A County-Based Mathematical Model for COVID: Incorporating Politically-Associated Vaccination Trends with United States COVID Dynamics in 2021

Chloe Comeau[1], Marcus Cruz[2], Nicholas Thompson[2], Tirthkumar Shah[2], Carlos Bustamante Orellana[2], and Research Supervisor: Paula Gonzalez Parra[3]

[1] *Vassar College*
[2] *Arizona State University*
[3] *Universidad Autónoma de Occidente*

July 2024

Abstract

COVID dynamics are influenced by vaccination rates. In the United States, a county's vaccination proportions are closely associated with that county's political leanings. Generally, Democratic counties tend to have higher vaccination proportions, and Republican counties tend to have lower vaccination proportions. We hypothesized that adjusting for variations in vaccination proportions between different types of counties would more accurately reflect COVID dynamics based on reported data. To test this, we classified counties into five categories, a_1 through a_5, using the percentage of Trump votes in a county during the 2020 presidential election as an indicator of that county's political leanings. a_1 represented counties with the lowest percentage of Trump votes ($<25\%$), and a_5 represented counties with the highest percentage of Trump votes ($>75\%$). Each category had a corresponding vaccination proportion. We created a county-based SEIAVR (Susceptible, Exposed, Infected, Asymptomatic, Vaccinated, and Recovered) model to represent the dynamics of each of those five types of counties between March 1st, 2021 and September 1st, 2021. We also hypothesized that daily movement would influence COVID dynamics, since people interact with individuals from other counties with different vaccination rates. We therefore proposed an SEIAVR model with Lagrangian mobility to simulate inter-county mobility. Our initial results indicate that a model incorporating politically-associated vaccination proportions accurately represents the dynamics of COVID based on existing cases and deaths data. This approach highlights the importance of considering political factors in epidemiological modeling.

1 Introduction

The COVID pandemic quickly spread worldwide in 2020 and still persists today. Symptoms of the first strain of COVID include cough, fever, fatigue, shortness of breath, and loss of taste or smell (Çalıca Utku et al., 2020). The disease is spread by infected particles emitted from an infected person's mouth or nose. These particles can vary in size from larger droplets to smaller aerosol particles. The best mitigation strategies are vaccination, masking, and social distancing (World Health Organization, 2024). The World Health Organization estimates that COVID may have caused upwards of 3 million excess deaths in 2020 alone (World Health Organization, 2021).

Various models have been formulated to model COVID, but none have been particularly successful in terms of data fitting or future predictions (Colonna et al., 2022).

This paper will begin with the literature review, which will cover relevant terminology, the relationships between politics and COVID, and existing COVID models. This will be followed by a methodology section outlining how we constructed our models. The models' results and implications will be addressed in the discussion. Future directions and limitations will be covered in the analysis section. Our findings will be summarized in the conclusion.

2 Literature Review

This literature review explores the relationship between political leaning and COVID trends in terms of cases, deaths, and mitigation strategies. It begins with an overview of relevant terminology in American politics. This is followed by an overview of existing research on how demographics affect COVID trends and shape Americans' willingness to adopt of mitigation strategies. Specific analyses of various COVID mathematical models will follow. Finally, the literature review will conclude with the gap analysis, which highlights the fact that no known research combines the qualitative political differences with a quantitative COVID model.

2.1 Relevant Terminology

The American political system is currently defined by its two main political parties, the Democratic party and the Republican party. Democrats tend to be more liberal, favoring larger government and higher taxes in exchange for government-funded social welfare programs. On social issues, Democrats favor greater freedoms and less government control. Republicans tend to be more conservative, prioritizing smaller government and lower taxation. They generally support military expenditure. With respect to social liberties, Republicans

tend to have more traditional, conservative views (The Editors of Encyclopaedia Britannica, 2022). The color most associated with the Democratic party is blue, and the color tied to the Republican party is red. For the purposes of this paper, terms including Democratic, liberal, and blue or, conversely, Republican, conservative, and red may be used interchangeably to classify the respective general political leanings of a subpopulation.

2.2 Politics and COVID Mitigation Strategies

In early 2020, near the beginning of the COVID pandemic, Desmet and Wacziarg observed a pattern where (largely due to population density) Democratic areas tended to have higher COVID mortality cases than predominantly Republican areas. This correlated with higher approval rates for lockdown and control measures in blue areas, whereas in red areas, these measures were deemed less necessary. There was also more support for Trump in those red, low death areas than in the high death, mainly blue areas. However, Desmet and Wacziarg also noted that once they added control measures to account for minority populations (Latino, Black, and other racial minorities in particular), regions with high support for Trump tended to be more heavily affected by COVID. The authors argued that "these results may help explain the observed political fault lines over the desirability of lockdown policies, with Republican-leaning locations seemingly much more eager to reopen early and suspend the lockdowns as compared to Democratic-leaning locations" (Desmet & Wacziarg, 2020, p. 14).

2.2.1 Vaccination

Once COVID vaccinations became widely available in the U.S., the correlation between political leaning and COVID cases began to shift. Albrecht's research attempted to explain differences in political views in terms of "race/ethnicity, educational attainment, and poverty" (Albrecht, 2022, p. 3). He compared data on COVID cases and deaths per 100,000 residents per county from 3112 counties from March 1, 2021 to September 1, 2021, a time frame in which vaccines were widely available. Data on dependent variables in the study were resourced from New York Times information on cumulative cases and deaths. Albrecht used values published in September 2020, March 2021, and September 2021. Albrecht noted that after "March 1, 2021 when vaccines were readily available," vaccination rates became less dependent on supply disparities and more closely associated with political leanings. He observed that "COVID19 deaths increased by 26.1 per 100,000 residents in counties where Trump received less than 25% of the votes, while the rate of increase was more than twice as great (54.8 per 100,000 residents) in counties where Trump received more than 75% of the vote" (Albrecht, 2022, p. 7). Moreover, "as vaccination rates increased, COVID-19 cases and deaths per 100,000 tended to decline. Most significantly, not only were political views strongly related to vaccination rates, but they also had important implications for COVID-19 cases and deaths. In Trump leaning counties, COVID19 cases and deaths were more extensive than in counties

where Trump received a lower percent of the vote" (Albrecht, 2022, p. 9).

Using statistics from the CDC's COVID data tracker, Brown et al. analyzed counties based on "the U.S. COVID-19 Community Vulnerability Index" which focuses on the following: "socioeconomic status (SES); minority status and language; housing type, transportation, household composition, and disability; epidemiological factors; healthcare system factors; high risk environments; and population density" (Brown et al., 2021, p. 4246). Generally, they concluded that "housing type, transportation, household composition, and disability" were associated with the greatest disparity in vaccination rates (Brown et al., 2021, p. 4245).

Alemi and Lee used vaccination rate data from 3109 counties with available data in the U.S. as of April 2022. They analyzed 36 possible explanations for hesitancy towards vaccination, "including demographic, social, economic, environmental, and medical variables known to affect vaccine hesitancy" (Alemi & Lee, 2023, p. 1). They found "a statistically significant relationship between the percentage of Republican supporters and rates of vaccine hesitancy" in which higher levels of Republican support were associated with lower vaccination rates (Alemi & Lee, 2023, p. 1). Moreover, they found that political leaning was not only closely correlated with vaccination rates, but it also functioned as "a mediator of the effects of other variables" (Alemi & Lee, 2023, p. 5).

Monrad et al. used a least-squares regression model to compare how early supply patterns and long-term demand factors affected vaccination rates during the first nine months of 2021. They concluded that both affected vaccination rates, but supply issues were mostly an issue during early 2021, and after that, other factors had greater impact. They discovered associations between COVID vaccination rates and other factors including "health expenditure, vaccine hesitancy, cost obstacles to care, Democratic voting, and elderly population share" (Monrad et al., 2022, p. 6528).

Sun and Monnat sourced data from a CDC report on vaccination rates for adults (aged 18 and older) as of August 11, 2020 from 2,869 counties. They then synthesized this data with "county-level data on demographic and socioeconomic composition, health care infrastructure, 2020 Trump vote share, and USDA labor market type" (Sun & Monnat, 2022, p. 916). Using regression models, they found that while 59.8 percent of adults in urban counties had been vaccinated fully, only 45.8 percent of those in rural counties had been fully vaccinated. Moreover, they found that average vaccination rates decreased as rurality increased. They explained this difference through "a combination of lower educational attainment and higher Trump vote share" and poorer health care infrastructure (Sun & Monnat, 2022, p. 916). They recommend that leaders in rural communities where the vaccine is accessible combat vaccine misinformation and promote vaccination (Sun & Monnat, 2022, p. 920).

Cole examines potential causes for vaccine hesitancy on both sides of the political spectrum. He argues that anti-vaccine beliefs on the right wing of the political spectrum are "often motivated by antipathy toward governmental authority and scientific expertise" (Cole, 2023, p. 286). They often view "vaccination requirements as government overreach, an infringement of individual civil liberties, bodily integrity, and parental rights" (Cole, 2023, p. 286). He also notes that there exist left-wing vaccine skepticists, but these tend to be overlooked in big data compilations because they tend to cluster in small "pockets of homogeneity" (Cole, 2023, p. 287). These concerns tend to be related more to the concerns regarding the overlooking of women's health issues in medication and vaccine production, whereas right-wing hesitancies lie more along the lines of reluctance to relinquish personal rights and freedoms.

Through these studies, it is clear that there is an association between political leaning and vaccination. And it is not just COVID vaccines that reflect these political differences. Rates of vaccination generally tend to be influenced by political ideology. Through an online survey, Rabinowitz et al. asked participants to rate on a scale from one to four how strongly they agreed or disagreed with twenty statements, ten of which were facts, such as "The earth revolves around the sun," the other ten of which were beliefs like "Sleeping with the windows open is good for you" (Rabinowitz et al., 2016, p. 6). Next, the participants were assigned twenty statements in random order, half of them anti-vaccination, half pro-vaccination. Respondents were then asked to rate how much they agreed with each statement, how strongly they believed the average American would agree with each statement, how strongly they believed the average conservative would agree with each statement, how strongly they believed the average liberal would agree with each statement, and if they regarded each statement as true or false. The survey concluded by asking if the participants had children and if they fully vaccinated their children before the age of two. They found that "liberals were significantly more likely to endorse pro-vaccination statements and to regard them as 'facts' (rather than 'beliefs'), in comparison with moderates and conservatives" (Rabinowitz et al., 2016, p. 1). Additionally, conservatives were more likely to overestimate the number of people who shared their opposition to vaccines, whereas liberals were more likely to underestimate the number of people who shared their support for vaccines.

Overall, political association is clearly correlated with attitudes toward vaccinations. The following subsection will review similar associations between political affiliation and masking.

2.2.2 Masking

In addition to vaccination, another mitigation measure for COVID was masking. Masking levels also tended to be influenced by political orientation.

Cunningham and Nite used public health data and self-reported masking rates

from a NYT poll by county. Counties were measured in terms of rates of mask wearing, health behaviors, clinical care, socioeconomic factors, and physical environment. They found no association between socioeconomic factors or access to clinical care and mask wearing. They did observe that "people who lived in counties marked by pollution, severe housing, and long commutes likely to be taken alone, are all more likely to wear a mask than their colleagues" (Cunningham & Nite, 2021, p. 6). Cunningham and Nite inferred that high population density and pollution led to "people in these counties [being] acutely aware of the need for and value of wearing a mask" (Cunningham & Nite, 2021, p. 6).

Young et al. used data from a national survey asking American adults to report their mask wearing behavior. Demographic data was weighted to reflect the overall U.S. population. They found that "mask-wearing was negatively associated with Republican party membership, conservatism (in the 50+ sample), Trump favorability, and positively associated with Biden favorability" (Young et al., 2022, p. 6).

Haischer et al. visited 36 retail locations in five counties in Wisconsin. Stores (grocery or large retail chains) were categorized as being in urban, suburban, or rural areas. Shoppers were observed over the course of several days in June, July, and August by the researchers, and their masking status (wearing, wearing incorrectly, or not wearing), gender, age, and location were recorded. The authors then performed a multiple logistic regression analysis. Haischer et al. found that roughly "41% of the June sample wore a mask," however, the proportion rose to roughly 90% in July and August as mask mandates were implemented. (Haischer et al., 2020, p. 1). Additionally, the probability of a shopper masking was "1.5x greater for females than males" four times greater in urban than rural areas (Haischer et al., 2020, p. 1).

In addition to the correlation between party allegiance and masking measures, there is also an association between masking behavior and the perception of others and their political leanings. Carey et al. conducted a multi-wave panel study in summer 2020 with 2982 participants. Participants were randomly assigned to one of four groups: a control group, the American norms treatment condition, the Democratic norms treatment condition, and the Republican norms treatment condition. Participants in the American norms treatment condition were informed of "the percentage of Americans (74%) who reported wearing masks 'all of the time' or 'most of the time'"(Carey et al., 2023, p. 380). Participants in the Democratic norms treatment were provided with "the figure for self identified Democrats and Democratic-leaning independents (89%)", and the Republican norms treatment were given "the figure for self-identified Republicans and Republican-leaning independents (56%)" who mask (Carey et al., 2023, p. 380). After the random treatment condition phase, participants were asked about their personal willingness to wear masks, mask effectiveness, and their attitudes toward Democratic and Republican voters. Using linear regres-

sion, they concluded that "learning that a majority of Americans report wearing masks regularly increases mask-wearing intentions and perceived effectiveness, though the effects of this information are not distinguishable from other treatments" (Carey et al., 2023, p. 377).

Additionally, it seems that Americans are more likely to be swayed in their mitigation strategies if a trusted political elite recommends a mitigation measure to them. In Pink et al.'s study, 1480 self-identified Republicans were randomly assigned to one of three groups: Republicans endorse, Democrats endorse, and control. The first group watched a speech in which prominent Republicans (among them Donald Trump) promoted the COVID vaccine. The second watched videos of prominent Democrats (including Joseph Biden) endorsed the vaccine. The third watched nothing. Overall, they found that endorsements from Republican elites had positive effects and Democratic elites' endorsements had negative effects on participants' attitudes toward the vaccine and their willingness to recommend vaccination to others (Pink et al., 2021, p. 1).

2.3 Analysis of Existing COVID Models

This literature review will examine other existing COVID models. Our model was developed with consideration for the strengths and weaknesses of these models in helping to answer our research question.

Among the various inputs for our model, one of the most useful was Colonna et al.'s 2022 retrospective assessment of COVID-19 model performance in the USA. They assessed 23 models, all of which were chosen based on inclusion criteria based on sufficient data reporting. Models were assessed based on a cumulative score from multiple assessments, including a calibration score of precision and accuracy for distributions of predicted versus observed values and an informativeness score based on the width of confidence interval versus overall range of observations. The baseline model outperformed 18/22 other models (Colonna et al., 2022).

In late 2021, Jayatilaka et al. compared the SIR Kermack-McKendrick, SIRmp, SEIR, and SEIRpqr models and discussed how 1918 pandemic models could have potential for COVID modeling.

The SIR Kermack-McKendrick model assumed a fixed population size, no incubation period, no social distancing, and a natural immunity parameter based on vaccination rates. The SIRmp model included interaction with public health measures, assumed infection rate and reproductive number varied with time, and incorporated a time mixing factor to reflect social distancing practices in different time periods. The SEIR model assumed an exposed population with an incubation rate parameter. It also referenced the Lambert W function and assumed each person had an average number of contacts. The SEIRpqr model used m, p, q, and r as mixing coefficients based on social distancing practices

where p, q, and r represented scale incubation, infection, and removal rates.

Jayatilaka et al. found that parameters had to be fit for multiple time windows within each wave of COVID in Canada. They concluded that level of adherence to prevention methods was a limiting factor in accuracy of SIR and SEIR models. Mixing coefficients had strong influence on disease spread projections. Generally, all models failed to account for unidentified asymptomatic cases (Jayatilaka et al., 2022).

In August 2020, Mollalo et al. constructed a GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. They collected and grouped data based on socioeconomic, behavioral, environmental, topographic, and demographic factors. The authors generated and compared SLM (spatial lag), SEM (spatial error), GWR (geographically weighted regression), and MGWR (multiscale geographically weighted regression).

The SLM model featured a spatial lag parameter and a spacial weight vector of weight matrix that considers neighboring counties. The SEM model had two error terms, one spatially-dependent and the other not. The GWR used parameters derived for individual locations rather than global ones. The MGWR varied bandwidth by county and used county-level regression. The collinearity analysis narrowed from 35 to 4 parameters: income inequality, median household income, percentage of nurse practitioners, and proportion of Black women in the total female population.

The GWR and MGWR models significantly outperformed the SLM and SEM in explaining COVID occurrences in U.S. by county; however, the GWR and MGWR performed poorly in the central and southern U.S. The authors acknowledged that limited data availability was an obstacle in these regions (Mollalo et al., 2020).

Maged et al. proposed an SEIR model in 2022 that included masking and vaccinations. They split the population into individual compartments for subgroups of mask wearers versus non-maskers. They tracked death count as a separate compartment. They modeled the effects of more masking as cases rise and attitudes change. Their mask to non-mask rate was determined by sigmoid function with a threshold number of infectious individuals (Maged et al., 2023).

Yang and Wang's 2020 research considered an SEIHRV (susceptible, exposed, infected non-hospitalized, infected hospitalized, concentration of COVID) model. They based their model off of Hamilton County, Tennessee, since it has demographic similarities to the U.S. as a whole and geographic diversity with different-sized towns and cities. Their base assumptions were that the rate of change of concentration of COVID depends on the E, I, H, and V compartments and deaths come from the H compartment only (Yang & Wang, 2021).

Pathak et al.'s model incorporated an incubation period parameter. There was a lack of data on vaccinations at the time, so values were determined through data fitting. The study used different time intervals for stay-at-home orders, transition to pre-COVID activity, and stable public activities to determine transmission rates. Overall, the data fitting wasn't ideal, especially for future predictions (Pathak et al., 2021).

Wynants et. al. compared COVID models from around the world. Out of the 606 selected models, 80 focused on the U.S. They found that these models were rushed and built on misleading data and assumptions. These models were found to be all very similar and none of them particularly accurate. This trend was explained by saying, "Prediction models for covid-19 entered the academic literature to support medical decision making at unprecedented speed and in large numbers. Most published prediction model studies were poorly reported and at high risk of bias such that their reported predictive performances are probably optimistic" (Wynants et al., 2020, p. 1).

Generally, these models were helpful in influencing our own model through their strengths and weaknesses. We strived to correct flawed assumptions, but we also drew inspiration from the various aspects of the models that succeeded.

2.4 Gap Analysis

Overall, there is a clear association between political ideology and vaccination rate in a county. Political leaning is not only the most influential factor on vaccination rates, but it is also closely associated with other demographic factors that influence vaccination rates. Politics also affect masking rates. Moreover, political leaning is also associated with level of belief in vaccine science and the effectiveness of masking mitigation measures. However, we failed to discover an existing mathematical model that synthesizes political vaccination patterns with epidemiology dynamics. Our model attempts to do this.

3 Methodology

We began with a simple SIR (susceptible, infectious, recovered) model and expanded into an SEIAVR (susceptible, exposed, infectious, asymptomatic, vaccinated, and recovered) model based on important characteristics of COVID we needed to incorporate.

We knew there would be an E compartment because there is an incubation period between exposure and infection of 2.8 days (Cortés Martínez et al., 2022). There needed to be an asymptomatic compartment as well as a symptomatic compartment because not all COVID cases were symptomatic (Buitrago-Garcia et al., 2020). Because we wanted to focus on COVID dynamics during a time frame in which vaccinations were first widely available, we needed to include a

vaccinated compartment.

We selected the time period of March 1st, 2021 to September 1st, 2021 because COVID vaccinations first became widely available to the public during this time frame (Albrecht, 2022). We limited our time window to six months to focus on the initial roll out of the vaccine and limit the number of COVID variants to consider.

The Johnson & Johnson vaccine is 76 percent effective after the first (only) dose (2 weeks until effective) (Lamb et al., 2023). The Pfizer vaccine is 91.3 percent effective 7 days after the second dose (Pfizer, 2021). The Moderna vaccine is 94.5 percent effective two weeks after the second dose (Moderna, 2020).

The time from vaccination to vaccine activation, 10.17 days, we calculated as a weighted average of the three vaccines (Mathieu et al., 2021). The average COVID recovery time is 9 days (Tamiru et al., 2023). The weighted average of vaccine efficacy is 0.92 (Mathieu et al., 2021). We defined vaccination rates by county as b_i (Albrecht, 2022). The rate of infection, β is

$$\frac{1}{t_{incubation}}.$$

The rate of susceptibles exposed by infected individuals, γ_I was calculated through data fitting to be

$$\frac{\rho_i}{N_i}.$$

The rate of susceptibles exposed by asymptomatic individuals, γ_A, is $0.8 \cdot \gamma_I$ (Peter et al., 2023). The probability of a symptomatic COVID case resulting in death in the U.S. from March 1st, 2021 to September 1st, 2021 is cumulative deaths divided by cumulative cases, or 0.017395 (Mathieu et al., 2020). For COVID death cases, the median time from symptom onset to death is 25 days (Elezkurtaj, S., Greuel, S., Ihlow, J. et al., 2021). 0.017395 divided by 25 days is 0.000696, the COVID death rate μ_C. Natural birth rate, is 11 births per 1,000 people, which was the 2021 natural birth rate (Korhonen, 2024). The pre-COVID natural death rate was 715.2 deaths per 100,000 in the U.S. in 2019 (Murphy et al., 2021). Converted into daily rates, the birth rate, θ, is 0.00003014, and the death rate, μ is 0.00001959. The vaccination rate, ν is 0.0057 (Mathieu et al., 2020). The rate of asymptomatic to recovered, λ_A is

$$\frac{1}{\text{Average COVID Recovery Time}}$$

(Peter et al., 2023). The rate of infected to recovered, λ_I, is

$$\frac{1}{\text{Average COVID Recovery Time}}$$

(Peter et al., 2023). The rate of vaccine activation, φ, is

$$\frac{1}{\text{Vaccine Activation Time}}.$$

The ratio of asymptomatic cases to all COVID cases, σ is 0.35 (Buitrago-Garcia et al., 2020).

We defined the symptomatic transmission rate, γ_I, as $\frac{\rho_i}{N_i}$. The asymptomatic transmission rate, γ_A is $0.8 \cdot \gamma_I$. The transmission rate from symptomatic to vaccinated individuals, γ_{VI}, is $(1 - \text{Vaccine Efficacy}) \cdot \gamma_I$. The transmission rate from asymptomatic to vaccinated individuals, γ_{VA}, we defined as $(1 - \text{Vaccination Efficacy}) \cdot \gamma_A$. The a_1 population vaccination proportion, b_1, is 0.553, making the vaccination rate for a_1 $\nu \cdot 0.553$, or 0.003152. We defined α, the vaccine effectiveness proportion, as 0.92. We defined the infection rate, β, as $\frac{1}{2.8}$. The vaccine activation rate, φ, is $\frac{1}{10.17}$. The proportion of asymptomatic individuals, σ, is 0.35. The symptomatic recovery rate, λ_I, and the asymptomatic recovery rate, λ_A, are both $\frac{1}{9}$. The population size, N, we defined as $2.9 \cdot 10^7$.

The key distinguishing feature of our model is the use of county-classification-based patches. Using data from Albrecht (2022), we created five county categories, denoted a_i, to group counties by percentage of Trump vote in the 2020 presidential election.

Table 1
Percentage Trump Vote and Corresponding County Category and Vaccination Proportion

Percentage Trump Vote	County Classification	Vaccination Proportion
<25%	a_1	55.3%
25-45%	a_2	53.1%
45-55%	a_3	49.0%
55-75%	a_4	42.2%
>75%	a_5	34.7%

We omitted Hawaii and Alaska because they are not part of the continental United States, and there was not enough data for cases, deaths, and vaccinations in either state.

The following maps show the continental United States shaded according to various factors. Figure 1 is a map of U.S. counties colored by their a_i category. Figure 2 is a map of proportion of vaccinated individuals by county. White counties are those with not enough data. Figure 3 maps COVID case rates by county as cumulative cases from March 1st through September 1st, 2021 out of total population. Figure 4 maps COVID death rates by county during the same time period. There is a visible association between all four factors. a_1 counties tended to have higher vaccination proportions, while a_5 counties tended to have lower vaccination proportions. a_1 counties tended to have lower COVID case rates, while a_5 counties tended to have higher COVID case rates. Similarly, a_1 counties tended to have lower COVID death rates, while a_5 counties tended to

have higher COVID death rates.

Figure 1
United States by County Classification Type a_i

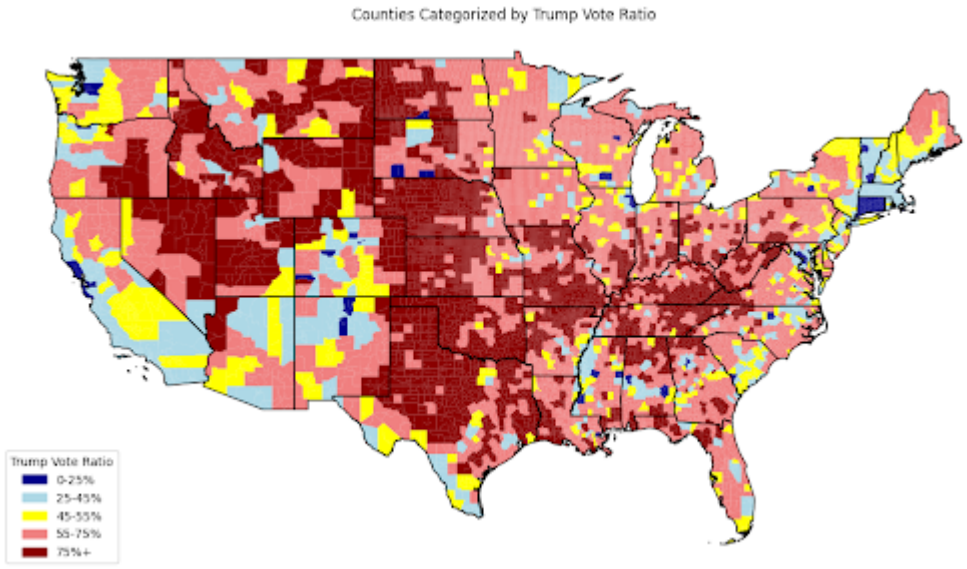


Figure 2
US Counties by Vaccinated Proportion

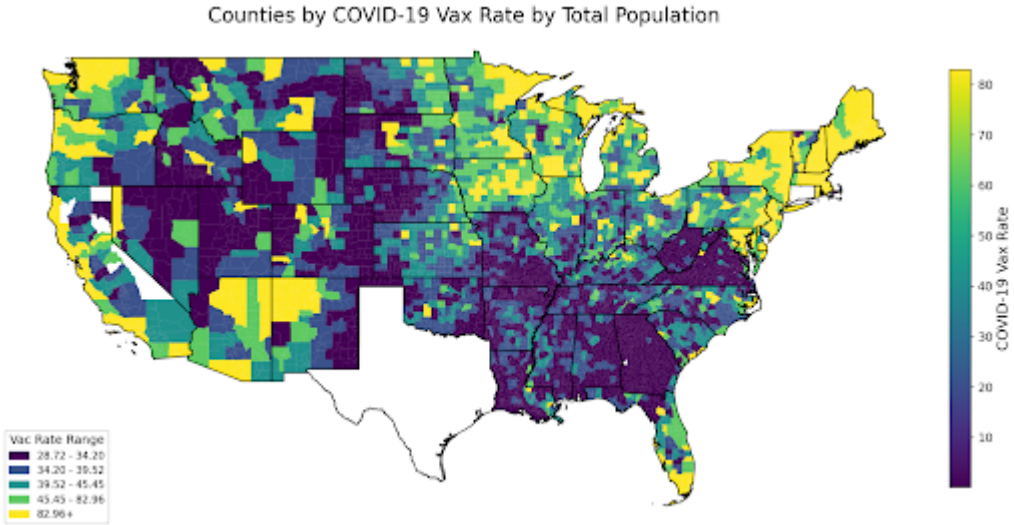


Figure 3
US Counties by Cumulative COVID Cases by Population

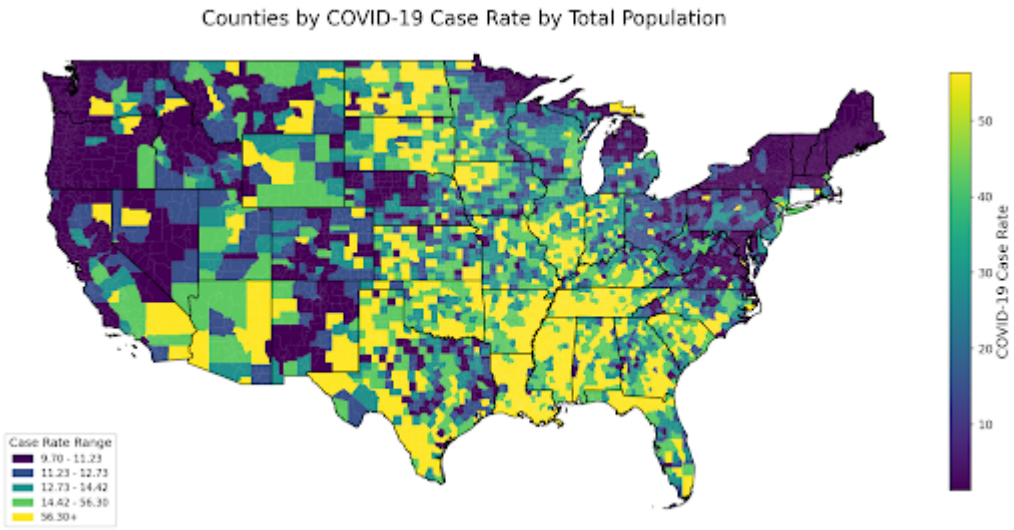
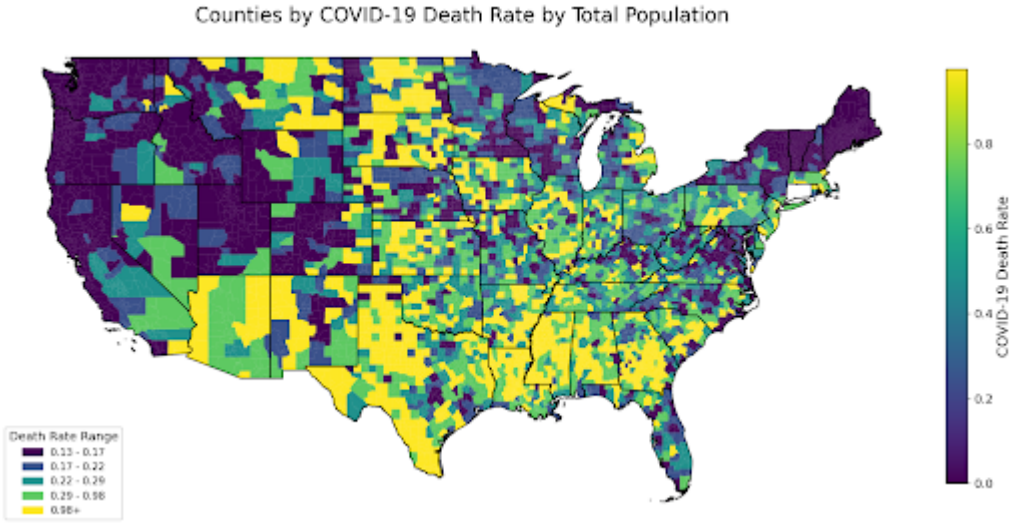


Figure 4
US Counties by Cumulative Deaths by Population



3.1 Model 1

For our initial model, we needed to data fit for values of γ_I for each of the five subpopulations.

The parameters estimated for this model were γ_I, γ_A, γ_{VI}, and γ_{VA}. We assumed γ_A, γ_{VI}, and γ_{VA} are all proportional to γ_I, so only γ_I needed to be generated directly from data fitting. Parameter estimation was done by systematically guessing a value for γ_I with proportional values for γ_A, γ_{VI}, and γ_{VA} to minimize the sum-of-squares error between the reported cumulative number of COVID-induced deaths and the model's predicted number of COVID-induced deaths. For the simple model, the transmission rate was estimated after being scaled to the size of each subpopulation, which allowed for more analytical comparisons between the transmission rates of the five subpopulations. For the mobility-considerate models, the transmission rate was estimated independent of population size, as the mobility component already considered the size of each subpopulation.

All other parameters were drawn from aforementioned data sources or were scalar multiples of γ_I.

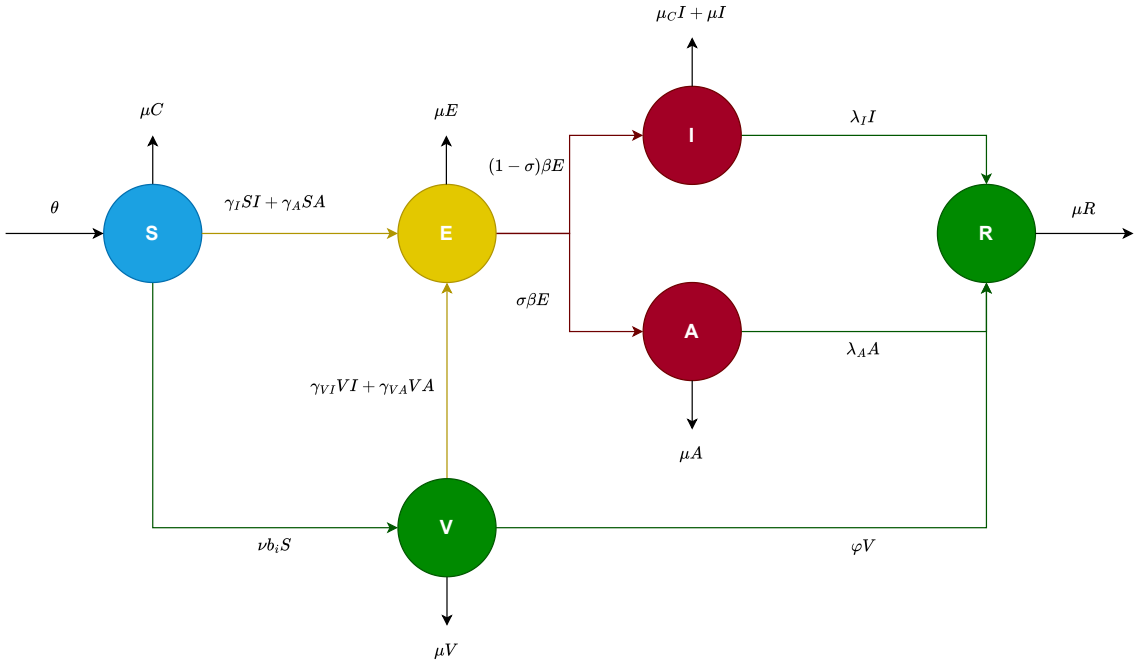
Consider the following table of parameters:

Parameter	Value	Parameter Definition
N_i	N_i	Subpopulation Size
γ_I	ρ_i / N_i	Rate of Infection (S Infected by I)
ρ_i	ρ_i	Transmission Rate Upon Encountering Infected Individual
γ_A	$0.8 \cdot \gamma_I$	Rate of Infection (S Infected by A)
γ_{VI}	$(1-\alpha) \cdot \gamma_I$	Rate of Infection (V Infected by I)
γ_{VA}	$(1-\alpha) \cdot \gamma_A$	Rate of Infection (V Infected by A)
ν	0.0057	Standardized Vaccination Rate
b_i	b_i	Proportion Vaccinated by County
α	0.92	Vaccine Effectiveness Proportion
β	1 / 2.8	Infection Rate
φ	1 / 10.17	Vaccine Activation Rate
σ	0.35	Proportion Asymptomatic
μ_C	0.0007	COVID Death Rate
μ	0.00001959	Natural Death Rate
θ	$0.00003014 \cdot N_i$	Natural Birth Rate
λ_I	1 / 9	Symptomatic Recovery Rate
λ_A	1 / 9	Asymptomatic Recovery Rate

Consider the following mathematical model:

$$
\begin{aligned}
\frac{dS}{dt} &= \theta - \nu b_i S - \gamma_I SI - \gamma_A SA - \mu S \\
\frac{dE}{dt} &= \gamma_I SI + \gamma_A SA + \gamma_{VI} VI + \gamma_{VA} VA - \beta E - \mu E \\
\frac{dI}{dt} &= (1-\sigma)\beta E - \lambda_I I - \mu I - \mu_C I \\
\frac{dA}{dt} &= \sigma\beta E - \lambda_A A - \mu A \\
\frac{dV}{dt} &= \nu b_i S - \gamma_{VI} VI - \gamma_{VA} VA - \varphi V - \mu V \\
\frac{dR}{dt} &= \lambda_A A + \lambda_I I + \varphi V - \mu R
\end{aligned}
\tag{1}
$$

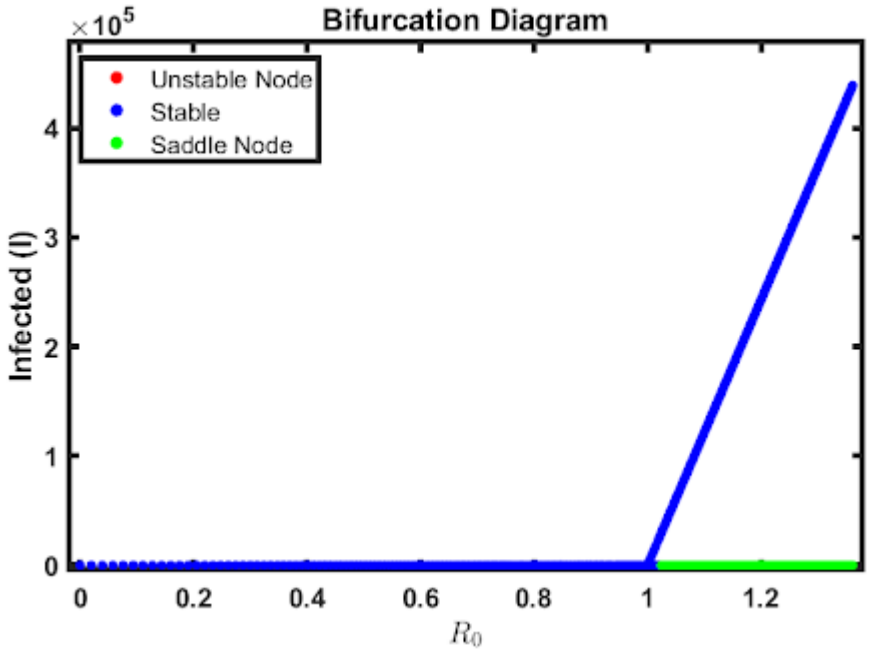
The following is a visualization of the model:
Figure 5
SEIAVR Model Visualization



γ_I	:	symptomatic transmission rate
γ_A	:	asymptomatic transmission rate
γ_{VI}	:	symptomatic transmission rate (I infecting V)
γ_{VA}	:	asymptomatic transmission rate (A infecting V)
ν	:	vaccination rate
b_i	:	population vaccination proportion
α	:	vaccine effectiveness proportion
β	:	infection rate
φ	:	vaccine activation rate
σ	:	proportion asymptomatic
μ_C	:	infected death rate
μ	:	natural death rate
θ	:	natural birth rate
λ_I	:	symptomatic recovery rate
λ_A	:	asymptomatic recovery rate

Through simulations, we identified that the SEIAVR model had two equilibria: the disease-free equilibrium and an endemic equilibrium. As can be seen in Figure 5, the disease-free equilibrium is stable when $R_0 < 1$ and is unstable when $R_0 > 1$. We were not able to solve for the endemic equilibrium, but the simulations show that it exists and is stable when $R_0 > 1$.

Figure 6
Bifurcation Diagram of R_0 versus Infected



The disease-free equilibrium exists when (S, E, I, A, V, R) is $(S_{eq}, 0, 0, 0, V_{eq}, R_{eq})$. Consider the following equations for S_{eq}, V_{eq}, and R_{eq}:

$$S_{eq} = \frac{\theta}{\nu b_i + \mu}$$

$$V_{eq} = \frac{\theta \nu b_i}{(\mu + \varphi)(\nu b_i + \mu)}$$

$$R_{eq} = \frac{\varphi \theta \nu b_i}{\mu(\mu + \varphi)(\nu b_i + \mu)}$$

At the disease-free equilibrium, the following is the expression for R_0:

$$\begin{aligned} R_0 &= (1-\sigma)\left(b_i\nu \, \tfrac{\gamma_{VI}\ \beta\theta}{(\beta+\mu)(\lambda_I+\mu_c+\mu)(\mu+b_i\nu)(\mu+\varphi)} + \tfrac{\gamma_I\ \beta\theta}{(\beta+\mu)(\lambda_I+\mu_c+\mu)(\mu+b_i\nu)}\right) \\ &+ \sigma\ \left(b_i\nu \, \tfrac{\gamma_{VA}\ \beta\theta}{(\beta+\mu)(\lambda_A+\mu)(\mu+b_i\nu)(\mu+\varphi)} + \tfrac{\gamma_A\ \beta\theta}{(\beta+\mu)(\lambda_A+\mu)(\mu+b_i\nu)}\right) \end{aligned}$$

The new infections due to symptomatic individuals are represented by $(1-\sigma)\left(b_i\nu \, \tfrac{\gamma_{VI}\ \beta\theta}{(\beta+\mu)(\lambda_I+\mu_c+\mu)(\mu+b_i\nu)(\mu+\varphi)} + \tfrac{\gamma_I\ \beta\theta}{(\beta+\mu)(\lambda_I+\mu_c+\mu)(\mu+b_i\nu)}\right)$, and the new infections due to asymptomatic individuals are represented by
$\sigma\ \left(b_i\nu \, \tfrac{\gamma_{VA}\ \beta\theta}{(\beta+\mu)(\lambda_A+\mu)(\mu+b_i\nu)(\mu+\varphi)} + \tfrac{\gamma_A\ \beta\theta}{(\beta+\mu)(\lambda_A+\mu)(\mu+b_i\nu)}\right)$.

3.1.1 Underreporting

Considering unreported COVID cases has made it more difficult to comprehend the dynamics of COVID spread. Our model accounts for case underreporting by fitting the transmission rate parameters to the reported number of cumulative COVID deaths, then setting the initial conditions to reflect the estimated present number of unreported cases. COVID deaths were also underreported by a factor of approximately 1.54 according to (Angulo et al., 2021), with out-of-hospital deaths contributing as discussed by (Pathak et al., 2021), but since cause-of-death is reported, and hospitals became more prepared to handle COVID cases as time went on, COVID cases resulting in deaths are more reliable to be tracked than those that did not require urgent treatment, especially within our time frame of interest.

Seroprevalence tests are a reliable indicator of underreporting, as antibodies are present in people who have been both symptomatically and asymptomatically infected with COVID. Antibodies wane faster for asymptomatic cases, however, suggesting that the number of unreported cases found through seroprevalence testing may be even lower than the true number. The initial lockdown phase of the U.S. pandemic contained the greatest current ratio of unreported cases, according to (Irons & Raftery, 2021), which could be due to lack of accessible testing at the time. Our time frame of interest starts with the largest period of daily vaccinations in the U.S., but according to (Angulo et al., 2021) and (Irons & Raftery, 2021), there were still roughly 2.3 times more COVID cases

present at the time than were reported, implying a greater number of COVID cases at the start of our time frame than the data suggests, and more accurately reflecting the dynamics of reported COVID deaths when considered. Our model suggests that subpopulations a_2, a_4, and a_5 were especially prone to underreporting based on how the model fit to the data for those subpopulations.

Eutsler et al. used Benford's Law to assess COVID case and death data for underreporting. They found that reported cases and deaths violated Benford's Law, implying underreporting. They also classified counties' political parties based 2016 presidential voting majorities crossed with 2020 gubernatorial political party of each county's state. For their sample period from January 21, 2020 to November 3, 2020, they found that "Democratic counties demonstrate the smallest departures from Benford's Law while Republican counties demonstrate the greatest departures" (Eutsler et al., 2023, p. 1). They argued that "public officials may have manipulated the reporting records in accounting for COVID-19 infection cases and deaths to validate the effectiveness of political party objectives" (Eutsler et al., 2023, p. 1).

3.1.2 Initial Conditions

Initial conditions for symptomatic cases were derived from incidence reporting of the 10 days prior to the start of our time frame, with each prior day's report multiplied by a fraction with the COVID recovery time in the denominator to reflect the probability of still having COVID during our initial time frame based on the day of the incidence reporting.
For asymptomatic cases, the initial symptomatic case number was multiplied by $\frac{\sigma}{1-\sigma}$ to reflect the proportion of asymptomatic cases.
For exposed people, the initial sum of symptomatic and asymptomatic cases was multiplied by the COVID incubation period.
For vaccinated people, the total subpopulation size was multiplied by 0.1, obtained from (Centers for Disease Control and Prevention, 2024), to reflect the proportion of that subpopulation fully vaccinated at the start of our time frame.
For recovered people, the cumulative number of deaths was subtracted from the cumulative number of cases.
The initial number of susceptible people is obtained by subtracting the initial number of people in the other compartments from the initial subpopulation size.

3.1.3 Data Fitting

We strived to estimate through fitting to data as few parameters as possible, using existing data whenever applicable. Our model only requires direct parameter estimation for one parameter, $gamma_I$ (symptomatic-to-susceptible transmission rate), with $gamma_A$, $gamma_{VI}$, $gamma_{VA}$ (other transmission rates) proportional to $gamma_I$ and all considered when minimizing sum-of-squares error. All other parameter values were identified from existing literature.

Each of the five subpopulations of interest, a_1 to a_5, has unique characteristics that contribute to the likelihood of an individual person transmitting the disease. The subpopulations were decided by their by-county 2020 Presidential Election voting results, but they also possess differences in population density and COVID response behavior that contribute to how likely an individual is to transmit the disease. Rather than making arbitrary assumptions about behavior such as masking, social distancing, and self-enforced quarantining, we decided it would be most representative of these factors to fit the transmission rate for each subpopulation to the cumulative reported COVID death count of each subpopulation. The transmission rate for each subpopulation was fitted based on the cumulative number of COVID deaths reported in each respective subpopulation as opposed to the cumulative number of COVID cases reported, as COVID deaths were less likely to go unreported than COVID cases.

Our initial fitting of our simple SEIAVR model saw challenges with the initial conditions based solely on the cumulative cases data. When accounting for underreporting with an initial present number of cases equaling 2.3x the *reported* initial number of cases, a conservative estimate corroborated by (Angulo et al., 2021), (Taylor et al., 2023), and (Irons & Raftery, 2021), the behavior of cases and deaths predicted by the model much more strongly represents the behavior of the cumulative reported number of cases and deaths. The model further suggests that subpopulations a_2, a_4, and a_5 show a greater proportion of unreported present cases than subpopulations a_1 and a_3, which could result from various subpopulation-specific factors such as population size, population density, tendency for a person to self-report, tendency for the local government to report cases, and possible proportion of asymptomatic cases.

3.2 Model with Eulerian Mobility

We can classify movement into two categories: migration and visitation. Migration is movement from one place to another permanently, without intent to return. Visitation is temporary travel to another place with intent to return to the original location (Sandip Mandal et al., 2011, p. 14). We will classify movement between various patches, each with sub-models to reflect localized COVID dynamics (Eikenberry & Gumel, 2018, p. 60).

We generated a matrix, K, to represent the average daily mobility between different types of county. The data values used to calculate K are from Kang et al.'s multiflow mobility project, which used SafeGraph's data from millions of anonymous cell phones' weekly mobility (Kang et al., 2020). K is a square matrix where K_{ij} represents the average daily movement from a county of type i to a county of type j. Notice the entries along the diagonal are all zero because we are focusing on inter-county, not intra-county mobility, so the movement from county i to county i is zero.

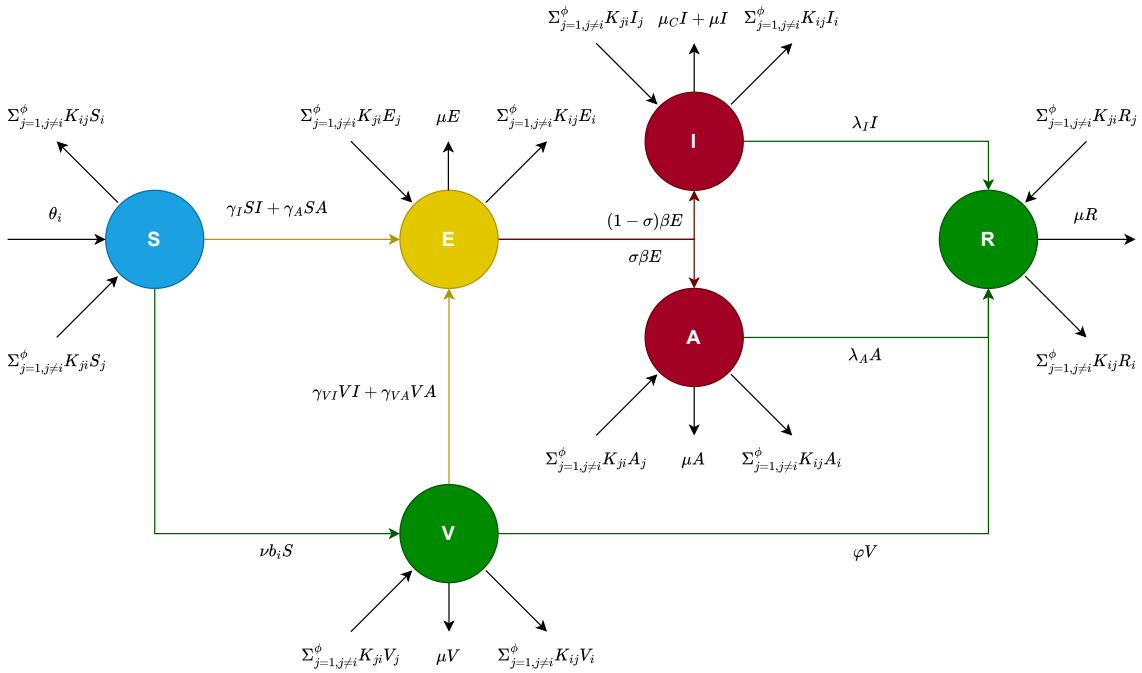
$$\begin{bmatrix} 0 & 0.171 & 0.043 & 0.033 & 0.004 \\ 0.041 & 0 & 0.057 & 0.072 & 0.011 \\ 0.019 & 0.129 & 0 & 0.100 & 0.016 \\ 0.011 & 0.147 & 0.098 & 0 & 0.052 \\ 0.004 & 0.080 & 0.070 & 0.356 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & K_{12} & K_{13} & K_{14} & K_{15} \\ K_{21} & 0 & K_{23} & K_{24} & K_{25} \\ K_{31} & K_{32} & 0 & K_{34} & K_{35} \\ K_{41} & K_{42} & K_{43} & 0 & K_{45} \\ K_{51} & K_{52} & K_{53} & K_{54} & 0 \end{bmatrix}$$

Consider the following mathematical model incorporating Eulerian mobility:

$$\begin{aligned}
\tfrac{dS}{dt} &= -\gamma_I SI - \nu b_i S - \gamma_A SA + \theta - \mu S + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} S_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} S_i \\
\tfrac{dE}{dt} &= \gamma_I SI + \gamma_A SA + \gamma_{VI} VI + \gamma_{VA} VA - \beta E - \mu E + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} E_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} E_i \\
\tfrac{dI}{dt} &= (1-\sigma)\beta E - \mu_C I - \lambda_I I - \mu I + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} I_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} I_i \\
\tfrac{dA}{dt} &= \sigma\beta E - \lambda_A A - \mu A + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} A_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} A_i \\
\tfrac{dV}{dt} &= \nu b_i S - \gamma_{VI} VI - \gamma_{VA} VA - \varphi V - \mu V + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} V_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} V_i \\
\tfrac{dR}{dt} &= \lambda_A A + \lambda_I I + \varphi V - \mu R + \textstyle\sum_{j=1,j\neq i}^{\phi} K_{ji} R_j - \sum_{j=1,j\neq i}^{\phi} K_{ij} R_i
\end{aligned} \tag{2}$$

The following is a visualization of the model:
Figure 7
SEIAVR Model with Eulerian Mobility Visualization



3.3 Model with Lagrangian Mobility

Consider the following mathematical model incorporating Lagrangian mobility:

$$
\begin{aligned}
\frac{dS_i}{dt} &= \theta_i - \nu b_i S_i - S_i\delta_{Ii} - S_i\delta_{Ai} - \mu S_i \\
\frac{dE_i}{dt} &= S_i\delta_{Ii} + S_i\delta_{Ai} + (1-\alpha)V_i\delta_{Ii} + (1-\alpha)V_i\delta_{Ai} - \beta E_i - \mu E_i \\
\frac{dI_i}{dt} &= (1-\sigma)\beta E_i - \lambda_I I_i - \mu I_i - \mu_C I_i \\
\frac{dA_i}{dt} &= \sigma\beta E_i - \lambda_A A_i - \mu A_i \\
\frac{dV_i}{dt} &= \nu b_i S - (1-\alpha)V_i\delta_{Ii} - (1-\alpha)V_i\delta_{Ai} - \varphi V_i - \mu V_i \\
\frac{dR_i}{dt} &= \lambda_A A_i + \lambda_I I_i + \varphi V_i - \mu R_i
\end{aligned}
\tag{3}
$$

Where

$$\begin{aligned} \delta_{Ii} &= \sum_{j=1}^{n} \gamma_{Ii} P_{ij} \frac{\sum_{k=1}^{n} P_{kj} I_k}{\sum_{n=1}^{n} P_{kj} N_k} \\ \delta_{Ai} &= \sum_{j=1}^{n} \gamma_{Aj} P_{ij} \frac{\sum_{k=1}^{n} P_{kj} A_k}{\sum_{k=1}^{n} P_{kj} N_k} \end{aligned} \tag{4}$$

Where γ_{Ij} and γ_{Aj} are the risk of a person who lives in patch i becoming infected by infected or asymptomatic individuals in patch j, respectively. P_{ij} is the proportion of time that residents in patch i spend in patch j.

3.3.1 Calculating the Time Proportion Matrix P

Although the K matrix, representing daily mobility between county types, is suitable for Eulerian mobility, finding a time proportion matrix for the model with Lagrangian mobility is a more complicated matter. Ideally, the Lagrangian model would use the proportion of time a person spends in each patch out of their entire lifetime, but data about this is both unavailable and unrealistic to obtain. Furthermore, there is no simple conversion from a K matrix about daily mobility to this sort of matrix. To generate such a matrix, we chose to use an agent-based model, for its ability to focus on individuals and assign different attributes to them, which we could use to account for multiple types of travel, such as work and vacations, as well as varying travel lengths. For this simulation, we first created a matrix T from K, restoring the values along the diagonal by subtracting the other values in each row from 1.

$$\begin{bmatrix} 0.749 & 0.171 & 0.043 & 0.033 & 0.004 \\ 0.041 & 0.820 & 0.057 & 0.072 & 0.011 \\ 0.019 & 0.129 & 0.736 & 0.100 & 0.016 \\ 0.011 & 0.147 & 0.098 & 0.691 & 0.052 \\ 0.004 & 0.080 & 0.070 & 0.356 & 0.489 \end{bmatrix}$$

$$\begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} & T_{15} \\ T_{21} & T_{22} & T_{23} & T_{24} & T_{25} \\ T_{31} & T_{32} & T_{33} & T_{34} & T_{35} \\ T_{41} & T_{42} & T_{43} & T_{44} & T_{45} \\ T_{51} & T_{52} & T_{53} & T_{54} & T_{55} \end{bmatrix}$$

Each row, summing to 1, represents the proportion of individuals in patch a_i who either stayed in their county or traveled to another patch, with the diagonal representing the proportion of individuals who remained in their county. Because T shows the proportions of each type of travel, T can be treated as the probabilities of an individual in a patch travelling to another patch or staying put. Using this, we simulated movement for 100,000 individuals, distributed according to the population sizes of each patch, over the span of 4 years, which is the average length of employment (U.S. Bureau of Labor Statistics, 2022).

Ideally, we would have been able to simulate movement for the entire U.S. population over 76.1 years, the average life expectancy for an American (Arias, E., Tejada-Vera, Betzaida, Kochanek, K. D., and Ahmad, F., 2022). However, due to our limited computational power, this was not possible. To compensate for this, we ran our smaller simulation 20 times.

We assigned the following behavior to each individual in our simulation: Every individual begins in their home patch and continues to have the same home patch over the entire simulation. Every day, they either travel once to another patch, or choose to stay within their current patch, with probability T_{ij}. Each agent is assigned a value $x \in [1, 10]$, representing the number of days that the agent is away from home, which accounts for different types of travel such as work or vacation. After day x, the agent is assigned a random probability between 0.51 and 1 of returning home. After running this simulation 20 times for a total of 2,000,000 agents, we calculated the proportion of time spent in each patch, creating the matrix P below:

$$\begin{bmatrix} 0.505 & 0.292 & 0.096 & 0.092 & 0.015 \\ 0.073 & 0.634 & 0.124 & 0.144 & 0.024 \\ 0.046 & 0.253 & 0.513 & 0.161 & 0.027 \\ 0.038 & 0.266 & 0.152 & 0.495 & 0.049 \\ 0.025 & 0.195 & 0.124 & 0.304 & 0.352 \end{bmatrix}$$

4 Results

4.1 Simple Model - No Subpopulations

Our simple model without subpopulations with parameters universal to the U.S. as a whole produces the following simulations:

Figure 7:
Compartment Simulations

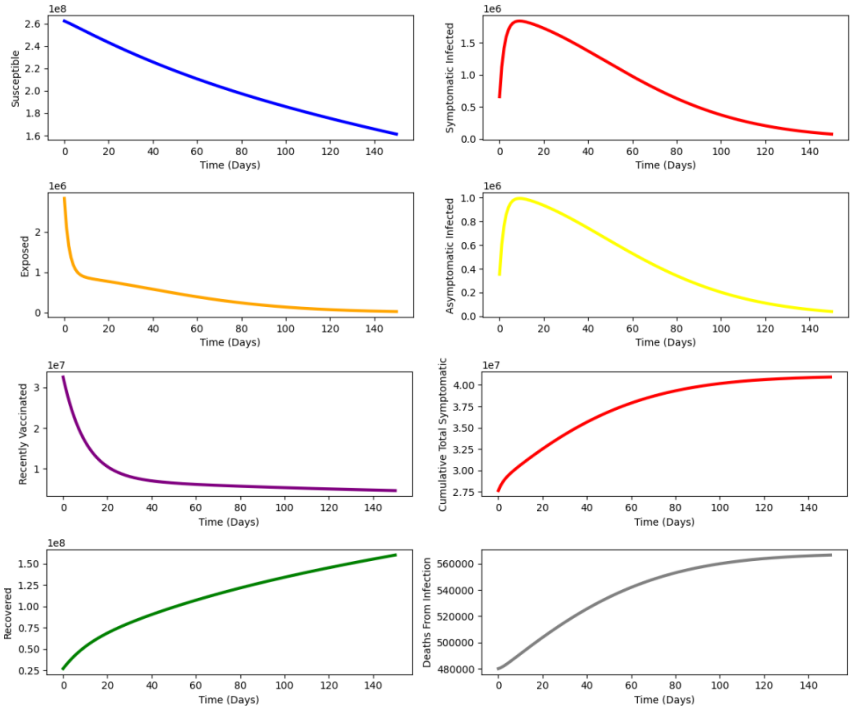
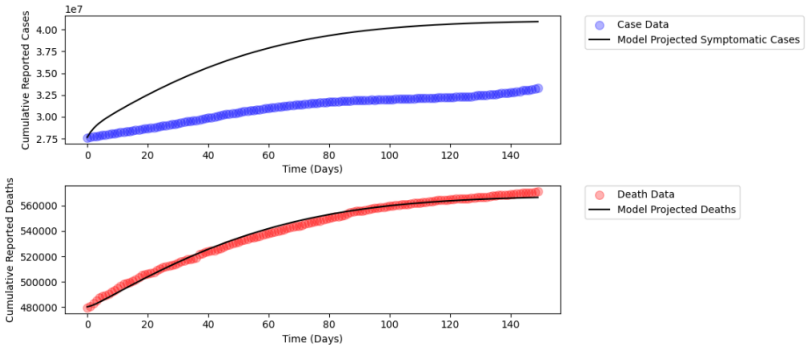


Figure 8:
Model vs. Data



Cumulative Deaths $r^2 = 0.99$

4.2 Simple Model - Subpopulations

Our simple model with our a_1, a_2, a_3, a_4, a_5 subpopulations and parameters specific to each one produces the following simulations:

Figure 9:
a_1 *Compartment Simulations*

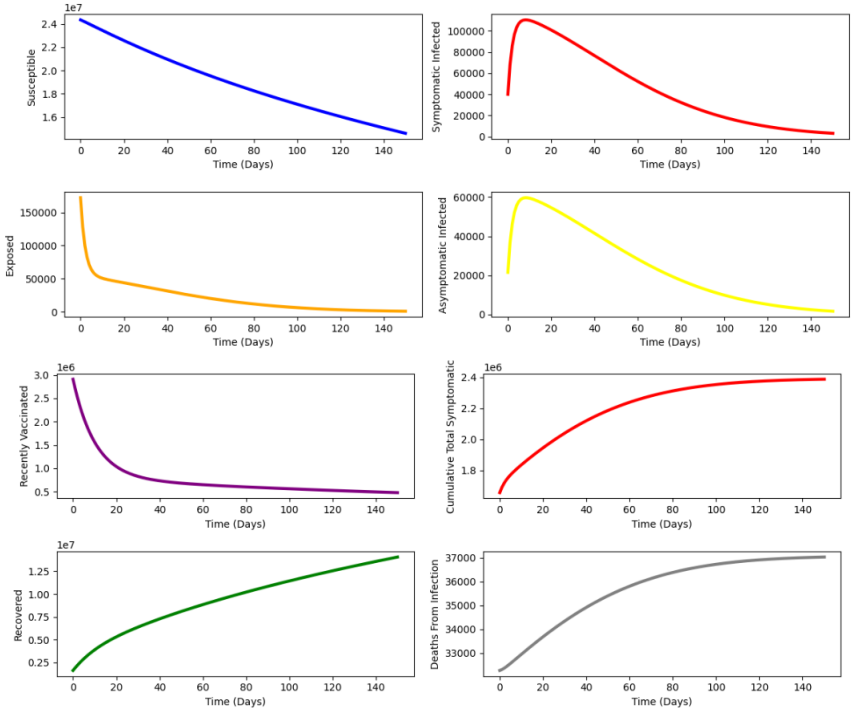
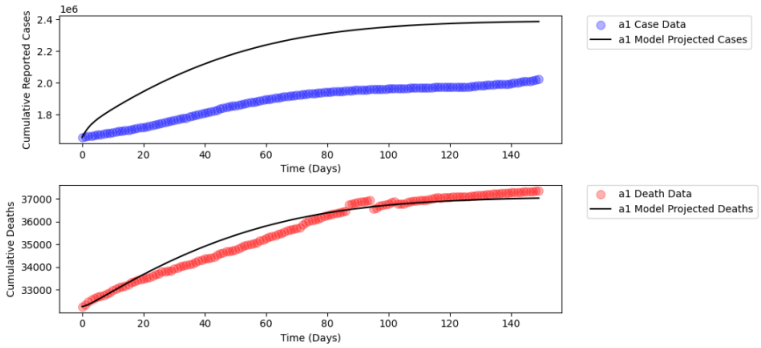


Figure 10:
a_1 *Model vs. Data*



Cumulative Deaths $r^2 = 0.95$

Figure 11:
a_2 *Compartment Simulations*

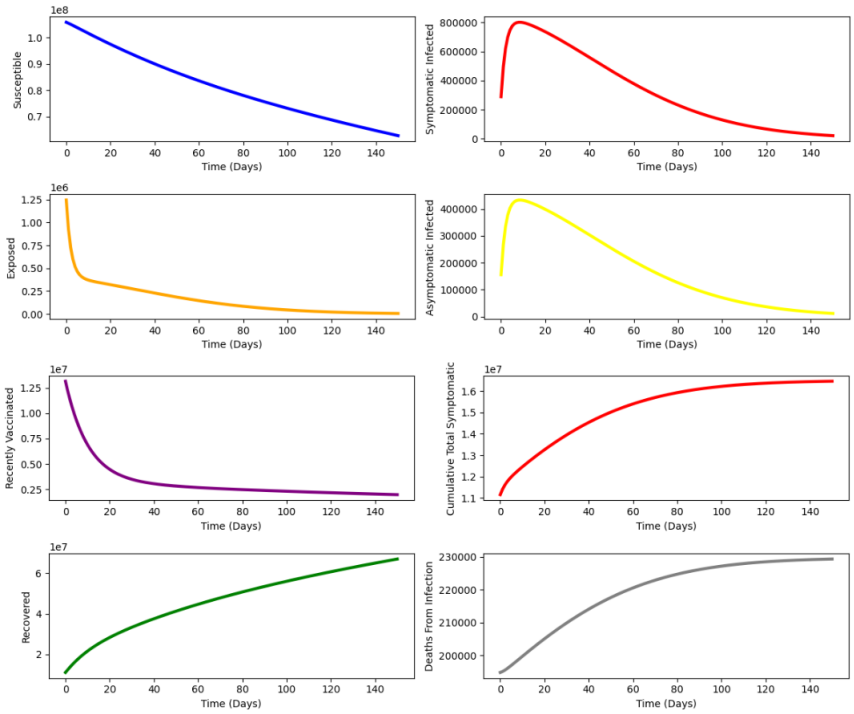
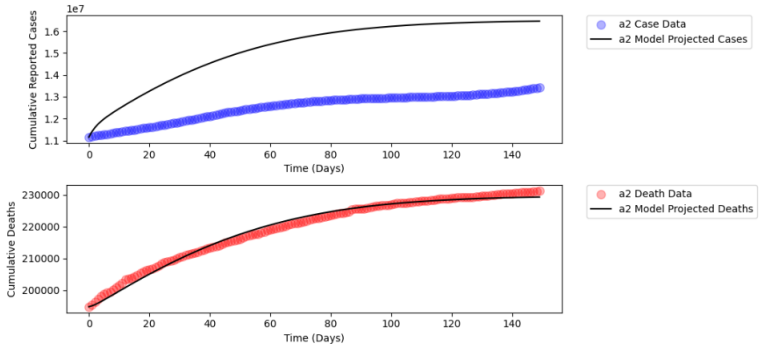


Figure 12:
a_2 *Model vs. Data*



Cumulative Deaths $r^2 = 0.99$

Figure 13:
a_3 *Compartment Simulations*

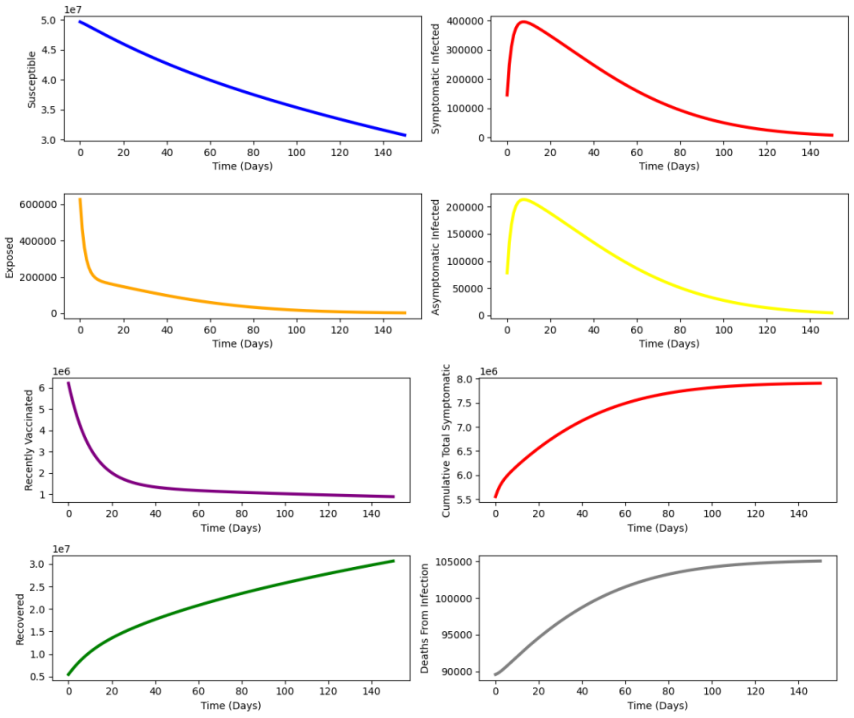
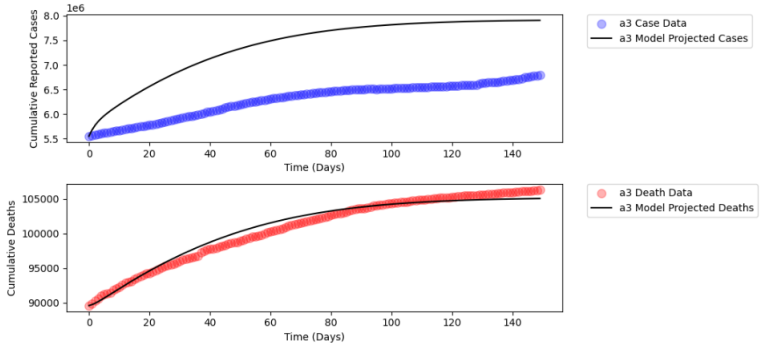


Figure 14:
a_3 *Model vs. Data*



Cumulative Deaths $r^2 = 0.97$

Figure 15:
a_4 *Compartment Simulations*

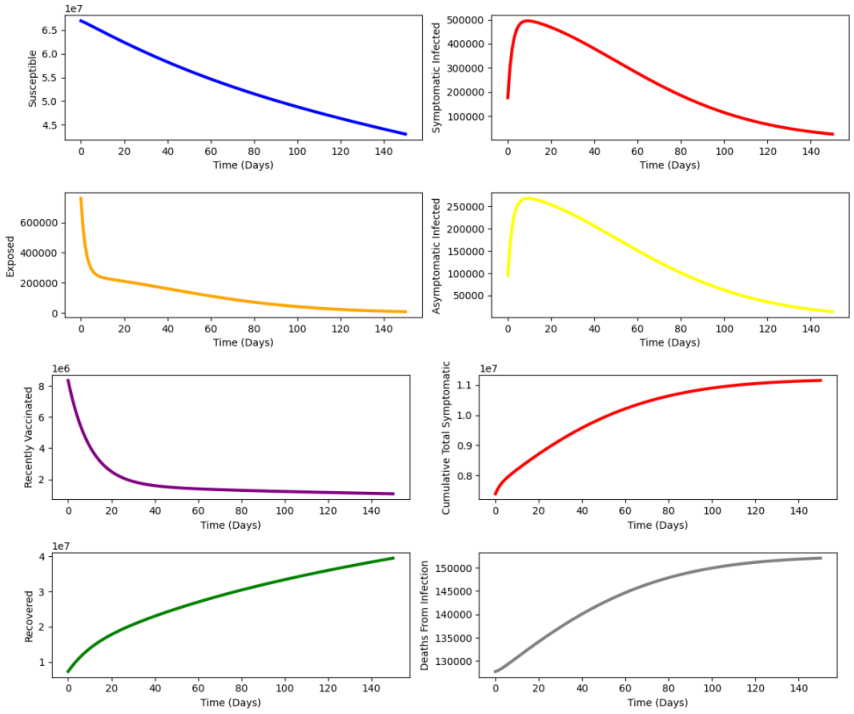
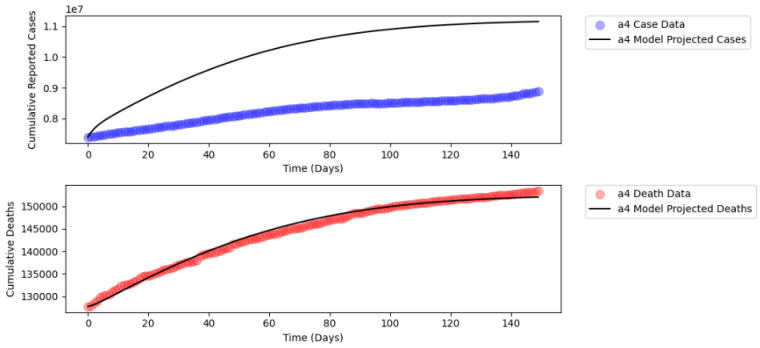


Figure 16:
a_4 *Model vs. Data*



Cumulative Deaths $r^2 = 0.99$

Figure 17:
a_5 *Compartment Simulations*

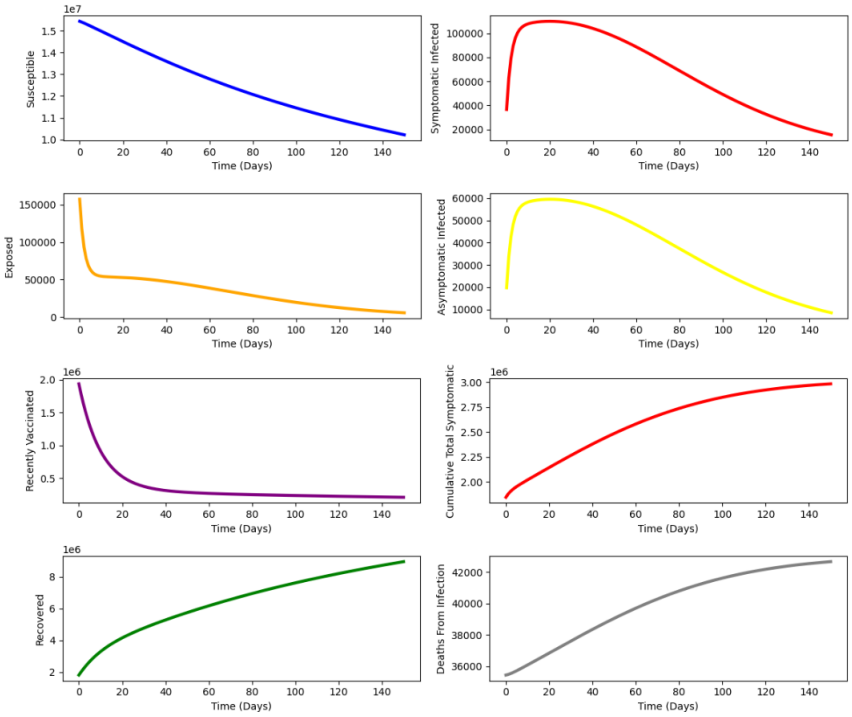
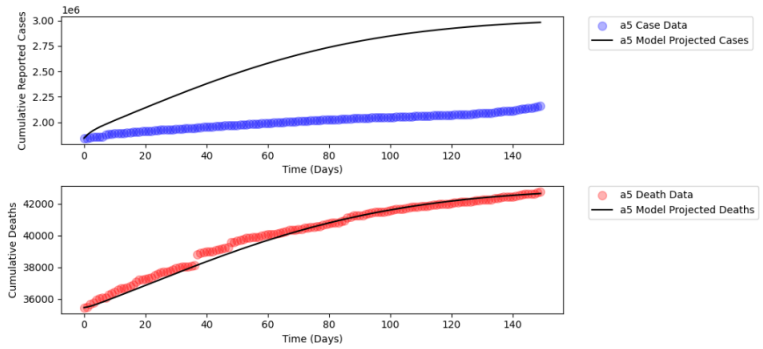
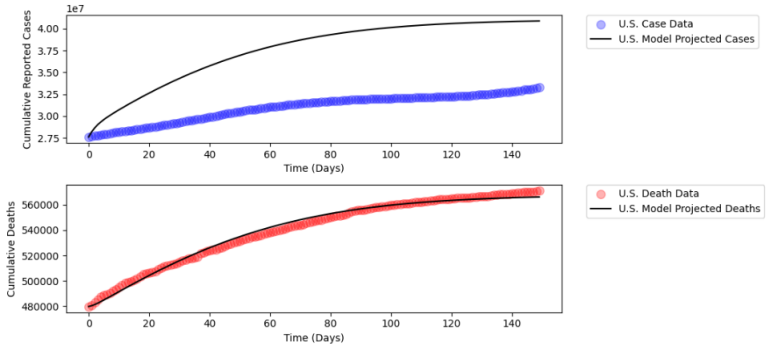


Figure 18:
a_5 *Model vs. Data*



Cumulative Deaths $r^2 = 0.98$

Figure 19:
U.S. Model vs. Data



Cumulative Deaths $r^2 = 0.99$

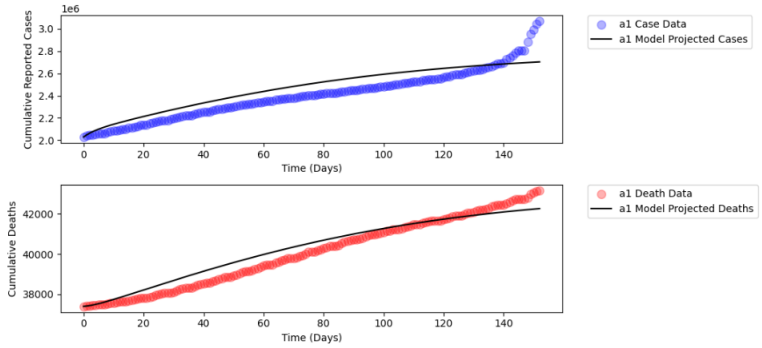
4.3 Simple Model - Delta Variant

Consider the following table of updated parameters for the Delta Variant:

Parameter	Value	Parameter Definition
γ_I	ρ_i / N_i	Rate of Infection (S Infected by I)
ρ_i	ρ_i	Transmission Rate Upon Encountering Infected Individual
γ_A	$0.8 \cdot \gamma_I$	Rate of Infection (S Infected by A)
γ_{VI}	$(1-\alpha) \cdot \gamma_I$	Rate of Infection (V Infected by I)
γ_{VA}	$(1-\alpha) \cdot \gamma_A$	Rate of Infection (V Infected by A)
ν	0.0024	Standardized Vaccination Rate
b_i	b_i	Proportion Vaccinated by County
α	0.8	Vaccine Effectiveness Proportion
β	1 / 5	Infection Rate
μ_C	0.0005	COVID Case Fatality Rate
λ_I	1 / 14	Symptomatic Recovery Rate
λ_A	1 / 14	Asymptomatic Recovery Rate

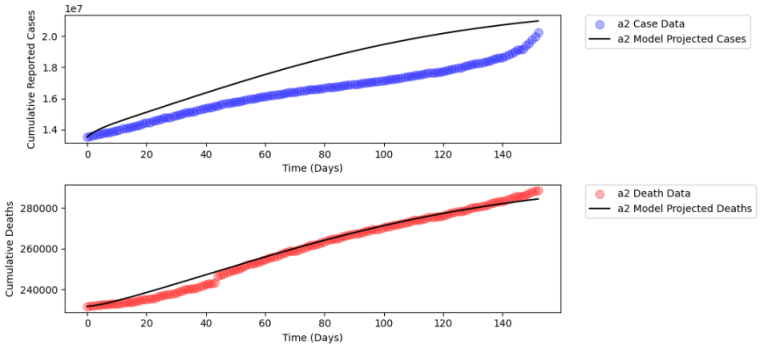
The parameter ν comes from (Mathieu et al., 2020). μ_C is derived as a proportion of the case fatality rate of the original time frame of interest, a proportion obtained from (Xia et al., 2024). α is estimated based on data from (Pormohammad, A., Zarei, M., Ghorbani, S., Mohammadi, M., Aghayari Sheikh Neshin, S., Khatami, A., Turner, D. L., Djalalinia, S., Mousavi, S. A., Mardani-Fard, H. A., Kasaeian, A., & Turner, R. J., (2021)). β is the reciprocal of the incubation period of the Delta variant, obtained from (Li et al., 2024). λ_I and λ_A are the reciprocal of the Delta variant recovery time, estimated from (Kumar et al., 2022) due to lack of specific data available. γ_I, γ_A, γ_{VI}, and γ_{VA} are again obtained through data fitting. b_i is unchanged.
Initial conditions were also updated from the same sources, but now reflecting the conditions of August 1 with Delta variant parameters.
Our simple model with our a_1, a_2, a_3, a_4, a_5 subpopulations and updated parameters specific to each one produces the following simulations over the time frame of August 1-January 1:

Figure 20:
a_1 *Model vs. Data*



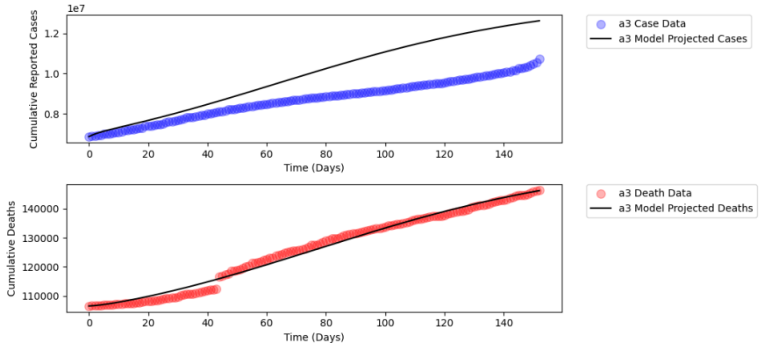
Cumulative Deaths $r^2 = 0.94$

Figure 21:
a_2 *Model vs. Data*



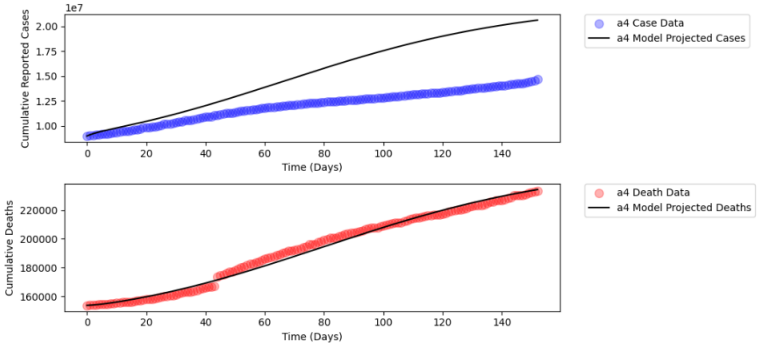
Cumulative Deaths $r^2 = 0.98$

Figure 22:
a_3 *Model vs. Data*



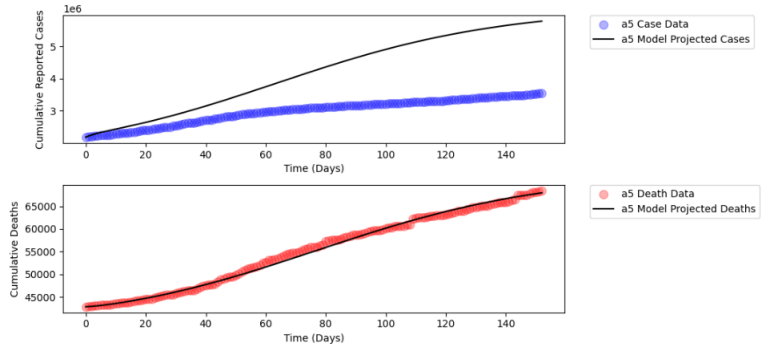
Cumulative Deaths $r^2 = 0.99$

Figure 23:
a_4 *Model vs. Data*



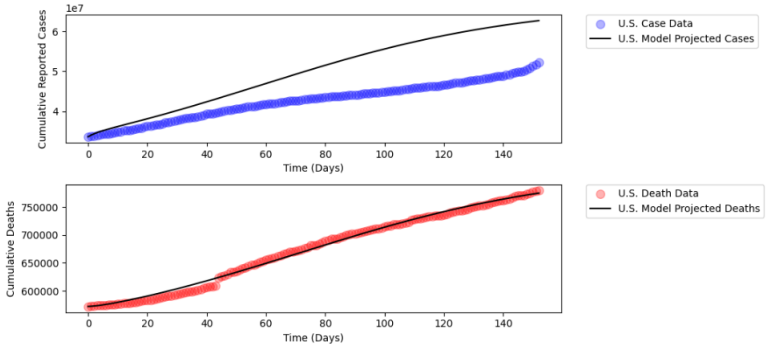
Cumulative Deaths $r^2 = 0.99$

Figure 24:
a_5 *Model vs. Data*



Cumulative Deaths $r^2 = 0.99$

Figure 25:
U.S. Model vs. Data



Cumulative Deaths $r^2 = 0.99$

4.4 Model with Eulerian Mobility

Figure 26
Data Modeling Simulations of Patch a_1 with Eulerian Mobility

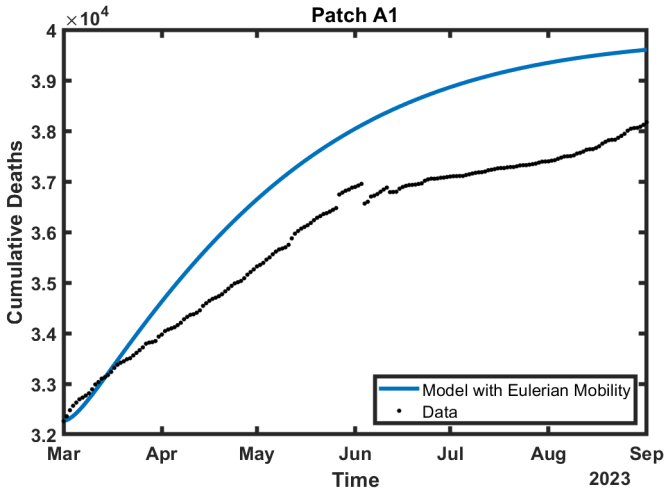


Figure 27
Data Modeling Simulations of Patch a_2 with Eulerian Mobility

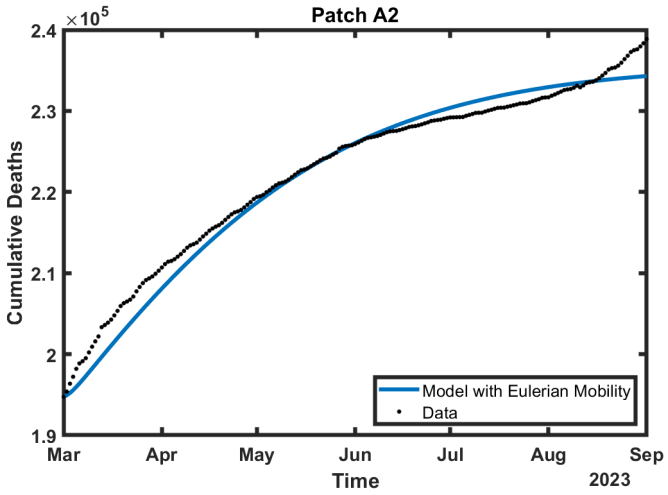


Figure 28
Data Modeling Simulations of Patch a_3 with Eulerian Mobility

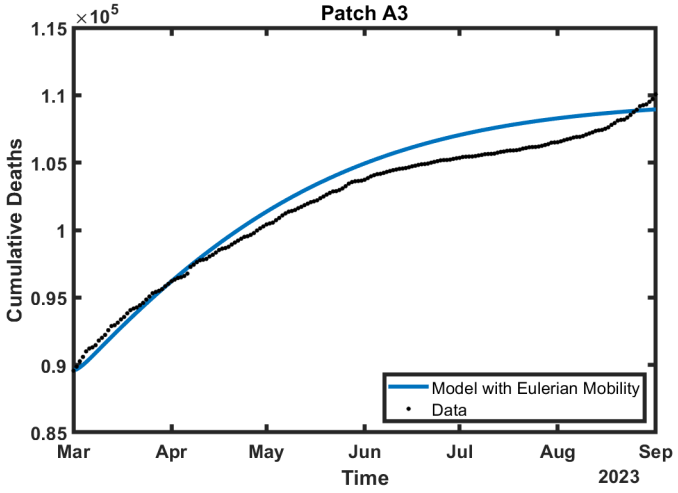


Figure 29
Data Modeling Simulations of Patch a_4 with Eulerian Mobility

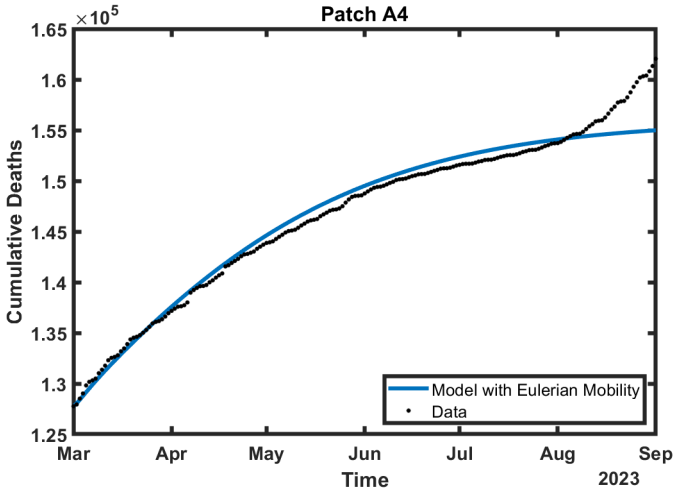
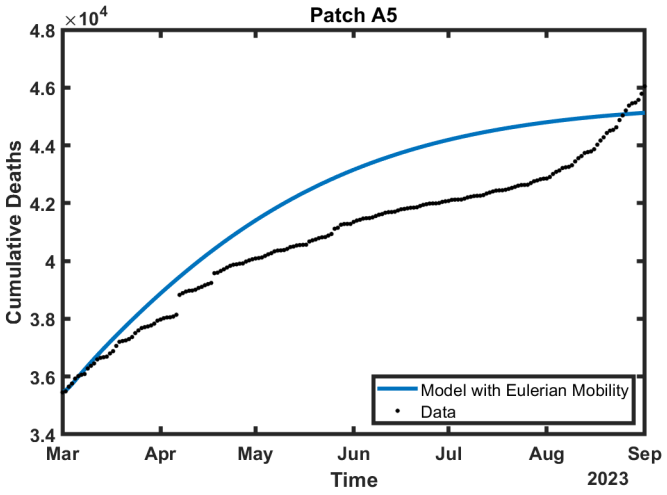


Figure 30
Data Modeling Simulations of Patch a_5 with Eulerian Mobility



4.5 Model with Lagrangian Mobility

In order to better understand Lagrangian mobility between patches, we first created simulations using two hypothetical patches. Patch 1 is similar to an a_1 county. It has a large population and high vaccination proportion. Patch 2 is similar to an a_5 county with a small population and low vaccination proportion.

Figure 31 shows a scenario in which there is low mobility from Patch 1 to Patch 2 and varying mobility from Patch 2 to Patch 1. The mobility from Patch 1 to Patch 2 remains low because urban counties are less likely to spend a higher proportion of time in rural counties.

Figure 31
Data Modeling Simulations of Two-Patch Mobility Scenario 1

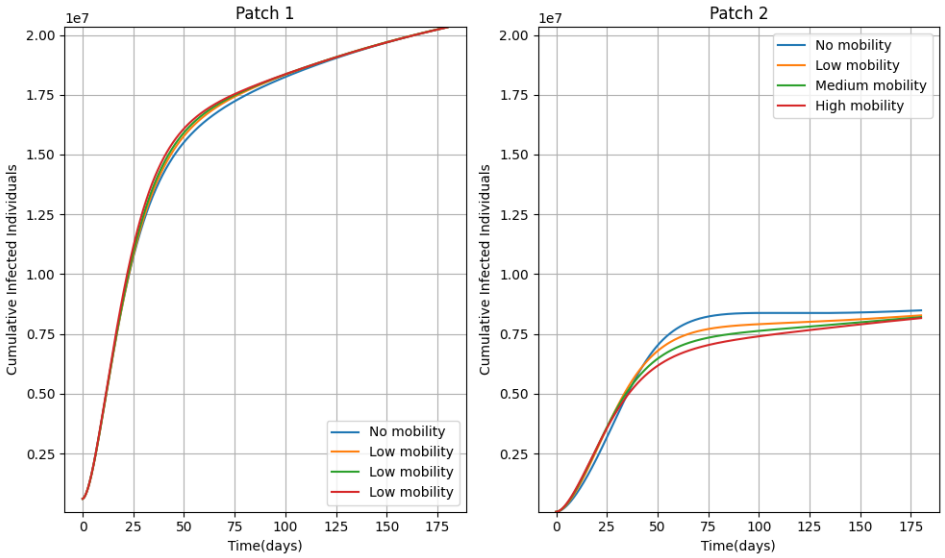
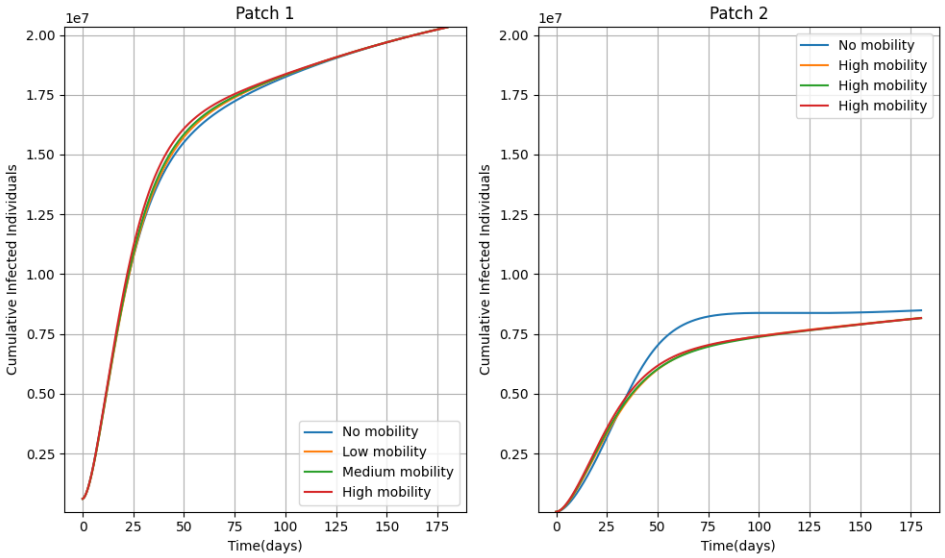


Figure 32 displays a scenario in which mobility from Patch 1 to patch 2 varies, but mobility from Patch 2 remains high. The mobility from Patch 2 to Patch 1 remains high because a rural county is more likely to have high mobility into an urban county than the inverse scenario.

Figure 32
Data Modeling Simulations of Two-Patch Mobility Scenario 2



Using the proportion of time matrix P, we data modeled patches a_i with Lagrangian mobility. Figure 33 models patch a_1, Figure 34 patch a_2, Figure 35 patch a_3, Figure 36 patch a_4, and Figure 37 patch a_5.

Figure 33
Data Modeling Simulations of Patch a_1 with Lagrangian Mobility

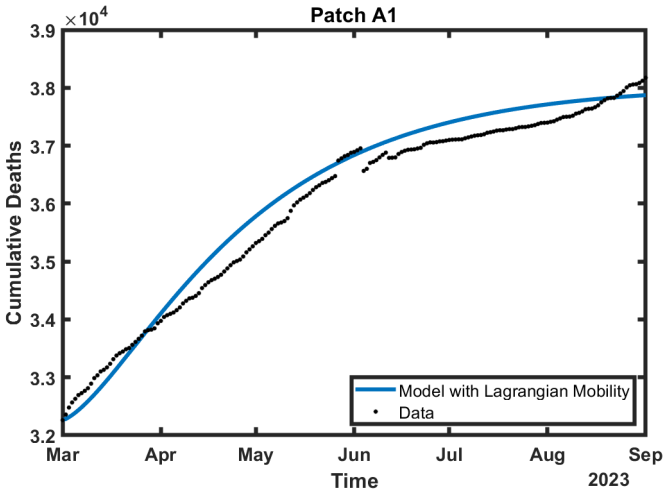


Figure 34
Data Modeling Simulations of Patch a_2 with Lagrangian Mobility

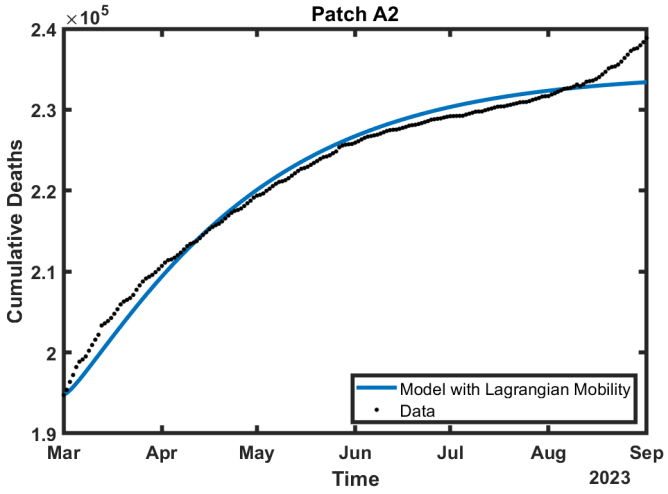


Figure 35
Data Modeling Simulations of Patch a_3 with Lagrangian Mobility

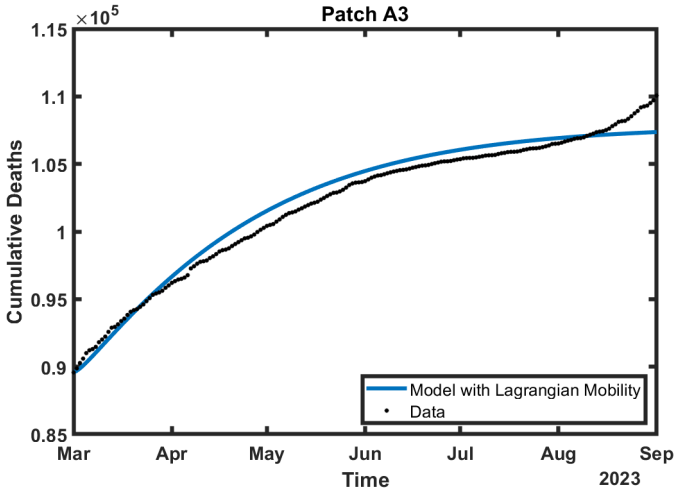


Figure 36
Data Modeling Simulations of Patch a_4 with Lagrangian Mobility

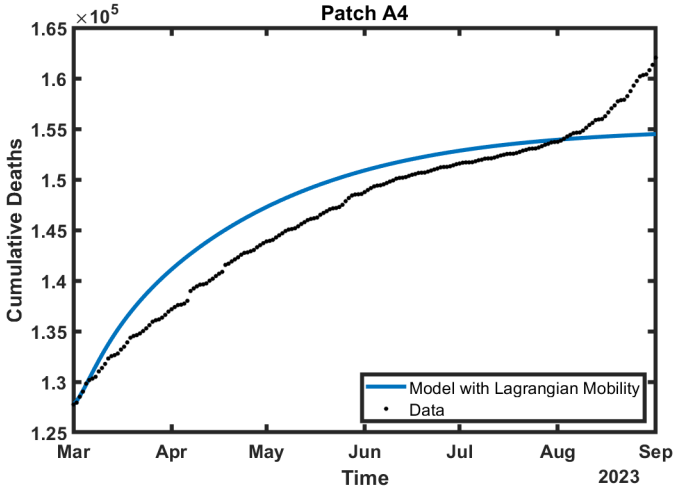
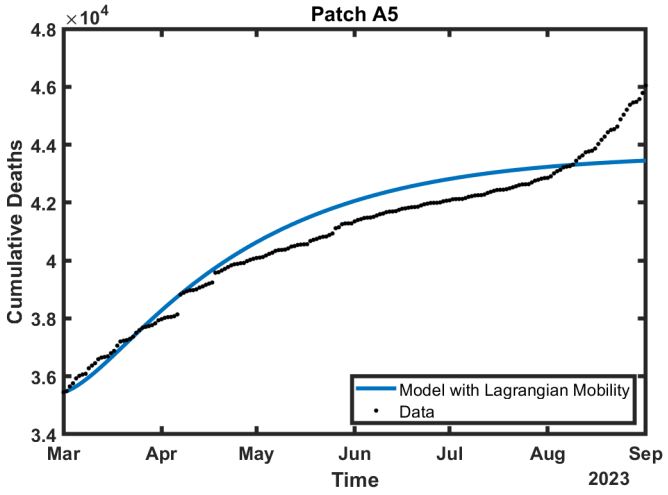


Figure 37
Data Modeling Simulations of Patch a_5 with Lagrangian Mobility



5 Discussion

5.1 Simple Model

The simple model performed well in fitting to cumulative deaths both by subpopulation and in the United States as a whole.

The r^2 of the one-population model is 0.990, while the r^2 of the sum-of-five-subpopulations model is 0.989. Given the individual subpopulations could have had higher or lower proportions of underreporting and symptomatic cases, more specific initial conditions could improve the fitting of the sum-of-five-subpopulations model to surpass the fitting of the one-population model. Regardless, we are encouraged by the strength of the fitting of the individual subpopulations and believe we can draw more thorough interpretations of how COVID dynamics in the United States through subpopulation modeling.

Table 2 shows the R_0 values for each patch a_i.

Table 2:

Subpopulation	R_0
a_1	0.011
a_2	0.011
a_3	0.012
a_4	0.015
a_5	0.020

We see these R_0 results as conservative due to assumptions about the longevity of vaccine effectiveness and antibodies that are applicable to only the time period we modeled. These R_0 numbers also do not factor in COVID variants that could change the rate of disease spread and vaccine efficacy. Mobility could also impact the R_0 of the U.S. beyond the R_0 of each subpopulation.

In most of the data plots, there are upticks in cases and deaths between August and September. We attribute this to the emergence of the Delta variant during this time. (Mathieu et al., 2021).

5.2 Simple Model - Delta Variant

The r^2 of the sum-of-five-subpopulations model for the Delta variant is 0.993. As a result, we are encouraged by our model's ability to model different time frames and COVID variants by changing parameters and initial conditions to reflect the data on each time frame and variant.

Table 3 shows R_0 values for each subpopulation a_i.

Table 3:

Subpopulation	R_0
a_1	0.052
a_2	0.061
a_3	0.071
a_4	0.076
a_5	0.084

We also see these R_0 results as conservative due to initial assumptions, but the difference between the R_0 values for the Delta variant time period and the original time period of interest shows that the Delta variant did impact COVID dynamics. Subsequently, we see initial spikes with each variant resulting from the different characteristics of each variant. Mobility could also impact the R_0 of the U.S. beyond the R_0 of each subpopulation.

5.3 Model with Eulerian Mobility

The model with Eulerian mobility performed poorly compared to the simple SEIAVR model. This is likely due to the K matrix we used to estimate mobility. The data we used from Kang et al. was daily mobility data. We used this to simulate migration between county patches. It seems unlikely that such high proportions of people are migrating between patches over a six-month time span. While the model with Eulerian mobility was not particularly successful, it does reinforce our assumption for the simple model that there was no immigration or emigration over this time period. Eulerian mobility might be more valuable over a longer time frame or on a national rather than county level, but it was not ideal for our modeling conditions.

5.4 Model with Lagrangian Mobility

To illustrate how Lagrangian Mobility impacts COVID dynamics, a two-patch model with Lagrangian mobility was derived. Higher mobility of the smaller patch members into the larger patch corresponds with fewer cases among the members of the smaller patch, while the number of cases among the members of the large population increases only slightly in comparison.

The full-scale, nationwide model performed very well for a_2 and a_3, and fairly well for a_1, a_4 and a_5. We feel the strong performance of the model indicates that incorporating Lagrangian mobility is useful in fine-tuning a U.S. epidemiology model.

As with the simple model and model with Eulerian mobility, the uptick in deaths between August and September 2021 is likely due to the emergence of the Delta variant.

5.5 Limitations

Our model is constrained by the six-month time window from March 1st, 2021 to September 1st, 2021. Future adaptations of this model would need to account for new variants of COVID, different transmission rates, decreasing vaccination effectiveness over time, and the addition of booster shots. This unfortunately means our model is not suitable for forecasting purposes. However, the core structure of the model and the county-based patches remain useful. Future COVID models can update which counties belong in each category a_i based on

the upcoming 2024 presidential election results to reflect ever-shifting political leanings in the U.S.

6 Conclusion

Altogether, we have created a county-based SEIAVR model that is good at modeling deaths for all subpopulations. Our model suggests underreporting factor of 30-100% for cases within our time frame across all subpopulations. Our model with Eulerian Mobility performs strongly for a_2 and a_4 but struggles with a_1 and a_5. The model with Eulerian mobility also performs more poorly than the simple model. We attribute this to flawed mobility assumptions given the data we had to work with. The model with Lagrangian Mobility generally performs well. It also illustrates how Lagrangian movement impacts COVID dynamics. The two-patch model is especially illuminating. Future models using a this structure will likely need to update initial parameters to account for decreasing vaccine effectiveness, booster shots, and new variants. Nevertheless, the county-based model provides a useful tool for analyzing subpopulation COVID dynamics in the U.S. and the impacts of politics on COVID dynamics.

Acknowledgement

We would like to thank Dr. Fabio Milner, Director of the Simon A. Levin Mathematical, Computational and Modeling Sciences Center (Levin Center), and Co-Director Dr. John Nagy for giving us the opportunity to participate in the Quantitative Research in the Life and Social Sciences program. We also recognize the work of the many administrative staff and tutors who supported this effort. This research was conducted as part of 2024 QRLSSP at the Levin Center (MCMSC) at Arizona State University (ASU) and has been partially supported by a grant from the National Science Foundation (NSF Grant FAIN-2150492), and by the Office of the President of ASU, and the Office of the Provost of ASU.

7 Code

`https://github.com/TIRTH4SHAH/QRLSSP-2024-A-County-Based-COVID-SEIAVR-Model`

References

Albrecht, D. (2022). Vaccination, politics and COVID-19 impacts. *BMC Public Health*, *22*(1), 96.

Alemi, F., & Lee, K. H. (2023). Impact of Political Leaning on COVID-19 Vaccine Hesitancy: A Network-Based Multiple Mediation Analysis. *Cureus*, *15*(8).

Angulo, F. J., Finelli, L., & Swerdlow, D. L. (2021, 01). Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys. *JAMA Network Open*, *4*(1), e2033706-e2033706. Retrieved from `https://doi.org/10.1001/jamanetworkopen.2020.33706` doi: 10.1001/jamanetworkopen.2020.33706

Arias, E., Tejada-Vera, Betzaida, Kochanek, K. D., and Ahmad, F. (2022). Provisional Life Expectancy Estimates for 2021. *National Center for Health Statistics*, *23*.

Brown, C. C., Young, S. G., & Pro, G. C. (2021). COVID-19 vaccination rates vary by community vulnerability: A county-level analysis. *Vaccine*, *39*(31), 4245–4249.

Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., Hossmann, S., Imeri, H., Ipekci, A. M., ... Low, N. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS medicine*, *17*(9), e1003346.

Çalıca Utku, A., Budak, G., Karabay, O., Güçlü, E., Okan, H. D., & Vatan, A. (2020). Main symptoms in patients presenting in the COVID-19 period. *Scottish medical journal*, *65*(4), 127–132.

Carey, J., Nyhan, B., Phillips, J. B., & Reifler, J. (2023). Partisanship Unmasked? The Role of Politics and Social Norms in COVID-19 Mask-Wearing Behavior. *Journal of Experimental Political Science*, *10*(3), 377–390.

Centers for Disease Control and Prevention. (2024). COVID Data Tracker. Retrieved from `https://covid.cdc.gov/covid-data-tracker`

Cole, W. M. (2023). Political Ideology and Childhood Vaccination in Cross-National Perspective, 1995 to 2018. *International Journal of Sociology*, *53*(4), 283–320.

Colonna, K. J., Nane, G. F., Choma, E. F., Cooke, R. M., & Evans, J. S. (2022). A retrospective assessment of COVID-19 model performance in the USA. *Royal Society open science*, *9*(10), 220021.

Cortés Martínez, J., Pak, D., Abelenda-Alonso, G., Langohr, K., Ning, J., Rombauts, A., ... Gómez Melis, G. (2022). SARS-Cov-2 incubation period according to vaccination status during the fifth COVID-19 wave in a tertiary-care center in Spain: a cohort study. *BMC infectious diseases*, *22*(1), 828.

Cunningham, G. B., & Nite, C. (2021). Demographics, politics, and health factors predict mask wearing during the COVID-19 pandemic: a cross-sectional study. *BMC Public Health*, *21*, 1–9.

Desmet, K., & Wacziarg, R. (2020). *Understanding spatial variation in COVID-19 across the United States* (Tech. Rep.). CEPR Discussion Papers.

Eikenberry, S. E., & Gumel, A. B. (2018). Mathematical modeling of climate change and malaria transmission dynamics: a historical review. *Journal of mathematical biology*, *77*(4), 857–933.

Elezkurtaj, S., Greuel, S., Ihlow, J. et al. (2021). Causes of death and comorbidities in hospitalized patients with COVID-19. Retrieved from `https://doi.org/10.1038/s41598-021-82862-5`

Eutsler, J., Harris, M. K., Williams, L. T., & Cornejo, O. E. (2023). Accounting for partisanship and politicization: Employing Benford's Law to examine misreporting of COVID-19 infection cases and deaths in the United States. *Accounting, Organizations and Society*, *108*, 101455.

Haischer, M. H., Beilfuss, R., Hart, M. R., Opielinski, L., Wrucke, D., Zirgaitis, G., ... Hunter, S. K. (2020). Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic. *PloS one*, *15*(10), e0240785.

Irons, N. J., & Raftery, A. E. (2021). Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proceedings of the National Academy of Sciences*, *118*(31), e2103272118. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.2103272118` doi: 10.1073/pnas.2103272118

Jayatilaka, R., Patel, R., Brar, M., Tang, Y., Jisrawi, N., Chishtie, F., ... Valluri, S. R. (2022). A mathematical model of COVID-19 transmission. *Materials Today: Proceedings*, *54*, 101–112.

Kang, Y., Gao, S., Liang, Y., Li, M., & Kruse, J. (2020). Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic. *Scientific Data*, 1–13.

Korhonen, V. (2024). *Birth rate in the United States from 1990 to 2021.* (`https://www.statista.com/statistics/195943/birth-rate-in-the-united-states-since-1990/` [Accessed: (27 June 2024)])

Kumar, N., Quadri, S., AlAwadhi, A. I., & AlQahtani, M. (2022). Covid-19 recovery patterns across alpha (b.1.1.7) and delta (b.1.617.2) variants of sars-cov-2. *Frontiers in Immunology*, *13*. Retrieved from `https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.812606` doi: 10.3389/fimmu.2022.812606

Lamb, C., Owens, C., Gamboa, W., & Lopez-Yunez, A. (2023). Evaluating Johnson and Johnson COVID-19 vaccination outcomes in a low-income hispanic population. *Vaccines*, *11*(1), 148.

Li, Y., Jiang, X., Qiu, Y., Gao, F., Xin, H., Li, D., ... Li, Z. (2024, March). Latent and incubation periods of Delta, BA.1, and BA.2 variant cases and associated factors: a cross-sectional study in China. *BMC Infect. Dis.*, *24*(1), 294.

Maged, A., Ahmed, A., Haridy, S., Baker, A. W., & Xie, M. (2023). SEIR Model to address the impact of face masks amid COVID-19 pandemic.

Risk analysis: an official publication for the Society of Risk Analysis, *43*(1), 129–143. doi: https://doi.org/10.1111/risa.13958

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., ... Rodés-Guirao, L. (2021). A global database of COVID-19 vaccinations. *Nature human behaviour*, *5*(7), 947–953.

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., ... Roser, M. (2020). Coronavirus Pandemic (COVID-19). *Our World in Data*. (https://ourworldindata.org/coronavirus)

Moderna. (2020). *Moderna's COVID-19 Vaccine Candidate Meets its Primary Efficacy Endpoint in the First Interim Analysis of the Phase 3 COVE Study.* (`https://investors.modernatx.com/news/news-details/2020/Modernas-COVID-19-Vaccine-Candidate-Meets-its-Primary-Efficacy-Endpoint-in-the-First-Interim-Analysis-of-the-Phase-3-COVE-Study/default.aspx` [Accessed: (25 June 2024)])

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the total environment*, *728*, 138884.

Monrad, J. T., Quaade, S., & Powell-Jackson, T. (2022). Supply, then demand? Health expenditure, political leanings, cost obstacles to care, and vaccine hesitancy predict state-level COVID-19 vaccination rates. *Vaccine*, *40*(45), 6528–6548.

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2021). Mortality in the United States, 2020.

Pathak, E. B., Garcia, R. B., & Menard, J. L., Janelle Mand Salemi. (2021, July). Out-of-Hospital COVID-19 Deaths: Consequences for Quality of Medical Care and Accuracy of Cause of Death Coding. *Am. J. Public Health*, *111*(S2), S101–S106.

Peter, O. J., Panigoro, H. S., Abidemi, A., Ojo, M. M., & Oguntolu, F. A. (2023). Mathematical model of COVID-19 pandemic with double dose vaccination. *Acta biotheoretica*, *71*(2), 9.

Pfizer. (2021). *Pfizer and BioNTech Confirm High Efficacy and No Serious Safety Concerns Through Up to Six Months Following Second Dose in Updated Topline Analysis of Landmark COVID-19 Vaccine Study.* (`https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-confirm-high-efficacy-and-no-serious` [Accessed: (25 June 2024)])

Pink, S. L., Chu, J., Druckman, J. N., Rand, D. G., & Willer, R. (2021). Elite party cues increase vaccination intentions among Republicans. *Proceedings of the National Academy of Sciences*, *118*(32), e2106559118.

Pormohammad, A., Zarei, M., Ghorbani, S., Mohammadi, M., Aghayari Sheikh Neshin, S., Khatami, A., Turner, D. L., Djalalinia, S., Mousavi, S. A., Mardani-Fard, H. A., Kasaeian, A., & Turner, R. J. ((2021)). Effectiveness of COVID-19 Vaccines against Delta (B.1.617.2) Variant: A Systematic Review and Meta-Analysis of Clinical Studies. *National Library of Medicine*. Retrieved from `https://doi.org/10.3390/vaccines10010023`

Rabinowitz, M., Latella, L., Stern, C., & Jost, J. T. (2016). Beliefs about Childhood Vaccination in the United States: Political Ideology, False Consensus, and the Illusion of Uniqueness. *PloS one*, *11*(7), e0158382.

Sandip Mandal, S. M., Sarkar, R., & Somdatta Sinha, S. S. (2011). Mathematical models of malaria-a review.

Sun, Y., & Monnat, S. M. (2022). Rural-urban and within-rural differences in COVID-19 vaccination rates. *The Journal of Rural Health*, *38*(4), 916–922.

Tamiru, D. H., Azene, A. G., Tsegaye, G. W., Mihretie, K. M., Asmare, S. H., Gete, W. A., ... Santangelo, O. E. (2023). Time to Recovery from COVID-19 and Its Predictors in Patients Hospitalized at Tibebe Ghion Specialized Hospital Care and Treatment Center, A Retrospective Follow-Up Study, North West Ethiopia. *Global Health, Epidemiology and Genomics*, *2023*, e4.

Taylor, K. M., Ricks, K. M., Kuehnert, P. A., Eick-Cost, A. A., Scheckelhoff, M. R., Wiesen, A. R., ... Sanchez, J. L. (2023). Seroprevalence as an Indicator of Undercounting of COVID-19 Cases in a Large Well-Described Cohort. *AJPM Focus*, *2*(4), 100141. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2773065423000780` doi: https://doi.org/10.1016/j.focus.2023.100141

The Editors of Encyclopaedia Britannica. (2022). *How is the Democratic Party different from the Republican Party?* (`https://www.britannica.com/question/How-is-the-Democratic-Party-different-from-the-Republican-Party` [Accessed: (24 June 2024)])

U.S. Bureau of Labor Statistics. (2022). Employee Tenure in 2022.

World Health Organization, T. (2021). *The true death toll of COVID-19: estimating global excess mortality.* (`https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality` [Accessed: (24 June 2024)])

World Health Organization, T. (2024). *Coronavirus disease (COVID-19).* (`https://www.who.int/health-topics/coronavirus#tab=tab_1` [Accessed: (2 July 2024)])

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... others (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*, *369*.

Xia, Q., Yang, Y., Wang, F., Huang, Z., Qiu, W., & Mao, A. (2024). Case fatality rates of COVID-19 during epidemic periods of variants of concern: A meta-analysis by continents. *International Journal of Infectious Diseases*, *141*, 106950. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1201971224000183` doi: https://doi.org/10.1016/j.ijid.2024.01.017

Yang, C., & Wang, J. (2021). Modeling the transmission of COVID-19 in the US – A case study. *Infectious Disease Modelling*, *6*, 195-211. Retrieved from `https://www.sciencedirect.com/science/article/pii/S246804272030110X` doi: https://doi.org/10.1016/j.idm.2020.12.006

Young, D. G., Rasheed, H., Bleakley, A., & Langbaum, J. B. (2022). The politics

of mask-wearing: Political preferences, reactance, and conflict aversion during COVID. *Social Science & Medicine*, *298*, 114836.