Community, docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html. Accessed 25 November 2022.
20. Leo, Giovanni Di and Francesco Sardanelli. "Statistical Significance: p Value, 0.05 Threshold, and Applications to Radiomics - Reasons for a Conservative Approach." European Radiology Experimental, vol. 4, no. 18, 11 Mar. 2020, doi.org/10.1186/s41747-020-0145-y.
21. Alimohamadi, Yousef et al. "Estimate of the Basic Reproduction Number for COVID-19: A Systematic Review and Meta-analysis." Journal of Preventive Medicine and Public Health, vol. 53(3), no. 151-157, 20 Mar 2020, doi:10.3961/jpmph.20.076.
22. "Coronavirus Recovery." WebMD, Internet Brands, www.webmd.com/lung/covidrecovery- overview#1. Accessed 11 January 2022.
23. "Coronavirus: Recovery After a Hospital Stay." Johns Hopkins Medicine, The Johns Hopkins Hospital, www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-recovery-after-a-hospital-stay. Accessed 22 November 2022.
24. Venkatesan, Priya. "European Congress of Clinical Microbiology & Infectious Diseases 2022." Lancet Respir Med., vol. 10(6), no. 56-57, 17 May 2022, doi:10.1016/S22132600(22)00183-7.
25. "Long COVID." Centers for Disease Control and Prevention, US Department of Health and Human Services, www.cdc.gov/nchs/covid19/pulse/long-covid.htm.
26. "GDP by County, Metro, and Other Areas." Bureau of Economic Analysis, U.S. Department of Commerce, www.bea.gov/data/gdp/gdp-county-metro-and-other-areas. Accessed 11 January 2022.
27. "Statewide COVID-19 Vaccines Administered by County." California Open Data Portal, data.ca.gov/dataset/covid-19-vaccine-progress-dashboard data/resource/c020ef6b-2116-4775-b11d-9df2875096ab. Accessed 23 November 2022.
28. "Trends in Number of COVID-19 Cases and Deaths in the US Reported to CDC by State/Territory." Centers for Disease Control and Prevention, US Department of Health and Human Services, covid.cdc.gov/covid-data-tracker/#trends_totalcases_select_00. Accessed 25 November 2022.