NBER WORKING PAPER SERIES

BEHAVIOR AND THE DYNAMICS OF EPIDEMICS

Andrew Atkeson

Working Paper 28760
http://www.nber.org/papers/w28760

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

This paper was prepared for the Spring 2021 Brookings Panel on Economic Activity. I would like to thank Michael Droste, Karen Kopecky, Ben Moll, Steven Shafer, James Stock, Tao Zha, and for many useful conversations. All errors are mine. The views expressed here are entirely my own and not official statements of the Federal Reserve Bank of Minneapolis or the Federal Reserve. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Behavior and the Dynamics of Epidemics
Andrew Atkeson
NBER Working Paper No. 28760
May 2021
JEL No. C0,I0

ABSTRACT

I use a model of private and public behavior to mitigate disease transmission during the COVID pandemic over the past year in the United States to address two questions: What dynamics of infections and deaths should we expect to see from a pandemic? What are our options for mitigating the impact of a pandemic on public health? I find that behavior turns what would be a short and extremely sharp epidemic into a long, drawn out one. Absent the development of a technological solution such as vaccines or life-saving therapeutics, additional public health interventions suffer from rapidly diminishing returns in improving long-run outcomes. In contrast, rapidly implemented non-pharmaceutical interventions, in combination with the rapid development of technological solutions, could have saved nearly 300,000 lives relative to what is now projected to occur.

Andrew Atkeson
Bunche Hall 9381
Department of Economics
UCLA
Box 951477
Los Angeles, CA 90095-1477
and NBER
andy@atkeson.net

A data appendix is available at http://www.nber.org/data-appendix/w28760

1 Introduction

During the first half of the Twentieth Century, Americans enjoyed tremendous gains in health and life-expectancy as infectious diseases were drastically curtailed thanks to major medical advances and significant investments in sanitation and public health. Annual mortality rates from infectious disease in the United States fell by an order of magnitude: from nearly 800 per 100,000 in 1900 to under 50 per 100,000 by 1960, in a steady downward trend, interrupted, dramatically, by the 1918-1919 Great Influenza.[1] But as the HIV/AIDS pandemic made evident, and the COVID-19 pandemic reinforced, infectious diseases are far from vanquished.[2] In fact, the risk that we experience another pandemic in the not too distant future is considerable. For example, according to a September of 2019 estimate by the President's Council of Economic Advisors, there is a four percent probability at an annual rate of a pandemic influenza that, at the high end, would cause in the United States nearly $4 trillion in economic damage and over half a million deaths.[3]

Given that we are likely to see significant outbreaks of infectious disease in the future, this moment, after a year of COVID-19, seems an opportune time to re-examine our models of disease dynamics and the policy options for disease control implied by these models. What dynamics of infections and deaths should we expect to see from a pandemic? What are our options for mitigating the impact of a pandemic on public health? How might this mitigation be done in a manner to reduce the nega-

[1]See Armstrong, Gregory L., Conn, Laura A., Pinner, Robert W. "Trends in Infectious Disease Mortality in the United States During the 20th Century" JAMA. 1999;281(1):61-66. doi:10.1001/jama.281.1.61 https://jamanetwork.com/journals/jama/fullarticle/768249. To place the mortality from COVID-19 in historical perspective, note that COVID mortality in the United States was roughly 100 in 100,000 in 2020 and may very well reach this level again in 2021. So while mortality from COVID will not reach the levels reached during the Spanish Flu, it will clearly be the most significant short term increase in mortality from infectious disease in the United States in at least 60 years.

[2]See Morens, David M. and Fauci, Anthony S. "Emerging Pandemic Diseases: How we got to COVID-19" Cell 182, September 3, 2020 https://doi.org/10.1016/j.cell.2020.08.021

[3]See this September 2019 report "Mitigating the Impact of Pandemic Influenza through Vaccine Innovation" from the President's Council of Economic Advisors on the potential public health and economic impact of pandemic influenza. https://www.hsdl.org/?view&did=831583

tive impact of a pandemic on the economy? These are questions that will provoke new research in the light of worldwide data from this COVID pandemic for years to come. But, after one year of data on COVID, one conclusion seems clear: the endogenous response of both public and private behavior to the prevalence of COVID-19 has transformed this epidemic from what standard epidemiological models predicted to be a short, but exceedingly intense, episode into a drawn out pandemic that will impact public health and the economy over several years, until, with luck, the technological solutions of vaccination an life-saving therapeutics[4] brings this disease under much greater control worldwide.

In this paper, I use a simple model of our experience with COVID-19 in the United States over the past year to explore how the interaction of disease and behavior changes the dynamics of an epidemic and constrains our options for mitigating the impact of a pandemic on public health absent a technological solution such as vaccines and life-saving therapeutics. Based on this model, I present four conclusions.

1. The behavioral responses that we have seen to COVID over the past year, both private and public, have had a powerful impact in "flattening the curve", reducing peak levels of daily infections and deaths by an order of magnitude relative to predictions of standard epidemiological models. These behavioral responses, however, are forecast to have only a modest impact in reducing the long-term death toll from COVID relative to predictions of standard epidemiological models in the absence of the development of technological solutions such as vaccines or life-saving therapeutics. Absent such technological solutions, the long run death toll in the United States would approach 1.25 million, even with the private and public efforts at mitigation that have been undertaken. Thus, without the success of vaccines that we have now experienced, here in the United States, we would be roughly halfway through this

[4]It is clear that the development of vaccines for COVID has been a technological marvel. There also appears to be considerable promise for the development of therapeutics that could substantially reduce the severity of the disease and thus complement vaccines in bringing the pandemic to an end worldwide. See https://science.sciencemag.org/content/371/6530/681 for a description of such therapeutics.

pandemic in terms of cumulative deaths. Moreover, absent the development of technological solutions, my model implies sharply diminishing returns to further non-pharmaceutical interventions in reducing the long-run death toll from COVID even if such measures were implemented early in 2020 and maintained for a long, but finite, length of time. Thus, absent a technological solution, we would be faced with few options for further mitigating the long run impact of COVID on public health.

2. Here in the United States, we have been very fortunate with our success in developing and now implementing effective vaccines against COVID. With vaccines, the long-run death toll from COVID is forecast to be roughly 600,000, or half the level forecast without such a technological solution. This forecast takes into account both the relaxation of private and public efforts at disease control that we have seen this Spring and the introduction of new, more contagious variants of the virus. Clearly, Operation Warp Speed and the associated research effort has been a scientific and public health achievement of historic importance.

3. In contrast to the case if no technological solution were developed, strong non-pharmaceutical interventions implemented early on are highly complementary with speedy development of vaccines and life-saving therapeutics in that they save lives by delaying illness and death until such technological solutions are available. This model forecast that plausible additional non-pharmaceutical interventions, applied early on and consistently over time on top of the policies that were implemented at state and local levels, could have reduced the long-term death toll from COVID in the United States to roughly 300,000. This forecast takes into account the likely countervailing relaxation of private and state and local mitigation efforts had such interventions been implemented at a Federal level. Based on this forecast, I conclude that here in the United States, over the course of the past fourteen months, we failed to take actions that would have saved hundreds of thousands of lives. Given the success of a number of countries in containing COVID over the past year while preserving

economic activity, it is entirely plausible that such non-pharmaceutical interventions would not have led to high economic costs and, in fact, might have led to better economic outcomes.

4. Finally, looking ahead, we face a future in which COVID will remain a threat as long as it is prevalent elsewhere in the world and in which new pandemic threats will likely arise. We should use the worldwide experience with COVID to guide investments in public health infrastructure that will allow us to rapidly identify and react to pandemic threats with more effective and less costly non-pharmaceutical interventions. Such investments have a strong "public good" rationale and would be highly complementary with increased investments in the scientific and clinical research infrastructure to rapidly develop technological solutions such as vaccines and life-saving therapeutics for future threats from infectious disease.

2 Epidemic dynamics with and without behavior

The public health policies enacted around the world to combat COVID-19 have been guided by standard epidemiological models built on the SIR framework developed by Kermack and McKendrick.[5] These models simulate disease transmission as arising when infectious individuals (corresponding to the I in SIR) interact with others. Through this interaction, a virus or other pathogen succeeds in infecting those who have no immunity and are thus susceptible (corresponding to the S in SIR), turning such agents into newly infectious individuals. Individuals who gain immunity from prior infections or vaccinations are said to be removed (corresponding to the R in SIR) as they no longer contribute to the transmission of the disease. The progress of the epidemic through the population is mechanical as the rate at which infectious people interact with others is assumed to be invariant to the current prevalence of

[5]Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." *Proc. Roy. Soc. Lond. A* 115, 700-721, 1927.

the disease.

When applied to COVID-19, three quantitative implications of this standard model stand out.[6] First, the model gives dire forecasts for the peak of the disease's first wave — 10 to 20% of Americans were predicted to be sick with COVID-19 simultaneously at the first peak of infections absent drastic efforts such as lockdowns to slow transmission. At current estimates of the infection fatality rate for COVID-19, this rate of infection would have corresponded to peak death rates on the order of 30,000 to 60,000 deaths per day.[7] Second, this standard model forecast that if efforts to slow transmission through lockdowns were applied early on but were only temporary, this dramatic first peak would be delayed but not prevented: cases and deaths would explode again once efforts to slow transmission were relaxed. Third, this standard model offered dramatic long-run predictions made famous by Angela Merkel in March of 2020[8] —more than 2/3 of the population were forecast to experience infections (if not vaccinated) before the pandemic would end through herd immunity. Again, applying current estimates of the average infection fatality rate for COVID-19 in the US, this implies a long-run death toll on the order of 1.49 million or more.[9]

These implication of a standard epidemiological model for the magnitude of the first peak and the long-run impact of COVID-19 in terms of infections are driven by a single parameter known as the basic reproduction number of the virus (the $\mathcal{R}_0$).[10]

[6]See https://www.nber.org/papers/w26867 and https://www.nber.org/system/files/working_papers/w26902/w26902.pdf for expositions of these predictions of standard SIR models from one year ago.

[7]This estimate for peak deaths is likely understated given that such a wave of infections would clearly have overwhelmed the health care system.

[8]See https://www.bundesregierung.de/breg-de/themen/coronavirus/statement-chancellor-1732296 and https://edition.cnn.com/world/live-news/coronavirus-outbreak-03-11-20-intl-hnk/h_ab9bb8236fa91a9bf63cdbc7a69e0f10

[9]This forecast for the cumulative death toll in this model scenario does not take into account that the infection fatality rate would likely have risen substantially due to congestion in the health care system if the first wave of infections had approached anything close to the levels forecast by this standard model.

[10]See https://www.cell.com/immunity/fulltext/S1074-7613(20)30170-9 for a description of the calculations and considerations involved.

The implications of these infections for deaths from COVID-19 are determined by the average infection fatality rate across the infected population. While we now know that the infection fatality rate from COVID-19 varies widely with age and other factors, estimates of the disease burden from COVID-19 from the CDC are consistent with an average infection fatality rate of 0.5% across the entire infected population in the United States for 2020.[11] The emergence of new, more transmissible, virus variants with higher basic reproduction numbers make the predictions of standard epidemiological models for peak infections and long run impact even more dire.

It is now clear that the first prediction of standard epidemiological models for the first peak of infections and deaths were off by at least an order of magnitude — it is unlikely that more than 2 percent of Americans have ever been infected simultaneously, and the peak of daily deaths in America from COVID-19 has fortunately stayed under 4000. Looking at data worldwide, it appears that the second prediction of standard epidemiological models is also off perhaps by an order of magnitude. While many locations within the United States and abroad have suffered severe second or third waves of COVID-19 deaths after relaxing costly public measures to control disease transmission, these waves have been much smaller than predicted by a standard SIR model. In contrast, the standard SIR model's third prediction, regarding long-run impact, looks to be closer to the mark. While the precise threshold of herd immunity—the fraction of the population that has to gain immunity through infection or vaccination before the pandemic can end—is not yet empirically resolved, available data from locations such as Manaus, in Brazil, which has experienced high rates of infection, and from Israel, the United Kingdom, and the United States, each of which now have high vaccination rates with effective vaccines, indicate that the predictions of a standard epidemiological model for the long-run impact of COVID-19 are likely correct: this pandemic will not resolve until high proportions of the population have acquired immunity either through infection or vaccination.[12]

[11]The CDC estimates that 83 million Americans had been infected by the end of December 2020. Total COVID deaths reached 445,000 30 days later, giving an average estimated infection fatality rate including the delay from infection to death of slightly over 0.005. See https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html

[12]See https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00183-5 and https:

2.1 Behavior regulates disease dynamics

How does consideration of the impact of behavior on the progression of a pandemic help us understand this relationship between the predictions of a standard SIR model and observed outcomes? Within economics, Tomas Philipson pioneered the study of the interaction of behavior and the spread of disease in his work on the HIV/AIDS pandemic. In a 1999 handbook chapter summarizing work on that pandemic[13], Philipson argued that epidemiological models should incorporate prevalence-elastic private demand for costly measures to prevent of the spread of infectious disease. Such models, he maintained, offered two fundamental economic insights.

The first insight is that costly private efforts to prevent disease transmission are self-limiting—as disease incidence falls, these costly efforts to control disease spread are relaxed and the disease re-emerges.

Within the United States, it appears that this observation holds for public policies aimed at COVID-19 as well — state and local disease control measures are often conditioned on measures of disease prevalence such as infections or hospitalizations, and these public measures aimed at the control of COVID-19 are relaxed as disease prevalence falls. In the model I present below, I interpret this correlation between public policies and disease prevalence as arising from a public behavioral response to shifting political calculations as disease prevalence rises and falls, that is, as a social-choice behavioral response that is conceptually similar to private behavioral responses. I thus interpret the reduced form behavioral response of transmission rates to disease prevalence in my model as resulting from a combination of private and public reactions to disease prevalence.

The second insight is that the private and public behavioral response to changing

//www.nature.com/articles/d41586-021-00316-4 regarding data from Manaus and Israel on the empirical herd immunity threshold. The CDC COVID data tracker site reports on vaccination rates in the United States. Data on vaccination rates in the United Kingdom are available here https://coronavirus.data.gov.uk/details/vaccinations

[13]NBER working paper 7037 published here https://www.sciencedirect.com/science/article/abs/pii/S1574006400800463

disease prevalence partially offsets the impact of additional non-pharmaceutical interventions aimed at disease control. In short, the effect of a specific non-pharmaceutical intervention is limited by its success as private and public efforts aimed at disease control are relaxed in response.

That both public and private prevalence-elastic demand for costly measures to control disease is self-limiting is a particularly powerful insight for understanding where the standard epidemiological model fails as a description of disease dynamics and where it succeeds. In joint work with Karen Kopecky and Tao Zha[14], I find that the data on the progression of the COVID-19 pandemic across many countries and U.S. states throughout 2020 conform strikingly well with a core prediction of the standard epidemiological model modified to include prevalence-elastic demand for disease prevention—that after the first phase of the pandemic in which disease grows rapidly, the growth rates of infections and deaths should remain in a relatively narrow band around zero until the pandemic is over.[15]

The intuition for this prediction regarding disease dynamics in the context of a model with prevalence-elastic demand for disease prevention is simple. If new infections and daily deaths from the disease grow too high, people and governments take costly efforts to avoid interaction and thus slow disease spread. Likewise, if the prevalence of the disease falls, people and government relax those costly efforts at disease prevalence and the prevalence of the disease rises again. The reaction of behavior, both public and private, to the prevalence of the disease regulates the equilibrium prevalence of the disease in the same way that a cruise control regulates the velocity of a car on the highway that winds up and down hills. The equilibrium level of daily deaths, corresponding in this analogy to the velocity of the car, remains within a relatively narrow band (relative to that predicted by a standard SIR model) in response to shocks impacting disease transmission because of the stabilizing role of endogenous prevalence-elastic public and private disease avoidance behavior. The

[14] https://www.minneapolisfed.org/research/staff-reports/behavior-and-the-transmission-of-covid19

[15] Joshua Gans reviews the implications of epidemiological models with a prevalence-elastic demand for costly measures to prevent disease transmission and much of the work by others on this topic in https://www.nber.org/papers/w27632.

impact of this behavior then is to transform what would otherwise be a short and sharp disease episode into a much more slowly evolving and drawn-out phenomenon.

What are the implications of a model with prevalence-elastic demand for disease prevention for the long run impact of an epidemic? Here the insight that the demand for disease prevention is self-limiting is particularly relevant. For an epidemic to end, the prevalence of the disease must fall towards zero. As disease prevalence falls towards zero, the demand for costly disease prevention efforts also falls towards zero, and hence the disease will come back unless the population has already achieved herd immunity measured at pre-pandemic levels of behavior. That is, the predictions for the long-run impact of COVID-19 using a standard epidemiological model should continue to hold.[16] Given estimates of the basic reproduction number in the range of 3 or now higher with new variants, this herd immunity threshold should kick in when significantly less than one-third of the population remains susceptible.

This logic implies that, absent a vaccine or the development of life-saving therapeutics, the implications of a model that includes a prevalence-elastic demand for disease prevention for the long-run impact of a pandemic in terms of cumulative infections and deaths should be similar to that of a standard epidemiological model. The slowing of the epidemic that results from a behavioral response to disease prevalence can reduce the cumulative death toll by reducing the extent to which cumulative infections in the long run overshoot the herd-immunity threshold, but this behavioral response does not reduce the cumulative impact of the epidemic to a point below this threshold. In the case of COVID-19 in the United States, in the model I present below, this would be a cumulative death toll on the order of 1.24 million.

[16]More complex models that emphasize heterogeneity and the network structure of human interaction potentially offer more optimistic implications for the long-run impact of COVID. See, for example https://www.nber.org/papers/w27373, https://www.nber.org/papers/w27374, https://www.nber.org/papers/w27741, and https://www.nber.org/papers/w28282.

2.2 A Quantitative Illustration

To illustrate these points regarding the predictions of a standard epidemiological model and one with a prevalence elastic demand for disease prevention for the dynamics of an epidemic, I turn to a model of the dynamics of deaths from the COVID-19 epidemic in the United States that I presented in a recent working paper[17] and which is included as an online appendix to this paper . This model accounts for the dynamics of deaths from COVID-19 in the United States over the past year with shocks to transmission rates due to seasonality, due to the emergence of a new, more transmissible, variant of the novel coronavirus, and due to potential changes in the prevalence-elasticity of demand for costly measures to mitigate disease transmission. (I refer to this third shock as "pandemic fatigue" as a short-hand description of a decline in the responsiveness of private and public demand for costly disease prevention measures to changes in disease prevalence. This shock is perhaps a reduced form for a more dynamic response of behavior as a pandemic wears on.)

This model accounts remarkably well for the pandemic's evolution in the United States over the past year. In the online appendix to this paper, I document that, in the model, a seasonal decline in transmission rates explains why the prevalence of COVID-19 dropped to relatively low levels in the summer of 2020. In the model, a decline in the strength of the behavioral response to disease prevalence in late fall — "pandemic fatigue" —explains the large waves of infections and deaths seen in the late fall and winter. The introduction of a more transmissible variant in early December together with the start of an aggressive vaccination program explain the progress of the epidemic this Spring.[18]

In Figure 1, I show the model's prediction (in a solid blue line) for daily deaths

[17] https://www.nber.org/papers/w28434

[18] In the online appendix to this paper, I document the specific features of this model that allow it to fit the pattern of daily deaths observed over the past 14 months with relatively few shocks and discuss the procedure used to choose the model parameters. The fit of the model to the data is serendipitous. Further research is needed to develop behavioral models that can fit the wide range of experiences with COVID seen across regions and countries of the world.

from COVID-19 in the United States from mid-February 2020 to mid-February 2022, and data (in a red dashed line) on the seven-day moving average of daily deaths in the United States over the past year downloaded from the Center for Disease Control's (CDC's) COVID data tracker website.[19] The behavioral model matches the data on deaths over the past year quite well, and it forecasts, absent vaccines, a continuation of the pandemic well into 2022. The predicted long plateau of daily deaths through 2021 shown in this figure is driven by the spread of the new, more contagious, virus variant in the model. The long-run cumulative death toll in this forecast run of the model in Figure 1 is 1.24 million. The forecast shown in this figure does not include any consideration of the impact of vaccines, both to permit comparison with projections from a standard epidemiological model, and to serve as a benchmark for the impact of vaccination efforts.

To illustrate the impact of the behavioral response to disease prevalence in shaping the growth rate of the epidemic, in Figure 2, I show the model-implied growth rate of daily deaths from the simulation of the model shown in Figure 1. We see in this figure that the growth rate of daily deaths starts out at a very high level — above 30% per day — and then falls rapidly towards zero and hovers around zero even with shocks due to seasonality in transmission, pandemic fatigue, and the introduction of new variants. In the model, the response of private and public behavior to the level of daily deaths acts to slam the brakes on the growth of the epidemic in its initial phase and then maintain that growth rate of daily deaths in a narrow band around zero in the face of shocks to transmission much as a cruise control regulates the acceleration of a car on the highway.

[19] https://covid.cdc.gov/covid-data-tracker Note that these data on daily deaths omit roughly 14,000 deaths included in the CDC estimate of the cumulative death toll from COVID available on the same site as these additional deaths were included retroactively due to reclassification of state and local death counts.

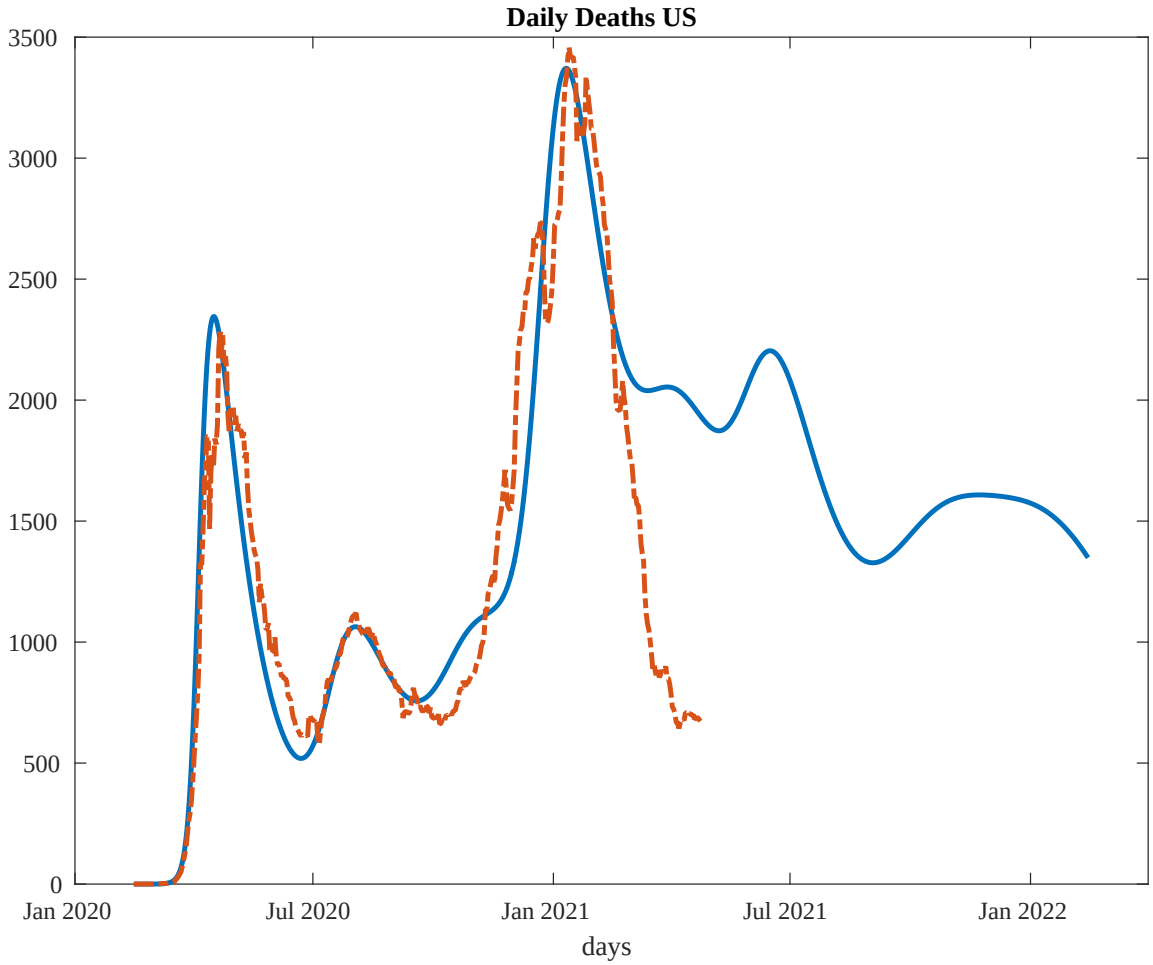

Figure 1: Behavioral model implications for daily deaths in the United States from mid-February, 2020 through mid-February 2022 are shown in a solid blue line. Transmission rates in the model are impacted by seasonal variation, the introduction of a more contagious variant in December of 2020, and prevalence elastic demand for costly measures to slow disease transmission. The onset of pandemic fatigue in late 2020 accounts for much of the peak of deaths late that year. The long plateau of daily deaths forecast to occur in 2021 is driven in the model by the introduction of the new, more contagious variant of the virus. Data on the seven-day moving average of daily deaths in the United States over the past year are shown in a red dashed line. The forecast for cumulative deaths over the long-run implied by this model is 1.24 million.

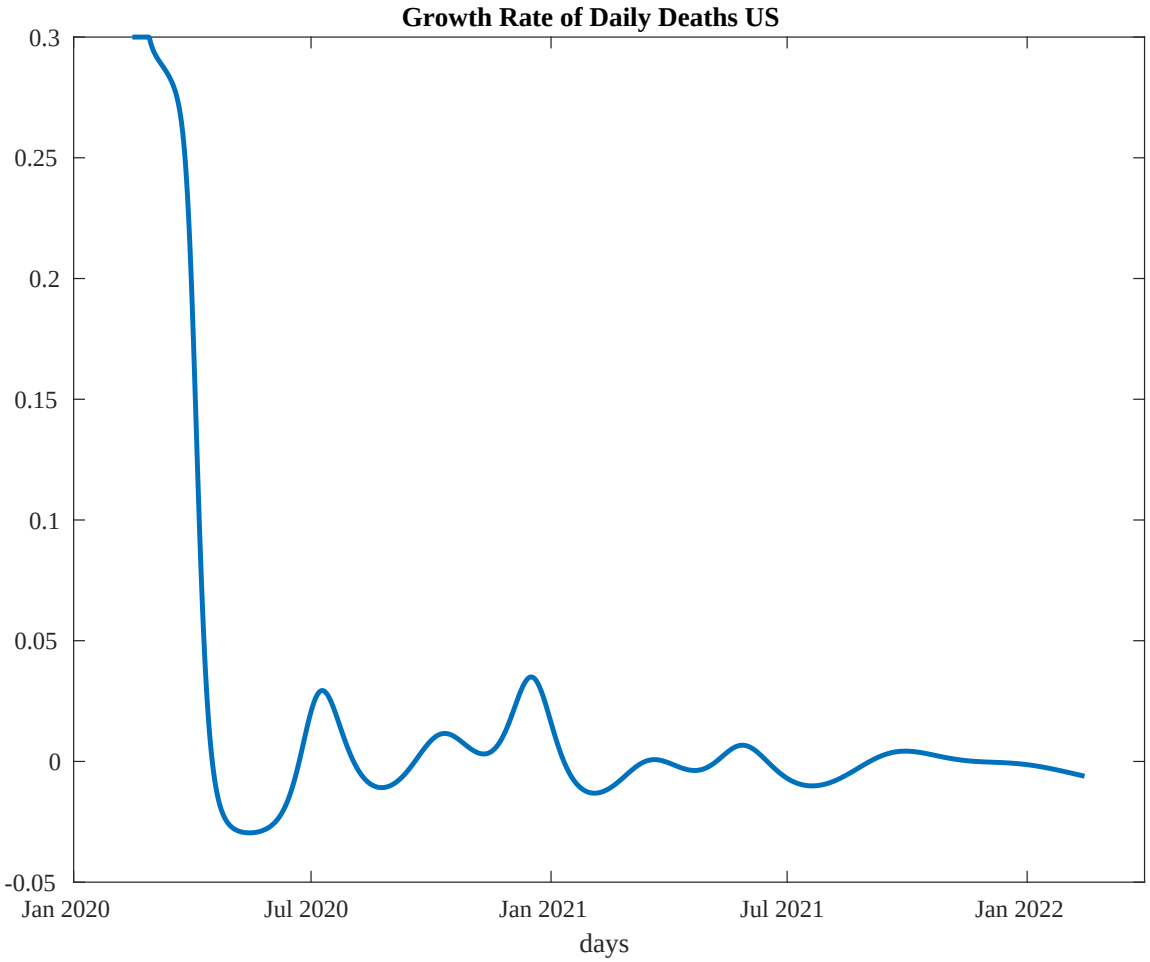

Figure 2: The growth rate of daily deaths implied by the solution of the model shown in Figure 1. In this figure we see that behavior closely regulates the growth rate of the epidemic after its initial phase of rapid growth.

To contrast the implications of this model with a behavioral response to disease prevalence with those of a standard model without such a response, in Figure 3, I show the prediction for daily deaths of the same model with the behavioral response to disease prevalence turned off (in a blue solid line), relative to data on the seven-day moving average of daily deaths (in a red dashed line). As we see in this figure, this standard epidemiological model without a behavioral response overstates the first peak of daily deaths by at least an order of magnitude (these peak at over

30,000/day), but then the pandemic comes quickly to an end in the fall of 2020. The cumulative death toll in this model forecast is 1.49 million. This prediction for the cumulative death toll is certainly larger than in the model with a behavioral response, but the gap between the two models in this dimension is much smaller than in their predictions for the initial peak and the time scale of the pandemic.[20]

What is evident from these figures is that incorporating a response of public and private behavior to disease prevalence gives a dramatically different forecast for the severity of disease peaks as well as for the speed with which this epidemic passes through the population. In this behavioral model, absent the introduction of vaccines, the pandemic takes two-and-a-half years to play out rather than six-to-nine months as forecast by the standard model without consideration of behavior. The model's implications, however, for the long-run impact of the disease are not much altered by the consideration of behavior. In both basic and behavior variations, the model forecasts that a substantial majority of the population must become immune through infection or vaccination for the pandemic to end.

[20]This difference between the cumulative death toll forecast in the model run in Figure 3 and that in Figure 1 is due to what is known as "overshooting" of herd immunity in the model without behavior in Figure 3. See https://www.nytimes.com/2020/05/01/opinion/sunday/coronavirus-herd-immunity.html for an explanation of this concept.

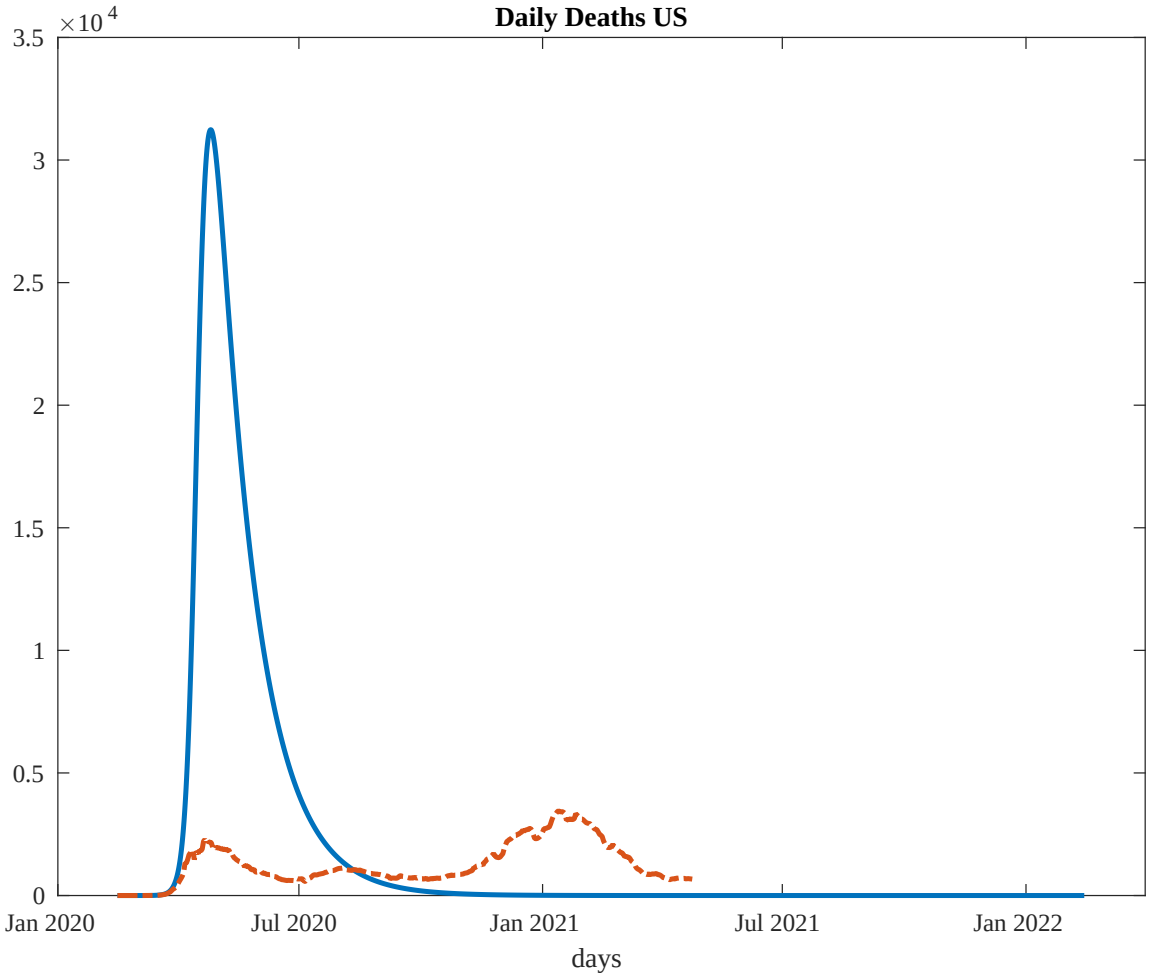

Figure 3: Standard model implications for daily deaths in the United States from mid-February, 2020 through mid-February 2023 are shown in a blue solid line. Transmission rates in the model are impacted by seasonal variation, the introduction of a more contagious variant in December of 2020, but this specification of the model has no prevalence elastic demand for costly measures to slow disease transmission. Data on the seven-day moving average of daily deaths in the United States over the past year are shown in a red dashed line. The forecast for cumulative deaths over this three-year period is 1.49 million.

3 Private Behavior and Constraints on Policy

Given these insights on the impact of prevalence elastic demand for disease prevention on the dynamics of an epidemic, what are our options for using public policy to mitigate the impact of a pandemic on public health? One insight that I have already mentioned is that there is likely to be an offsetting private behavioral response to public measures that limit the spread of disease — that is, that additional non-pharmaceutical interventions to control an epidemic may well be partially undone by private responses and the responses of other government actors to declining disease prevalence. The other insight is that public measures at disease prevention must be essentially permanent to result in a meaningful reduction of the long-run impact of an epidemic absent a technological solution such as a vaccine or the development of life-saving therapeutics.

We can use our simple behavioral model to illustrate the quantitative implications of these two insights. Imagine that through public policies facilitating a wide range of disease control measures such as masking and social distancing protocols, testing and contact tracing with isolation of the infectious, and other measures, it was possible to significantly reduce the transmission rate of COVID-19, holding fixed seasonality and the level of costly disease control measures undertaken by both private agents and state and local authorities. Imagine that these policy interventions are undertaken for a fixed period of time independent of disease prevalence. In this sense, I imagine that these interventions are undertaken independently of the political process that leads currently observed public interventions to rise and fall with disease prevalence. Here, for purposes of illustration, I imagine these interventions as being carried out by the Federal government.

In Figure 4, I show a simulation of the model with such measures put in place for a two-year period from May 1, 2020 through May 1, 2022. Here I assume that these additional mitigation measures are put in place independently of the level of daily deaths and that they act to reduce disease transmission by 40% (a factor of $\exp(-0.5)$) on top of whatever reductions in transmission are brought about by

private and public changes in behavior undertaken in response to disease prevalence. I show the model implications for daily deaths over a four-year period as a solid blue line and the data on the seven-day moving average of daily deaths in a dashed red line. As we see in this figure, these disease control measures, when imposed on top of those arising in equilibrium from the prevalence-elastic demand of both private agents and public authorities for costly measures to control disease, have a significant impact in reducing deaths from the disease in the first year. Then, in this simulation, in early 2021, the arrival of the new variant and, in mid 2022 the abandonment of these disease control measures, leads to significant spikes in forecast deaths. Over the long run, the cumulative death tool is 1.22 million — almost exactly what we found in the simulation in Figure 1 that had no such additional disease control measures imposed. This simulation indicates sharply diminishing returns to additional non-pharmaceutical interventions absent a technological solution such as a vaccine or life-saving therapeutics.

3.1 Waiting for a technological solution

We saw in Figure 4 that additional but temporary disease control measures do not significantly reduce the long run public health impact of the epidemic in the absence of a technological solution such as vaccines or life-saving therapeutics. How does the analysis of the impact of such measures change when there is a good prospect that a vaccine or therapeutics might arrive? Here I use the model to show that such measures can have a significant long-run public health benefit in reducing deaths from disease while waiting for the arrival of that technological solution.

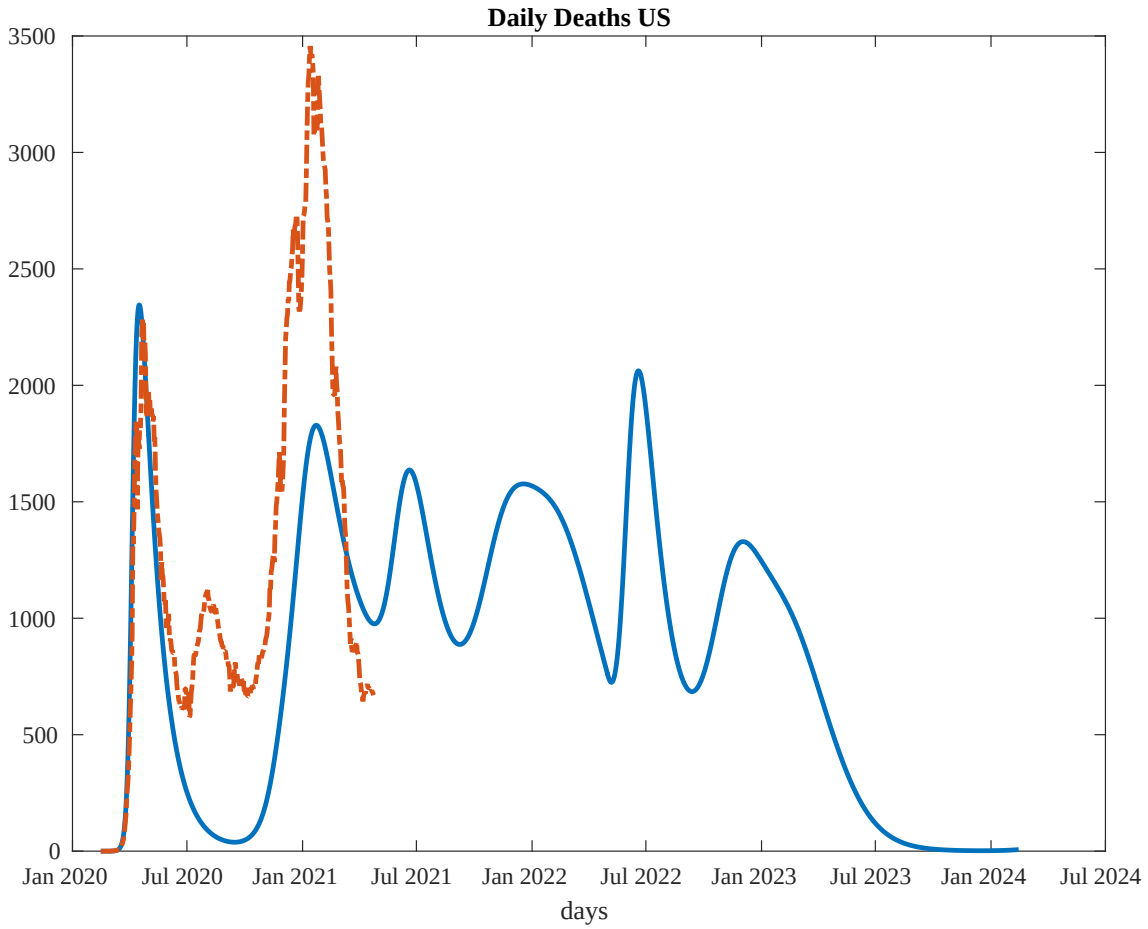

Figure 4: Predictions of the model for the evolution of daily deaths from COVID-19 in a version of the model in which disease control measures such as masks, social distancing, testing with contact tracing and isolation of the infected cut the transmission rate of the disease in by 40% (a factor of $\exp(-0.5)$) holding fixed the level of private and state and local disease control efforts undertaken in response to the prevalence of the disease. These measures are assumed to be in place for two years from May 1, 2020 to May 1, 2022. While these disease control measures are effective in reducing deaths in the first year, they do not succeed in later years. The data on the seven-day moving average of daily deaths are shown in red. The model-implied cumulative death toll is 1.22 million — nearly the same as we found in the simulation in Figure 1.

In Figure 5, I show the implications of the model for the evolution of daily deaths (in a solid blue line) when a program of vaccination starts on January 1, 2021 at a pace sufficiently fast to succeed in protecting half of the United States population by July 1, 2021. This vaccine is assumed to prevent both illness and disease transmission

by the vaccinated. The data on the seven-day moving average of daily deaths is again shown in a dashed red line. To see the model-implied impact of this vaccination program on the epidemic, one can compare the blue lines in Figures 1 and 5. Here we see that in the model, this vaccination program significantly reduces the forecast impact of the new variant later this Spring and brings the epidemic to an end late this summer or fall. Note that here the vaccination program succeeds despite the model-implied relaxation of public and private efforts at disease prevention in response to falling daily deaths.

The model predictions for the long-run death toll with this vaccination program is 595 thousand, less than half of what is forecast in the absence of a vaccine (in the simulations in Figures 1 and 4). In this sense, the vaccination program succeeds in substantially reducing cumulative deaths in a manner that a two-year program of disease mitigation absent a vaccine does not.

But now consider the model-implied scenario for cumulative deaths if the temporary disease mitigation measures used in the simulation in Figure 4 had been imposed starting May 1, 2020 and the same vaccination program applied in the simulation in Figure 5 had started on January 1, 2021. With this combination of temporary disease mitigation measures and a successful vaccination program, the cumulative death toll implied by the model would have been only 302 thousand. Clearly, the combination of temporary disease control measures applied while waiting for a technological solution can save many lives. The lesson here is that there are tremendous complementarities between early and aggressive mitigation and the development of a technological solution such as vaccines or life-saving therapeutics in terms of reducing the public health impact of a pandemic.

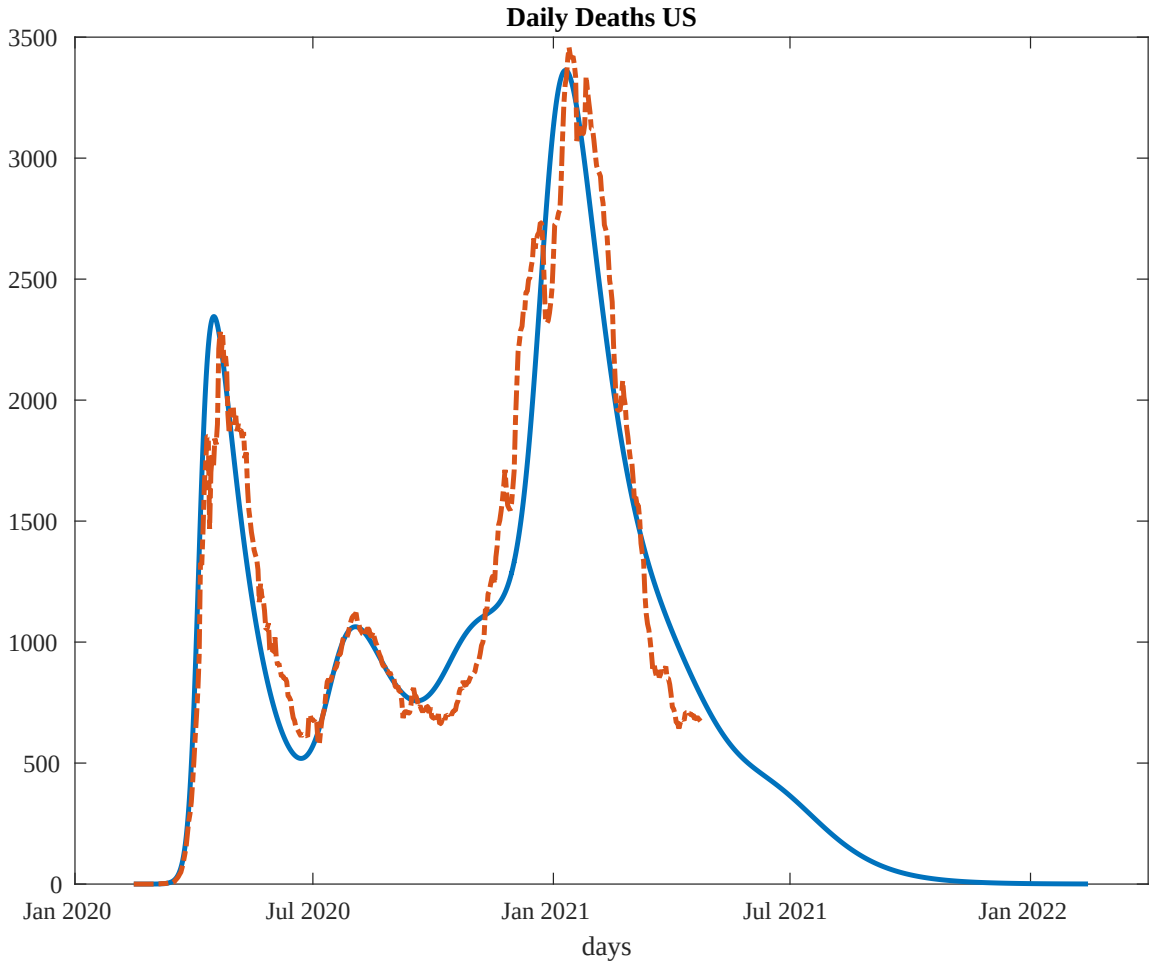

Figure 5: Predictions of the model for daily deaths from COVID-19 (in a solid blue line) in a simulation with a vaccination program starting on January 1, 2021 that proceeds at a rate fast enough to protect half of the population by July 1, 2021. In this simulation, the vaccine is assumed to protect against illness and to prevent disease transmission by the vaccinated. The data on the seven-day moving average of daily deaths are shown in a red dashed line. To see the predicted impact of the vaccine on the dynamics of the epidemic, compare the blue line for model-implied daily deaths 2021 in Figure 1 to the solid blue line here. The long-run cumulative death toll in this simulation is 595 thousand.

4 Conclusion

The global COVID pandemic has clearly demonstrated that the risks from the emergence of new infectious diseases that epidemiologists have been speaking about for years is terribly real. This pandemic has also posed a severe test of public health strategies and capabilities worldwide. In many countries, the associated economic impact has been as severe as any downturn seen since the Great Depression. How might we do better next time?

Based on the lessons about the interaction of behavior and disease dynamics discussed here, I suggest the following three-part strategy to improve our public health and economic response to emerging infectious disease.

First, we need to invest in our disease surveillance capabilities worldwide, perhaps using the infrastructure developed for worldwide influenza surveillance as a model.[21] It is certainly worth a lot of money to have the capacity to identify the threat from a new infectious disease anywhere in the world before it gets going so as to buy time to mount a public health and scientific response.

Second, we must consolidate all that has been learned about the implementation of non-pharmaceutical public health measures for disease control over the past year so that we might be able to quickly implement those measures that have been proven to effectively slow the spread of an infectious disease with the least cost to the economy. Given the widespread discussion of "pandemic fatigue" in the population, we should also look at policies for infectious disease control that have low personal costs and thus have a greater change of enjoying widespread voluntary adherence. A number of countries, many of them in Asia, have been able to keep COVID infections and deaths to low levels over the course of the past year with effective public health interventions based on travel restrictions, testing, contact tracing, and isolation of infected individuals while preserving considerable economic activity and personal autonomy. Several universities in the United States have also succeeded at control

[21] https://www.who.int/influenza/gisrs_laboratory/en/

of COVID infections with extensive testing and isolation regimes.[22] As we have seen from the model simulations in this paper and these real-world experiences, public health measures that allow us to wait for the development of a technological solution to a global pandemic with minimal loss of life and economic damage can be extremely valuable. Given the public good nature of infectious disease surveillance and public health system preparedness to implement rapidly scalable countermeasures, it seems a high priority to fund these capabilities at the Federal level.

Third, we need to invest in new models for accelerating the development, financing, and distribution of vaccines and life-saving therapeutics for emergent disease. In the end, it is these technological solutions that will allow us to contain the long run impact of new pandemics once they become global.[23]

To illustrate the urgency of addressing these public health priorities now, consider one final model scenario. As long as COVID-19 remains prevalent worldwide, new mutations of the virus are likely to emerge and there is increasing evidence that such mutations might allow COVID-19 to evade the immunity conferred by prior infection and vaccines. In such a scenario, COVID-19 could be an endemic, seasonal, disease that might require essentially permanent efforts at disease control.[24] To illustrate how such a scenario might play out, I simulate the model with vaccines shown in Figure 5 with a version of the virus circulating that is 2/3 more transmissible than

[22]See, for example, the experience of the University of Illinois at Champagne Urbana documented here https://covid19.illinois.edu/on-campus-covid-19-testing-data-dashboard/ and that of Georgia Tech documented here https://health.gatech.edu/coronavirus/health-alerts

[23]See "Mitigating the Impact of Pandemic Influenza through Vaccine Innovation" by the President's Council of Economic Advisors, September 2019 for a careful analysis of the economic and public health rationale for a large Federal investment in such technologies. See Angus, et. al. "Emerging Lessons From COVID-19 for the US Clinical Research Enterprise" JAMA editorial February 26, 2021 https://jamanetwork.com/journals/jama/fullarticle/2777058 for a discussion of current difficulties in conducting rapid clinical trials of new treatments in the United States.

[24]See for example Christopher Murray and Peter Piot "The Potential Future of the COVID-19 Pandemic: Will SARS-CoV-2 Become a Recurrent Seasonal Infection?" Journal of the American Medical Association (JAMA) , published online March 3, 2021 doi:10.1001/jama.2021.2828. See also Lavine et. al. "Immunological characteristics govern the transition of COVID-19 to endemicity" Science, February 2021, Vol. 371, Issue 6530, pp. 741-745 https://science.sciencemag.org/content/371/6530/741

the original virus, but in which immunity from infection and/or vaccination lasts on average for only 18 months. I show the resulting forecast path of daily deaths from COVID over a five-year period in Figure 6. In this simulation, I assume that the vaccination program continues at a constant rate of roughly 1.3 million vaccinations per day throughout the forecast period. As one can see in this figure, the epidemic is forecast in this scenario to settle into a regular seasonal pattern killing over 100,000 Americans per year even with new vaccines and a response of public and private behavior to the changing prevalence of the disease. Clearly, in such a scenario, we would benefit greatly from finding ways to mitigate this disease on an ongoing basis at a lower economic cost.

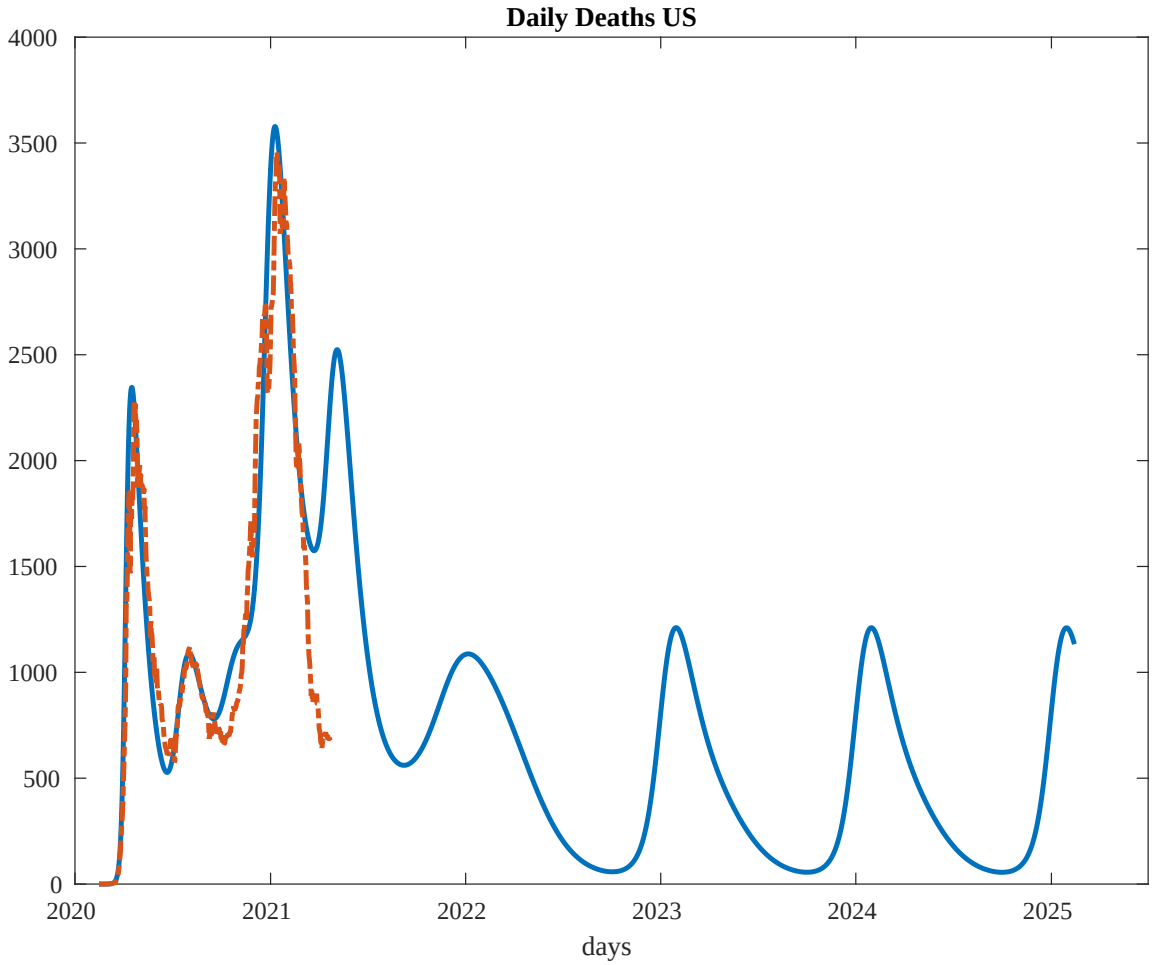

Figure 6: Predictions of the model for daily deaths from COVID-19 (in blue) in a simulation with a vaccination program starting on January 1, 2021 that proceeds at a rate fast enough to protect half of the population by July 1, 2021. The data on the seven-day moving average of daily deaths are shown in red. In this simulation, immunity acquired from prior infection or vaccination is assumed to last 18 months on average. The vaccination program is assumed to continue at a constant rate throughout the entire period with presumably new booster shots conferring immunity against new variants as they occur. Even with this program of booster vaccines and continued prevalence elastic behavior, in this simulation, roughly 100,000 Americans die each year from COVID on a persistent basis.

5 Appendix

This appendix presents the model and parameters used in "Behavior and the Dynamics of Epidemics" by Andrew Atkeson for the Brookings Panel on Economic Activity Spring 2021. This model is based closely on that presented in "A Parsimonious Behavioral SEIR Model of the 2020 COVID Epidemic in the United States and United Kingdom" which is available as NBER working paper 28434 and as Federal Reserve Bank of Minneapolis Staff Report 619. This appendix discusses the model extended to include vaccines and the potential for waning immunity. It is applied to the United States.

This model is a an SEIR model (with compartments for agents who are susceptible, S, exposed, E, infectious, I, and recovered and hence removed R) modified to include a compartment for those infected agents who end up with serious disease. I refer to this compartment as H, for hospitalized. Agents who die from COVID are assumed to transition from infection I to death, D, through this compartment H. The expected time that agents spend in this compartment is set to 30 days to capture the delay between serious illness, death, and the reporting of that death.

Behavior in this model is assumed to respond to daily death rates as they are reported. It is assumed that behavior does not respond immediately to new infections as these are not directly observed. As discussed by John Cochrane[25] and Weitz et. al. 2020[26] the delay between infection and death introduced by this compartment H implies that this simple behavioral model has oscillatory endogenous dynamics that are helpful in allowing the model to reproduce the data with only a few shocks.

In the United States, we have seen public authorities tighten and loosen COVID mitigation policies in tandem with changes in the prevalence of this disease. I inter-

[25]See https://johnhcochrane.blogspot.com/2020/05/an-sir-model-with-behavior.html

[26]Joshua Weitz, Sang Woo Park, Ceyhun Eksin, and Jonathan Dushoff, "Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations" , *Proceedings of the National Academy of Science*, vol. 117, no. 51, December 22, 2020

pret this correlation between public policies and disease prevalence as arising from a public behavioral response to shifting political calculations as disease prevalence rises and falls, that is, as a social choice behavioral response. I interpret private behavioral responses as arising from rising concern over personal infection risk with rising disease prevalence. As described below, I thus interpret the reduced form behavioral response of transmission rates to disease prevalence as resulting from a combination of private and public reactions to disease prevalence. I do not attempt to distinguish the relative importance of these two responses.

The four shocks considered in this paper are as follows. First, I add a standard seasonal variation in the baseline transmission rate of the virus from a winter peak to a low in midsummer. Second, I introduce a one-time change in behavior modeled as a reduction in the semi-elasticity of the transmission rate with respect to the daily death rate from an initial level to a new, permanently lower level. I refer to this second shock as the onset of pandemic fatigue. When I fit the model to the United States, I assume that pandemic fatigue sets in late in 2020. Third, I introduce a more contagious variant of COVID to the United States through a single individual on December 1, 2020. The transmissibility of this variant is set to be 50% higher than the original variant. The model implies that this new variant becomes the dominant variant circulating in the United States by summer of 2021. The fourth shock is the introduction of vaccines starting on January 1, 2021. I assume an aggressive vaccination program that proceeds at a rate to vaccinate just over 50% of the total population by July 1, 2021 and that continues indefinitely.

This appendix has three parts. In section 5.1, I present the equations of the model. I discuss the interpretation of the reduced form behavioral response of the transmission rate to disease prevalence measured by daily deaths as a combination of endogenous private and public responses to disease prevalence. I also compare the structure of this model with that of a simpler behavioral SIRD model as analyzed in

Atkeson, Kopecky, and Zha (2021)[27] and Droste and Stock (2021)[28] .

In section 5.2, I discuss the values of the parameters. The model is fit to US data on daily deaths from COVID. Several parameters are set to match recommendations from the Center for Disease Control for modeling of COVID-19. Other parameters, in particular, the basic reproduction number of the original variant of the virus, the original semi-elasticity of the transmission rate to the level of daily deaths, and the size of the seasonal fluctuation in transmission, are chosen through a process of trial and error. To illustrate this fitting process, I show the sensitivity of the model predictions to variation in these three parameters around their baseline values.

In section 5.3, I discuss the shocks to the model. I illustrate the role that these shocks play in allowing the model to match the data on daily deaths from COVID by solving the model with each of these shocks turned off.

5.1 Model

The model is as follows.

The SEIHR model extends the SIR model by adding both the exposed state E and the hospitalized state H. In this version of the model the total population N is given by the sum of susceptible agents in state S, exposed in state E, infected in I, hospitalized in H, recovered in R, and dead in D.

The rate at which agents leave the E compartment for both the normal and more transmissible variants is σ and the rate at which agents leave the I compartment for both variants is γ. The mean generation time for the model is then $1/\sigma + 1/\gamma$. As discussed below, the choice of these parameters is guided by CDC recommendations for these disease parameters.

[27] Atkeson, Kopecky, and Zha "Behavior and the Transmission of COVID-19" forthcoming, *American Economic Review Papers and Proceedings* with the longer version available here https://www.minneapolisfed.org/research/staff-reports/behavior-and-the-transmission-of-covid19

[28] Droste and Stock "Adapting to the Pandemic" forthcoming, *American Economic Review Papers and Proceedings*

As agents leave the I compartment, fraction η go into the hospitalized compartment H and $1-\eta$ transition directly to the recovered compartment R. The rate at which agents leave the H compartment is ζ. Of those leaving the H compartment fraction ν die and $1-\nu$ survive. Thus, the overall infection fatality rate is given by $\eta\nu$ and the mean time in the H compartment corresponding to illness and delays in reporting deaths is $1/\zeta$.

To model the introduction of a new variant, I add separate compartments E_v and I_v for those exposed to and infectious with the new variant. The transmission rate of the original variant is denoted by $\beta(t)$. That for the new variant is denoted by $\beta_v(t)$. I will assume that the new variant is always 50% more transmissible than the original variant, That is, I set $\beta_v(t) = 1.5\beta(t)$ for all t. The new variant is introduced by setting $\bar{E}_v(t) = 1/population$ in the equations below for one day on a specified date t_v and equal to zero otherwise. This corresponds to the introduction from abroad of a single individual carrying this more transmissible variant. Note that this quantity is subtracted off of the change in the R compartment simply to keep the population constant. Since this shift is only one person for one day, it does not impact the quantitative implications of the model for large populations.

I model the impact of vaccines as moving agents from the susceptible compartment S directly to the removed compartment R at a rate λ per day. With this assumption, I impose that the vaccine blocks both transmission by the vaccinated and disease in the vaccinated. These assumptions are clearly optimistic. I model waning immunity as a movement of agents from the removed compartment R back to the susceptible compartment at a rate ξ per day. I do not consider population growth in the model.

The dynamics of the model are given by

$$\frac{dS(t)}{dt} = -(\beta(t)I(t) + \beta_v(t)I_v(t))S(t) - \lambda(t)S(t) + \xi R(t)$$

$$\frac{dE(t)}{dt} = \beta(t)I(t)S(t) - \sigma E(t)$$

$$\frac{dE_v(t)}{dt} = \beta_v(t) I_v(t) S(t) - \sigma E_v(t) + \bar{E}_v(t)$$

$$\frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t),$$

$$\frac{dI_v(t)}{dt} = \sigma E_v(t) - \gamma I_v(t)$$

$$\frac{dH(t)}{dt} = \eta\gamma(I(t) + I_v(t)) - \zeta H(t)$$

$$\frac{dR(t)}{dt} = (1-\nu)\zeta H(t) + (1-\eta)\gamma(I(t) + I_v(t)) - \bar{E}_v(t) + \lambda(t) S(t) - \xi R(t)$$

$$\frac{dD(t)}{dt} = \nu\zeta H(t),$$

The reduced-form for the behavioral response of the transmission rate to the level of daily deaths is given by

$$\beta(t) = \bar{\beta} \exp(-\kappa(t) \frac{dD(t)}{dt} + \psi(t)) \tag{1}$$

$$\beta_v(t) = \bar{\beta}_v \exp(-\kappa(t) \frac{dD(t)}{dt} + \psi(t))$$

where the parameters $\bar{\beta}$ and $\bar{\beta}_v$ control the baseline transmissibility of the normal and variant of COVID, the parameter $\psi(t)$ is used to introduce seasonality in transmission, and $\kappa(t)$ is the semi-elasticity of transmission with respect to the level of daily deaths.

To model seasonality in the transmission of the virus, we set

$$\psi(t) = seasonalsize * (\cos((t + seasonalposition) * 2\pi/365) - 1)/2$$

where *seasonalsize* controls the magnitude of the seasonal fluctuations in transmissibility holding behavior fixed and *seasonalposition* controls the location of the seasonal peak in transmission. Note that t is indexed to $t = 0$ on February 15, 2020.

To model pandemic fatigue, I set

$$\kappa(t) = \bar{\kappa} * (1 - normcdf(t, fatiguemean, fatiguesig))+$$

$$fatiguesize * \bar{\kappa} * normcdf(t, fatiguemean, fatiguesig)$$

where $\bar{\kappa}$ sets the initial semi-elasticity of transmission with respect to daily deaths, $fatiguesize$ sets the percentage reduction in this semi-elasticity in the long run, $normcdf$ is the normal CDF, $fatiguemean$ sets the date at which the transition in $\kappa(t)$ from its initial to new long run level is halfway complete, and $fatiguesig$ sets the speed with which that transition occurs.

For all model forecasts, I leave the behavioral parameter κ which determined the semi-elasticity of the transmission rate with respect to daily deaths fixed at its final value at the end of the model estimation period of one year. Thus, I assume that there are no further changes in behavioral parameters going forward.

Initial conditions for all simulations are $E(0) > 0$, $E_v(0) = I(0) = I_v(0) = R(0) = H(0) = D(0) = 0$, $S(0) = 1 - E(0)$. For the United States, $E(0) = 33$ on February 15 out of a population of 330 million. The new more contagious variant is introduced with $\bar{E}_v(t) = 1/330,000,000$ for one day on December 1, 2020.

The MATLAB code to run this model and create all of the figures in this document and the main text is comprised of the following files. *BPEAAtkesonSpring2021.mlx* is a Matlab live script that is the main file. This file calls three Matlab scripts *bpeaodefile.m* is a function with all the differential equations described above, *bpearunthemodel.m* contains the code to solve these differential equations starting from the initial conditions described above with various parameter configurations, and *usdata.m* contains the CDC data on daily COVID deaths in the United States used in the plots.

5.1.1 Behavior as a combination of private and public responses to disease prevalence

The reduced form response of the transmission rates of the original and new variants of the virus to daily deaths assumed in the model can be understood as arising from the combined responses of private and public actors to disease prevalence as follows.

This reduced form response of transmission to daily deaths can be obtained as a result of a two-equation system in which the transmission rate is given as a function of activity $Y(t)$ with

$$\beta(t) = \bar{\beta} Y(t)^{\alpha} \exp(v(t) + \psi(t))$$

and activity is given as a declining function of daily deaths

$$Y(t) = \exp(-\frac{\kappa_p(t)}{\alpha} \frac{dD(t)}{dt} + u(t))$$

In the first of these two equations, the parameters $\bar{\beta}$ and $\bar{\beta}_v$ are fixed coefficients that capture features of the virus and the population determined prior to the epidemic that might impact transmission of the original and more transmissible variants. Factors considered in the literature include population density, modes of transportation (subway vs. car, etc.), household and demographic structure, cultural norms (bowing vs. shaking hands or kissing), temperature and humidity, etc.

The parameter α captures the elasticity of transmission with respect to activity.

The parameter $v(t)$ represents a potentially time-varying shock to the region-specific relationship between activity and transmission that may represent the impact of policy over time. I normalize $v(0) = 0$. When interpreting variation in $v(t)$ as representing the impact of policies, here I am thinking about policies such as mask-wearing, ventilation, physical distancing, redesign of workspaces, or other measures implemented after the start of the epidemic that reduce transmission given a fixed level of activity.

I normalize the level of activity at the start of the pandemic to $Y(0) = 1$. Given this normalization, the parameter $\bar{\beta}$ sets the transmission rate of the virus at the start of the epidemic. Specifically, $\bar{\beta}/\gamma$ and $\bar{\beta}_v/\gamma$ correspond to the *basic reproduction number* of the original and more transmissible variants at the peak of the seasonal cycle of transmissibility.

In the second of these two equations, I model individuals' decisions to engage in activity at time t, $Y(t)$, as a declining function of the level of daily deaths, $\dot{D}(t)$. The parameter $\frac{\kappa_p(t)}{\alpha}$ is the semi-elasticity of private behavior with respect to daily deaths. The variable $u(t)$ in this second equation represents a time-varying shock to the region-specific relationship between deaths and activity. I interpret $u(t)$ as reflecting public policies such as lockdowns or closures that would reduce activity below what agents might choose in a decentralized fashion.

I consider government policies that are "behavioral" in that they are responsive to the prevalence of the disease as measured by daily deaths as well as policies that depend only on time. To model the behavioral component of government policies, I assume that public policies impacting the relationship between activity and transmission $v(t)$ and between daily deaths and activity $u(t)$ are responsive to the level of daily deaths. Specifically, I assume that

$$v(t) = -\eta_v \frac{dD(t)}{dt}$$

$$u(t) = -\frac{\eta_u}{\alpha} \frac{dD(t)}{dt}$$

I assume that all government mandated mitigation that is dependent only on time enter the model through $\psi(t)$.

With these assumptions, we have that the behavioral decline in activity and disease transmission with daily deaths can be interpreted as arising either from a change in private behavior or public mandates that are conditioned on the prevalence of the disease. Specifically, with these assumptions, we obtain the reduced-form relationship between daily deaths and disease transmission in equation (1) with the overall semi-

elasticity of transmission with respect to daily deaths given as a compound parameter capturing both private and public responses to daily deaths

$$\kappa = \kappa_p + \eta_v + \eta_u$$

The parameter $\psi(t)$ corresponds to changes in disease transmission that are unrelated to the level of daily deaths. In the baseline model, this variation is pure seasonal variation in disease transmission. When I consider additional government efforts at mitigation of transmission such as through testing and tracing and isolation of the infected, I model these interventions as a shift in $\psi(t)$ as well. In this case, this component of public policy should be interpreted as an intervention that is not directly responding to disease prevalence.

5.1.2 Comparison with behavioral SIRD models

Note here that the addition of the compartments E and H to this model together with the assumption that behavior responds to the level of daily deaths distinguishes the behavioral component of this model from those in the SIRD models in Atkeson, Kopecky, and Zha (2021) and Droste and Stock (2021). These additional compartments substantially alter the endogenous dynamics implied by this model relative to these models. The difference between the implied dynamics can be understood as follows.

In an SIRD model, the level of daily deaths is directly proportional to the current level of the infected population $I(t)$, so, an assumption that behavior reacts to disease prevalence measured by the level of daily deaths is equivalent to an assumption that behavior reacts to the current level of active infections in the population. In turn, since the growth of the infected population is directly governed by the transmission rate $\beta(t)$, then the assumption, as above, that behavior reacts to daily deaths is equivalent to assuming a tight relationship between the level and growth rate of active infections. One way to interpret the findings in Atkeson, Kopecky, and Zha

(2021) and Droste and Stock (2021) that large and frequent shocks are needed to fit a behavioral SIRD model to the data on daily deaths from the United States is as a finding that this simple behavioral model is "over-controlled". That is, for the simple behavioral SIRD model, the the model's endogenous dynamics do not permit significant fluctuations in the growth rate of daily deaths after the initial phase of the pandemic and hence such a model requires large and frequent shocks to match the observed fluctuations in the growth rates of daily deaths.

In contrast, in the model presented here, the level of daily deaths in not directly proportional to the infected population $I(t)$, but instead it is directly proportional to the level of hospitalizations $H(t)$, and this level of hospitalizations is, in turn, equivalent to a distributed lag of past levels of infections $I(t)$, with ζ controlling the length of these lags. This lag between infections and the behavioral response gives the model oscillatory dynamics. I show below how changes in this parameter ζ change the endogenous dynamics of the model in such a way to substantially reduce the need for shocks to the model.

That the addition of the E and H compartments allow for a good fit to the deaths data for the United States and United Kingdom is serendipitous. The COVID epidemic has displayed different dynamics in different locations. I have not had similar success in fitting this model to some other locations that I have tried. I leave it future research to determine whether further refinement of the structure of the model or additional shocks are needed to fit the wide range of COVID outcomes observed around the world.

5.2 Parameters

In this section I discuss the choice of parameters and shocks to the model. To illustrate how parameter values and shocks impact the implications of the model, I use the simulation of the model shown in Figure 5 of the main text as a baseline for comparison. This simulation of the model has the baseline parameter values

described below and the shocks of seasonality in transmission, pandemic fatigue, and the new variant, and the introduction of vaccines. To ease comparison, I reproduce Figure 5 from the main text here in Figure A7.

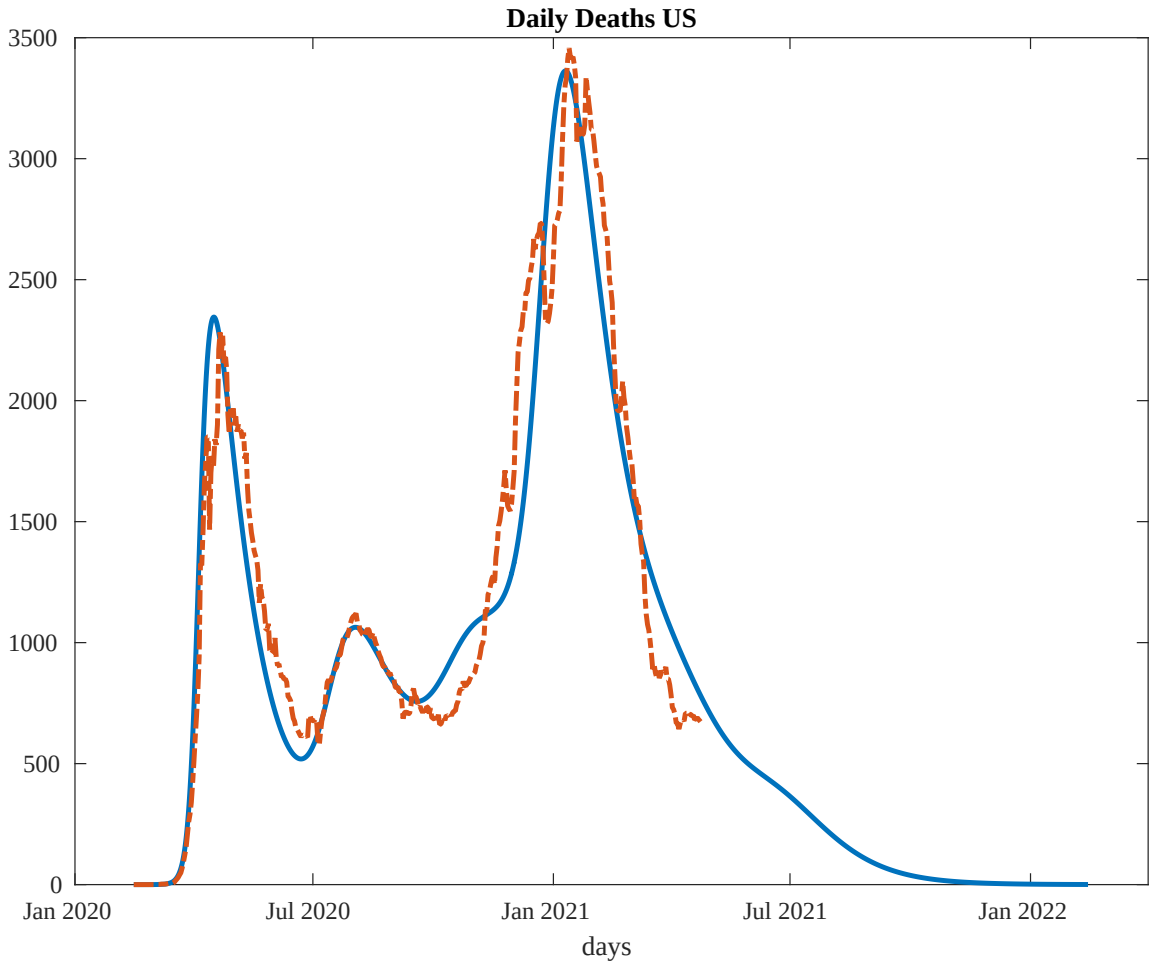

Figure A7: Baseline model implications corresponding to Figure 5 in the main text. This simulation has the baseline parameter valeus, the shocks due to seasonality, pandemic fatigue, and the introduction of a new, more transmissible variant, as well as the introduction of vaccines.

I group the parameters of the model into three sets.

The first set are those set from CDC model recommendations for the generation time of infections and the infection fatality rate. As is standard in an SIR model,

the basic reproduction number, which is a ratio of two rates and is thus unitless, determines the overall shape of the epidemic. The generation time then determines the time-scale over which the epidemic plays out.

This first set of parameters is set as follows: $\gamma = 0.4$, $\sigma = 0.425$, $\eta = 0.025$, $\nu = 0.2$. The parameter σ corresponds to an expected time before an exposed agent becomes infectious of 2.35 days and the parameter γ corresponds to an expected time for which an infected individual is infectious of 2.5 days. The generation time is defined as the average time between which one infected individual shows symptoms and a person infected by that individual shows symptoms. These two parameters together imply an average generation time of $1/\sigma + 1/\gamma = 4.85$ days.[29] As mentioned above, this generation time sets the time-scale of the epidemic implied by the model.

The infection fatality rate implied by these parameters is $\eta\nu = 0.005$. This parameter is key for translating the model's implications for infections into implications for deaths. While this model is simplified in that it does not consider the age distribution of infections, it does do fairly well in matching the overall estimates of the burden of COVID provided by the CDC.[30] The decomposition of this infection fatality rate has fraction $\nu = 0.2$ of those hospitalized dying (again see the CDC model scenario page cited above for data on the fatality rate conditional on hospitalization) and thus fraction $\eta = 0.025$ resulting in hospitalizations. The decomposition of the infection fatality rate into a rate of hospitalization and a fatality rate conditional on hospitalization does not impact the prediction of the model for daily deaths, but is relevant if one were to attempt to use the model to match data on hospitalizations.

The second set of parameters is comprised of the rate ζ at which those hospitalized flow either to death or recovery. This rate is chosen to have an average stay in

[29] See https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html. On that webpage, the CDC notes a mean time of approximately six days between symptom onset in one person to symptom onset in another person infected by that individual.

[30] The CDC estimates that 83 million Americans had been infected by the end of December 2020. Total COVID deaths reached 445,000 30 days later, giving an average estimated infection fatality rate including the delay from infection to death of slightly over 0.005. See https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html

compartment H of 30 days, which corresponds to an average stay in the hospital of two weeks for those with serious illness and a reporting delay of deaths of two weeks.[31] This assumption combined with the average generation time assumed above implies that roughly six generations of infections pass on average before reported daily deaths rise provoking a behavioral response.

The delay between infection and reported death introduced by this compartment H implies that this simple behavioral model has oscillatory endogenous dynamics that are helpful in allowing the model to reproduce the data with only a few shocks. To illustrate the role of this delay in generating endogenous model dynamics, in Figure A8, I show the implications of the model with all the parameters set as in Figure A7 except that I set $\zeta = 10$ so that the average stay in the H compartment is only a few hours. In comparing Figure A8 below and Figure A7, we see that if we have only a short stay in the H compartment, the model does not match the waves in the first six months of the epidemic while the baseline model shown in Figure A7 does.

[31] See https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html. On that webpage, the CDC notes a median time from symptom onset to death of approximately two weeks and a median time from death to reporting just under three weeks.

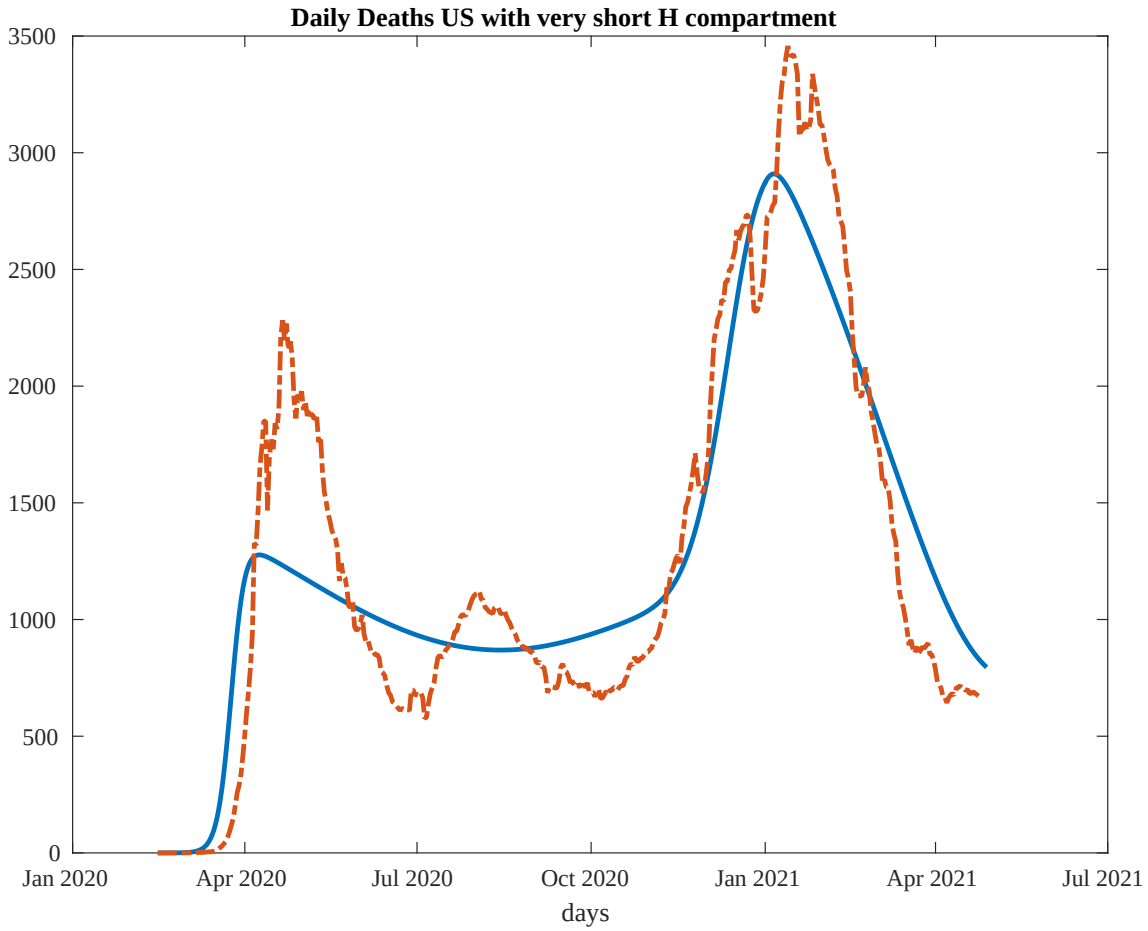

Figure A8: Model implications with a very short duration of the H compartment

The third set of parameters are those that I choose through a process of trial and error to fit the model to data from the first six months of the epidemic in the United States. These are the basic reproduction number, the original semi-elasticity of transmission with respect to daily deaths, and the size of the seasonal fluctuation in transmission rates. Note from Figure A7 that the model fits the path of daily deaths quite well over the first six months of the epidemic with the only shock being the seasonality in transmission. The assumed shocks for pandemic fatigue, the new variant, and vaccines, do not impact the model's predictions over these first six months.

The basic reproduction number of the virus at peak transmissibility is $\mathcal{R}_0(t) = \bar{\beta}/\gamma$ for the original virus and $\mathcal{R}_0(t) = \bar{\beta}_v/\gamma$ for the new variant. For the United States, I set $\bar{\beta} = 3\gamma$ giving a peak basic reproduction number at pre-pandemic levels of behavior in Winter of 3. This number is well within the range of estimates of this parameter from the early phase of the pandemic.[32] As discussed above, this parameter was chosen by trial and error. In Figure A9 I show how the model implications for the path of daily deaths varies if this parameter is increased or decreased by 20% (from 3.6 to 3 to 2.4). As can be seen in this figure, increasing this basic reproduction number increases the speed and height of the first wave of deaths.

[32]See https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html.

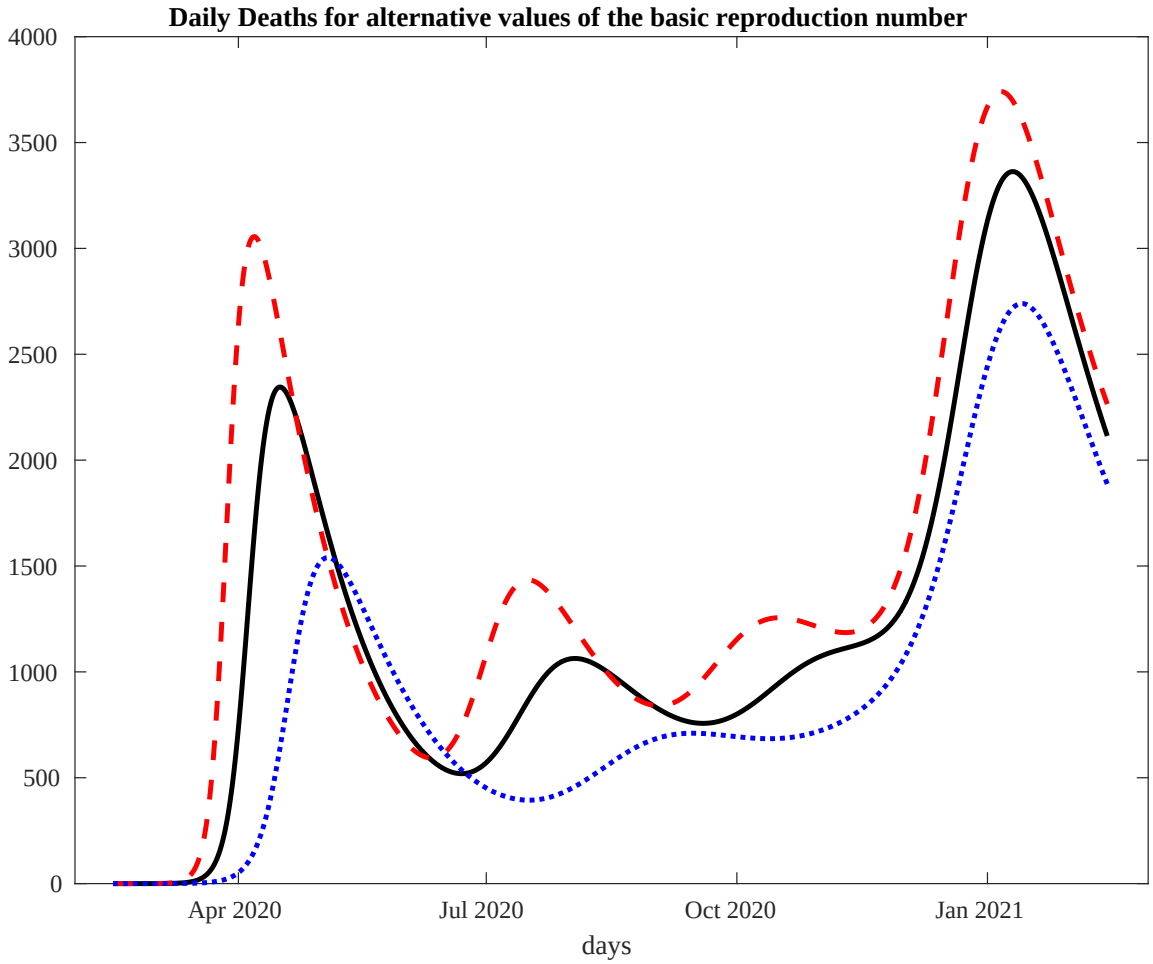

Figure A9: Implied path for daily deaths with the basic reproduction number $\bar{\beta}/\gamma = 3$ and with this parameter set 20% above and below this baseline value.

The initial semi-elasticity of transmission with respect to daily deaths (measured as a fraction of the population) for the United States is $\bar{\kappa} = 250000$. This parameter value was also chosen through trial and error. In Figure A10, I show the implications of the model with this initial semi-elasticity set to this baseline value and to values 20% above and below this baseline value. As can be seen in this figure, this semi-elasticity primarily impacts the level of the daily deaths in a roughly parallel manner.

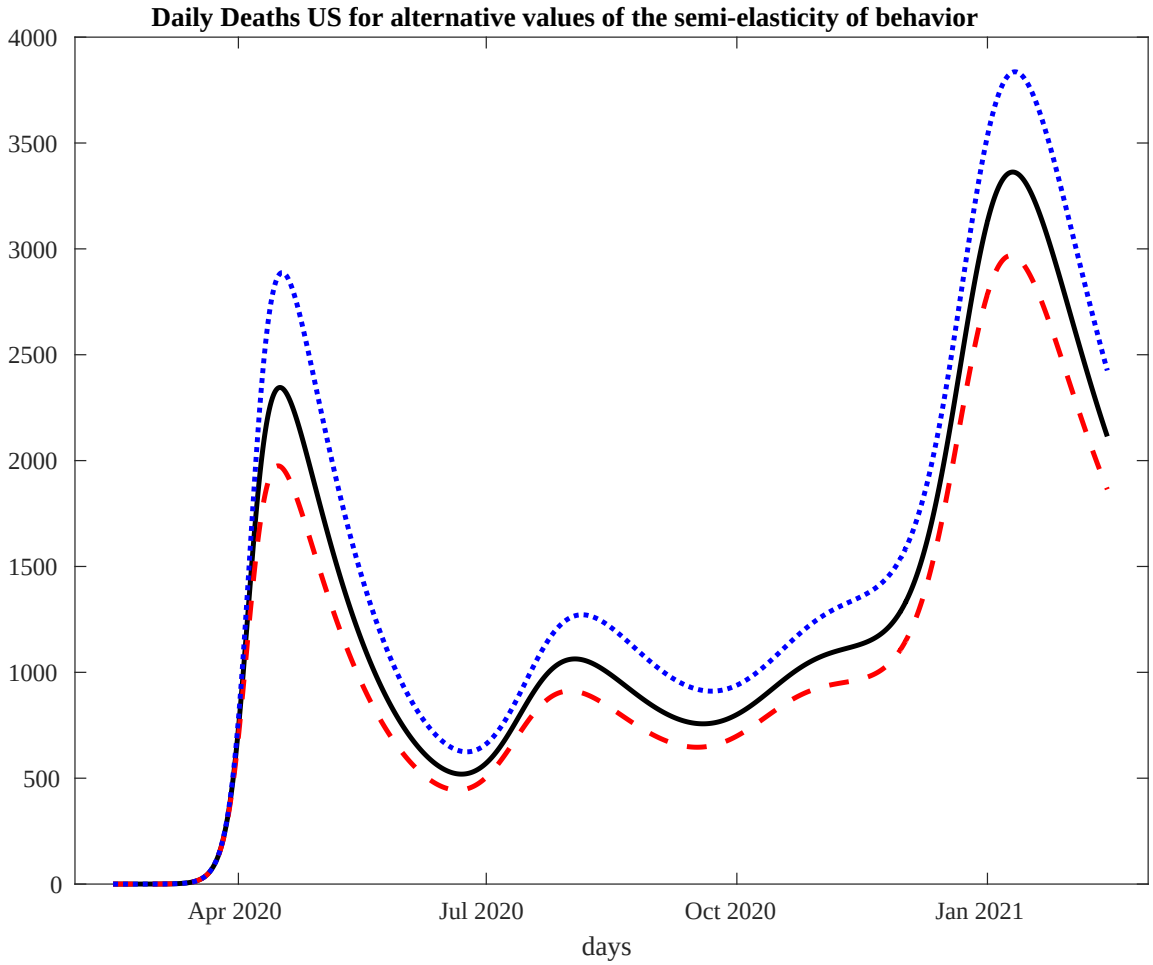

Figure A10: Implied path for daily deaths with the semi-elasticity $\bar{\kappa} = 250000$ and with this parameter set 20% above and below this baseline value.

To model seasonality of transmission in the United States, I set $seasonalsize = 0.35$ and $seasonalposition = 20$. This seasonal variation in the parameter $\psi(t)$ leads to variation over time in the basic reproduction number of the virus. Figure A11 shows the basic reproduction number for the original variant corresponding to no reduction in transmission due to a behavioral response (the transmissibility of the virus with behavior at pre-pandemic patterns of behavior). We see that the assumed pattern for seasonality introduces a reduction in transmissibility of the virus by a factor of $\exp(-0.35)$ holding behavior fixed from the winter peak in late January and

the summer low in late July.

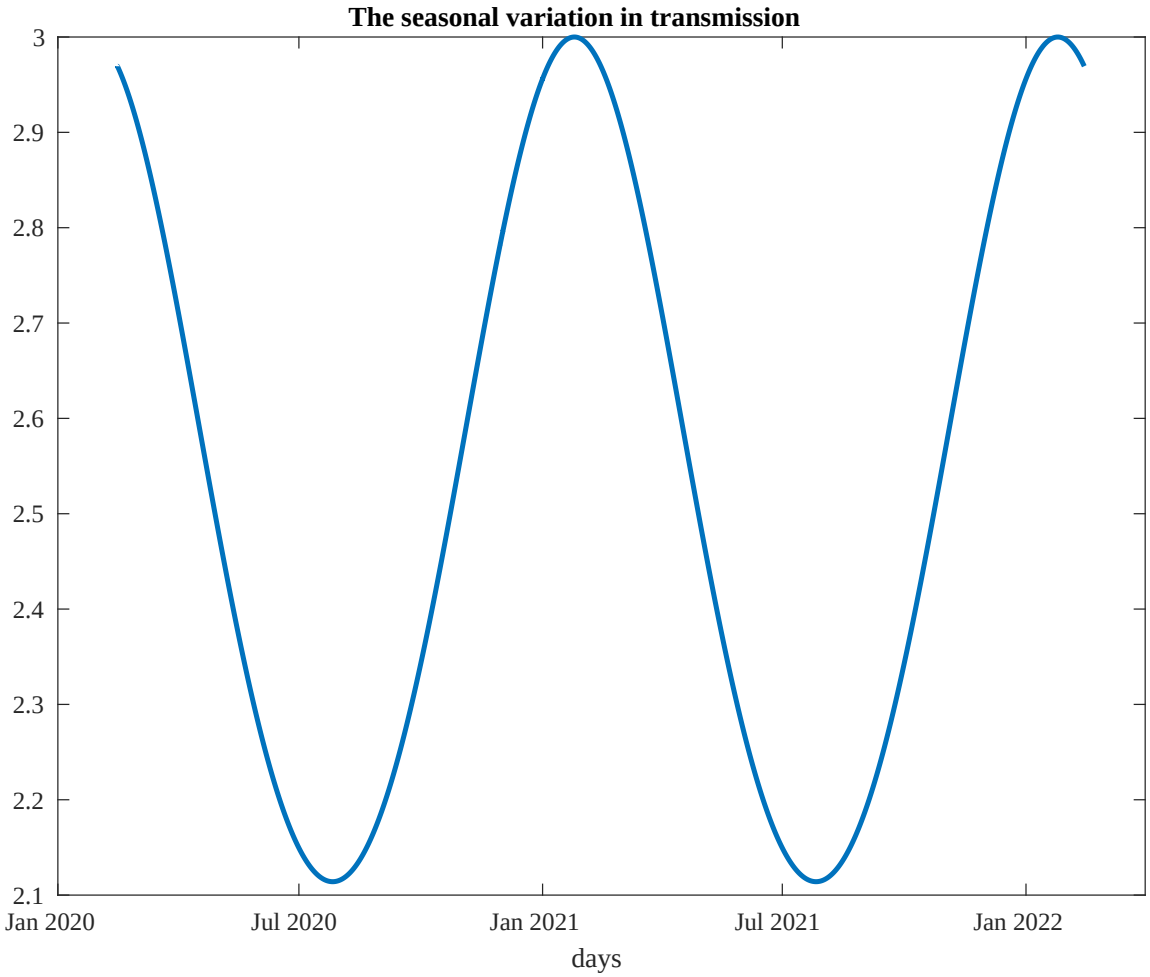

Figure A11: Assumed seasonality in the basic reproduction number.

In Figure A12 I show the implications of the baseline model from Figure A7 but with no seasonality in transmission. By comparing this Figure A12 to Figure A7, we see that seasonal variation in transmission is important for matching the drop in daily deaths observed over the summer.

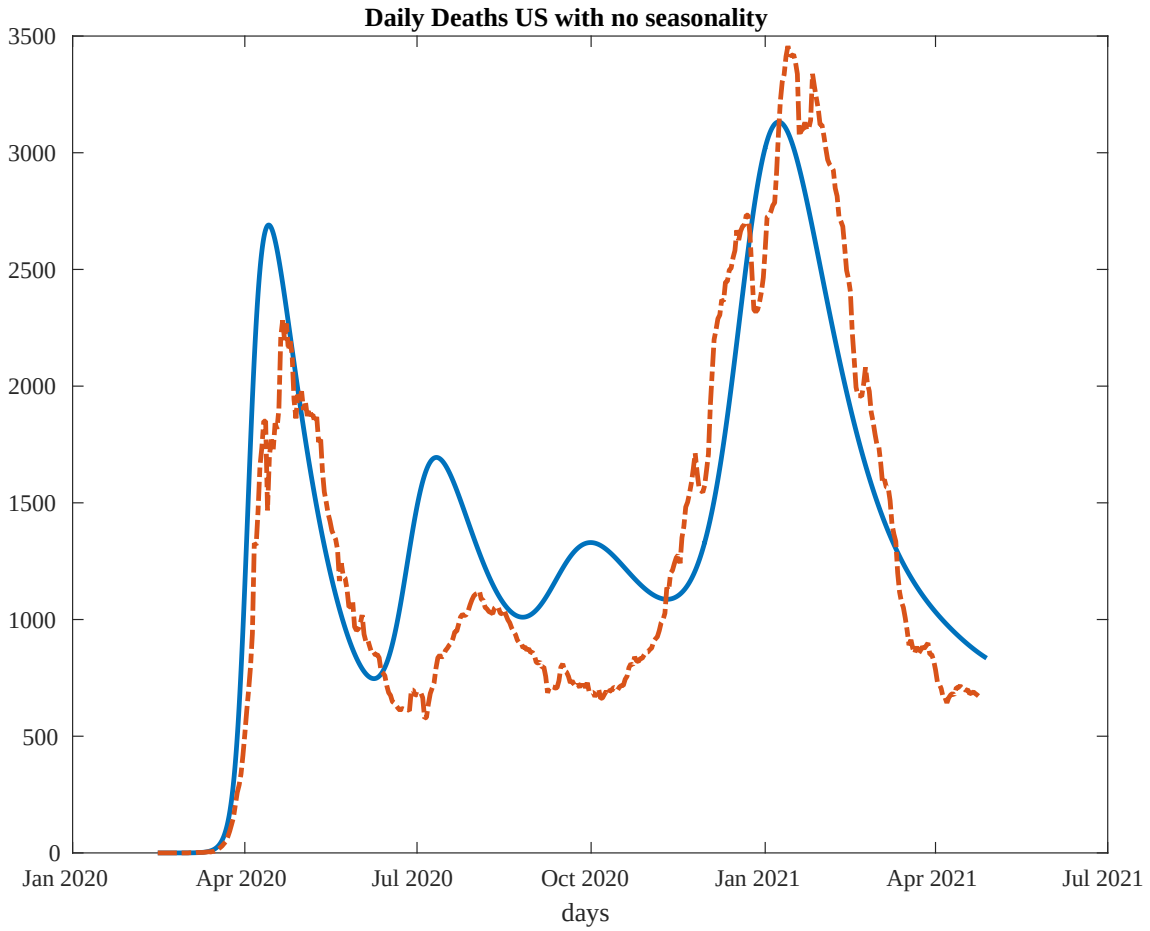

Figure A12: Implications of the model with vaccines, the new variant, and pandemic fatigue, but no seasonality in transmission.

The magnitude of the seasonal fluctuation in transmission was chosen by trial and error. While the seasonal variation in transmission is important for matching the pattern of daily deaths in the data, the model's implications are fairly insensitive to the precise magnitude of the seasonal. In Figure A13, I show the implications of the model for daily deaths with the size of the seasonal variation in transmission set to its baseline value and 20% above and below this value.

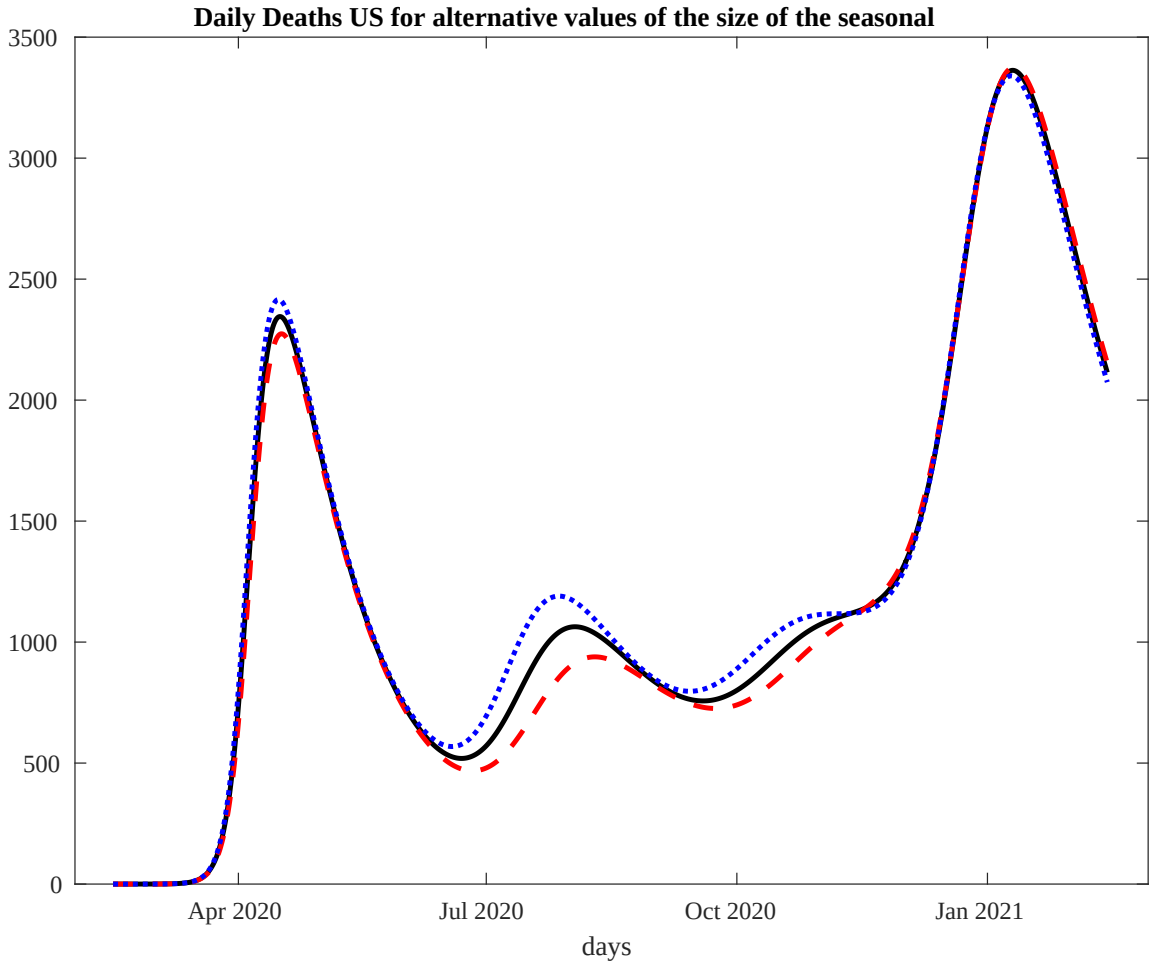

Figure A13: Implications of the model with the size of the seasonal variation in transmission 20% above and below its baseline value.

5.3 Shocks

The model has several shocks in addition to the seasonal fluctuation in transmission rates. These include pandemic fatigue, the introduction of a new variant on December 1, 2020, the introduction of vaccines, and the counterfactuals of extra mitigation measures that are not dependent on disease prevalence and of waning immunity. I describe the specification and impact of these shocks and counterfactual experiments

here.

The initial semi-elasticity of transmission with respect to daily deaths (measured as a fraction of the population) for the United States is $\bar{\kappa} = 250000$. To model the onset of pandemic fatigue in the United States, I set $fatiguesize = 0.375$, $fatiguemean = 285$ and $fatiguesig = 15$. Figure A14 shows the the semi-elasticity of the transmission rate with respect to the level of daily deaths. We see in that figure that this semi-elasticity is assumed to fall to 37.5% of its original level late in the year.

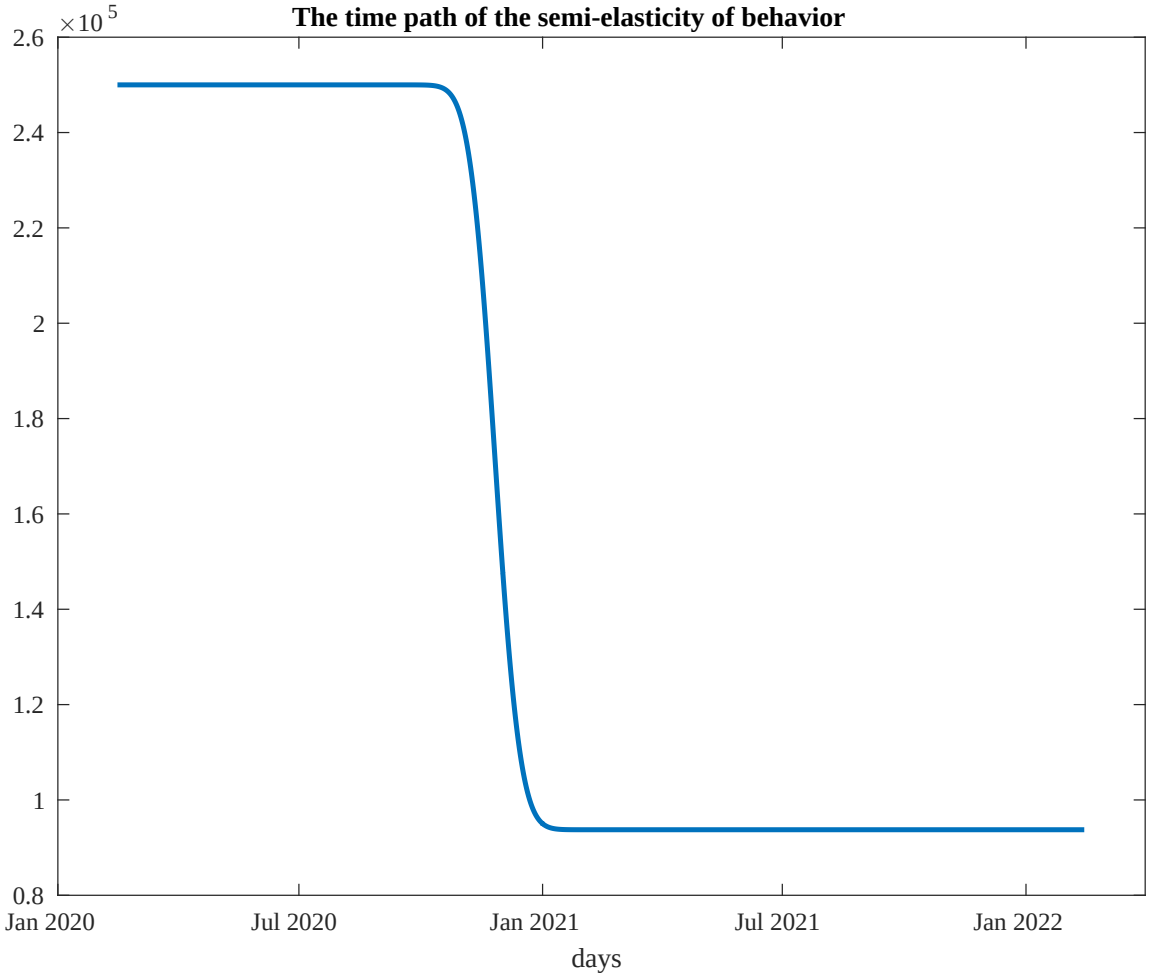

Figure A14: Assumed pandemic fatigue. The blue line shows the evolution of the semi-elasticity of transmission with respect to daily deaths relative to its initial level.

In Figure A15 I show the implications of the baseline model with no pandemic fatigue. By comparing this figure with Figure A7, we see that pandemic fatigue is quite important in generating the large wave of deaths in the winter of 2020. Essentially, the initial behavioral response in the model is too strong to allow for a significant winter wave even with the seasonal rise in transmission.

Note from Figure A14 that I leave the semi-elasticity of transmission with respect to behavior at its new, lower level, throughout the remainder of the Spring of 2021.

The match between the model in this time period shown in Figure A7 suggests that the behavioral reaction of transmission to the reduction in daily deaths induced by the introduction of vaccines has remained as it was at the end of 2020 through the first three months of 2021.

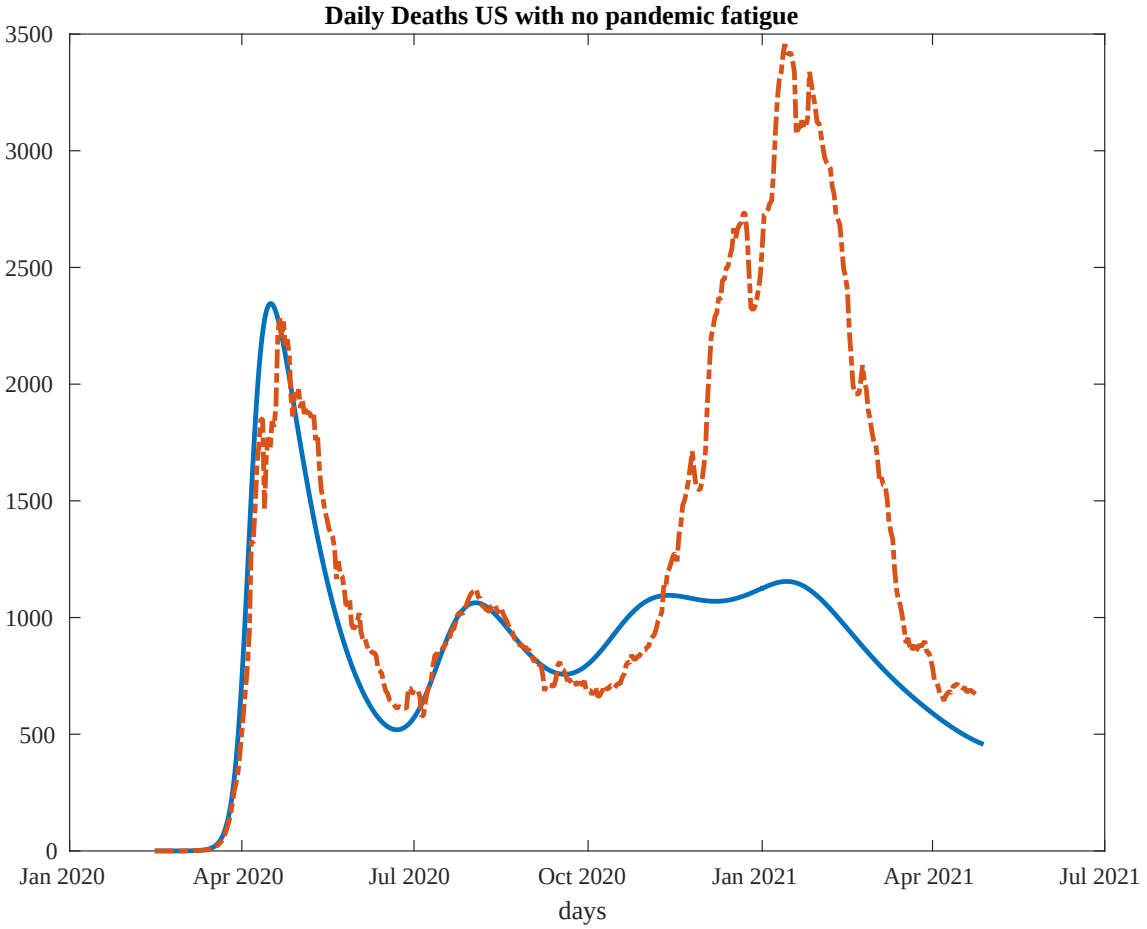

Figure A15: Implications of the model with no pandemic fatigue

To model the transmissibility of the new variant of the virus, I set $\bar{\beta}_v = 4.5\gamma$ giving a basic reproduction in Winter of 4.5. This implies that the new variant is 50% more transmissible than the original variant at any given level of activity. Note that the other epidemiological parameters associated with this new variant are assumed to stay the same, including the infection fatality rate.

The assumption that the new variant of the virus is 50% more transmissible than the original variant of the virus controls how fast it takes over as the dominant variant among those currently infected. In Figure A16 I show the model implications for the fraction of current infections that are due to the new variant of the virus. As can be seen in this figure, under our baseline simulation of the model with vaccines, the new variant becomes dominant quickly in the spring of 2021 after starting from a single infection on December 1, 2020.[33]

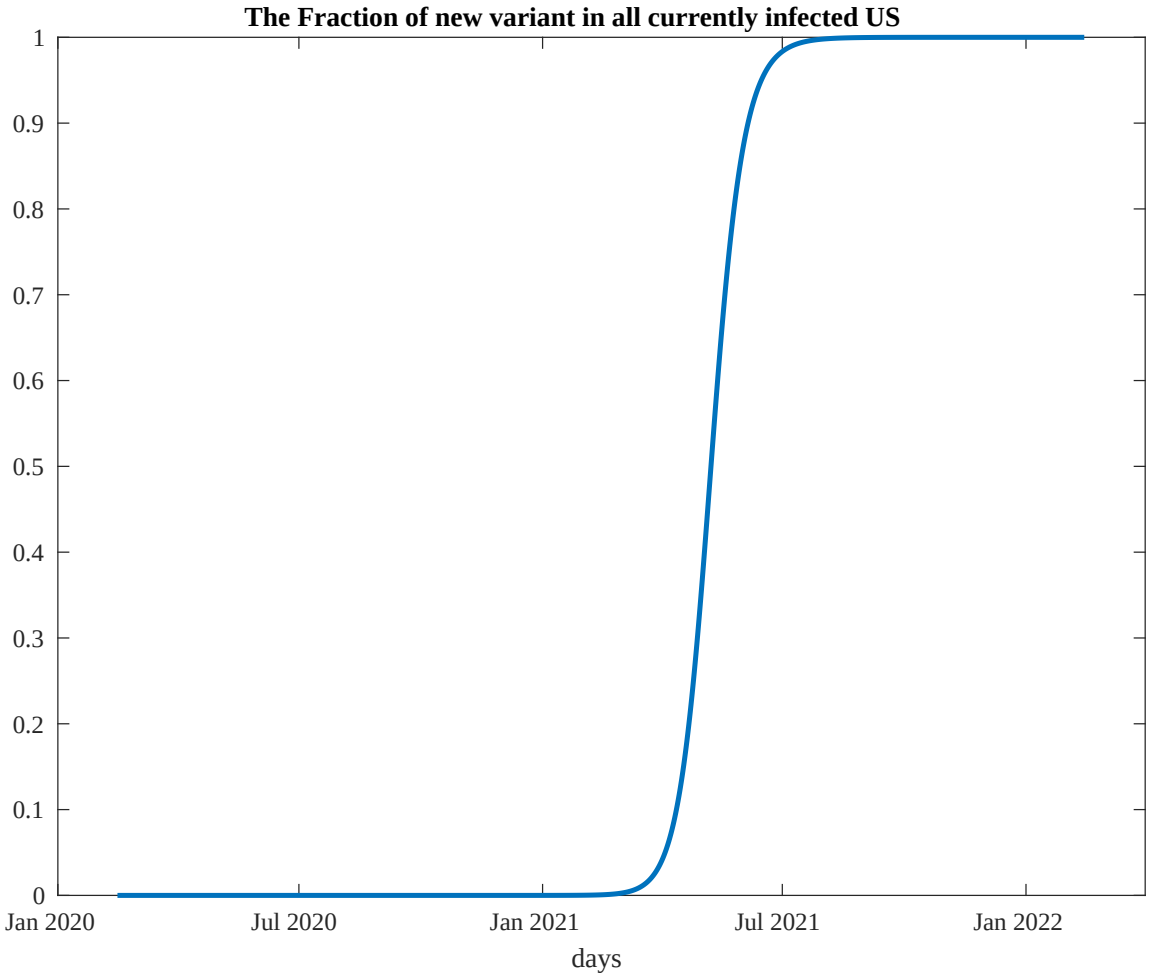

Figure A16: Current infections with the new variant as a fraction of total current infections

[33]CDC estimates of the prevalence of variants this Spring is available here https://covid.cdc.gov/covid-data-tracker under the page labelled "Genomic Suvelliance: variant proportions".

In simulations in which I include a vaccine, I set $\lambda(t) = 0.004$ starting on January 1, 2021. I assume that vaccinations are offered to the general population. This implies that in a population of 330 million, the daily number of vaccines administered is close to 1.3 million. In comparing this number to data on vaccinations, one must take into account that most of the vaccines administered require two doses for full effect. This assumption implies that roughly 51% of the population is fully vaccinated by July 1, 2021. Note that Figure 1 in the main text shows the implications of the model with the baseline parameters but no vaccines. A comparison of Figures 1 and 5 in the main text suggest that the introduction of vaccines has had an important impact on the death toll from COVID in March and April of 2021.

In simulations in which I assume that the Federal government takes steps to mitigate the transmission of the virus through a program such as testing and tracing that is not contingent on the level of daily deaths but instead is contingent on fixed dates, I subtract 0.5 from the seasonal pattern in $\psi(t)$ for those dates. The implications of the model with this additional mitigation are shown in the main text in Figures 3 in a simulation with no vaccines. The implications of the model with additional mitigation and the introduction of vaccines are shown in Figure A17. As is evident in this figure, the extra mitigation substantially reduces the death toll in the first year of the epidemic. As discussed in the main text, in this simulation, the predicted cumulative death toll over a five year period is 302,000.

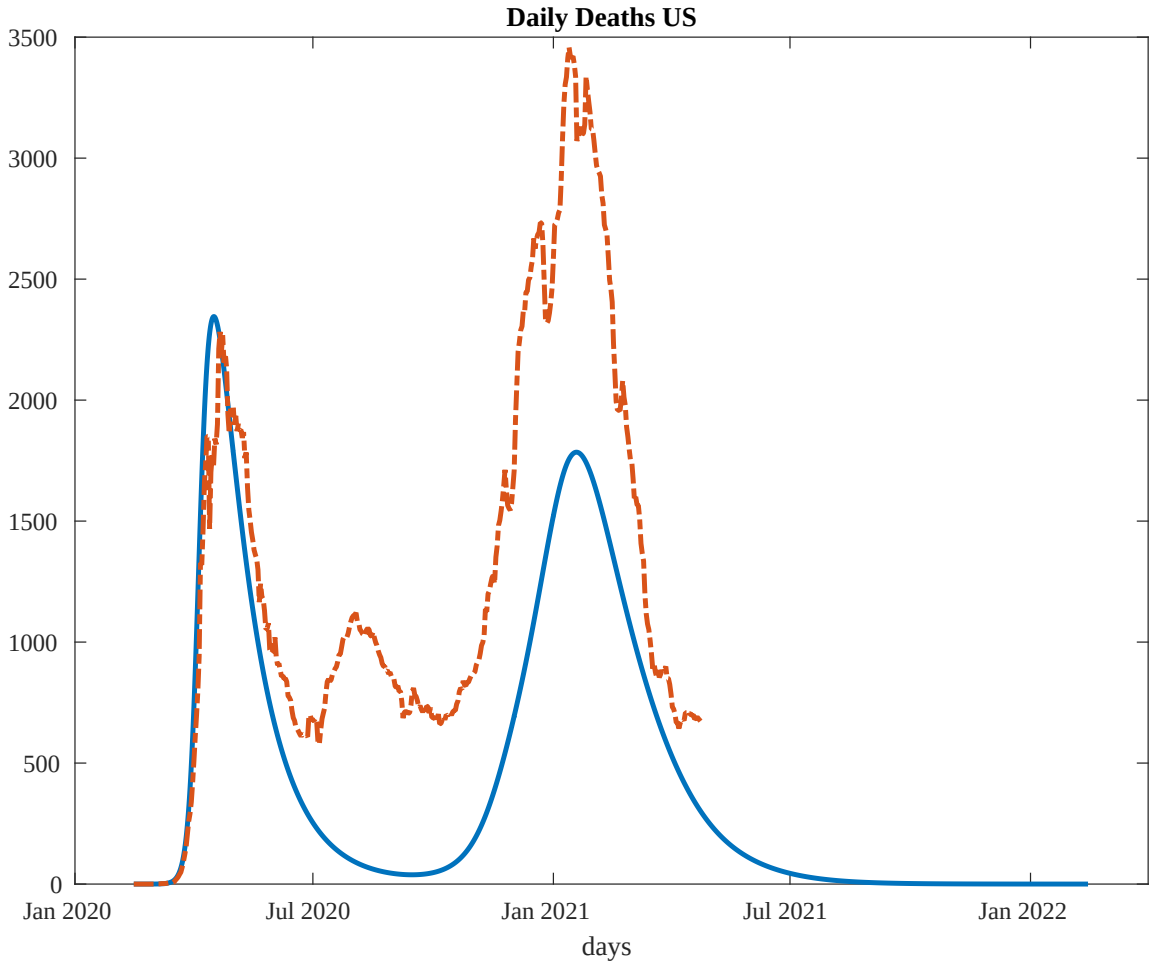

Figure A17: Implications of the model with extra mitigation and vaccines

In simulations in which I assume waning immunity, I set $\xi = 1/547.5$ which corresponds to an expected time to loss of immunity of 18 months. In these simulations, I also assume a higher rate of transmission for the new variant. In this simulation shown in Figure 5 in the main text, I set $\bar{\beta}_v/\gamma = 5$ rather than 4.5.