



HAL
open science

Balancing economic and epidemiological interventions in the early stages of pathogen emergence

Andy Dobson, Cristiano Ricci, Raouf Boucekkine, Giorgio Fabbri, Ted Loch-Temzelides, Mercedes Pascual

► To cite this version:

Andy Dobson, Cristiano Ricci, Raouf Boucekkine, Giorgio Fabbri, Ted Loch-Temzelides, et al.. Balancing economic and epidemiological interventions in the early stages of pathogen emergence. *Science Advances*, 2023, 9 (21), pp.eade616. 10.1126/sciadv.ade6169. hal-04150117

HAL Id: hal-04150117

<https://hal.science/hal-04150117>

Submitted on 4 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

BALANCING ECONOMIC AND EPIDEMIOLOGICAL INTERVENTIONS IN THE EARLY STAGES OF PATHOGEN EMERGENCE

ANDY DOBSON^{1,*}, CRISTIANO RICCI², RAOUF BOUCEKKINE³, FAUSTO GOZZI⁴,
GIORGIO FABBRI⁵, TED LOCH-TEMZELIDES⁶, AND MERCEDES PASCUAL⁷

ABSTRACT. The global pandemic of Covid-19 has underlined the need for more coordinated responses to emergent pathogens. These responses need to balance epidemic control in ways that concomitantly minimize hospitalizations and economic damages. We develop a hybrid economic-epidemiological modelling framework that allows us to examine the interaction between economic and health impacts over the first period of pathogen emergence when lockdown, testing, and isolation are the only means of containing the epidemic. This operational mathematical setting allows us to determine the optimal policy interventions under a variety of scenarios that might prevail in the first period of a large scale epidemic outbreak. Combining testing with isolation emerges as a more effective policy than lockdowns, significantly reducing deaths and the number of infected hosts, at lower economic cost. If a lockdown is put in place early in the course of the epidemic, it always dominates the “laissez faire” policy of doing nothing.

Keywords: Emerging pathogen, Covid-19, economics, epidemiology, lockdown, testing, optimization.

¹EEB, Princeton University, Princeton, NJ 08544, USA, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA and Smithsonian Tropical Research Institute, Panama City, Panama. dobson@princeton.edu

²Department of Economics and Management, Università di Pisa, Italy

³Rennes School of Business, 35000 Rennes, France

⁴Department of Economics and Finance, Luiss University, Roma, Italy

⁵Univ. Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL, 38000 Grenoble, France

⁶Department of Economics and Baker Institute for Public Policy, Rice University, 6100 Main St, Houston, TX 77005, USA

⁷Department of Ecology and Evolution, University of Chicago, Chicago IL 6-637, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA

*Corresponding author: dobson@princeton.edu

1. INTRODUCTION

Covid-19 has now infected nearly half a billion people globally and has led to over 6 million fatalities, although the true total may be closer to 20 million (see Adam 2022). Part of the economic damages associated with the pandemic stem directly from people’s inability to work when sick. A significant proportion is also driven by “lockdown” policies put in place to minimize disease transmission through contact between infected hosts and potential susceptible hosts. A template for initial response to future pandemics is needed that will avoid the mistakes and ambiguities that occurred in the initial response to Covid-19. This is particularly important as at least 20 pathogens have emerged as threats to human populations over the last fifty years (see Bernstein et al. 2022). In the one to two year period before vaccines become widely available to control a novel pathogen, the only public health measures to control outbreaks are isolation of susceptible hosts and testing for infection once tests have been developed and distributed. Both of these measures can generate non-trivial financial costs associated with a reduction in employment and income. Policy makers have the task of determining the best way to offset these costs against those caused by sickness and possibly death of infected hosts.

There are increasing calls for better foresight of how to respond to future pandemics (Jones 2022; Sridhar 2022). In a recent article, Persad and Pandya (2022) argue that combining epidemiological studies with economic cost-benefit analysis is essential for an effective policy response. Policymakers must be equipped with tools to rigorously compare various interventions, both when evaluating individual policies and when determining which policies to include in a regulatory package. They argue that Covid-related policy decisions require considering not only trade-offs between health outcomes and the direct costs of interventions, but additional dimensions related to economic activity, social justice, and individual liberty. Decision making always uses some type of mental model to weigh the pros and cons of different policy options. Rigorous economic evaluation formalizes this process. Value judgments will still be present, but economic evaluation can make the decision-making process more systematic, comprehensive, and transparent. A particularly relevant quote in Persad and Pandya (2022) states: “Critics of this type of approach might argue that in the midst of a pandemic that is still killing thousands of people globally every day, we don’t have time to engage in economic evaluation — that we should do the best we can, without fully weighing the costs and benefits of the options under consideration. In contrast, we believe the severity of the pandemic makes the need for evaluation all the more urgent. Choosing optimal interventions is associated with a bigger payoff when risks are higher.” Economic evaluations of Covid-related policies must consider non-linear effects, as policies might have different results in combination than they do independently.

We have developed a model that combines existing, widely-used economic models and an expansion of the SEIR (Susceptible-Exposed-Infected-Recovered) epidemiological framework. The resultant hybrid epi-econ model introduces novel features and allows us to develop optimization methods to concomitantly minimize deaths

and damages to the economy. Our model is innovative in two ways: first, the expanded SEIR model accounts for a larger set of possible states for the hosts in an attempt to consider the dynamics of isolating and testing contacts of infected hosts, that may themselves develop infections. Second, we take into account that realistic epidemic control policies are subject to inefficiencies resulting from “economic frictions” inherent in the implementation of such policies. A partial list of these frictions includes incomplete information, transactions costs associated with initiating multiple rounds of lockdowns in rapid succession, incomplete enforcement, and costs associated with transitions in and out of lockdown. By improving the realism of both epidemic dynamics and policy modelling we can better understand the structure of the mechanisms through which public health and economic factors interact.

Prior to the current health crisis, economic research on epidemics has been limited and has mostly attempted to connect epidemics with economic development in the global South (see for example, Cuddington and Hancock 1994 or Young 2005). One of the first significant contributions connecting economics and epidemic compartmental modeling (hereafter, epi-econ modeling) traces back to Gersovitz and Hammer in 2004 (Gersovitz and Hammer 2004). The current Covid-19 crisis has stimulated additional research at this intersection. Most studies in the area use standard compartmental epidemic models ranging from SIS to SEIRAD (“A” stands for asymptomatic), and typically investigate appropriate controls to target specific aspects of the Covid-19 crisis. Examples include Acemoglu et al. (2021), Alvarez et al. (2021), Gollier (2020), Fabbri et al. (2021), and Hritonenko et al. (2021). The policies examined in these papers tend to focus on lockdowns as a control strategy, in line with Acemoglu et al. (2021). Testing and/or social distancing are less frequently addressed (see Alvarez et al. (2021) or Piguillem and Shi (2022)). These interventions are often modeled in a stylized way (for example, lockdown controls are assumed to be continuous-time control variables; Aspri et al. 2021 provide a notable exception). They suggest that targeted isolation avoids the sharp economic decline of lockdown by creating incentives for infectious individuals to isolate, while allowing unexposed individuals to continue to consume and work. Ash et al. (2021) calibrate a dynamic economic model to COVID-19 related epidemiological data and use it to evaluate the effects of different scenarios, including voluntary isolation, targeted isolation, and blanket lockdowns.

Unlike the majority of the recent epi-econ literature, we use a finite-horizon model in order to concentrate on short-term outcomes. More precisely, we build a framework that provides insights on how policy makers should respond in the first two years of a novel emergent pathogen for which (A) there is very limited epidemiological information, (B) there are no available specific drugs or vaccines, and (C) tests for infectivity are in the early stages of development. Given these constraints, a lockdown is one of the main tools available to policy-makers. We specifically depart here from the common assumption that lockdown policies can be adjusted in continuous time and consider that they take place in a finite number of phases, the lockdown parameters (intensity of the lockdown and duration) being optimally chosen at each phase. In

addition, we incorporate technological implementation delays (e.g. in efficient testing) and capacity constraints (e.g. in test/mask production). This in turn allows to compare optimal control policies for countries at different levels of development, or governments with different levels of concern for the welfare of their citizens/workforce.

2. RESULTS

Our model has two main components: an epidemiological sub-model and an economic one. They are coupled together by the population of hosts who can either work, or are restricted from working to full capacity by exposure to infected hosts or by being unable to work when ill and incapacitated. In both cases we will assume an underlying well-mixed “mean-field” structure, a framework that has been widely used in both epidemiology and economics. The flow diagram and state variables of the epidemiological model are described in Fig. 1. The corresponding equations are given in the Methods.

The economic analysis is based on a standard intertemporal production model (Methods). We postulate that a policy-maker solves an “optimal control problem.” They maximize a function representing the society’s well-being, or “welfare,” over time, subject to the set of differential equations describing the evolution of the epidemiological variables. The welfare function is a linear combination of economic and epidemiological objectives.

The policy-maker’s objective is to optimally balance between two components. The first involves maximizing economic well-being associated with the flow consumption of goods and services produced in the economy. The second is to minimize the society’s direct costs associated with deaths resulting from the pathogen. In addition to reducing output as a result of declining labor supply, deaths impose several other costs to society. Our modeling of these costs is meant to capture the intrinsic value of lives lost, as well as the resulting social and psychological effects on families and affected communities more broadly. As is standard in economic models, we assume that utility from consumption increases at a decreasing rate as consumption increases. The power function specification is also standard, it features constant elasticity of substitution in consumption over time, which captures the willingness of the policy-maker to switch consumption/production over time. In addition, we assume that disutility from deaths increases at an increasing rate with the number of deaths. Finally, we use the parameter θ to weight the importance of deaths relative to consumption of goods and services in the evaluation of the policy maker. At one extreme, $\theta = 0$ corresponds to a purely economic model, where costs from lives lost are not directly taken into consideration (they reduce economic well-being indirectly, but only as a result of lost production). As θ increases, the epidemiological objective (lives saved) increases in importance. Different values of θ can thus trace a Pareto frontier between economic performance and lives lost.

We use the model to examine the concomitant response of the economy and the pathogen to three different non-pharmaceutical interventions (NPI’s) that are likely to be available before a vaccine or anti-viral drugs can be developed. Until that

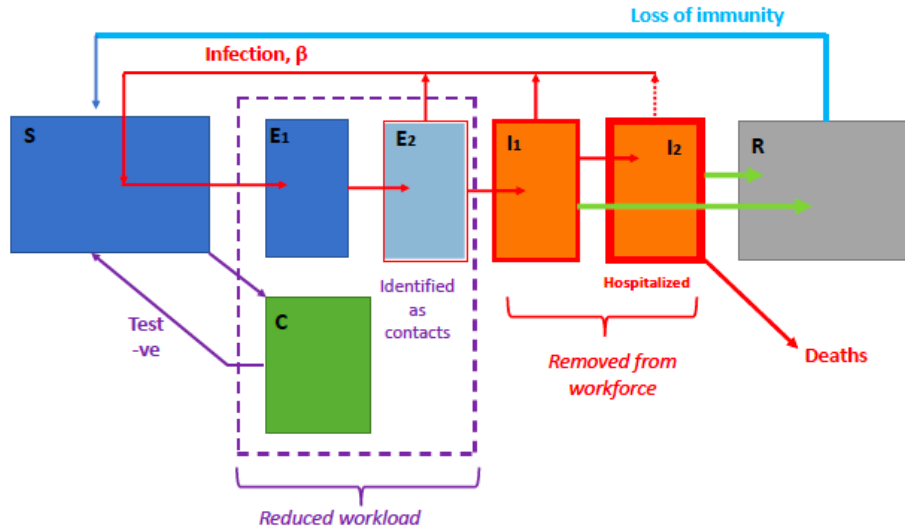


FIGURE 1. We have modified the basic SEIR formulation by dividing the exposed (E) and infectious (I) classes into two sequential classes, E_1 and E_2 and I_1 and I_2 . Exposed hosts, who are not yet infectious are classified as E_1 , while asymptomatic, contagious hosts are classified as E_2 . We assume that E_1 individuals transform to E_2 at an exponential rate determined by φ_1 . The pre-symptomatic hosts, E_2 , transform to symptomatic infected hosts, I_1 , at a rate φ_2 . Both E_2 and I_1 are infectious. This rate largely determines the duration of time during which exposed hosts are able to transmit infection before they show symptoms of infection. If φ_2 is large (~ 365) (around one day), then exposed hosts quickly exhibit signs of symptoms and can be identified as infectious (as occurred with SARS). In contrast, if φ_2 is slower ($\sim 365/7$) (a week), then asymptomatic hosts may transmit the disease for up to a week before showing symptoms, as in the case of Covid-19 (or many years in the case of HIV or TB, when φ_2 may range from 0.1 to 0.5). In a similar way, infected hosts, I_1 , may become sick and get hospitalized, I_2 . These hosts have a higher mortality rate, but are assumed to be in relative isolation and are thus unable to transmit the pathogen, except to unprotected health care workers. The majority of the pathogen-induced mortality occurs in the I_2 class. We also include an additional class, C , into our model structure, these are contacts of infectious hosts who do not develop infection. Contact tracing identifies $C + E_1 + E_2$ as contacts of infected hosts, testing is used to differentiate uninfected contacts, C , from exposed hosts (E_1 and E_2); the former can return to work, the latter remain in isolation and go on to develop infection.

time we assume that the only interventions available are (A) a lockdown of a fraction of the economy, (B) isolation of contacts of people known to be infected, and (C) wearing of surgical masks. In what follows we initially ignore mask-wearing and assume this activity is a voluntary adjunct to A and B. We discuss this further in the conclusions (and Supplementary Material). We assume that tests will be developed quite quickly that allow identification of infected people before symptoms appear. This is particularly important for Covid-19, when significant levels of transmission are undertaken by hosts who do not yet show symptoms. The efficacy of these tests will also improve as they are more widely used. We can modify our model to include improving levels of specificity and sensitivity in the accuracy of tests used to identify recently infected hosts. Here we simplify by assuming average recorded values for specificity and sensitivity for the tests used for Covid-19 (see Böger et al., 2021 and Bastos et al., 2020).

Transient dynamics under different epidemic control policies. In this section, we study the dynamics generated by our model under four possible scenarios: the first benchmark we consider is to do nothing to control the epidemic (Case 0 below). This “laissez-faire” policy is always available to policymakers, particularly if the costs of intervention are high. This essentially represents the policies pursued against Covid-19 in Sweden and Tanzania. The second case (Case 1) is one in which the policy-maker responds to the presence of the pathogen only with a lock-down policy (sensu China and North Korea). We characterize the optimal duration and intensity of the lock-down in this context, and discuss its impact on economic dynamics. Symmetrically, the following sections (Case 2 and 3) are devoted to the case where the policy-maker copes with the epidemic only by testing individuals; this can be done at random, or by focusing on those who have been in contact with hosts who have developed symptoms that progress to illness. Finally, we investigate the case where, after some initial delay, the policy-maker may use both tools at their disposal (Case 4). In all scenarios we characterize the optimal sequential combination of lock-down followed by testing.

We first consider the case where no formal attempt is made to initiate a lockdown of the economy, nor to restrict interactions between infected and susceptible hosts, or to isolate those who have contacted the pathogen and those who have developed symptoms. For this scenario we also assume that no tests are available. Such a laissez-faire scenario could, for example, be favored by those entities whose economic interests would be most severely impacted by reductions in economic activity under lockdowns. We use this case as a benchmark for our initial examination of the model’s dynamics, as it provides a comparison with the different forms of intervention we introduce later. This case also allows us to identify the conditions under which minimal policy intervention might indeed be desirable from the perspective of maximizing social welfare.

One view that has been put forward in support of the laissez-faire case is that the pathogen will generate levels of “herd-immunity” among survivors of infection that will slow the further spread of infections while allowing uninfected hosts to continue

to operate in the economy. These arguments assume relatively low rates of mortality and prolonged periods of immunity (Fine et al. (2011); Fine (1993); Randolph and Barreiro (2020); Fontanet and Cauchemez (2020)). Crucially, loss of immunity over time always leads to a resurgence of the epidemic. We explore this possibility by allowing immunity to wane at four different average rates: 6 months, and 1, 2, and 3 years. As we are only focusing on the first 3 years of the epidemic, the slowest rate of loss of immunity is essentially equivalent to the classical studies of measles which assume life-long immunity and form the basis of most SEIR modelling approaches.

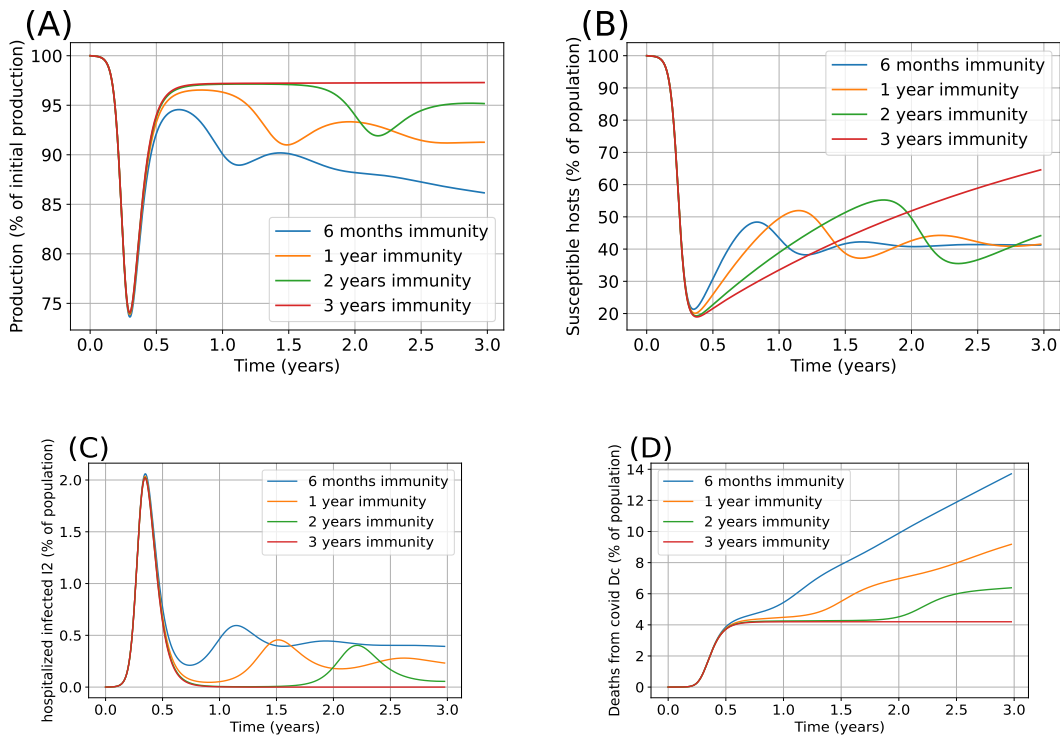


FIGURE 2. The *laissez-faire* case: impact of duration of immunity on epidemic dynamics and the economy in the absence of intervention. The upper two figures illustrate (A) the impact of the pathogen on the economy for different levels of duration of immunity, and (B) the reduction in the numbers of susceptible hosts in the population. The lines are coloured to reflect different rates of loss of immunity (blue 6 months; orange 1 year; green 2 year; red 3 years). The lower figure (C) illustrates the numbers of hospitalized patients. Values of the economic and epidemiological parameters are described in Table 1.

Case 0: The *laissez-faire*. In addition to providing a benchmark, this scenario also quantifies the economic damages that result from reduction of the workforce due to illness and through loss of life (Figure 3). In all cases, we emphasize that the pathogen does not die off, but continues to generate new deaths and a continuous reduction in economic activity. It is worth pointing out that our *laissez-faire* scenario is restrictive

in the sense that it abstracts from people’s endogenous decisions to change their behavior, for example, by isolating or wearing face-masks. There is increasing evidence from epidemiological models that such changes in behavior do have a significant impact on reducing rates of pathogen transmission, particularly when these are driven by reduced levels of aggregation in the host population Hollingsworth et al. 2011; Eichenbaum et al. 2021; Fenichel et al. 2011; Yan et al. 2021. Indeed, the large differences observed between different countries in the early months of the Covid-19 pandemic reflect differences in both government mandated and individual choice behaviors Anderson et al. (2020); Eichenbaum et al. (2021).

The rate at which immunity is lost has profound effects on both the dynamics of the epidemic and it’s influence on the economy. The upper two graphs illustrate the impact on the economy and on the number of susceptible hosts as sequential waves of infection pass through the population (Figure 3). If immunity lasts for 3 years, the epidemic is experienced as a short outbreak that peaks when ‘herd immunity’ is attained. As the duration of immunity shortens, it becomes progressively easier for the pathogen to reassert itself, causing repeated waves of infection resulting in corresponding sequential impacts on the economy. Deaths initially peak at around 4% of the population. They continue to rise after the short duration of immunity provided by transient levels of protection begins to diminish and those who have recovered from infection become susceptible and potentially reinfected. The deaths are matched by a prolonged 4% reduction in economic output, as births have not yet had time to enter the workforce. Notice that loss of economic productivity continues in repeated waves, as each cohort loses their immunity, become sick, and transiently leave the workplace. When immunity lasts for less than a year, economic productivity experiences a steady long-term decline.

We have assumed that those who receive 2nd or 3rd bouts of infection have no residual immunity to infection after their immunity has waned. This is pessimistic in that it is likely that recovered hosts do receive some residual immunological benefits from prior infection, but incorporating these would require additional stages to our model that acknowledge a second and third class of resistant host.

Case 1: Lockdowns. Next, we consider the case where the policy-maker responds to the presence of the pathogen by initiating a lockdown that closes a proportion of the economy for a sequence of time intervals, eventually leading to the eradication of the pathogen from the local population. Our optimization algorithm minimizes economic losses by setting the level and duration of the lockdown, in principle for up to ‘n’ distinct time intervals. In each case we characterize the combined optimal duration and depth required to maximize the total social welfare functional.

Policy-makers have to make crucial decisions at the beginning of an outbreak. These decisions trade off the political expediency of being seen to act promptly, against the cost of slowing down economic activity, which might expose them to claims of over-reaction, particularly if reports of an epidemic run the risk of being false alarms.

We assume that two weeks after the epidemic emergence is the fastest time when lockdowns can be put in place. This allows for identification of sufficient initial cases before it is concluded that something needs to be done, and a lockdown is the only possible response as no treatments or vaccines are available in this scenario. We examine the economic and human costs resulting from more extended delays (Figure 3). We do this by starting the lockdown process at sequentially later dates: 2 weeks, 4 weeks, 8 weeks, 16 weeks after initial detection of transmission.

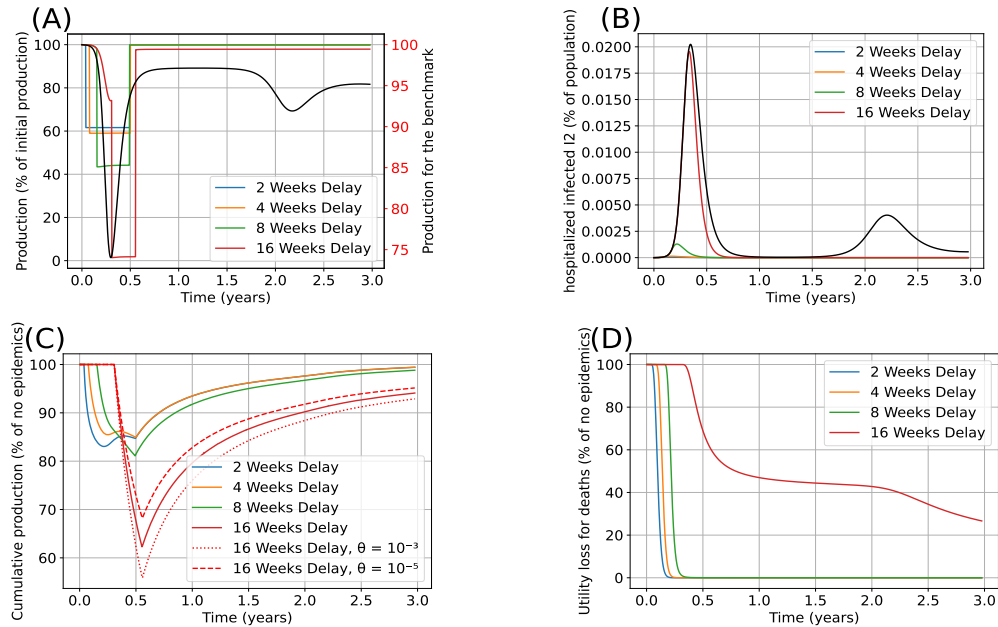


FIGURE 3. Lock-down without testing. Depth and duration of optimal lockdown and the related impact on the economy and on the epidemic of varying the delay in the policy response: 2 weeks (blue), 4 weeks (orange), 8 weeks (green) and 16 weeks (red). In black we illustrate the benchmark “laissez faire” case (calibrated on the right hand axis). The table shows (A) the optimal duration and intensity of the lock-down, (B) the related flow of hospitalizations, (C) the cumulative production (from time 0 to each period) in terms of the cumulative production in an epidemic-free dynamics and (D) the utility loss (from time 0 to each period) due to epidemic-related deaths weighted by the parameter θ . The dashed and dotted lines around the 16-week delay illustrate the sensitivity of lockdown to θ , the index that parameterizes the value of human life. The values of the economic and epidemiological parameters are described in Table 1.

The first main message emerging from this analysis is quite intuitive: the longer the delay in initiating lockdown, the deeper its impact on the economy. However, the lockdown does not necessarily have to last longer, particularly when the delay extends towards the natural underlying first peak of the epidemic. In practice it is not easy to know *a priori* when an epidemic wave of infections will peak. A second important finding is that lockdown levels do not have to match the levels of herd immunity.

Eradication can be achieved by a 33 to 36% reduction in economic activity for a period of up to 10 months. The relative insensitivity of the duration of the lockdown is not matched by differences in initiating a lockdown: 8 times as many people are hospitalized and 5 times as many people die if the lockdown is delayed from two to four weeks. Delay always leads to more deaths up until the time when a lockdown corresponds to the time when the epidemic has peaked. With delays of this duration, transient levels of herd immunity in the recovered section of the population reduce the number of future deaths, but not the economic cost of achieving eradication. Perhaps a more relevant way to measure the relative efficiency of a lockdown policy with respect to “laissez faire” is to compare cumulative costs over the whole (3 years long) period. Figures 3(C and D), point to some important related factors. First, while the lockdown policy is clearly outperformed in terms of production by the laissez faire after one year (from slightly more than 5% below in the 2 weeks delay scenario to 20% in the 16 weeks delay case), this gap tends to vanish at the end of the 3-year long period (with the notable exception of the 16 weeks delay scenario). Provided that the lockdown starts early enough, the economic rebound offsets the initial drop in production relative to “laissez faire” after 3 years. This is largely due to the lives saved and the infections prevented as a result of the lockdown policy. Second, our model asserts that a lockdown policy is more effective in reducing cumulative deaths than the “laissez faire.” Even in the worse lockdown case, the 16 weeks delay scenario, the cumulative death rate over the whole period is half the one generated by the “laissez faire.” It is worth pointing out that the balance between the economic cost and the welfare losses due to aversion to deaths in our model depends on the value of the parameter θ . As reflected in Figure 3(C), the cumulative economic losses (relative to “laissez faire”) also depend on this parameter: as θ rises, more human lives are saved under a longer optimal lockdown, which in turn increases the cumulative production losses due to the lockdown.

Our results are sensitive to the two key parameters that determine the pathogen’s transmission efficiency: (1) the average rate of transmission per contact, β , and (2) the duration of time for which an asymptomatic host is infectious before symptoms appear and the host is isolated. Figure 4 A-D illustrates the effect of variation in transmission. When transmission is relatively low, the pathogen can be eliminated by a short, deep lockdown. As transmission efficiency increases, the depth of lockdown increases, but the duration is relatively constant. Contrasting results emerge if the duration of infectiousness changes. Initially, the pathogen can be controlled by deepening the lockdown level. However, when duration of infectivity is significantly prolonged, then the lockdown has to last longer in order to eliminate infections. This implies that the optimal depth of a lockdown is partly driven by the pathogen transmission rate and the lockdown’s duration is dependent on the duration of infectivity.

Case 2: Random Testing and Isolation. The development of tests to identify people who are infectious but not yet showing symptoms is potentially a powerful tool to contain an emerging epidemic (see Mina et al., 2020, Mina and Andersen, 2021, Larremore et al., 2021). The speed with which these tests can be developed is

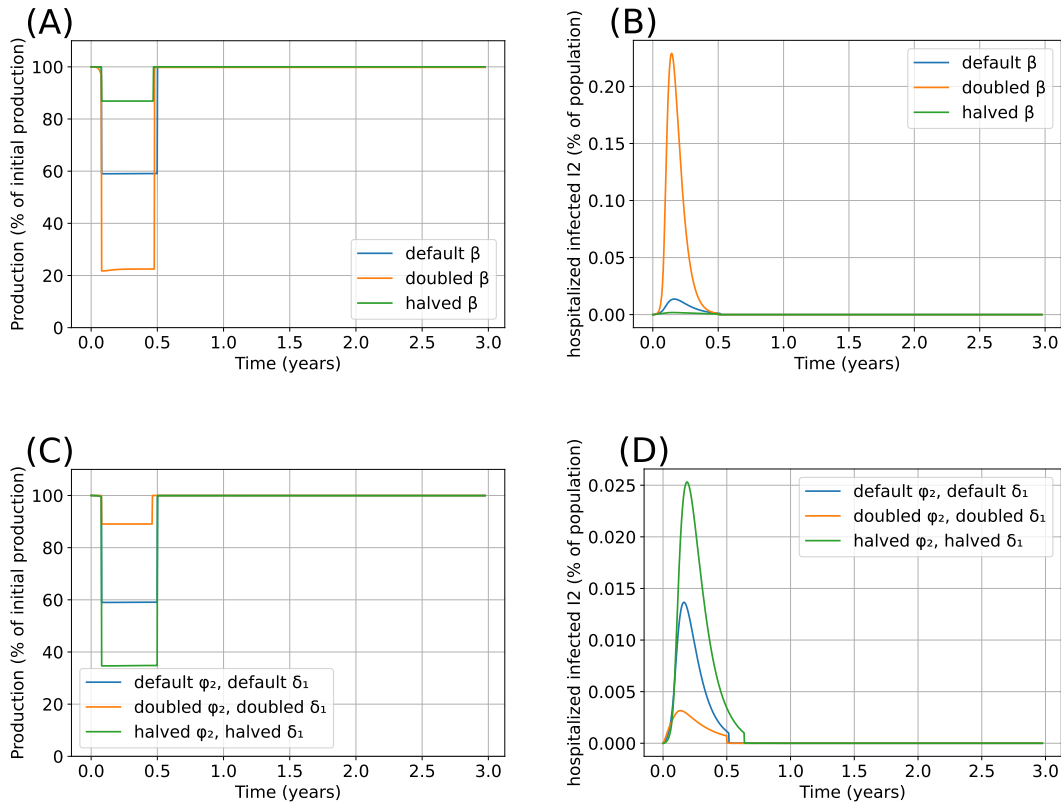


FIGURE 4. Lock-down without testing: effects of variations in infectious rate and in the transition rates. Impact of varying the transmission rate and duration of infectivity rates (from E_2 to I_1 and from I_1 to R) on the size of the epidemic and the optimal lockdown duration. The upper two figures illustrate (A) the changes in production (relative to initial pre-epidemic production) and (B) the number of hospitalized individuals for different values of the transmission rate β . The lower figures shows the same variables for different durations of infectivity ϕ_2 (transition from E_2 to I_1) and δ_1 (transition from I_1 to R). Values of the rest of the economic and epidemiological parameters are described in Table 1.

dependent upon obtaining viral material from infected hosts. It may then take time for tests to be manufactured and be made widely available. There are a variety of methods that have been used to identify contacts of infected hosts. These methods vary in accuracy: some may simply be based on self-recognition of contacts; others may be based on cell-phone based associations that identify when a potential contact has been within the vicinity of a cell-phone bearing exposed host; see Ferretti et al. (2020). We attempt to capture this range of efficiencies using a parameter, c . When $c = 0$ there is “perfect knowledge” and only hosts known to be exposed and infected, E_1 and E_2 , are identified. As c increases ($c > 0$), a larger pool of potential contacts are identified, a smaller proportion of whom are actually infected. This creates a pool

of individuals, $(C+E1+E2)$, who are placed in partial isolation from both susceptible and infected hosts. Concomitantly, this reduces their contributions to the economy by a factor ϵ_C . It is important to notice that we lump C, E1 and E2 together for the purposes of our economic calculations, as they have all been exposed to infection, but we cannot differentiate between their infectious status without testing. Once they have been tested, they can either progress to the E2 classes of infection, if positive, or be returned to the susceptible class, S, if they have a negative test. Larger values of c reflect a larger degree of caution which is synonymous with a larger impact on the economy.

We initially assume two types of testing: (1) random testing of a proportion of uninfected hosts, and (2) contact testing of people who have been placed in the C, E1, and E2 classes. We note that the rate at which tests identify infected hosts is a function of the type of tests used. In the case of Covid, PCR tests would identify all four categories of hosts as infected (E1,E2, I1, I2), although it may take several days for the laboratory to return the results of the tests. In contrast, contacts can run laminar flow tests for themselves and have results in 15 minutes. These tests will likely only identify E2, I1, and some I2 hosts as positive, but these are the hosts that are most likely to be transmitting the virus (particularly E2). Plainly, tests that can be self-administered and that produce rapid results will minimize absences from work and whence the impact of the pathogen on the economy.

Case 3 : Contact identification, testing and isolation. It is natural to assume that tests are likely to be imperfect, at least initially. They will give rise to both false positives and false negatives. We have included two parameters for sensitivity and specificity that reflect measured value of false positive and negatives; see Larremore et al., 2021. Contacts who have tested negative are returned to the susceptible class, S, where they continue to work and mix at the same rate as other susceptible hosts, S.

Our results consistently suggest that testing and isolation of infected hosts are considerably more effective than lockdowns in controlling the epidemic outbreaks (Figure 5 part 1). In the case of purely random testing, we again see that the longer the policy is delayed, the larger numbers of people are identified as asymptotically infected and isolated from work until tested for infection status. Isolation always reduces the number of people hospitalized and dying. However, these increase at a more rapid rate than they do when lockdown is delayed. Crucially, the economic costs of testing and isolation are always lower than when using a lockdown as a control measure, particularly when testing is started early. In addition, there are significantly fewer deaths when testing is initiated early.

If contact tracing is perfect and all contacts of infected hosts are identified and isolated then testing is unsurprisingly highly efficient at both containing the epidemic and in minimizing economic costs (Figure 5 part 2). However, as mentioned earlier, it is unlikely that contact tracing is perfect, so we allow our model to test different numbers of potential contacts when an ever-widening net of contacts is offset by a reduction in efficiency in identifying infected contacts (we essentially assume this is

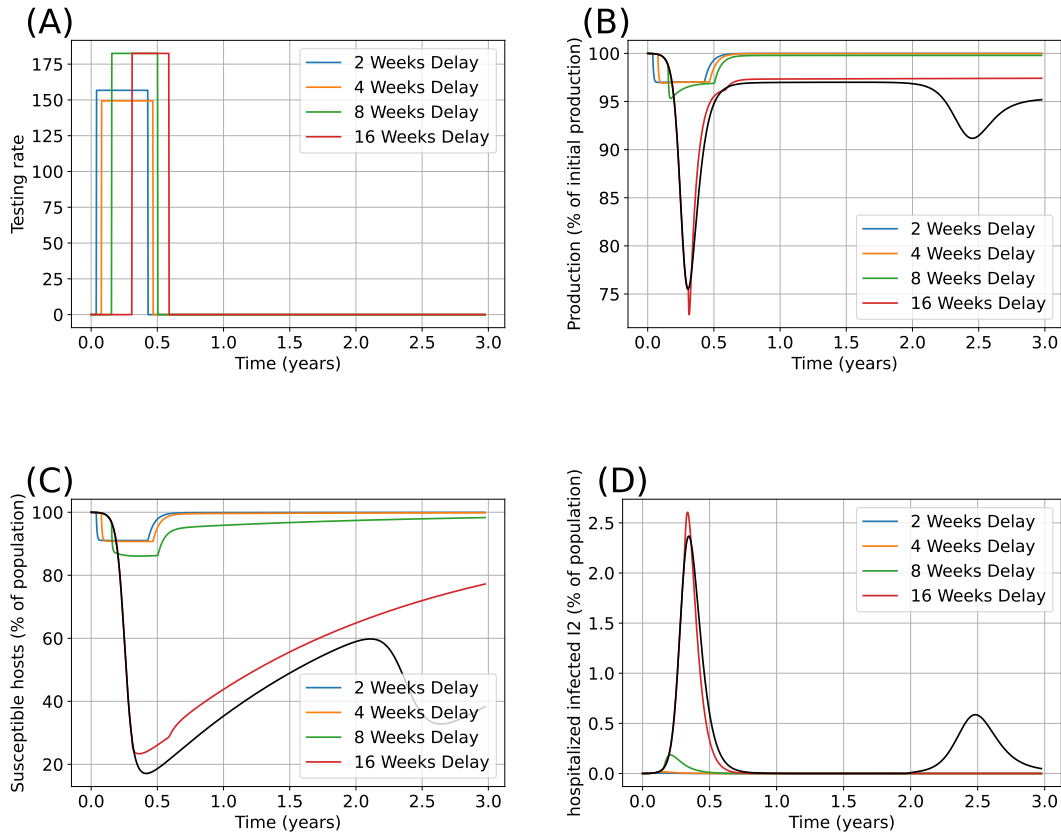


FIGURE 5 PART 1. Random testing without lock-down. Strength and duration of optimal random testing and its impact on the epidemic. The figures illustrate (A) the intensity and duration of optimal random testing, (B) resultant production compared to initial pre-epidemic production, (C) the relative size of the susceptible population and (D) the numbers of hospitalized individuals. In B, C, and D the black line illustrates the “laissez-faire” case of zero testing. In each figure we vary the initiation of testing: 2 weeks (blue), 4 weeks (yellow), 8 weeks (green) and 16 weeks (red). Values of the economic and epidemiological parameters are described in Table 1, $r = 1$, $c = 0$.

linear as the time taken to test and identify all infected costs increases linearly with the number of people tested). Delays in starting contact testing again increase the levels of mortality and economic damages, as more people have to be isolated as contacts when the epidemic has progressed and more hosts are infectious (Figure 5 part 2). Testing reduces the time that healthy contacts, C, are removed from the workforce, but enhances the background force of infection as rapid testing returns them to the pool of susceptible hosts, S . Testing not only reduces the mortality and hospitalizations associated with the epidemic, it also minimizes the epidemic’s overall size. Even delays of up to a month result in only 5 percent reductions in economic

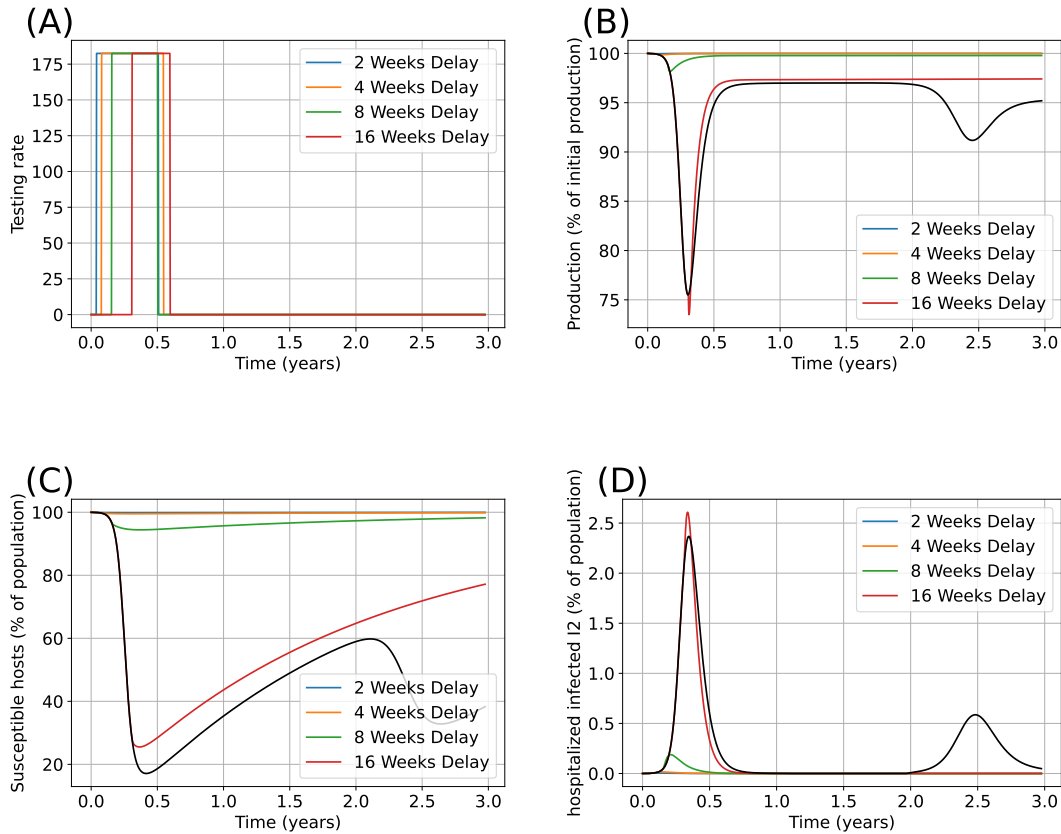


FIGURE 5 PART 2. Targeted (“perfect”) testing without lock-down. Strength and duration of the optimal random testing and its impact on the epidemic in case of target testing. The figures illustrate (A) the intensity and duration of the targeted testing, (B) resultant production with respect to initial pre-epidemic production, (C) the relative size of the susceptible population, and (D) the number of hospitalized individuals. In each case we again vary the initiation of testing: 2 weeks (blue), 4 weeks (yellow), 8 weeks (green) and 16 weeks (red). Values of the economic and epidemiological parameters are described in Table 1, $r = 0$, $c = 0$.

productivity, for periods of up to 2 years. In contrast, starting testing within 2 weeks, generates a 3% reduction in economic growth for a period of around a year.

Case 4: Lockdown followed by testing. We now consider joint policies where lockdown and testing/tracing are jointly optimally determined. We concentrate on the realistic scenario where, while lockdown can be started at any point in time after the epidemic begins, testing at full capacity and maximal efficiency requires additional time. Accordingly we focus on the case where full-capacity testing can only be implemented with some delay. In the results below, we have assumed that a lockdown starts after 4 weeks and testing after 16 weeks. Our optimization algorithm then determines the optimal length and intensity of the lockdown and of testing under

these constraints. An interesting question concerns whether optimal lockdown and testing policies will overlap for certain time intervals or not. We shall see that this depends not only on the maximal testing capacity, but also on the efficiency of testing and tracing.

The first panel presents the results for a developed country with a maximal testing capacity $\tau_{max} = 365/2$, which is our reference value (see Table 1), we then consider the case of a developing country with more limited access to testing. In addition to the benchmark “laissez faire” case (black curves), we have considered four different types of testing/tracing ranging from the most efficient combination of both ($r = 0, c = 0$, blue curves) to less efficient combinations (e.g. $r = 0, c = 2$, green curves). Figures A and B display the optimal lockdown and testing policies for these four types of testing/tracing. In all cases, we find that the lockdown is optimally stopped as soon as testing starts, regardless of the type of testing (random or targeted) and for all the values of tracing efficiency considered. Even when contact-tracing casts a broad net, $c = 2$, which generates higher level of an initial lockdown, it is still stopped as soon as testing becomes available. This strongly illustrates the superiority of testing/tracing over lockdown when testing is efficient and capacity is high. It should be also noted that, although the optimal duration of testing may differ, optimal testing always reaches full capacity under all parameterizations we consider. We find that the policy leading to the highest social welfare outcome is obtained when $r = c = 0$. Interestingly, we also find that (optimal) targeted testing is not always superior to random testing from the total social welfare point of view, for example, if tracing efficiency is sufficiently low, random testing ($r = 1, c = 0$) dominates contact tracing ($r = 0, c = 2$) from a social welfare perspective.

As mentioned earlier, different countries have different amounts of resources to deal with an epidemic and this includes their testing capacity. To mimic the case of developing countries, with a smaller capacity to produce or import effective tests, we consider the case where the testing capacity parameter is reduced by a factor of one-third (which implies setting $\tau_{max} = 365/6$). This single modification leads to significantly altered results. Under this scenario we no longer observe that lockdown is outperformed by testing as soon as the latter is available. Indeed, as testing costs are now more binding in relative terms, a lockdown remains in place for an initial period after testing is introduced. The only exception is when tracing is highly efficient at identifying contacts ($r = c = 0$). When tracing is less efficient, the lockdown is optimally extended to about 6 months, which is more than two months after testing becomes available. This finding serves as an important reminder for policy-makers. Optimal epidemic policies might be different in developed versus developing countries and what works well in one country might not necessarily work well in a different context.

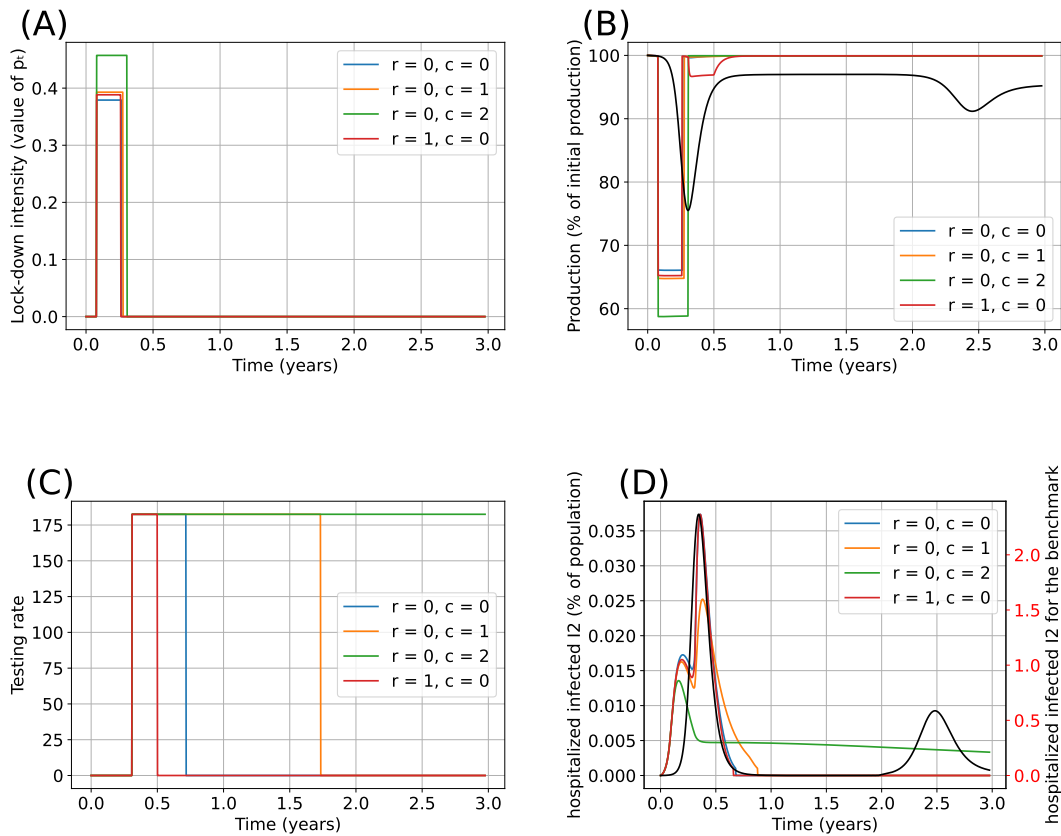


FIGURE 6 PART 1. Optimal combination of lockdown and testing in the high testing capacity, s . Strength and duration of the optimal lockdown, strength and duration of the optimal testing and their impact on the epidemic varying r and c : $r = 0, c = 0$ (blue), $r = 0, c = 1$ (yellow), $r = 0, c = 2$ (green), and $r = 1, c = 0$ (red) when the capacity of the lock-down is high. The figures illustrate (A) the strength and duration of the optimal random lockdown, (B) the production level relative to initial production, (C) the strength and duration of the optimal random testing and (D) the share of hospitalized individuals. The black line again provides a comparison with the laissez-faire case. The value of the economic and epidemiological parameters are described in Table 1.

3. DISCUSSION

In his classic volume on pandemic prevention, Chris Dye quotes “Everything we do before a pandemic will seem alarmist. Everything we do after a pandemic will seem inadequate” (Dye 2021). The primary motivation of this paper was to develop a quantitative framework that provides guidance for policy makers when faced with this dilemma. Different countries pursued different strategies in their attempts to control and minimize the damage caused by the recent Covid-19 pandemic. At the

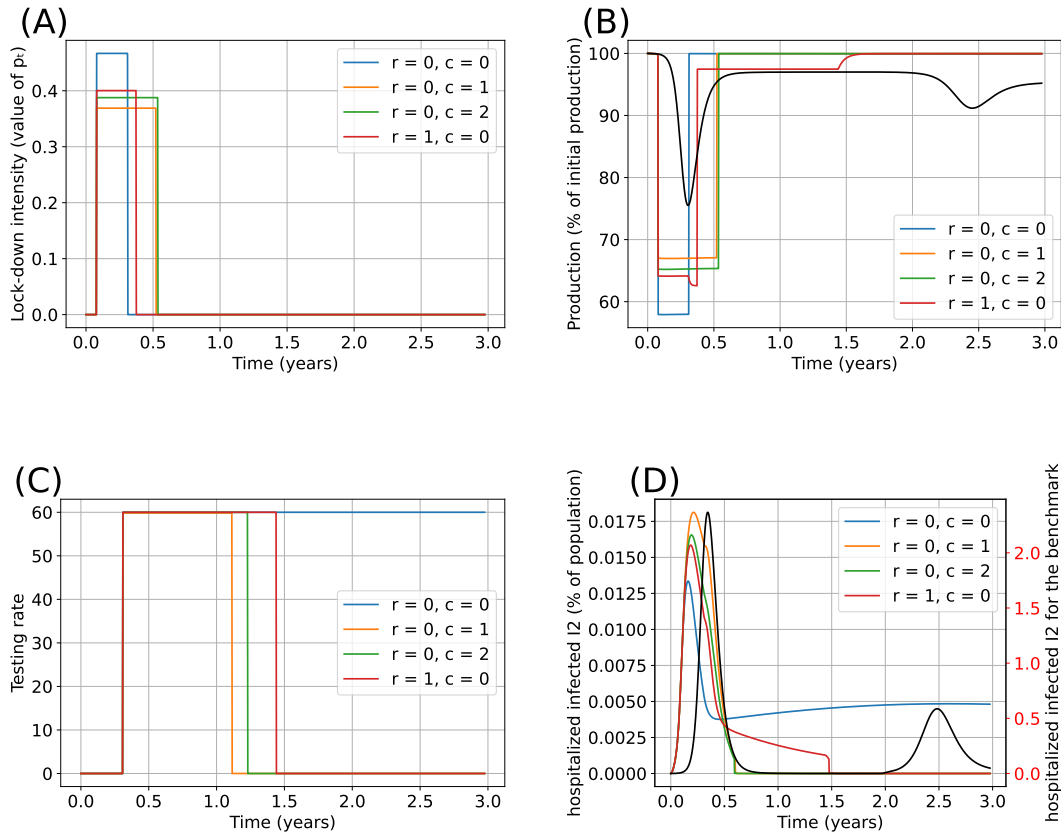


FIGURE 6 PART 2. Optimal combination of lockdown and testing under reduced testing capacity. Strength and duration of the optimal lockdown, strength and duration of the optimal testing and their impact on the epidemic varying the parameters r and c : $r=0, c=0$ (blue), $r=0, c=1$ (yellow), $r=0, c=2$ (green), and $r=1, c=0$ (red), when the capacity of the lockdown is reduced. The figures illustrate (A) the strength and duration of the optimal random lockdown, (B) the production relative to initial production level, (C) the strength and duration of the optimal random testing, and (D) the share of hospitalized individuals. The value of the economic and epidemiological parameters are described in Table 1.

earliest stages, the only control tool available was either a lockdown or the “laissez faire” response of doing nothing. The primary purpose of lockdowns is to “flatten the epidemic curve” and reduce the peak levels of hospitalizations (Koelle et al. (2022)). Our results suggest that, when we optimize taking into account economic as well as public health/mortality objectives, both economic damages and pathogen-induced mortality can be significantly reduced relative to the laissez faire case if a lockdown is put in place as swiftly as possible. This is true even if we take into account the significant adverse economic consequences resulting from a lockdown.

Ash et al. (2021) calibrate a dynamic economic model to COVID-19 related epidemiological data in order to evaluate the effects of different scenarios, including voluntary isolation, targeted isolation, and blanket lockdowns. Their numerical findings assert that voluntary isolation or blanket lockdowns suppress the epidemic nearly as effectively as targeted isolation, but impose higher economic costs. Similar results have been found in earlier models for influenza pandemics (Hollingsworth et al. 2011). Our results also suggest that lockdown policies could be replaced by contact-tracing and testing as soon as viable tests become available. Delays in initiating this transition in policy will always lead to higher economic damages and enhanced mortality. Looking into how to best prepare for the future, this suggests that developing a genetic library of potential pathogens that may cross over to humans in the future would provide an important safety-net for minimizing the potential economic and human costs of epidemics, particularly if they can be used to rapidly develop effective tests. Investment in developing these tests from material gathered in broad-scale surveys of potential novel viruses in wild reservoir hosts is likely to prove economically valuable (Bernstein et al. 2022 and Dobson et al. 2020).

The potential magnitude of voluntary changes in behavior can be partially gauged by considering Sweden, when our “laissez-faire” case would predict around 400,000 deaths in a population of around 10 million people. In contrast, Sweden only experienced around 20,000 deaths, strongly suggesting that people’s personal response to the epidemic may have had a pronounced impact on buffering transmission. In this paper, we have also chosen to study in depth the interaction of an extensive epidemiological model and a comprehensive (frictional) epidemic control policy menu without endogenizing individual responses. In the recent epi-econ literature, several authors have tried to account for these endogenous responses. For example, Eichenbaum et al. (2020), among others, calibrate their model to the US economy and assume that agents endogenously choose to reduce their consumption and labor supply by taking the probability of infection into consideration. However, as documented by Chetty et al. (2020), crucial aspects of individual behavior derive from heterogeneous characteristics, as well as strategic reactions to epidemic control policies and local considerations. Modelling strategic behavior by both heterogeneous individuals in the economy and the policy maker(s) is beyond the scope of our analysis. Rather than incorporating simple ad hoc specifications of individual behavior, we focused our present analysis on the implications of the added epidemiological factors we consider and on a more realistic set of epidemic control policies.

Another important open question for future research concerns the connections between different pathogen characteristics and the optimal duration/depth of the lockdown policy. The desirable lockdown depth is clearly a function of the pathogen’s transmissibility. Our analysis pointed to the conclusion that more transmissible pathogens require deeper lockdowns (Figure 4). We also suspect that the duration of the lockdown would increase with the duration of infectivity. Establishing such connections between pathogen characteristics and policy response is an important topic for future research.

We have not included age-, nor sex-, structure in our model; both of these are potentially important extensions, as mortality from Covid-19 increases significantly with age. Men also seem to suffer higher mortality rates than women (Davies et al. 2020a; Promislow 2020). We are currently developing an extended version of our model that incorporates these aspects, but our primary focus in this paper was to examine epidemiological and economic interactions between a generic ‘emerging pathogen’ and a homogeneous workforce. As the sharpest increase in Covid-19 mortality occurs in age-classes that are usually post-retirement and no longer part of the active workforce, we do not think the addition of these important aspects of population structure will have a major impact on our main conclusions. We do explicitly acknowledge that older retired people will place significant additional pressure on health care services in the early stage of a pandemic and this may lead to increased mortality within the workforce when the capacity of the healthcare system is reduced by large numbers of elderly sick people (Hollingsworth et al. 2011, Anderson et al. 2020).

Our model assumes asymptomatic individuals eventually exhibit symptoms of infection. This is not always the case for Covid-19 where asymptomatic patients may play a significant role in transmission. Modification of the basic structure of the epidemiological model can incorporate this assumption (e.g. Subramanian et al. (2021)). Because our model was parameterized to obtain levels of transmission and hospitalization consistent with those reported for Covid-19, such an extension would effectively shift some of the contribution to the force of infection from pre-symptomatics to asymptomatics. Thus, we do not expect the major results to be modified except for further emphasizing the importance of testing efforts.

Arguably the most economical control strategy during the Covid-19 pandemic has been the wearing of surgical masks at work and in public places (Davies et al. 2020b). We have not incorporated the use of masks in our analysis, partly because we see it as an important “safety play” that is unlikely to stop the pandemic. Masks may provide some protection to those that wear them, while also significantly reducing transmission from asymptomatic infected hosts. An important limitation here is that we do not have good estimates of the efficacy of masks (neither for transmission blocking, nor for protection). The best available methods suggest an average efficiency of around 45 percent (averaged across both transmission blocking and susceptible protection) (Eikenberry et al. 2020). Similarly, the proportion of people wearing masks varies widely and may follow current levels of infection and perceived risk in the population. A simple static analysis can compare the efficacy of face masks with that of a vaccine (Supplementary Material). A high proportion of people have to wear very efficient masks if they are to be effective in reducing R_0 below unity. They will still serve the useful function of slowing the epidemic and reducing the pressure on health care services. Ultimately, we see them as a useful adjunct to the other forms of NPI discussed in more detail below. If more people wear masks, then lockdown can be shorter and impact a smaller proportion of the economy.

We have also not considered “long-Covid” in our model framework, even though there is increasing evidence that this is a significant problem. Current estimates suggest that between 1 and 5 percent of people who acquired Covid in the early stages

of the epidemic continue to feel debilitated and unable to work. These symptoms last between 6 months and the three years for which we have data. They could be included in our model as an additional equation that diverts a proportion of people from the I_2 class into a long-Covid class, rather than the recovered and immune class. Hosts would remain in this class for a number of years and would make a reduced contribution to the economy during this period. All of the additional terms required to include this into our current model structure are linear, whence the additional cost to the economy is essentially the product of the proportion of people who develop long-Covid and their reduced ability to work. All of which makes our calculations optimistic. The spectre of long Covid always increases the economic value of interventions that minimize the number of hosts ever infected (as the number of people with long Covid will always scale with the number of people ever infected).

The development of vaccines for Covid-19 have had a significant impact on the dynamics of the pandemic. Vaccines have increased levels of personal protection and permitted the beginnings of an economic recovery from the significant initial collapse in the spring and summer of 2020. Unfortunately, vaccines have not provided us with a perfect solution, as vaccinated people seem still able to transmit the pathogen, while exhibiting considerably reduced levels of morbidity. Immunity also seems disconcertingly short lived (6 to 8 months). Everything we have described above assumes we do not yet have a vaccine for the hypothetical emergent pathogen. In a second paper we will expand our framework to consider the role of vaccines and their efficacy, and resistance to vaccine adoption within the host population.

Our framework is quite flexible and allows for various additional extensions. A few illustrative examples are worth mentioning. Adding hospitalization costs and a capacity constraint on the health system, requiring that the amount of hospitalized people I_2 remain under a given congestion threshold. This would provide an additional realistic constraint on optimal policies. We could also consider additional policy variables, for example, regarding the modality/intensity of contact tracing. We could further investigate the robustness of our results to the use of different utility/disutility functions. For example, we could use linear U and V as in Alvarez et al. (2021). We believe that the qualitative aspects of our findings related to the relative effectiveness of lock-downs versus testing would remain intact. It is also possible to use our framework in order to study the optimal number of lockdown phases, given certain adjustment costs associated with introducing additional phases. In our simulations, comparing the optimal strategies and the welfare function under different scenarios, we observed that in most cases one lockdown phase is sufficient, even if the policy maker could in principle implement additional lock-downs. Finally, we could pursue a more detailed sensitivity analysis, for example, in connection to the choice of the parameter θ . This parameter establishes the relative weights between the utility from production and the disutility from deaths during the epidemic. An extension of the model could vary θ in order to characterize the Pareto frontier; i.e., the policies which cannot lead to improvements in both economic and public health objectives.

Concerns about future pandemics motivated this work (Bernstein et al. 2022 and Dobson et al. 2020). We have therefore included sufficient flexibility within the model

structure to allow us to adjust parameters in order to consider novel pathogens with different characteristics such as SARS or a new influenza strain, where symptoms do not appear until a host is infectious, or pathogens with long asymptomatic phases (such as HIV). Our framework could also be expanded to consider vector-borne infections such as those caused by the Dengue or Zika virus.

Covid-19 was not the first pandemic and it will not be the last. A huge proportion of the impact of Covid has been driven by different governments responding to the pandemic in an 'ad-hoc' fashion as political pressures to maintain economic activity clash with epidemiological advice. The contrasting responses of different nations reflect the relative magnitude of economic and epidemiological forces and national levels of expertise in these areas. We believe that combining economic modeling, which highlights incentive constraints, with epidemiological modeling, which focuses on public health considerations, will be increasingly relevant in designing policy interventions that are both effective and attainable during future public health crisis. The structure described here provides crucial initial steps in this direction.

4. METHODS

Epidemiological Model. Our epidemiological model is based on the standard SEIR framework (Anderson and May 1992; Keeling and Rohani 2011). We have modified the basic framework by dividing the exposed (E) and infectious (I) classes into two sequential classes, E1 and E2 and I1 and I2 (Figure 2). Exposed hosts, who are not yet infectious are classified as E1, while asymptomatic, contagious hosts are classified as E2. We assume that E1 individuals transform to E2 at an exponential rate determined by φ_1 . The pre-symptomatic hosts, E2, transform to symptomatic infected hosts, I1, at a rate φ_2 . Both E2 and I1 are infectious. This rate largely determines the duration of time during which exposed hosts are able to transmit infection before they show symptoms of infection. If φ_2 is large (~ 365) (around one day), then exposed hosts quickly exhibit signs of symptoms and can be identified as infectious (as occurred with SARS). In contrast, if φ_2 is slower ($\sim 365/7$) (a week), then asymptomatic hosts may transmit the disease for up to a week before showing symptoms, as in the case of Covid-19 (or many years in the case of HIV or TB). In a similar way, infected hosts, I1, may become sick and get hospitalized, I2. These hosts have a higher mortality rate, but are assumed to be in relative isolation and are thus unable to transmit the pathogen, except to unprotected health care workers. The majority of the pathogen-induced mortality occurs in the I2 class.

We also include an additional class, C, into our model structure, these are contacts of infectious hosts who do not develop infection. Contact tracing identifies C+E1+E2 as contacts of infected hosts, testing is used to differentiate uninfected contacts, C, from exposed hosts (E1 and E2); the former can return to work, the latter remain in isolation and go on to develop infection.

The main equations of our SCEEIIIR models are the following

$$\begin{aligned} \dot{S} = & \mu N - \mu S - (1-p)^2 \beta \frac{S(E_2(1-\gamma) + \gamma I_1)(1+c)}{N} \\ & + \phi R + \tau \left(r + \frac{1-r}{1+c} \right) C s_p + \delta C - \tau r S(1-s_p) \end{aligned} \quad (1)$$

$$\begin{aligned} \dot{C} = & (1-p)^2 \beta (cS - (1-\epsilon_C)C) \frac{E_2(1-\gamma) + \gamma I_1}{N} \\ & + \tau r S(1-s_p) - \mu C - \tau \left(r + \frac{1-r}{1+c} \right) C s_p - \delta C \end{aligned} \quad (2)$$

$$\begin{aligned} \dot{E}_1 = & (1-p)^2 \beta (S + (1-\epsilon_C)C) \frac{E_2(1-\gamma) + \gamma I_1}{N} \\ & - (\mu + \varphi_1)E_1 - \tau \left(r + \frac{1-r}{1+c} \right) s_e E_1 \end{aligned} \quad (3)$$

$$\dot{E}_2 = \varphi_1 E_1 - (\mu + \varphi_2)E_2 - \tau \left(r + \frac{1-r}{1+c} \right) s_e E_2 \quad (4)$$

$$\dot{I}_1 = \varphi_2 E_2 + \tau s_e \left(r + \frac{1-r}{1+c} \right) (E_1 + E_2) - (\mu + \mu_{I_1} + (1-\eta)\delta_1 + \eta\delta_1)I_1 \quad (5)$$

$$\dot{I}_2 = \eta\delta_1 I_1 - (\mu + p_M\delta_2 + (1-p_M)\delta_2)I_2 \quad (\text{where } \mu_{I_2} = p_M\delta_2) \quad (6)$$

$$\dot{R} = (1-\eta)\delta_1 I_1 + (1-p_M)\delta_2 I_2 - (\mu + \mu_R + \phi)R + \nu V_w \quad (7)$$

Here N denotes the size of the host population, all of whom contribute to economic output when healthy. Once the pathogen has established, N is the sum of all possible types of host.

$$N = S + C + E_1 + E_2 + I_1 + I_2 + R \quad (8)$$

As the death rates of the epidemics is low, we will approximate N with N_0 (the initial size of the population, see e.g. Anderson and May (1992) and Keeling and Rohani (2011)). In the absence of the pathogen, the total number of deaths at any time is given by $\dot{D} = \mu N$; when the pathogen is present the additional deaths due to the pathogen are given by $\dot{D}_C = \mu_{I_1} I_1 + p_M \delta_2 I_2 + \mu_E E_2 + \mu_R R$, where severe infections I_2 last an average duration $\frac{1}{\delta_2}$ with a fraction p_M of individuals leaving this class resulting in death, whereas a fraction $(1-p_M)$ are able to recover. Similarly, only a fraction $(1-\eta)$ of individuals leaving class I_1 recovers, whereas η proceeds to severe infection and hospitalization.

Lock-down intensity at time t is parameterized by $p(t)$, the probability that an individual, susceptible or infected, is protected from contact and therefore excluded from transmission events. Per-capita transmission rate is given by β , and for each infection, a number c of uninfected contacts is generated. Contacts are restricted in

their ability to work, with only a fraction ϵ able to do so and under isolated conditions, for example from home, which also reduces their exposure to infection.

Testing occurs at rate τ . The indicator variable r takes values of 1 or 0 to control the respective implementation of random testing vs. contact tracing. The specificity and sensitivity of testing are given respectively by s_p and s_e .

Structure of Economic Model. Flow output, Y , is produced through a production function using labor, L , as the only production factor. Abstracting from capital accumulation seems reasonable, since our analysis concentrates on a short time horizon. The short horizon of the epidemic also prevents a large-scale substitutability between capital and labor. Infections affect the economy by reducing labor supply and by making the labor used less productive. The latter could result from disruptions to supply chains or related shortages in necessary intermediate inputs. To model these effects we assume that infections reduce both the labor used in the production process and the total factor productivity parameter, A . We assume that the costs, Φ , associated with the number of people tested, x , are subtracted from total output. The specific form imposed on the function Φ ensures that when x is very small, the costs of testing increase linearly with x . We assume that it is forbiddingly costly to test the entire population. Thus, the maximum rate of testing is constrained.

The proportion of people $p(t)$ who are in lockdown are unable to contribute to the economy. We define the production function as:

$$Y(t) = \underbrace{A(t) [(1 - p(t))L(t)]^\alpha}_{\text{production function}} - \underbrace{\Phi(x(t))}_{\text{testing cost}} \quad (9)$$

where

$$\begin{aligned} A(t) &= A_0(1 - p(t))^\Delta \\ L(t) &= S(t) + \epsilon_C [C(t) + E_1(t) + E_2(t)] + R(t) \\ \Phi(x) &= \rho_0 x + \exp\left(\frac{\rho_1}{N - x}\right) - \exp\left(\frac{\rho_1}{N}\right) \\ x(t) &= \tau [C(t) + E_1(t) + E_2(t) + rS(t)], \end{aligned}$$

In the above expression, $A(t)$ stands for the (capital adjusted) total factor productivity (TFP). This parameter determines the effectiveness of the labor input in producing the consumption good. The flow labor supply consists of individuals who are susceptible or recovered from the pathogen, as well as those who are exposed or potentially exposed (contacts). The variable x indicates the flow of people tested. The form of the testing cost function, $\Phi(x)$, is meant to capture the property that at low levels of testing the costs increase linearly in the number of tests, while at the same time it is prohibitively expensive to test the entire population. More precisely, when ρ_1 and x are small, the marginal cost of testing; i.e., the cost of administering one additional test, is approximately ρ_0 . To calibrate ρ_0 , we convert the dollar value of a test to units of daily US per capita GDP. As a benchmark, we use the per capita US GDP value of 63,416 dollars in 2020, and we assume an average cost of testing

of 174 dollars. Following Piguillem and Shi, 2022, we set $\rho_1 = 0.1$, implying that the cost of testing increases relatively slowly as large numbers of people are tested.

The objective of the policy-maker is to maximize the Total Social Welfare (TSW) function, this is given by:

$$TSW(T) = \int_0^T U(Y(t)) - \theta V(D_c(t)) dt \quad (10)$$

with:

$$U(Y) = \frac{Y^{1-\sigma}}{1-\sigma}$$

$$V(D_c) = \frac{D_c^\omega}{\omega}$$

In the above expressions, $U(Y)$ stand for the satisfaction (utility) from consuming goods and services, while $V(D_c)$ stands for the direct utility loss of lives lost.

Calibration. We calibrate the model using the parameters described in the Supplementary Material, Table 1, these are based on those used in published studies (Ferguson et al. (2020); Flaxman et al. (2020); Subramanian et al. (2021); Lavine et al. (2021)) .

Optimization. The optimization problem we study is nontrivial and cannot be treated with the standard tools used in the current literature. In the supplementary Material, we describe in detail our approach. Here we provide a brief summary. First, our problem falls into the class of deterministic optimal control with exit time. A description of this class of problems is provided, e.g., in (Cannarsa and Sinestrari, 2004, Chapter 8). The main ingredients of this class of problems are:

- the time horizon of the problem;
- the state/control variables and the space where they belong (the state/control space);
- the state equation, which provides the dynamic behavior of the state variables as a function of their initial data and of the choice of the control strategies;
- the set of admissible control strategies;
- the objective function to optimize over all admissible control strategies;
- the target set \mathcal{O} where the epidemic ends.

The main difference with respect to the papers in the existing literature are the following:

- the “exit time feature;” i.e., the fact that the epidemic stops when all associated variables (namely E_1, E_1, I_1, I_2) fall below 1;

- the "discrete control strategies;" i.e., the fact that the control strategies are piecewise constant with a given finite number of switching times.

Both features are crucial in making the model more realistic, the first in connection to the behavior of the epidemic, and the second in order to take account of the constraints faced by policymakers. In this context, the existence of optimal strategies can be demonstrated using standard arguments, however, the uniqueness of the optimal solution is not guaranteed. Concerning the numerical approximations, in order to compute the objective function we rely on classical numerical methods for ODEs. However due to the lack of regularity of the controls, we need to employ a numerical method outside the class of those used for "stiff problems," such as implicit Runge-Kutta methods of high order. Instead, we approximate the continuous-time integration using a Gauss-Kronrod quadrature rule.

The process of numerical optimization is challenging due to the lack of convexity in the objective function. In the absence of convexity, there is no guarantee that a local optimum will also be a global one. Therefore during the numerical optimization, we have to rely on a global optimization algorithm, which is numerically more demanding than a local numerical optimizer. Global optimization suffers from the *curse of dimensionality*, that is the number of function evaluations required for a thorough search in the state space grows exponentially with the dimension of the problem. For the optimization procedure we employed both the DIRECT and DIRECT-L algorithm proposed in Jones et al. (1993) and Gablonsky and Kelley (2001), respectively. Both algorithms are deterministic procedures based on a subdivision of the domain in iteratively smaller rectangles until convergence is reached. To ensure the correctness of the results, we also performed additional tests based on a combination of a brute-force approach on a very fine grid and a local refinement based on a local optimizer. An additional confirmation of the accuracy of the optimization procedure is that the maximum value of the objective function shifts in the expected direction when changing some of the parameters (e.g. when increasing the delay in the control policies the objective function decreases).

Acknowledgements. . We thank the editor and referees for insightful comments that significantly improved the exposition.

Funding sources?

Code availability statement. All the code for numerical optimization and model simulation is available at <https://github.com/crricci/Epi-Econ>

REFERENCES

- Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2021). Optimal targeted lockdowns in a multigroup sir model. *American Economic Review: Insights*, 3(4):487–502.
- Adam, D. (2022). The pandemic’s true death toll: millions more than official counts. *Nature*, 601(7893):312–315.

- Alvarez, F., Argente, D., and Lippi, F. (2021). A simple planning problem for covid-19 lock-down, testing, and tracing. *American Economic Review: Insights*, 3(3):367–82.
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., and Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the covid-19 epidemic? *The lancet*, 395(10228):931–934.
- Anderson, R. M. and May, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.
- Ash, T., Bento, A., Kaffine, D., Rao, A., and Bento, A. (2021). Disease-economy trade-offs under alternative pandemic control strategies. *medRxiv*.
- Aspri, A., Beretta, E., Gandolfi, A., and Wasmer, E. (2021). Mortality containment vs. economics opening: optimal policies in a seiard model. *Journal of mathematical economics*, 93:102490.
- Bardi, M. and Capuzzo-Dolcetta, I. (1997). *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA. With appendices by Maurizio Falcone and Pierpaolo Soravia.
- Bastos, M. L., Tavaziva, G., Abidi, S. K., Campbell, J. R., Haraoui, L.-P., Johnston, J. C., Lan, Z., Law, S., MacLean, E., Trajman, A., et al. (2020). Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *bmj*, 370.
- Bernstein, A. S., Ando, A. W., Loch-Temzelides, T., Vale, M. M., Li, B. V., Li, H., Busch, J., Chapman, C. A., Kinnaird, M., Nowak, K., Castro, M. C., Zambrana-Torrel, C., Ahumada, J. A., Xiao, L., Roehrdanz, P., Kaufman, L., Hannah, L., Daszak, P., Pimm, S. L., and Dobson, A. P. (2022). The costs and benefits of primary prevention of zoonotic pandemics. *Science Advances*, 8(5):eabl4183.
- Böger, B., Fachi, M. M., Vilhena, R. O., Cobre, A. F., Tonin, F. S., and Pontarolo, R. (2021). Systematic review with meta-analysis of the accuracy of diagnostic tests for covid-19. *American journal of infection control*, 49(1):21–29.
- Cannarsa, P., Pignotti, C., and Sinestrari, C. (2000). Semiconcavity for optimal control problems with exit time. *Discrete & Continuous Dynamical Systems*, 6(4):975.
- Cannarsa, P. and Sinestrari, C. (2004). *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, volume 58 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA.
- Cuddington, J. T. and Hancock, J. D. (1994). Assessing the impact of aids on the growth path of the malawian economy. *Journal of Development Economics*, 43(2):363–368.
- Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., and Eggo, R. M. (2020a). Age-dependent effects in the transmission and control of covid-19 epidemics. *Nature medicine*, 26(8):1205–1211.
- Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., O’Reilly, K., Endo, A., Hellewell, J., Nightingale, E. S., et al. (2020b). Effects of non-pharmaceutical interventions on covid-19 cases, deaths, and demand for hospital services in the uk: a modelling study. *The Lancet Public Health*, 5(7):e375–e385.
- Dietz, K. and Heesterbeek, J. (2002). Daniel bernoulli’s epidemiological model revisited. *Mathematical Biosciences*, 180:1–221.
- Dobson, A. P., Pimm, S. L., Hannah, L., Kaufman, L., Ahumada, J. A., Ando, A. W., Bernstein, A., Busch, J., Daszak, P., Engelmann, J., Kinnaird, M. F., Li, B. V., Loch-Temzelides, T., Lovejoy, T., Nowak, K., Roehrdanz, P. R., and Vale, M. M. (2020). Ecology and economics for pandemic prevention. *Science*, 369(6502):379–381.

- Dye, C. (2021). *The Great Health Dilemma: Is Prevention Better Than Cure?* Oxford University Press.
- Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2021). The macroeconomics of epidemics. *The Review of Financial Studies*, 34(11):5149–5187.
- Eikenberry, S. E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., Kostelich, E., and Gumel, A. B. (2020). To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. *Infectious disease modelling*, 5:293–308.
- Fabbri, G., Gozzi, F., and Zanco, G. (2021). Verification results for age-structured models of economic–epidemics dynamics. *Journal of Mathematical Economics*, 93:102455.
- Fenichel, E. P., Castillo-Chavez, C., Ceddia, M. G., Chowell, G., Parra, P. A. G., Hickling, G. J., Holloway, G., Horan, R., Morin, B., Perrings, C., et al. (2011). Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311.
- Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., et al. (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand.
- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., and Fraser, C. (2020). Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936.
- Fine, P., Eames, K., and Heymann, D. L. (2011). “herd immunity”: a rough guide. *Clinical infectious diseases*, 52(7):911–916.
- Fine, P. E. (1993). Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302.
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whitaker, C., Zhu, H., Berah, T., Eaton, J. W., et al. (2020). Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261.
- Fontanet, A. and Cauchemez, S. (2020). Covid-19 herd immunity: where are we? *Nature Reviews Immunology*, 20(10):583–584.
- Gablonsky, J. M. and Kelley, C. T. (2001). A locally-biased form of the direct algorithm. *Journal of Global Optimization*, 21(1):27–37.
- Gersovitz, M. and Hammer, J. S. (2004). The economical control of infectious diseases. *The Economic Journal*, 114(492):1–27.
- Gollier, C. (2020). Cost–benefit analysis of age-specific deconfinement strategies. *Journal of Public Economic Theory*, 22(6):1746–1771.
- Hansen, L. P. and Sargent, T. (2006). *Robustnes*, volume 58 of *Progress in Nonlinear Differential Equations and their Applications*. Princeton University Press.
- Hartl, R. F., Sethi, S. P., and Vickson, R. G. (1995). A survey of the maximum principles for optimal control problems with state constraints. *SIAM review*, 37(2):181–218.
- Hodson, R. et al. (2022). Preparing the world for the next pandemic. *Nature*, 610(7933):33–33.
- Hollingsworth, T. D., Klinkenberg, D., Heesterbeek, H., and Anderson, R. M. (2011). Mitigation strategies for pandemic influenza a: balancing conflicting policy objectives. *PLoS computational biology*, 7(2):e1001076.
- Hritonenko, N., Yatsenko, O., and Yatsenko, Y. (2021). Model with transmission delays for covid-19 control: Theory and empirical assessment. *Journal of Public Economic Theory*.

- Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of optimization Theory and Applications*, 79(1):157–181.
- Jones, S. (2022). How to eradicate the next pandemic disease. *Nature*, 610(7933):S48–S49.
- Keeling, M. J. and Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.
- Koelle, K., Martin, M. A., Antia, R., Lopman, B., and Dean, N. E. (2022). The changing epidemiology of sars-cov-2. *Science*, 375(6585):1116–1121.
- Larremore, D. B., Wilder, B., Lester, E., Shehata, S., Burke, J. M., Hay, J. A., Tambe, M., Mina, M. J., and Parker, R. (2021). Test sensitivity is secondary to frequency and turnaround time for covid-19 screening. *Science advances*, 7(1):eabd5393.
- Lavine, J. S., Bjornstad, O. N., and Antia, R. (2021). Immunological characteristics govern the transition of covid-19 to endemicity. *Science*, 371(6530):741–745.
- Mina, M. J. and Andersen, K. G. (2021). Covid-19 testing: One size does not fit all. *Science*, 371(6525):126–127.
- Mina, M. J., Parker, R., and Larremore, D. B. (2020). Rethinking covid-19 test sensitivity—a strategy for containment. *New England Journal of Medicine*, 383(22):e120.
- Persad, G. and Pandya, A. (2022). A comprehensive covid-19 response — the need for economic evaluation. *New England Journal of Medicine*, 386:2449–2451.
- Piguillem, F. and Shi, L. (2022). Optimal covid-19 quarantine and testing policies. *The Economic Journal*.
- Promislow, D. E. (2020). A geroscience perspective on covid-19 mortality. *The Journals of Gerontology: Series A*, 75(9):e30–e33.
- Randolph, H. E. and Barreiro, L. B. (2020). Herd immunity: understanding covid-19. *Immunity*, 52(5):737–741.
- Sara Biagini, Fausto Gozzi, M. Z. (2022). Robust portfolio choice with sticky wages. *Sia Journal on Financial Mathematics*, 13(4):1–40.
- Sridhar, D. (2022). Five ways to prepare for the next pandemic. *Nature*, 610(7933):S50–S50.
- Subramanian, R., He, Q., and Pascual, M. (2021). Quantifying asymptomatic infection and transmission of covid-19 in new york city using observed cases, serology, and testing capacity. *Proceedings of the National Academy of Sciences*, 118(9):e2019716118.
- Svoboda, E. (2022). Mastering the art of persuasion during a pandemic. *Nature*, 610(7933):S34–S36.
- Yan, Y., Malik, A. A., Bayham, J., Fenichel, E. P., Couzens, C., and Omer, S. B. (2021). Measuring voluntary and policy-induced social distancing behavior during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(16):e2008814118.
- Yong, J. and Zhou, X. Y. (1999). *Stochastic controls: Hamiltonian systems and HJB equations*. Springer, Berlin.
- Young, A. (2005). The gift of the dying: The tragedy of aids and the welfare of future african generations. *The Quarterly Journal of Economics*, 120(2):423–466.

Supplementary Material

MODEL PARAMETERS

TABLE 1. Parameters used in the simulations, these are based on those used in published studies (Ferguson et al. (2020); Flaxman et al. (2020); Subramanian et al. (2021); Lavine et al. (2021)) .

Variable	Value	
SEIR		
μ	1/70	birth and death rate
μ_{I_1}	0.02	death rate of I_1
μ_R	1/70	death rate of R
μ_E	0.012	death rate of E_2
p_M	1/3	probability of death of an individual during severe infection
γ	0,05	proportional contribution to infection of E_2 and I
ϕ	1/2	transition rate from R to S
φ_1	365/5	transition rate from E_1 to E_2
φ_2	365/5	transition rate from E_2 to I
δ	365/18	transition rate from C to S
δ_1	365/18	duration ⁻¹ of I_1
δ_2	365/13	duration ⁻¹ of I_2
β	150.00	per-capita transmission rate
η	0.15	hospitalization fraction during infection period
testing		
c	variable > 0	number of contacts per infection generated (E_2 and I_1)
s_p	0.90	specificity of test
s_e	0.90	sensitivity of test
τ	365/12	default testing rate
τ_{max}	365/2	maximum testing capacity for a developed country
r	0 or 1	contact or random testing respectively
testCost	50,00	cost for one test in \$
yearly_US_GDP	21.5T	yearly US total GDP in 2019 in \$
Y_0	$A_0 N^\alpha$	yearly model total GDP with no disease spread
ρ_0	$Y_0 \frac{testCost}{yearly_US_GDP}$	cost for one test in model unit
ρ_1	0.1	testing cost parameter
N	1e5	total number of individuals
production		
A_0	1e5	(capital adjusted) total factor productivity
α	2/3	curvature of production function
Δ	1/5	short-run elasticity of capital utilization
ϵ_C	0.8	work efficiency of C-class (fraction of contacts that work and do so under isolation)

utility parameters		
σ	1/2	curvature of utility from consumption
ω	3	curvature of disutility from deaths
θ	1e-4	weight of utility from consumption versus disutility from deaths

OPTIMIZATION AND NUMERICAL ANALYSIS

Here we describe the main features of our optimization procedure, from both the theoretical and the numerical approximation viewpoints. Although our optimization problem falls into the class of the so-called Optimal Control Problems (OCP) in continuous time, it is not standard in some aspects. In the next subsections, we describe the problem, its main features, and the methods we use to find solutions and to approximate them. We pay particular attention to the points which are non-trivial and technically involved.

Basic ingredients of the optimal control problem. In an OCP one identifies the key variables which describe the state of the system, or the “state” variables, and the ones which are under the control of the optimizing agent, the “control” variables. The state variables belong to a set Z , the “state space.” This is typically a subset of \mathbb{R}^n , where n is the number of state variables. The control variables belong to a set K , the “control space.” This is typically a subset of \mathbb{R}^m , where m is the number of control variables. Both state and control variables may change with time, t , where $t \in [0, T]$. We refer to 0 as the initial time, while T is the final time, also called the “horizon” of the problem. We term a “control strategy” a function

$$k(\cdot) : [0, T] \rightarrow K$$

and a “state trajectory” a function

$$z(\cdot) : [0, T] \rightarrow Z.$$

Clearly not all control strategies are suitable for a given problem. Hence, later, we will introduce the set of admissible control strategies. Given the above, the main ingredients of an OCP are:

- the state equations which describe the time evolution of the state variables; i.e., the state trajectory $z(\cdot)$, depending on the choice of the control strategies $k(\cdot)$;
- the objective functional to be optimized on the time span $[0, T]$ over all admissible control strategies.

The state/control variables and the state equations. The control variables in our model are

- the lockdown intensity $p \in [0, p_{max}]$, where p_{max} is its maximum value;

- the testing intensity $\tau \in [0, \tau_{max}]$, where τ_{max} is the maximum rate of testing (or testing capacity).

Hence in our setting $k := (p, \tau)$ and $K = [0, p_{max}] \times [0, \tau_{max}]$. We will usually set $p_{max} = 1$, assuming that complete lockdown is possible, and $\tau_{max} = 365/2$ in days⁻¹, which is a reasonable value for the maximum testing capacity of a developed country. Clearly such parameters can be chosen differently, as we do for example in Section 3 where we consider the case where the maximum testing capacity is reduced by a factor 1/3.

The state variables are the eight epidemiological variables as discussed in the main text: $S, C, E_1, E_2, I_1, I_2, R, D_c$ which, for brevity, will be denoted by the vector $z \in \mathbb{R}^8$, with the components in the above order. All state variables are required to be positive and smaller than the initial size of the population N_0 . Hence, the state space, Z , is the compact subset of \mathbb{R}^8 given by $Z := [0, N_0]^8$. For brevity we will often denote the control strategies at time t by $k(t)$ and the state trajectories at time t by $z(t)$.

The state equations form a system of eight Ordinary Differential Equations (ODEs) which we rewrite here emphasizing the time dependent variables:

$$\begin{aligned} \dot{S}(t) = & \mu N(t) - \mu S(t) - (1 - p(t))^2 \beta \frac{S(t) (E_2(t)(1 - \gamma) + \gamma I_1(t)) (1 + c)}{N(t)} \\ & + \phi R(t) + \tau(t) \left(r + \frac{1 - r}{1 + c} \right) C(t) s_p + \delta C(t) - \tau(t) r S(t) (1 - s_p) \end{aligned} \quad (11)$$

$$\begin{aligned} \dot{C}(t) = & (1 - p(t))^2 \beta (c S(t) - (1 - \epsilon_C) C(t)) \frac{E_2(t)(1 - \gamma) + \gamma I_1(t)}{N(t)} \\ & + \tau(t) r S(t) (1 - s_p) - \mu C(t) - \tau(t) \left(r + \frac{1 - r}{1 + c} \right) C(t) s_p - \delta C(t) \end{aligned} \quad (12)$$

$$\begin{aligned} \dot{E}_1(t) = & (1 - p(t))^2 \beta (S(t) + (1 - \epsilon_C) C(t)) \frac{E_2(t)(1 - \gamma) + \gamma I_1(t)}{N(t)} \\ & - (\mu + \varphi_1) E_1(t) - \tau(t) \left(r + \frac{1 - r}{1 + c} \right) s_e E_1(t) \end{aligned} \quad (13)$$

$$\dot{E}_2(t) = \varphi_1 E_1(t) - (\mu + \varphi_2) E_2(t) - \tau(t) \left(r + \frac{1 - r}{1 + c} \right) s_e E_2(t) \quad (14)$$

$$\begin{aligned} \dot{I}_1(t) = & \varphi_2 E_2(t) + \tau(t) s_e \left(r + \frac{1 - r}{1 + c} \right) (E_1(t) + E_2(t)) \\ & - (\mu + \mu_{I_1} + (1 - \eta) \delta_1 + \eta \delta_1) I_1(t) \end{aligned} \quad (15)$$

$$\dot{I}_2(t) = \eta \delta_1 I_1(t) - (\mu + p_M \delta_2 + (1 - p_M) \delta_2) I_2(t) \quad (16)$$

$$\dot{R}(t) = (1 - \eta) \delta_1 I_1(t) + (1 - p_M) \delta_2 I_2(t) - (\mu + \mu_R + \phi) R(t) \quad (17)$$

$$\dot{D}_c(t) = \mu_{I_1} I_1(t) + p_M \delta_2 I_2(t) + \mu_E E_2(t) + \mu_R R(t) \quad (18)$$

where

$$N(t) := S(t) + C(t) + E_1(t) + E_2(t) + I_1(t) + I_2(t) + R(t), \quad \forall t \in [0, T] \quad (19)$$

with $N(0) = N_0$. It is useful to observe that in the state equations above we have, for all $t \in [0, T]$,

$$N'(t) + D_c'(t) = 0 \quad \iff \quad N(t) + D_c(t) = N_0.$$

Since $D_c(\cdot)$ is always positive and increasing, it follows that $N(t)$ is decreasing.

REMARK 1 If the mortality rates due to the epidemic are low, as is the case for COVID-19 (which is our driving example), then $N(t)$ remains very close to N_0 . In that case, (see e.g. Anderson and May (1992) and Keeling and Rohani (2011)), we can substitute $N(t)$ with N_0 in the above equations (and also below in the objective functional). While not necessary for our numerical procedure, this step simplifies the numerical approximations.

For later use, we introduce the notation for the state equations. We call $F : Z \times K \rightarrow \mathbb{R}^8$ the function defined by the right hand side of the state equations; i.e.,

$$F(S, C, E_1, E_2, I_1, I_2, R, D_c; p, \tau) = \quad (20)$$

$$\left(\begin{array}{l} \mu N - \mu S - (1-p)^2 \beta \frac{S(E_2(1-\gamma) + \gamma I_1)(1+c)}{N} + \phi R + \tau \left(r + \frac{1-r}{1+c} \right) C s_p + \delta C - \tau r S(1-s_p) \\ (1-p)^2 \beta (cS - (1-\epsilon_C)C) \frac{E_2(1-\gamma) + \gamma I_1}{N} + \tau r S(1-s_p) - \mu C - \tau \left(r + \frac{1-r}{1+c} \right) C s_p - \delta C \\ (1-p)^2 \beta (S + (1-\epsilon_C)C) \frac{E_2(1-\gamma) + \gamma I_1}{N} - (\mu + \varphi_1) E_1 - \tau \left(r + \frac{1-r}{1+c} \right) s_e E_1 \\ \varphi_1 E_1 - (\mu + \varphi_2) E_2 - \tau \left(r + \frac{1-r}{1+c} \right) s_e E_2 \\ \varphi_2 E_2 + \tau s_e \left(r + \frac{1-r}{1+c} \right) (E_1 + E_2) - (\mu + \mu_{I_1} + (1-\eta)\delta_1 + \eta\delta_1) I_1 \\ \eta\delta_1 I_1 - (\mu + p_M \delta_2 + (1-p_M)\delta_2) I_2 \\ (1-\eta)\delta_1 I_1 + (1-p_M)\delta_2 I_2 - (\mu + \mu_R + \phi) R \\ \mu_{I_1} I_1 + p_M \delta_2 I_2 + \mu_E E_2 + \mu_R R \end{array} \right) \quad (21)$$

With this notation, the state equations can be rewritten more concretely as

$$z'(t) = F(z(t); k(t)). \quad (22)$$

REMARK 2 The model can be studied for any choice of the initial conditions, z_0 , for the 8 state variables. Since our focus is the study of an emerging pathogen, our initial conditions will be those typical of the beginning of an epidemic; i.e.,

$$S(0) \sim N_0; \quad E_1 > 0 \text{ small}; \quad E_2, C, I_1, I_2, R, D_C = 0.$$

See e.g. Anderson and May (1992) Keeling and Rohani (2011), Dietz and Heesterbeek (2002)

Choosing the admissible control strategies. The set of admissible control strategies, \mathcal{K}_{ad} , must be chosen in a way that the above system of state equations admits a unique solution for any initial conditions $z(0) \in Z$ and remains in Z at all times. Moreover such a set of admissible strategies should be realistically implementable by the policymakers. The typical choice (see Section III.5 of Bardi and Capuzzo-Dolcetta, 1997) is to consider control strategies which are simply Lebesgue integrable¹; i.e.,

$$\mathcal{K}_{ad} := \{k(\cdot) : [0, T] \rightarrow K, \text{ Lebesgue integrable}\}.$$

For any such strategy, $k(\cdot) \in \mathcal{K}_{ad}$, and any initial condition $z_0 \in Z$ the state equation has a unique solution. We refer to this as the state trajectory corresponding to $k(\cdot)$ and we denote it by $z^{z_0, k(\cdot)}(t)$.

¹This class of functions is usually chosen in similar contexts, since it contains the piecewise constant policies commonly employed by decision makers in practice. Note that piecewise constant functions are not continuous, hence the class of continuous control strategies would not be an appropriate choice in our context.

The above set \mathcal{K}_{ad} consists of admissible policies that vary continuously with time. However, since in practice our control policies cannot be changed continuously, for our simulations in the main text we employ control policies that assume only a finite number of values. We call an n -phase policy for $n > 1$ any control function that is piecewise constant and assumes only n different values on contiguous time intervals, excluding the value 0. Moreover, we also impose that our control policy is null (inactive) before a fixed time, which we call *delay* in the main text, and which we denote hereafter by t_0 . Therefore we define the set of admissible controls with $n > 1$ phases (assuming only one between p and τ is considered)

$$\mathcal{K}_{ad}^n = \{k(\cdot) : [0, T] \rightarrow K, \exists(t_1, \dots, t_n) \in (t_0, T]^n, (k_1, \dots, k_n) \in K^n \text{ s.t.} \\ k(t) \equiv 0 \text{ for } t \leq t_0, k(t) \equiv k_i \text{ for } t \in (t_{i-1}, t_i] \text{ for } i = 1, \dots, N, k(t) \equiv 0 \text{ for } t > t_n\},$$

where we implicitly assumed the vector of times (t_1, \dots, t_n) to be monotonically increasing. It is clear from the previous definition that any function in \mathcal{K}_{ad}^n can be represented with $2n$ parameters, n for the vector of times when the policy changes, and n for the vector of intensity of the control policy in each phase. For $n = 0$ we define $\mathcal{K}_{ad}^0 = \{k : [0, T] \rightarrow K, k(t) \equiv 0\}$, the space consisting only of the zero function; that is, in this case the control does not affect the system.

Even if the function space \mathcal{K}_{ad}^n is not finite-dimensional as a vector space, the parameterization adopted here allows us to represent each element of \mathcal{K}_{ad}^n using a finite number of parameters ($2n$). However, notice that this representation is not unique, since two different combinations of the $2n$ parameters may describe the same function (in particular in the case where two consecutive i_i are equal).

In the case where both (p, τ) are present, we introduce the sets $\mathcal{K}_{ad}^{n,p}$ and $\mathcal{K}_{ad}^{M,\tau}$ for each of the controls and we define

$$\mathcal{K}_{ad}^{n,M} = \mathcal{K}_{ad}^{n,p} \times \mathcal{K}_{ad}^{M,\tau}.$$

Notice in particular that the number of lockdown phases, n , can be different from the number of testing interventions, M .

In what follows, the above sets of strategies will be denoted generically by \mathcal{K}_{ad} or, in case when the specification of the number of phases is needed, by $\mathcal{K}_{ad}^{n,M}$.

The objective functional to optimize: the starting problem (P_0). To define the objective functional we set

$$Y(t) = A_0(1 - p(t))^{\Delta+\alpha} L(t)^\alpha - \Phi(x(t), N(t)) \quad (23)$$

with $N(t)$ as in (19) and

$$L(t) = S(t) + \epsilon_C [C(t) + E_1(t) + E_2(t)] + R(t)$$

$$x(t) = \tau(t) [C(t) + E_1(t) + E_2(t) + rS(t)], \quad \text{the total number of people tested,}$$

where the function Φ for the cost of testing is given by

$$\Phi(x, N) = \rho_0 x + \exp\left(\frac{\rho_1}{N-x}\right) - \exp\left(\frac{\rho_1}{N}\right)$$

The objective functional to maximize is the Total Social Welfare (TSW) function specified in (10) over all admissible control strategies, i.e. over all $k(\cdot) \in \mathcal{K}_{ad}$. Here we denote it by $J_0(k(\cdot))$ to underline its dependence on the control strategies chosen by the policy-maker.

$$J_0(k(\cdot)) = \int_0^T U(Y(t)) - \theta V(D_c(t)) dt \quad (24)$$

with:

$$U(Y) = \frac{Y^{1-\sigma}}{1-\sigma} \quad \sigma \in (0, 1)$$

$$V(D_c) = \frac{D_c^\omega}{\omega} \quad \omega > 1;$$

Note that the utility from production is strictly concave while the disutility from deaths is strictly convex. Our framework allows us to consider different types of utility/disutility functions.

In what follows, we refer to “Problem (P_0)” as the problem of maximizing the functional $J_0(k(\cdot))$ over all $k(\cdot) \in \mathcal{K}_{ad}$, under the state equations (22). Note that, using the above setup, we can write the so-called running cost (i.e. the function appearing inside the integral in J_0) as a function $f_0(z(t); k(t))$ where $f_0 : \mathbb{R}^8 \times \mathbb{R}^2 \rightarrow \mathbb{R}$, more precisely (for brevity we write, as usual, N for $S + C + E_1 + E_2 + I_1 + I_2 + R$)

$$f_0(S, C, E_1, E_2, I_1, I_2, R, D_c; p, \tau) = \quad (25)$$

$$U(A_0(1-p)^{\Delta+\alpha}(S+R+\epsilon_C[C+E_1+E_2])^\alpha - \Phi(\tau[C+E_1+E_2+rS], N)) - \theta V(D_c)$$

The objective functional to optimize: the real exit time problem (P). As formulated above, problem (P_0) has an important drawback due to the fact that, even if all the infected compartments (E_1, E_2, I_1, I_2) are close to zero, the epidemic may restart. To avoid this artifact of continuous numbers of individuals, we introduce the following modification.

We first define for any given admissible state trajectory $z^{z_0, k(\cdot)}(\cdot)$ (associated to an initial datum $z_0 \in Z$ and a control strategy $k(\cdot) \in \mathcal{K}_{ad}$), and any closed region $\mathcal{O} \subset Z$, the first entry time $T_{\mathcal{O}}$ as

$$T_{\mathcal{O}}(z^{z_0, k(\cdot)}(\cdot)) := \inf \{t \geq 0 : z(t) \in \mathcal{O}\}$$

For brevity we will write $T_{\mathcal{O}}(z_0, k(\cdot))$, or simply $T_{\mathcal{O}}$ when no confusion is possible.

In our problem, as explained in the main text (see Section 3.2), we set²

$$\mathcal{O} := \{z \in \mathbb{R}^8 : 0 \leq z_3, z_4, z_5, z_6 \leq 1\}$$

The idea is that after time $T_{\mathcal{O}}$ the epidemic is no longer present and so both control variables and all the state variables E_1, E_2, I_1, I_2 are set equal to 0. The system still

²Note that to apply the standard optimality conditions for exit time problems (Cannarsa and Sinestrari, 2004, Section 8.4), the target set \mathcal{O} should not have corners like the one we use here. However, this does not present a problem here, as we use a discrete set of policies.

moves in the components S, R, D_c following equations

$$\begin{aligned}\dot{S}(t) &= \mu N(t) - \mu S(t) + \phi R(t) \\ \dot{R}(t) &= -(\mu + \mu_R + \phi)R(t) \\ \dot{D}_c(t) &= \mu_R R(t)\end{aligned}\tag{26}$$

with initial conditions given by their values at $T_{\mathcal{O}}$.

If $T_{\mathcal{O}} < T$, these components still have an influence on the utility/disutility functions so we have to take them into account. Hence, we consider the following modification of the above functional J_0 :

$$J_1(k(\cdot)) := \int_0^{T \wedge T_{\mathcal{O}}} U(Y(t)) - \theta V(D_c(t)) dt + \int_{T \wedge T_{\mathcal{O}}}^T U(Y(t)) - \theta V(D_c(t)) dt \tag{27}$$

where in the second integral, $Y(t) = A_0(S(t) + R(t))^\alpha$ and where the variables $S(t), R(t), D_c(t)$ solve (26) with initial conditions given by their values at $T_{\mathcal{O}}$.

Note that the second integral, when $T_{\mathcal{O}} < T$ can be expressed as a given function Ψ (which can be explicitly computed) of $T_{\mathcal{O}}$ and $z(T_{\mathcal{O}})$ so we can rewrite the functional J_1 as follows³

$$J_1(k(\cdot)) := \int_0^{T \wedge T_{\mathcal{O}}} U(Y(t)) - \theta V(D_c(t)) dt + \Psi(T_{\mathcal{O}}, z(T_{\mathcal{O}})) \mathbb{1}_{[0, T)}(T_{\mathcal{O}}) \tag{28}$$

Our Problem (P) is then maximizing J_1 with the above change of state equations.

This is a standard optimal control problem with exit time (see Chapter 8 in Cannarsa and Sinestrari (2004)). We have the following:

DEFINITION 1 *Admissible policy $k^*(\cdot)$ is an optimal control policy if it maximizes J_1 ; i.e.,*

$$J_1(k^*(\cdot)) \geq J_1(k(\cdot)) \quad \forall k(\cdot) \in \mathcal{K}_{ad}.$$

We define an optimal state trajectory $z^(\cdot)$ as the unique solution of the state equation associated to an optimal control strategy.*

We define the optimal $(k^(\cdot), z^*(\cdot))$ as an ordered pair of functions where the first is an optimal control strategy and the second is its associated state trajectory.*

In some cases it is useful to emphasize in the notation the fact that the objective functional depends, through the state equations, also on the initial condition $z(0) = z_0$, by writing $J_1(k(\cdot); z_0)$ in place of $J_1(k(\cdot))$. We then define the value function as the function

$$W(z_0) := \sup_{k(\cdot) \in \mathcal{K}_{ad}} J_1(k(\cdot); z_0)$$

defined for all $z_0 \in Z$. In economic terms this will be called the welfare function.

³The notation $\mathbb{1}_A$ stands for the indicator function of a set A ; i.e., a function which is 1 on the set A and 0 elsewhere.

Optimization in a finite dimensional setting (fixed number of phases). Under a fixed number of control phases corresponding to the values taken by the piecewise constant controls (p, τ) , see subsection 4, the problem can be reduced to a finite-dimensional optimization problem with inequality constraints.

Fix N to be the number of phases of the control p , and assume $M = 0$; i.e., the control τ representing the testing policy is inactive at all times. The general case where $N, M > 0$ can be treated in the same manner by using a slightly more cumbersome notation. For every control policy $p \in \mathcal{K}_{ad}^{N,p}$, we identify by I_p and t_p as the vectors of \mathbb{R}^N that represent intensities and times of phase change associated with the function p . We use the notation p_{I_p, t_p} to denote such a policy.

Therefore, our finite dimensional optimization problem becomes

$$\begin{aligned} \max_{I_p, t_p} \quad & J_{0, P'}(p_{I_p, t_p}(\cdot)) \\ \text{s.t.} \quad & 0 \leq i_j \leq 1 \quad \forall j = 1, \dots, N, \\ & t_j \leq t_{j-1} \quad \forall j = 1, \dots, N, \\ & t_N \leq T. \end{aligned}$$

This is a standard finite dimensional optimization problem in \mathbb{R}^{2N} under $2N + 1$ inequality constraints. The problem here is that the objective function cannot be computed explicitly given the parameters of the controls. Hence we have to rely on a numerical approximation to compute the objective function, as well as to perform the numerical optimization. To compute the objective function we rely on classical numerical methods for ODEs. However, due to the lack of regularity of the controls, we employ a numerical method which is suitable for stiff problems, such as implicit Runge-Kutta methods of high order. To approximate the continuous time integration we employed a Gauss-Kronrod quadrature rule.

The process of numerical optimization is challenging due to the lack of convexity in the objective function. In the absence of such convexity, there is no guarantee that a local optimum will also be a global one. Therefore, during the numerical optimization, we have to rely on a global optimization algorithm, which is numerically more demanding than a local numerical optimizer. In figure 7, we analyze the case of a single lockdown phase, showing the shape of the objective function for many combinations of the parameter describing the control policy using a very fine grid. In this case we see that concavity is indeed a feature of the objective with respect to this parameterization. A general result concerning the shape of the objective function is however still lacking.

It is well known that global optimization is one of the many problems suffering from the *curse of dimensionality*, that is the number of function evaluations required for a thorough search in the state space grows exponentially with the dimension of the problem. For the optimization procedure we employed both the DIRECT and DIRECT-L algorithm proposed in Jones et al. (1993) and Gablonsky and Kelley (2001). Both algorithms are deterministic procedures based on a subdivision of the domain in iteratively smaller rectangles until convergence is reached. To ensure

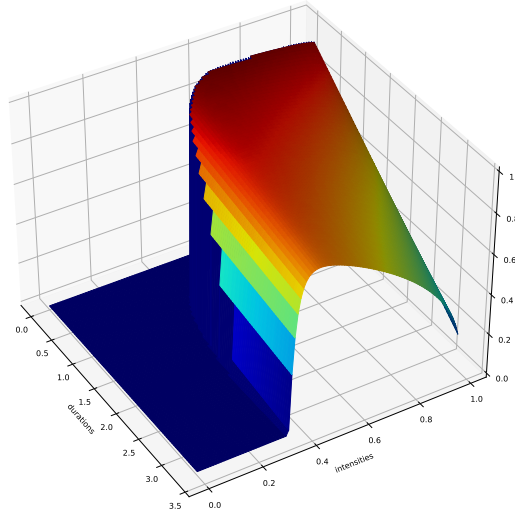


FIGURE 7. Normalized objective function $J_{0,P'}$ for the control policy consisting of a single phase lockdown (values below zero have been cut for graphical purposes). The x-axis “durations” refers to the length of the single lockdown phase, while the y-axis “intensities” refers to the strength of the policy. In this case the objective function is convex with respect to the parameterization adopted.

robustness of our results, we also performed some additional tests based on a combination of a brute-force approach on a very fine grid with a local refinement based on a local optimizer. An additional confirmation of the accuracy of the optimization procedure is that the maximum value of the objective function shifts in the expected direction when changing some of the parameters (e.g. when increasing the delay in the control policies the objective function decreases).

THE ADDED BENEFITS OF WEARING MASKS

Arguably the most economical control strategy during the Covid pandemic has been the wearing of surgical masks at work and in public places. We have not incorporated the use of masks in our analysis, partly because we see it as an important ‘safety play’ that is unlikely to stop the pandemic; masks may provide some protection to those that wear them, while also significantly reducing transmission from asymptomatic infected hosts. An important limitation here is that we do not have good estimates of the efficacy of masks (neither for transmission blocking, nor for protection). The best available methods suggest an average efficiency of around 45 percent (averaged across both transmission blocking and susceptible protection). Similarly, the proportion of people wearing masks varies widely and may follow current levels of infection and perceived risk in the population.

We have not explicitly examined the dynamic consequences of wearing face masks (see Eikenberry et al. (2020)). Instead we present a simple static analysis that compares the efficacy of face masks with that of a vaccine (Appendix 2). We do this as there has been resistance to wearing face masks and wearing them is subject to asymmetric and poorly quantified levels of efficacy. Specifically, when an infected individual wears a mask it may trap a high proportion of the viral infection stages they produce, e_i , but not all of them. Similarly, uninfected hosts who wear masks reduce their level of exposure but are unlikely to completely exclude viral particles, their masks have a protective efficiency of e_p . The efficacy of both processes will vary with the type of mask and how it is worn. We can examine the static implications of this by modifying the standard equation for the proportion of people who need to be vaccinated, p_c , to reduce the Basic Reproductive Number of a pathogen, R_0 , below unity Fine (1993); Fine et al. (2011); Anderson and May (1992),

$$p_c = (1 - 1/R_0) \quad (30)$$

The proportion of people who have to wear masks at different levels of R_0 is given by

$$p_c = (1 - \sqrt{e_p e_i}/R_0) \quad (31)$$

We coarsely quantify mask efficacy as the geometric mean of the two types of mask efficacy, this is plainly a conservative calculation as there are always more uninfected than infected hosts. Figure 8 illustrates the relationship between the Basic Reproductive Number, R_0 , of the pathogen and the proportion of people who have to wear masks at different average levels of mask efficacy as protective devices. A key point to notice from the earlier figure is that even when masks are 90 percent efficient, they are always less efficient than vaccination and the proportion of people wearing them has to increase rapidly as their efficacy declines. Nonetheless, the figure serves to underline our main point, a high proportion of people have to wear very efficient masks if they are to be effective in reducing R_0 below unity. They will still serve the useful function of slowing the epidemic and reducing the pressure on health care services. Ultimately, we see them as a useful adjunct to the other forms of NPI discussed in more detail below. If more people wear masks, then lock-downs can be shorter and impact a smaller proportion of the economy.

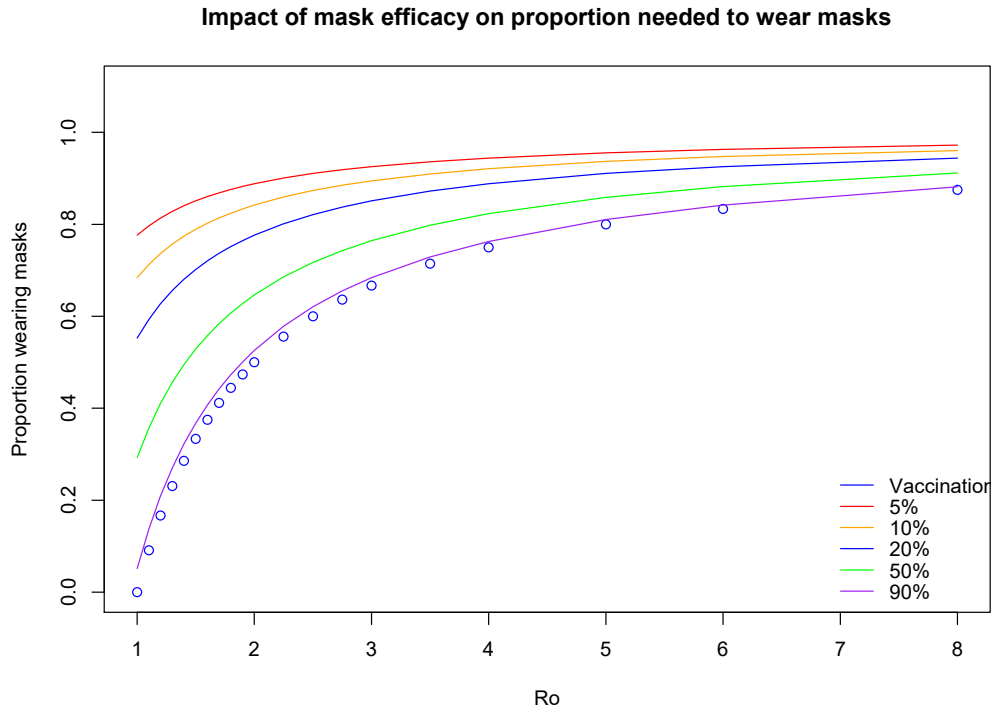


FIGURE 8. The relationship between the Basic Reproductive Number of a pathogen, R_0 and the proportion of people who need to wear masks so that R_0 is reduced below unity and the pathogen contained. The dotted lines illustrate the "classical" curve for a totally efficient vaccine that generates lifetime immunity. The colored lines illustrate the cases for different levels of protective efficiency for mask-wearers; this efficiency is assumed to be the geometric mean efficacy of protecting the wearer from infection and preventing transmission from an infectious person. The purple line signifies 90 percent mean efficacy, green 50 percent, etc.