UNIVERSITY OF CAMBRIDGE

Bennett Institute for Public Policy Cambridge

BENNETT INSTITUTE WORKING PAPER

WANING IMMUNITY AND THE SECOND WAVE: SOME PROJECTIONS FOR SARS-CoV-2

AUTHORS

Chryssi Giannitsarou[1]

Stephen Kissler[2]

Flavio Toxvaerd[3]

ABSTRACT

This paper offers projections of future transmission dynamics for SARS-CoV-2 in an SEIRS model in which immunity is waning over time. In a stylized optimal control setting calibrated to the USA, we show that the optimal social distancing policy is to reduce contact rates in the population for the first half year, and then to slightly relax these policies and switch to a permanent management of an endemic disease state. The optimal policy has three distinct features, namely mitigation, delay and non-eradication. We simulate how the optimal policy depends on how fast immunity wanes and on the infection fatality rate.

JEL Classification: E61, I18.

Keywords: COVID-19, economic epidemiology, social distancing, waning immunity, infection control, SEIRS.

Published: June 2020

Publication from the Bennett Institute for Public Policy, Cambridge
www.bennettinstitute.cam.ac.uk

[1] Faculty of Economics, University of Cambridge and CEPR
[2] Harvard School of Public Health
[3] Faculty of Economics, University of Cambridge

The ongoing COVID-19 crisis has had devastating effects around the world, not just in terms of health outcomes but also in terms of economic and social outcomes at large. This has led to an unprecedented surge of interest in and research on different measures that can be taken to mitigate the damage caused by the epidemic going forward. To this end, much research has resorted to modelling the economic control of the infection through variations of the classical *susceptible-infected-recovered* framework in which recovery from the disease confers permanent immunity. Yet according to the WHO, *"There is currently no evidence that people who have recovered from COVID-19 and have antibodies are protected from a second infection."*[1] This is an important observation, because much of the current policy thinking is based squarely on the assumption that we can rely on herd immunity setting in once societies have navigated the worst stages of the epidemic. The single-minded focus on getting through the first wave of the epidemic potentially leaves us exposed and unprepared to what comes after, in case immunity wanes over time with no effective pharmaceutical measures forthcoming. At the time of writing, we lack a vaccine or any treatments that clearly reduce mortality from SARS-CoV-2 and there are good reasons to believe that such pharmaceutical interventions could only become available several years from now.[2] This means that we may have to rely on some combination of non-pharmaceutical interventions and innovate on how we structure work and social life for the foreseeable future.

To better understand possible scenarios that we may have to face over the next few years, in this paper we offer a comparison of optimal social distancing measures, in the context of a stylized yet flexible SEIRS epidemic model, in which immunity wanes over time. We deliberately keep the model as simple as possible, in order to highlight the effects that waning immunity has on disease dynamics and the conduct of optimal policy. These effects will be qualitatively present in more elaborate models in which the economy is modeled in more detail. We calibrate the epidemic model and optimal control problem to the USA and perform comparative analysis with different assumptions about the waning period and disease-induced death rates.

An important feature of the epidemiological model is that whenever there is waning immunity (SEIRS model) the disease becomes *endemic*. This means that if left uncontrolled, it will never be eradicated, in sharp contrast to a model in which recovery confers permanent immunity (SEIR model). In the SEIR model, the disease will be eradicated in the long run once herd immunity is achieved, even when left uncontrolled. This observation has very important implications for the design of optimal policy while we wait for a vaccine or other pharmaceutical interventions to become available.

Indeed, we find that the optimal social distancing measures with waning immunity are qualitatively different from those when immunity is permanent. In the SEIR model, the optimal policy is to significantly reduce the number of infected individuals around the peak (to flatten the curve). To achieve this, the optimal policy imposes strict social distancing several weeks into the epidemic, when the number of infected individuals starts to pick up. The measures

[1] See https://www.who.int/news-room/commentaries/detail/immunity-passports-in-the-context-of-covid-19

[2] An article that appeared in the New York Times on April 30, 2020 examines how quickly a vaccine for SARS-CoV-2 can be made available to the public. Using their interactive tool and examining various scenarios, we believe that a realistic time frame is at some time in 2025 or 2026. (https://www.nytimes.com/interactive/2020/04/30/opinion/coronavirus-covid-vaccine.html)

are then completely lifted, well before the end of the first year. In the SEIRS framework, the optimal policy is to have less strict social distancing measures, but these are imposed at the very onset of the epidemic and maintained for at least six months. At that point, the measures are temporarily relaxed before being brought back at the start of the second wave. This policy achieves two goals: it *suppresses* the initial and more severe wave of infection and *delays* all future waves of reinfection as much as possible.

The fundamental difference between the models with and without waning immunity is the following. When immunity is permanent, any individual who is infected early in the epidemic will quickly contribute towards herd immunity through recovery. This means that once recovered, an individual becomes a permanent source of benefits to society in reducing the amount of additional infection. When immunity wanes, this is no longer the case. An infected individual that recovers only serves this beneficial role for a limited amount of time and will eventually re-enter the pool of susceptibles, creating the potential for additional waves of infection to take place. This relaxes the planner's incentive to very sharply increase social distancing at any stage of the epidemic. In contrast, the planner will now prefer to intertemporally smoothing the control of the disease. In short, the optimal policy under waning immunity is to implement an essentially permanent but relatively modest policy of social distancing that features *mitigation*, *delay* and *non-eradication*.[3]

In the last section we perform sensitivity analysis with respect to the parameter that governs the loss of immunity. We find that the more quickly immunity wanes, the stricter and longer lasting are the optimal social distancing measures. In addition, we examine how the optimal policy changes for different values of the infection fatality rate (IFR). We do this because estimates for the IFR vary widely across the medical literature, and its value will depend sensitively upon the age distribution and rate of underlying health conditions in the population. Similarly to waning immunity, the higher the IFR is, the stricter and longer lasting are the optimal social distancing measures. It is worth highlighting that for both these changes, the uncontrolled dynamics of the first wave with and without waning immunity are very similar. In contrast, the optimal social distancing policy responds substantially to changes in these parameters and yield substantial improvements in societal welfare and reductions in cumulative disease induced deaths.

The recent literature on the economics of the COVID-19 epidemic is too vast to do justice in this space; for reference, we therefore mention just a few examples of such research that are based on SEIR type frameworks. Eichenbaum et al. (2020), Alvarez et al. (2020) and Krueger et al. (2020) all consider macroeconomic models in the SIR mold. In contrast, Rowthorn and Toxvaerd (2020) consider optimal control of an epidemic with reinfection in an SIS type model. The SEIRS model we consider here nests both the SIR and the SIS models, as it features both immunity (like the SIR model) and the possibility of reinfection (like the SIS model).

[3] By construction, the SEIRS modelling framework does not allow for the possibility of eradication. Moreover, epidemiologists do not believe that eradication is likely, given how widely SARS-CoV-2 has now spread. Yet the theoretical possibility of eradication remains, if the disease were analyzed in a stochastic modeling framework.

1. A Model of Waning Immunity and Optimal Social Distancing

To model the possible future transmission of COVID-19 under optimal control, we resort to a flexible framework that allows for the possibility of presymptomatic infection, waning immunity, vital statistics (births and natural deaths) and disease-induced mortality. Time is continuous and for notational simplicity, we drop the time subscripts t from all time varying variables N, S, E, I, R, D and d, to be defined below.

Our framework is an extended SEIRS model and works as follows. The population is of size N and is subdivided into a number of compartments. Susceptible individuals in compartment S become exposed to the disease by interacting with other individuals who are either exposed (those in compartment E) or have the full-blown disease (those in compartment I). This happens at rate $\beta\left(I+\varepsilon E\right)/N$ where $\beta>0$ is the rate of infection transmission upon contact with someone who has the virus and $0<\varepsilon\leq 1$ is a factor that denotes reduced infectiousness from people who have themselves been exposed but are still not symptomatic (Brauer and Castillo-Chavez, 2012, ch. 9). Exposed individuals transition to the infected and fully symptomatic state I at some rate $\kappa>0$. Once infected, an individual recovers and transitions to compartment R at rate $\gamma>0$ and while in this state, enjoys temporary immunity to new infection. While recovered individuals cannot become infected, the immunity they enjoy wanes over time. We model this by assuming that those in compartment R transition back to the class of susceptible individuals at rate $\alpha\geq 0$. We allow for inflows of susceptible individuals into the population through births at rate $\nu\geq 0$, for outflows though natural death at rate $\mu\geq 0$ from any health state and though disease-induced mortality at rate $\delta\geq 0$. We denote by D those individuals who die either from natural causes or from the disease.

For any policy choice d, to be optimally chosen as described below, the dynamics of the epidemic model we describe here can be summarized by the following equations:

$$\dot{S}=\nu-(1-d)\beta\frac{\left(I+\varepsilon E\right)S}{N}+\alpha R-\mu S, \tag{1}$$

$$\dot{E}=(1-d)\beta\frac{\left(I+\varepsilon E\right)S}{N}-\left(\kappa+\mu\right)E, \tag{2}$$

$$\dot{I}=\kappa E-\left(\gamma+\delta+\mu\right)I, \tag{3}$$

$$\dot{R}=\gamma I-\left(\alpha+\mu\right)R, \tag{4}$$

$$\dot{N}=\nu-\mu N-\delta I, \tag{5}$$

$$\dot{D}=\delta I+\mu N, \tag{6}$$

for some initial state N_0, S_0, E_0, I_0 and R_0.

Turning to welfare considerations, we start our analysis based on the assumption that at some future date T, some combination of pharmaceutical interventions such as antivirals or vaccines will become available. After this point, it is no logner necessary to mitigate infection by implementing non-pharmaceutical interventions.

We assume that susceptible, exposed (i.e. presymptomatic), infected and recovered individuals earn instantaneous flow payoffs y_S, y_E, y_I, y_R and y_D per instant, respectively. If we interpret this model as a reduced-form macroeconomic framework, these parameters can be thought of as units of output produced or income generated per individual in the economy. We assume that

$$y_S = y_E \geq y_R > y_I > y_D = 0 \tag{7}$$

in order to capture the following: first, exposed individuals are not yet symptomatic and thus have the same income as those who are susceptible. That is, since they have no symptoms, they do not experience any drop in productivity, or quality of life, from being exposed. Second, infected individuals experience a decrease in income from becoming infected, e.g. due to a drop in productivity, and an increase once they recover, which could be at the same or lower levels than before becoming infected. Last, deaths constitute a welfare loss counted as the opportunity cost of the income earned while still alive, and we assume that $y_D = 0$.

Next, we introduce the policy instrument d, which we think of as capturing any broad contact-reducing measure that can be scaled from a complete lockdown of all social activity $d = 1$ to a laissez-faire outcome without any restrictions at all $d = 0$. Within the planning horizon $[0, T]$, the social planner chooses the degree of such social distancing $0 \leq d \leq 1$ at each moment t to optimally reduce the infectiousness parameter to $(1 - d) \times \beta$. At the two extremes, full social distancing completely eliminates additional infection, while the complete cessation of social distancing delivers the uncontrolled time path of the disease. For simplicity, we assume that the aggregate cost of social distancing measures for the society as a whole is given by

$$C(d) = \theta \frac{d^{1+\eta}}{1+\eta}, \tag{8}$$

where θ, $\eta > 0$ are parameters.

We consider the problem of a benevolent utilitarian social planner who chooses a time-path of social distancing that maximizes the net present value of welfare across the population during the period $[0, T]$. The planner's problem is given by

$$\max_{d \in [0,1]} \int_0^T e^{-\rho t} \left(y_S S + y_E E + y_I I + y_R R - \theta \frac{d^{1+\eta}}{1+\eta} \right) dt, \tag{9}$$

subject to (1)-(6) and the following set of equations,

$$N = S + E + I + R, \tag{10}$$

$$0 \leq d \leq 1, \tag{11}$$

for given initial N_0, S_0, E_0, I_0, R_0. In the Online Appendix, we show that the optimal social distancing policy is given by

$$d^* = \max \left\{ 0, \ \min \left\{ 1, \ \left(\frac{e^{\rho t}}{\theta} (\lambda_S - \lambda_E) \beta \frac{(I + \varepsilon E) S}{N} \right)^{1/\eta} \right\} \right\}, \tag{12}$$

where λ_S and λ_E are the co-state variables for constraints (1)-(2), respectively.

This expression simply equates the marginal cost of social distancing with its marginal benefit. Note that social distancing reduces additional transition from susceptible to exposed at rate $\beta (I + \varepsilon E) S/N$. In turn, this increases social welfare at rate $(\lambda_S - \lambda_E)$, which is the net cost, measured in terms of shadow prices, of additional exposed people. In the Online Appendix, we also state the additional necessary conditions for an optimal policy, namely the differential equations governing the evolution of the co-state variables and transversality conditions.

2. Projected Post-Epidemic Scenarios

We consider the dynamics of the uncontrolled and controlled SEIRS models and compare them to the dynamics of their SEIR counterparts. At the end of this section, we also perform some sensitivity analysis with respect to the waning immunity parameter α and the death rate δ.

2.1. Calibration and solution method. The dynamics under optimal control are solved for numerically by using a forward-backward sweep method, as described in Lenhart and Workman (2007, chapter 12). We parameterize our model to the USA. First, the model is calibrated to weekly frequency as is conventional in studies of epidemics. The planning horizon is set to six years, i.e. $T = 6 \times 52$, with a view that a vaccine or cure will become available by then. The estimated horizon for a vaccine to be available for individuals is in the most optimistic cases at least 1.5 years from around the start of the epidemic, but most sources and studies suggest that at least 5 years are needed, while there is a good chance that it will take more than a decade for a vaccine to be developed. We measure the population and epidemic compartment sizes in millions of individuals, so that the initial conditions for the model are $N_0 = 330$ (millions of people) and we assume that initially, there are approximately one in ten million individuals infected, so $I_0 = 0.0000033$. Moreover, we assume that $E_0 = 3 \times I_0$, i.e. that there are initially some exposed individuals who are still presymptomatic. These numbers correspond to those used by Atkeson (2020).

The epidemic parameters are taken from Kissler et al. (2020) and amended where necessary. In particular, we set the natural birth rate to $\nu = 3.8/52$, to match the fact that there are approximately 3.8 million babies born in the USA per year. The natural death rate is set to $\mu = 1/(80 \times 52)$ to capture an average expected lifespan of 80 years. The transition rate from infected to recovered is set to $\gamma = 7/14$, capturing initial estimates that it can take up 14 days for infected individuals to recover. We set $\beta = 3 \times \gamma$, to match the standard SIR calibration which assumes that the basic reproduction rate for SARS-CoV-2 is somewhere between 2.5 and 3.5. The medical and biology literature suggest that the median incubation period is five days so we set $\kappa = 7/5$ (Kissler et al., 2020 and references therein).

There is considerable uncertainty regarding the remaining epidemic parameters. First, the parameter ε captures the infectivity of individuals that are exposed but not yet symptomatic. We assume as a benchmark that $\varepsilon = 0.5$, i.e. that presymptomatic individuals are half as infective as the symptomatic ones, and we note that the results are not very sensitive to small changes in this parameter.[4] The parameter δ captures the death rate of infected individuals due to SARS-CoV-2 . The Imperial College COVID-19 Response Team (Report 9, 16 March 2020) estimates the IFR to be 0.9% on average, so we set $\delta = 0.009 \times \gamma$. Current estimates for the IFR for SARS-CoV-2 vary quite a lot, but some recent research with data from China (Verity e. al., 2020) and France (Salje et al., 2020) estimate the IFR at 0.66% and 0.7% respectively. Finally turning to the main parameter of interest α, we start by assuming as a benchmark that $\alpha = 1/52$, so that immunity wanes in about a year from from the time of infection. While

[4] This value for SARS-CoV-2 is also used by Davies et al. (2020).

Parameter	Value	Source
ν	$3.8/52$	USA average 3.8 births per week
μ	$1/(80 \times 52)$	USA life expectancy 80 years
γ	$7/14$	Kissler et al. (2020)
β	$\gamma \times 3$	Kissler et al. (2020)
κ	$7/5$	Kissler et al. (2020) and references therein
ε	$1/2$	N/A
δ	$\gamma \times 0.009$	IFR, Imperial College COVID-19 Response Team, Report 9
α	$1/52$	Kissler et al. (2020)
ρ	4%	Standard macro discount rate
y_S	1	Normalized 1 unit of income
y_E	1	Equal to y_S
y_I	0.9	10% reduction in productivity due to infection
y_R	1	Equal to y_S
η	1	Quadratic costs
$\theta/2$	330×0.165	Aggregate income loss (Strong and Wellbourn, 2020)
T	6×52	Estimated time until arrival of vaccine or cure (REF)
N_0	330	US population, 330m
I_0	0.000033	1 in 10m initially infected, Atkeson (2020)
E_0	$3 \times I_0$	Atkeson (2020)

Table 1: Parameter values for benchmark.

at this stage it is not possible to know whether immunity wanes and if so, how quickly, the epidemiology literature suggests that related coronavira, e.g. HCoV-HKU1 and HCoV-OC43, often have immunity that wanes between less than one year to two years after infection (Callow et al., 1990, Kissler et al., 2020 and Galanti and Shaman, 2020).[5]

Finally, we turn to the economic parameters of the model. First, we set $\eta = 1$ so that the costs of social distancing are quadratic. We normalise the production per individual to a unit, and set the corresponding parameters to $y_S = 1$, $y_E = y_S$, $y_I = 0.9$ and $y_R = 1$. This reflects that all but infected individuals have the same productivity, but the infected produce at 90% of the normalized value. Since these numbers are viewed as income per individual, we calibrate the parameter θ so that the cost of distancing d in the optimization problem can be interpreted as the cost of social distancing per individual in the initial population. With this in mind, we set $\theta = 330 \times 0.33$, so that the factor multiplying $d^{1+\eta}$ is $330 \times 0.33/2$, where the number $0.33/2 = 0.165$ is interpreted as the social distancing cost per person. This is consistent with the magnitudes reported by Strong and Wellbourn (2020), who estimate that the decline in income due to social distancing across a variety of sectors and lockdown scenarios is between 4.6 and 25.6%. The parameters values we use in our simulations are listed in Table 1.

2.2. The effects of optimal social distancing on infection. Figures 1 and 2 show the disease dynamics for the model with permanent immunity (SEIR) and waning immunity (SEIRS) respectively, for the benchmark calibration. In each of the two figures, we plot the dynamics of both the uncontrolled epidemic and those under optimal social distancing. The uncontrolled epidemics are shown with dotted lines, while the controlled dynamics are shown in solid lines. The lower left panel shows the differences in cumulative disease induced deaths

[5] For HCoV-HKU1 and HCoV-OC43, Kissler, et al. (2020) calibrate $1/\alpha$ to be 45 weeks.

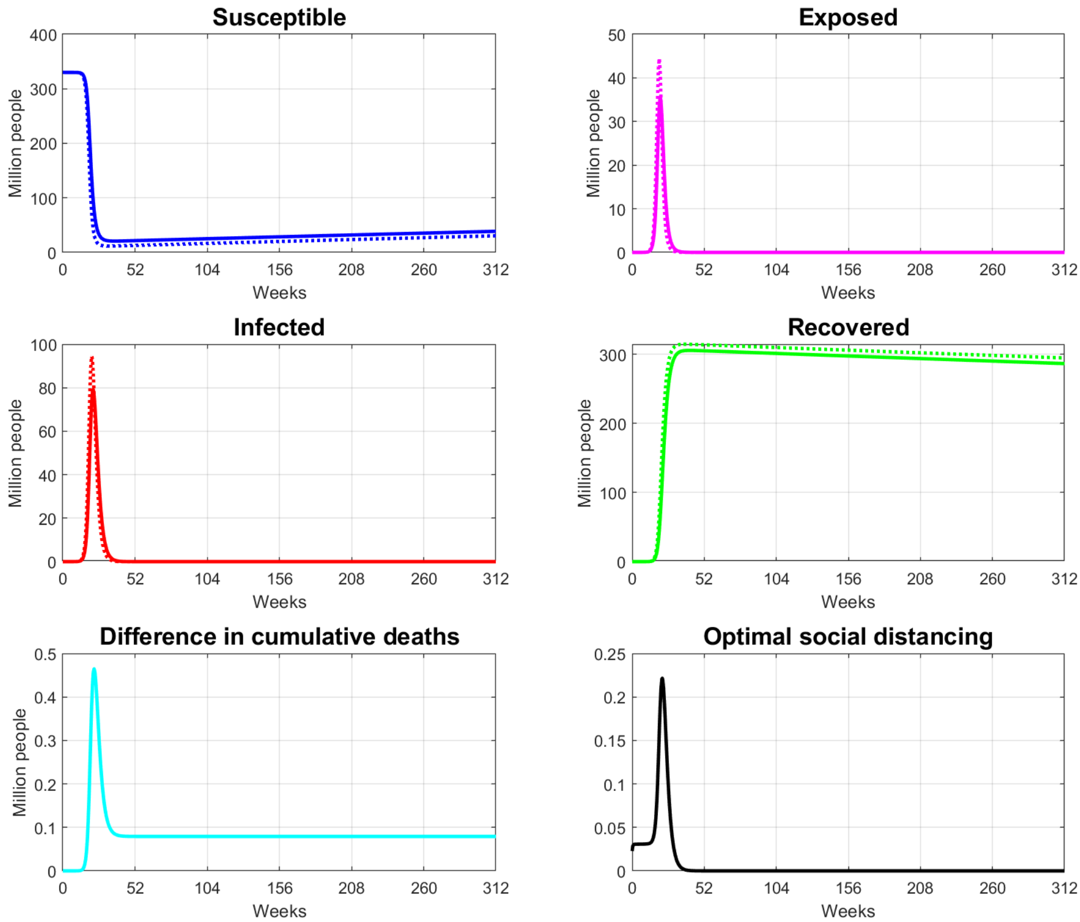

Figure 1: Dynamics for the SEIR model ($\alpha = 0$) under the benchmark parameterization. Dotted lines: dynamics under uncontrolled (epidemic) model. Solid lines: dynamics under optimal social distancing. Lower left panel: difference in cumulative disease-induced deaths under controlled and uncontrolled dynamics. Lower right panel: optimal social distancing.

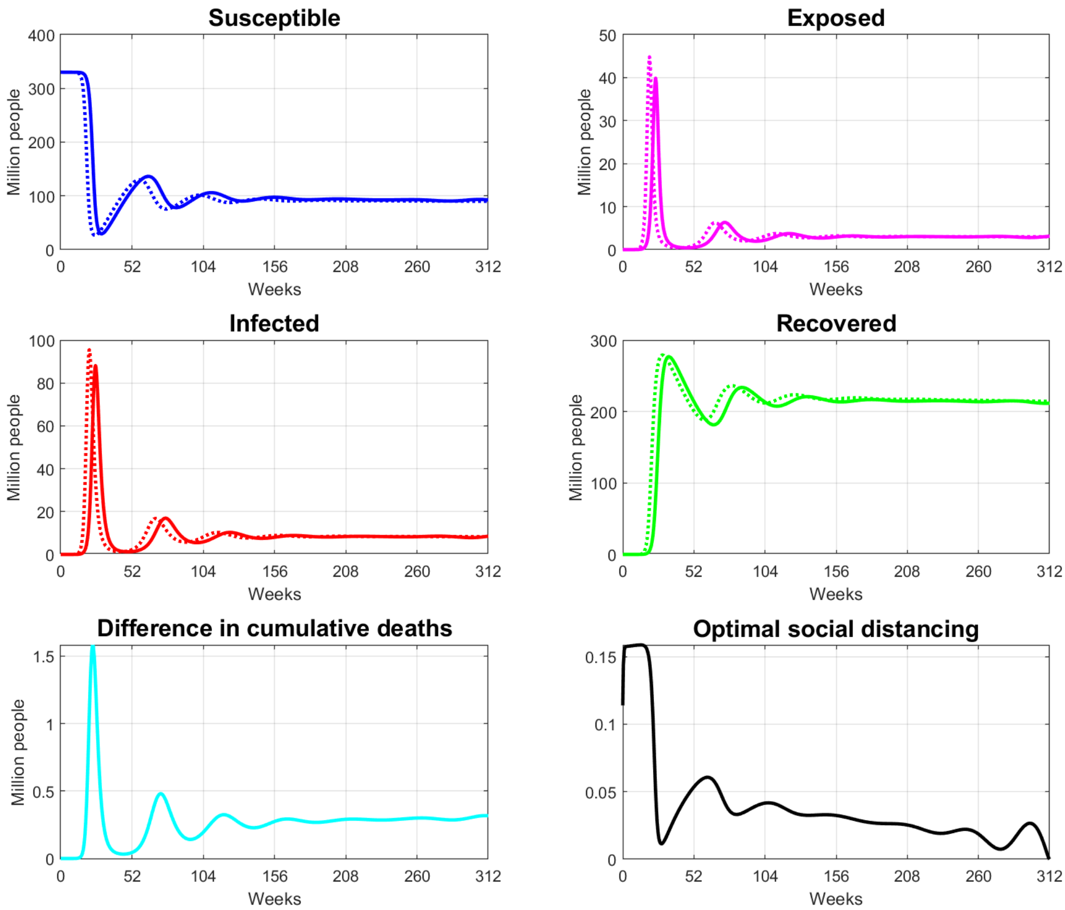

Figure 2: Dynamics for the SEIRS model under the benchmark parameterization. Dotted lines: dynamics under uncontrolled (epidemic) model. Solid lines: dynamics under optimal social distancing. Lower left panel: difference in cumulative disease-induced deaths under controlled and uncontrolled dynamics. Lower right panel: optimal social distancing.

and the lower right panel the path of optimal social distancing.

We first describe how the dynamics of the two uncontrolled models compare. The uncontrolled SEIRS dynamics initially share some properties of the uncontrolled SEIR counterparts. In particular, susceptible individuals initially decrease while recovered individuals increase. Infection first picks up and then decreases, an effect of susceptibles becoming scarcer and some measure of herd immunity setting in. In the SEIR model, this essentially completes the description of the dynamics. In the SEIRS model however, immunity is only temporary and wanes over time. This means that the large number of recovered individuals that accumulate at the initial stages of the epidemic now start migrating back to the pool of susceptibles, forming the basis of an additional wave of infection. This new reservoir of susceptibles creates the potential for a smaller second wave of infection, with the concomitant effects on susceptibles and recovered individual. Thus, in the uncontrolled SEIRS model, the dynamics can display dampened oscillations. In contrast to the SEIR dynamics, in the SEIRS system the disease is endemic, which means that a fraction of the population remains infected in perpetuity. While infected people do recover, waning immunity means that they eventually find their way back to the susceptibles, at which point they can become reinfected.

Next we turn to discussing the optimal social distancing policy for the models with and without waning immunity. In the SEIR framework, when immunity never wears off, the path of optimal social distancing, is straightforward and looks a lot like the policies that have been used so far across the world to control the evolution of COVID-19: in the beginning of the epidemic there is little social distancing, which then sharply increases as the peak of the epidemic approaches, and then sharply decreases as infected individuals recover, gain immunity and remain healthy thereafter. We note that at around the peak of the epidemic, sometime in week 20 after the epidemic starts, optimal social distancing in the SEIR framework reaches its maximum level of about 25.5%, however this lasts for a short period of time, and is essentially phased out by week 40. The effects of the social distancing policy are as expected a flattening of the curve and the slightly lengthier epidemic. The number of infected individuals under the optimal social distancing policy at the peak of the epidemic is about 20% smaller than that in the uncontrolled model.

In the SEIRS framework, the optimal social distancing policy is quite different. While restrictions start out at a higher level than in the SEIR model, the level never reaches more than 16.5%. These measures are sustained for about half a year and then dropped temporarily, only to be increased again intermittently when the second and following smaller waves arise. We note several interesting features. First, optimal control delays the different phases of the epidemic, all the while dampening disease incidence. It is noteworthy that even under optimal social distancing, the different trajectories remain non-monotone and so optimal policy still allows for varying disease intensity over time, i.e. it is part of the optimal policy to allow infection to increase temporarily. The overall pattern of the optimal policy with waning immunity is to impose and maintain some social distancing measures for the duration of the time horizon. As can be seen from the plots, the optimal control of social distancing has two main phases, namely an initial phase with significant suppression and a secondary phase characterized by varying management of subsequent infection waves. When immunity is waning, all those individuals who are infected in the initial wave of the epidemic go though the phases of the disease only to become susceptible to infection again. This is exactly what creates the second and subsequent waves. Under optimal control, the second and subsequent waves are still present, but dampened and postponed, relative to the uncontrolled case. Here, the number of infected individuals under the optimal social distancing policy at the first peak of the epidemic is about 8.5% smaller than that in the uncontrolled model, but the optimal policy now also flattens future waves of the epidemic and ensures that the long run endemic level of infected individuals is lower compared to the uncontrolled model.

Notably, when comparing the difference in cumulative disease induced deaths between controlled and uncontrolled models, we see that by the end of the horizon there are approximately 85,000 deaths difference for the SEIR model, while for the SEIRS model with waning immunity, the optimal policy reduces the deaths by approximately 315,000 people.

An important observation when comparing the two frameworks is that under the SEIR dynamics, where immunity is permanent, there would never be a second wave if the optimal social distancing policy were followed. In other words, were immunity not waning, a second wave would be a tell-tale sign that restrictions were lifted too early, unless of course there are other effects present like environmentally determined changes in infectiousness, such as seasonal variation, etc. But this is not the case under waning immunity; it is in fact not socially optimal to eradicate the disease through social distancing, as it would entail an essentially permanent closing of society at a cost that is not justifiable. A similar result is shown formally in Toxvaerd (2019) and in Rowthorn and Toxvaerd (2020). In an SIS setting, they show that a path that completely eradicates the disease through social distancing measures does not satisfy the required transversality condition for an optimal path. This does not imply that full eradication in the presence of waning immunity should not be pursued as a worthwhile goal; but that in order to achieve eradication, other policies must be used, such as vaccines and antivirals.

Another interesting outcome is that the optimal suppression policies under the SEIR and the SEIRS dynamics are qualitatively very different at the early stages of the epidemic. Without waning immunity, the best policy is to delay intervention for several weeks until infection picks up and then impose restrictions on social interaction. In contrast, when immunity is waning, the optimal policy is to immediately impose restrictions from the outset and to keep these in place for a period of about half a year. The intensity of the measures is lower, but more prolonged. From the perspective of a policy maker, it is therefore important to have a sense of which of the two models best describes the ongoing COVID-19 pandemic. While it is of course impossible to know at this early stage whether immunity wanes in the medium run, various medical studies seem to suggest that the relation of SARS-Cov-2 to other coronavira with waning immunity may be a good indicator for this.

We close by noting that the figures for the SEIR and SEIRS models suggest that the numbers of infected individuals at about four months into the epidemic are substantially higher than those currently projected for the USA.[6] A variety of factors may contribute to explaining this. First, epidemic parameters such as the IFR, the incubation period, or the infectivity of

[6] Our assumption is that the epidemic in the USA is initiated some time in early February, which means that at the time of writing, the USA is in week 16-17 of the epidemic.

presymptomatic individuals are not yet fully understood. In practice, they may differ from the ones used here. Second, it is thought that there are many more cases of infected individuals than those confirmed by testing, since a large fraction of the population are tested only when they exhibit symptoms; numerous reports since February 2020 suggest that more than one third of the world's population could end up being infected. In a recent unpublished paper, Swanson and Cossman (2020) suggest that the estimated number of cases in the USA is somewhere between 7.5 and 14 times higher than the reported number of infected cases. At the time of writing, on June 1, 2020, there are about 1.8 million reported/confirmed cases, which implies that the actual number of cases of infection may currently be between 13.5 and 25 million individuals. Our benchmark calibration generates numbers of comparable order for weeks 16 and 17 of the epidemic, for the uncontrolled model. Third, the epidemic models and the optimal control problem itself are rather stylized and the parameters governing the optimal social distancing policy are normalized to an arbitrary one unit of income per individual. Last, the calculation of the optimal social distancing policy does not factor in any spontaneous social distancing that individuals may be engaging in as a result of a different personal optimization problem than that of the social planner (see Toxvaerd, 2020).

3. Sensitivity Analysis

3.1. The effects of waning immunity. In the baseline parameterization, we assume that immunity wanes over a period of one year. In this section, we consider the effects of changes in waning time on the optimal social distancing policy and the concomitant path of infection over time. In Figure 3, we show the prevalence of infection and the optimal social distancing policy for waning periods $1/\alpha = 26$ weeks to 312 weeks (six years) and also include $\alpha = 0$. This corresponds to gradually changing the model from an SEIRS to an SEIR model in which there is effectively no return to susceptibility.

Looking at the top panel, we first note that the prevalence of infection in the uncontrolled model always exhibits a high first peak at around 20-22 weeks into the epidemic. These first peaks seem almost identical irrespective of the waning period. As the first wave subsides, a new wave occurs, whose peak is a lot lower than that of the first wave. In addition, its arrival time also depends on how quickly immunity wanes. The peak size of the second wave is smaller for longer-lasting immunity. In the limit, as α approaches zero, we revert to the SEIR model, which does not generate a second wave.

The middle panel shows the prevalence of infection under the optimal social distancing policy. Qualitatively these dynamics look like those in the uncontrolled system. With waning immunity, the controlled infection dynamics display dampened oscillations, a behavior inherited from the underlying dynamics of the uncontrolled SEIRS model. But the optimal policy has two additional effects. First, all the waves are dampened and second, they are pushed forward in time relative to the waves in the uncontrolled system. Moreover, the extent to which the first large wave is dampened and pushed forward now depends on the waning parameter, in contrast to the uncontrolled dynamics. Interestingly, this dependence is not monotonic. This may stem from the fact that the planning horizon is finite: when increasing the waning period above some threshold, later waves become less likely to occur within the planning horizon and so it becomes less important for the policy maker to keep the total number of infected individuals down.

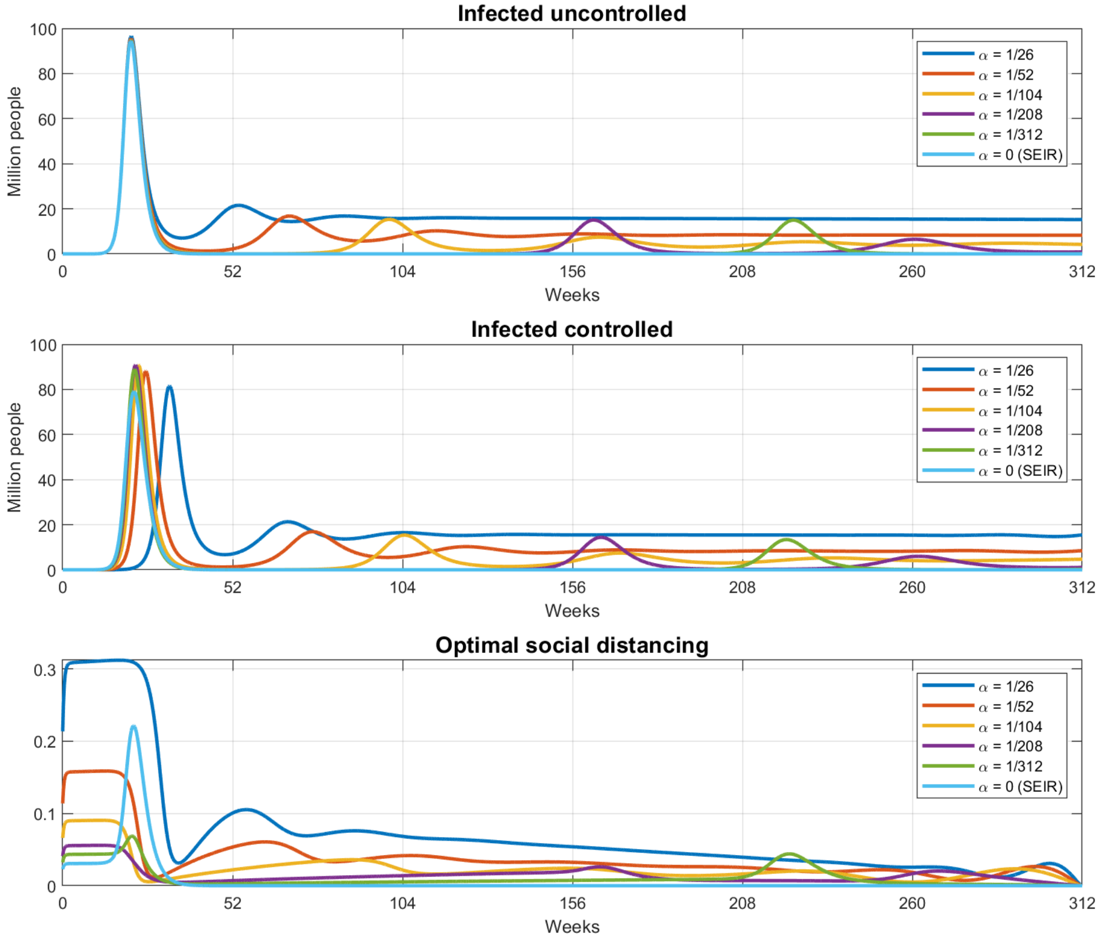

Figure 3: *Effects of waning immunity.* Top panel shows the number of infected individuals in the epidemic (uncontrolled) model, middle panel shows the infected for in the optimally controlled model and the bottom panel shows the optimal social distancing policy. All panels are plotted for $1/\alpha = 26,\ 52,\ 104,\ 208,\ 312$ and ∞ weeks.

In the bottom panel, we show the optimal social distancing policy for the different values of the waning parameter. We observe that there is a qualitative shift in the optimal policy as we make this change. For sufficiently long waning periods, the optimal policy prescribes an immediate and substantial reduction in social interactions lasting from three to six months, followed by intermittent periods of heightened control in order to manage subsequent waves of infection. As waning is reduced, the optimal policy increasingly resembles the optimal policy under SEIR dynamics. At initial stages, there is modest social distancing but as infection picks up, the optimal policy ratchets up rapidly, only to be subsequently phased out completely. Combining these observations with the evolution of infected individuals under the optimal policy, we can explain why we observe this non-monotonic flattening of the first peak described above.

We also observe another interesting effect of the optimal social distancing policy: Aggressive social distancing at the start of the epidemic for larger α has the effect of pushing the first peaks further down the time line. This happens because as immunity wanes more rapidly, the policy maker hopes to suppress as many of the potential peaks as possible before the end of the planning horizon.

3.2. The importance of disease-induced mortality. Our last experiment involves sensitivity analysis with respect to the rate of disease-induced death. Estimates for this parameter for COVID-19 seem to vary substantially across studies, countries and time. We plot the dynamics of infected individuals for three estimates of the IFR, namely 0.3%, 0.9% (benchmark) and 1.8%. The results are presented in Figure 4. We observe that the effects of these changes in the death rate δ are imperceptible for the uncontrolled system (top panel). However, the controlled dynamics of infected individuals differ significantly for the three parameters: at the top of the first peak, there are 3%, 8.5% and 19.4% fewer infected individuals for IFRs 0.3%, 0.9% and 1.8% respectively. At the end of the six year time horizon, the reduction in cumulative disease induced deaths relative to the uncontrolled model is 43,000, 316,500 and 1,550,000 for IFRs 0.3%, 0.9% and 1.8% respectively. The takeaway from this exercise is that a correctly designed social distancing policy can make a substantial difference in terms of the total number of deaths in the context of SEIRS models with waning immunity.

The optimal policy for social distancing is intuitive: the higher the fatality rate is, the stricter are the measures for social distancing at the start of the epidemic, and the longer they last. Following the initial relaxation of the measures, the social distancing increases again to suppress the second wave, in a monotonic way, i.e. by more and for longer the higher the fatality rate.

4. Conclusion

In this paper, we consider the possible future scenario in which individuals who recover from COVID-19 only gain temporary immunity. With such waning immunity, the dynamics of the disease, both under optimal control and under no control, are qualitatively different from a setting in which recovered individuals gain permanent immunity. In particular, with waning immunity, the dynamics of the disease can displays dampened oscillations in which additional waves of infection hit the population. These are brought about because those individuals who

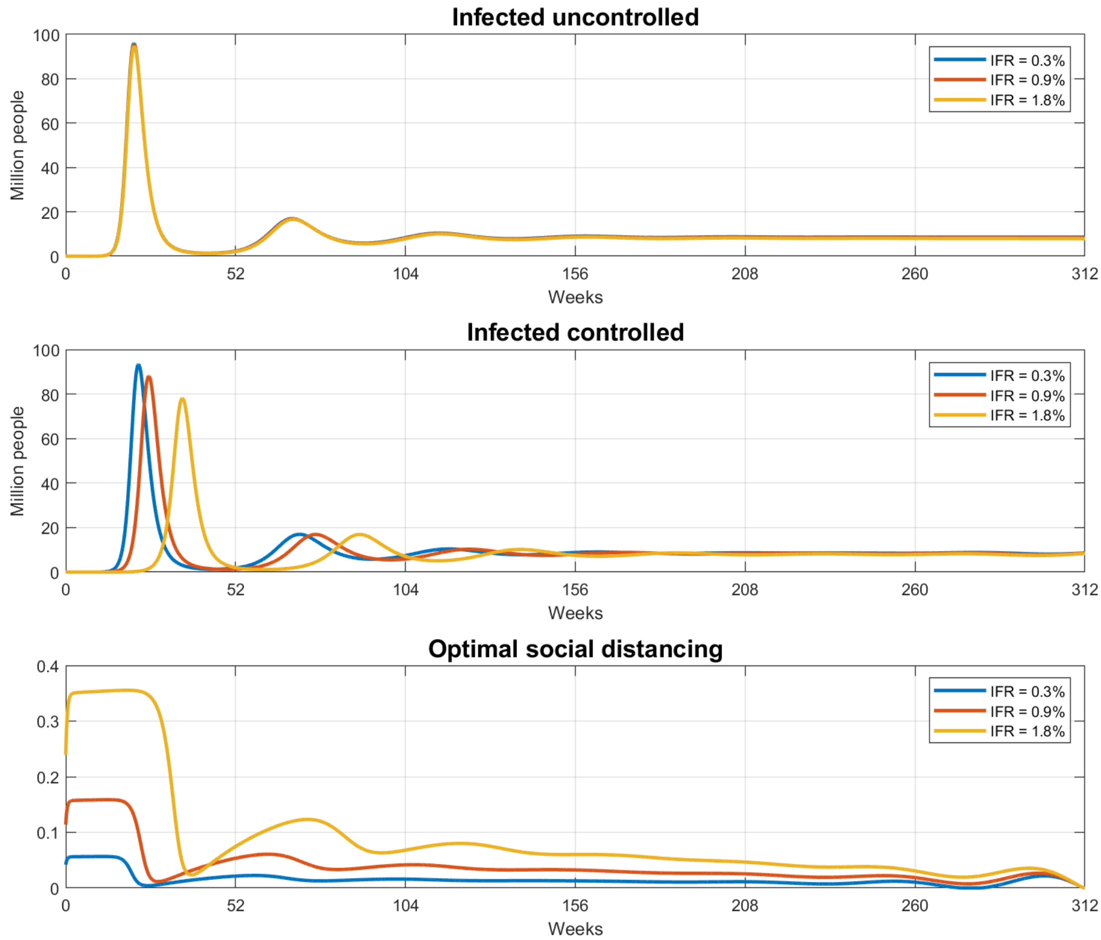

Figure 4: *Sensivity with respect to the infection fatality rate.* Top panel shows the number of infected individuals in the epidemic (uncontrolled) model, middle panel shows the infected in the optimally controlled model and the bottom panel shows the optimal social distancing policy. All panels are plotted for infection fatality rate (IFR) 0.3%, 0.9% and 1.8%.

survive infection eventually become susceptible to reinfection, thus setting the stage for recurrent waves. These waves are dampened under optimal infection control policies, but not eliminated. In addition, under the possibility of reinfection, the disease becomes endemic and it is not socially optimal to eradicate it through the permanent imposition of social distancing.

As a practical matter, it is too early to tell whether immunity to reinfection is temporary or permanent. This is something that can only be established through longitudinal studies, which may take years. Experience from related vira suggest that there is a realistic prospect that some waning may take place, in which case societies may be forced to find ways to live with the disease, rather than return to pre-epidemic norms. For this reason, it is important to start planning for that possibility now. The present research is a first step in this direction. We show that research that relies significantly on the emergence of permanent immunity for individuals (and thus eventual herd immunity in the population) may give incomplete policy recommendations for how to navigate the current epidemic.

References

[1] Alvarez, F., D. Argente and F. Lippi (2020). A Simple Planning Problem for COVID-19 Lockdown, *NBER Working Paper* 26981.

[2] Atkeson, A. (2020): What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios, *NBER Working Paper* 26867.

[3] Brauer, F. and C. Castillo-Chavez (2012): Mathematical Model in Population Biology and Epidemiology, *Springer*.

[4] Callow, K. A., H. F. Parry, M. Sergeant and D. A. Tyrrell (1990). The Time Course of the Immune Response to Experimental Coronavirus Infection of Man, *Epidemiology and Infection*, 105(2), 435–446. doi: 10.1017/s0950268800048019.

[5] Davies, N.G., P. Klepac, Y. Liu, K. Prem, M. Jit, CMMID COVID-19 working group, R. M. Eggo (2020). Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics, unpublished manuscript, medRxiv doi: https://doi.org/10.1101/2020.03.24.20043018.

[6] Eichenbaum, M., S. Rebelo and M. Trabrandt (2020). The Macroeconomics of Epidemics, *NBER Working Paper* 26882.

[7] Imperial College COVID-19 Response Team (2020). Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand, Report 9.

[8] Kissler, S. M., C. Tedijanto, E. Goldstein, Y. H. Grad and M. Lipsitch (2020). Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period, *Science*, 368(6493), 860–686. doi: 10.1126/science.abb5793

[9] Krueger, D., H. Uhlig and T. Xie (2020). Macroeconomic Dynamics and Reallocation in an Epidemic, *NBER Working Paper* 27047.

[10] Lenhart, S. and J. T. Workman (2007). Optimal Control Applied to Biological Models. Mathematical and Computational Biology Series, *Chapman & Hall/CRC*, London.

[11] Rowthorn, R. and F. Toxvaerd (2020). The Optimal Control of Infectious Diseases Via Prevention and Treatment, *Cambridge-INET Working Paper Series* No: 2020/13.

[12] Salje, H., C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hoze, J. Richet, C-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P-Y. Boelle, S. Cauchemez (2020). Estimating the burden of SARS-CoV-2 in France, *Science*, 13 May 2020, doi: 10.1126/science.abc3517.

[13] Strong, A. and J. W. Welburn (2020). An Estimation of the Economic Costs of Social-Distancing Policies, Research Report RR-A173-1, *RAND Corporation*.

[14] Swanson, D. and R. Cossman (2020). A Simple Method for Estimating the Number of Unconfirmed COVID-19 Cases in a Local Area that Includes a Confidence Interval: A

Case Study of Whatcom County, Washington, unpublished manuscript, medRxiv, doi: 10.1101/2020.04.30.20086181.

[15] Toxvaerd, F. (2019). Rational Disinhibition and Externalities in Prevention, *International Economic Review*, 60(4), 1737-1755.

[16] Toxvaerd, F. (2020). Equilibrium social Distancing, *Cambridge-INET Working Paper Series* No: 2020/08.

[17] Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunube, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson (2020). Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis, *The Lancet Infectious Diseases*, 20(6), 669-677, https://doi.org/10.1016/S1473-3099(20)30243-7.

A. Flow Chart of the SEIRS Model with Demographics

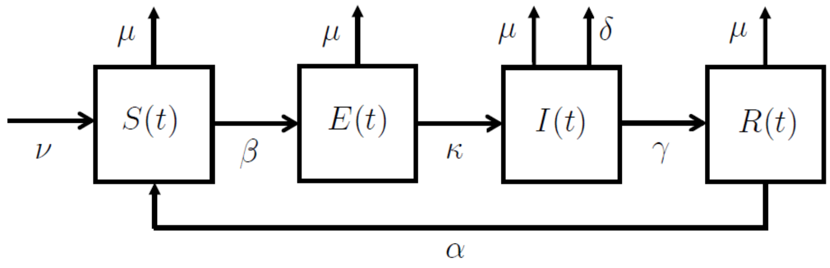

Figure 5: States and flows in the SEIRS model.

B. Optimal Social Distancing Policy

For the derivation, we use directly that $\eta = 1$. We start with the five differential equation constraints that describe the dynamics of the system:

$$\dot{S} = \nu - (1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N} + \alpha R - \mu S \tag{13}$$

$$\dot{E} = (1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N} - (\kappa + \mu)\,E \tag{14}$$

$$\dot{I} = \kappa E - (\gamma + \delta + \mu)\,I \tag{15}$$

$$\dot{R} = \gamma I - (\alpha + \mu)\,R \tag{16}$$

$$\dot{N} = \nu - \mu N - \delta I \tag{17}$$

Because of the accounting equation, the social planner's problem can be reduced to one with only four differential equation constraints. We do so by eliminating R and substituting in $R = N - S - E - I$. The constraints thus become

$$\dot{S} = \nu + \alpha N - (\alpha + \mu)\,S - \alpha E - \alpha I - (1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N}, \tag{18}$$

$$\dot{E} = (1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N} - (\kappa + \mu)\,E, \tag{19}$$

$$\dot{I} = \kappa E - (\gamma + \delta + \mu)\,I, \tag{20}$$

$$\dot{N} = \nu - \mu N - \delta I. \tag{21}$$

Letting the co-state variables for the constraints be denoted by λ_S, λ_E, λ_I and λ_N, the planner's Hamiltonian equation is given by

$$\begin{aligned} H \;=\;& e^{-\rho t}\left[(\pi_S - \pi_R)\,S + (\pi_S - \pi_R)\,E + (\pi_I - \pi_R)\,I + N\pi_R - \frac{\theta}{2}d^2\right] \\ &+\lambda_S\left[\nu + \alpha N - (\alpha + \mu)\,S - \alpha E - \alpha I - (1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N}\right] \\ &+\lambda_E\left[(1-d)\beta\,(I + \varepsilon E)\,\frac{S}{N} - (\kappa + \mu)\,E\right] + \lambda_I\left[\kappa E - (\gamma + \delta + \mu)\,I\right] \\ &+\lambda_N\,(\nu - \mu N - \delta I) \end{aligned} \tag{22}$$

The first order condition with respect to d is

$$\frac{\partial H}{\partial d} = -e^{-\rho t}\theta d + (\lambda_S - \lambda_E)\,\beta\,(I + \varepsilon E)\,\frac{S}{N} = 0. \tag{23}$$

The laws of motion for the co-state variables are given by

$$\dot{\lambda}_S = \lambda_S\left[\beta(1-d)\frac{(I+\varepsilon E)}{N} + \alpha + \mu\right] - \lambda_E\beta(1-d)\frac{(I+\varepsilon E)}{N} - e^{-\rho t}\left(\pi_S - \pi_R\right), \tag{24}$$

$$\dot{\lambda}_E = \lambda_S\left[\alpha + \beta\varepsilon(1-d)\frac{S}{N}\right] + \lambda_E\left[\kappa + \mu - (1-d)\beta\varepsilon\frac{S}{N}\right] - \kappa\lambda_I - e^{-\rho t}\left(\pi_S - \pi_R\right), \tag{25}$$

$$\dot{\lambda}_I = \left[\alpha + \beta\,(1-d)\,\frac{S}{N})\right]\lambda_S - \beta(1-d)\frac{S}{N}\lambda_E + (\gamma + \delta + \mu)\,\lambda_I + \lambda_N\delta - e^{-\rho t}\left(\pi_I - \pi_R\right), \tag{26}$$

$$\tag{27}$$

$$\dot{\lambda}_N = -\lambda_S\alpha - \lambda_S(1-d)\beta\,(I+\varepsilon E)\,\frac{S}{N^2} + \lambda_E(1-d)\beta\,(I+\varepsilon E)\,\frac{S}{N^2} + \mu\lambda_N - e^{-\rho t}\pi_R. \tag{28}$$

Last, we also need following transversality conditions to be satisfied:

$$\lambda_S\,(T) = \lambda_E\,(T) = \lambda_I\,(T) = \lambda_R\,(T) = 0.$$