

Sequential Lifting of COVID-19 Interventions with Population Heterogeneity *

Adriano A. Rampini[†]

April 17, 2020

Abstract

This paper proposes a sequential approach to lifting interventions in the COVID-19 pandemic taking heterogeneity in the population into account. The population is heterogeneous in terms of the consequences of infection (need for hospitalization and critical care, and mortality) and in terms of labor force participation. Splitting the population in two groups by age, a less affected younger group that is more likely to work, and a more affected older group less likely to work, and lifting interventions sequentially – for the younger group first and the older group later on – can substantially reduce mortality, demands on the health care system, and the economic cost of interventions.

Keywords: COVID-19; pandemic; epidemiological model

The COVID-19 pandemic has wreaked havoc on life around the globe. The severe interventions that suppression strategies entail have resulted in massive collateral damage to society and the world economy. Policies enacted by fiscal and monetary authorities are designed to blunt the impact of the dramatic reduction in economic activity by providing transfers and extending credit. But even drastic fiscal and monetary measures can only address disruptions of the scale experienced in many countries temporarily. Suppression strategies that result in lockdown for the better part of an 18-month period or until a vaccine is found as suggested by Ferguson et al. (2020) do not seem economically and politically feasible even for governments with substantial fiscal capacity, let alone for the typical government. In the best-case scenario, suppression strategies are so effective and testing capacity raised sufficiently quickly that epidemic spread is stopped around the world in a few months and does not rebound once the drastic interventions are lifted.

*First version. Preliminary. The author is the William and Sue Gross Professor of Financial Economics at Duke University, an NBER Research Associate, and a CEPR Research Fellow, and is currently on sabbatical leave at Princeton University and NYU; their hospitality is gratefully acknowledged.

[†]Duke University; rampini@duke.edu

In a more likely scenario, the epidemic spread will rebound as interventions are eased, especially given the global interconnectedness of the world economy. An evaluation of approaches to progressively lifting interventions and switching from suppression to mitigation strategies is hence urgently needed.

We focus on the health and economic consequences of a sequential approach that lifts interventions for the less vulnerable fraction of the population first and for the more vulnerable fraction of the population second. We show that this approach reduces mortality while increasing economic activity (relative to delayed lifting of interventions) by allowing the fraction of the population, which participates in the labor market to a greater extent, to return to work earlier. That said, technological progress in testing and pharmaceutical interventions, including development of treatments and a vaccine, continue to be a priority, as is increasing the capacity of the health system in terms of testing, hospital beds, and critical care. In fact, a mitigation approach makes such progress more urgent not less. An added benefit of imposing non-pharmaceutical interventions on the more vulnerable population for longer is that, by the time interventions are lifted on this population, health system capacity has increased and treatments, and even a vaccine, may be available.

This paper considers a deliberately simple discrete time version of the Kermack and McKendrick (1927) Susceptible-Infected-Recovered (SIR) model with a heterogeneous population to study the implications for mortality, the demands on the health system, and the economy. We consider a population comprised of two groups (but the approach can be extended to multiple groups in a straightforward matter). The two populations differ in the consequences of infection and in their participation in labor force. Specifically, one group is comprised of younger people who are less affected by infection and have a higher labor force participation rate. The other group is comprised of older people who are more affected by infection and have a lower labor force participation rate.

We show that a sequential approach to lifting interventions, in which the less affected younger group is released earlier and the more affected older group is released later, can substantially reduce mortality, the demands on the health care system, and the drop in economic activity. In our baseline specification, mortality is reduced by close to 40% and peak load hospitalizations and critical care demand by about 75% and 80%, respectively, compared to lifting interventions on both groups at the same time. Mortality is reduced because by the time restrictions are lifted for the more vulnerable population a sizable part of the population has recovered reducing the infectiousness of the pandemic. A staggered lifting of interventions reduces peak load demand for health care directly, but importantly also indirectly, if the more vulnerable group is released only once a sizable fraction of the

less vulnerable group has recovered which reduces the speed of infection. Most noteworthy is that the sequential approach makes a substantial reduction in mortality possible and herd immunity is achieved with a lower fraction of the population ever infected. The fact that the less affected group participates in the labor force to a greater extent is a fortunate coincidence, as the economic consequences of interventions can be reduced with the sequential approach compared to a delayed lifting of interventions for both groups. The output loss in our baseline sequential strategy is about 16% (of annual output), which is more than the 10% output loss if interventions are lifted early for both groups, but drastically less than the 40% output loss associated with delayed lifting of interventions for both groups.

Chikina and Pegdeny (2020) consider the consequences of similar age-targeted mitigation strategies for mortality and the demands on the health system.

The paper proceeds as follows. Section 1 describes the SIR model with a heterogeneous population. Section 2 evaluates the various strategies for lifting interventions. Section 3 considers various extensions and provides further discussion. Section 4 concludes. Appendix A reviews the basic discrete-time SIR model with a homogeneous population.

1 SIR model with heterogeneous population

We consider a SIR model with heterogeneity building on Allen (1994). Consider a population of size normalized to one. The population comprises two groups, labeled 1 and 2, of size ω_1 and ω_2 where $\omega_j \in (0, 1)$, $\sum_{j \in \mathcal{J}} \omega_j = 1$, and $\mathcal{J} = \{1, 2\}$. Denote the fraction of group j which is susceptible, infected, recovered, and deceased by s_j , i_j , r_j , and d_j , respectively, where $s_j + i_j + r_j + d_j = 1$. The fraction susceptible in the population overall is then $s = \sum_{j \in \mathcal{J}} \omega_j s_j$, the fraction infected in the population is $i = \sum_{j \in \mathcal{J}} \omega_j i_j$, and analogously for other variables. The contact rate at which type j meets type k is $\beta_{jk} > 0$, where by symmetry $\beta_{jk} = \beta_{kj}$. The fraction that recover or die each period is $\gamma \in (0, 1)$, of which fraction $\delta_j \in (0, 1)$ die.

We consider γ a fixed parameter determined by the disease. However, the rate at which the population meets is considered a policy variable determined by intervention, which may vary over time, that is, $\beta(t)$, but we suppress this dependence on time t henceforth.

Susceptible individuals in group j contact individuals in group k at rate β_{jk} . Contact with an infected individual leads to infection and the probability of contact with an infected individual of type k is $\omega_k i_k$; the rate of contact with an infected individual of any type is hence $\sum_{k \in \mathcal{J}} \beta_{jk} \omega_k i_k$. A key assumption is that recovered individuals are no

longer susceptible to infection. The dynamics of the fraction of group j that is susceptible, infected, recovered, and deceased are

$$s'_j = s_j - s_j \sum_{k \in \mathcal{J}} \beta_{jk} \omega_k i_k \quad (1)$$

$$i'_j = i_j + s_j \sum_{k \in \mathcal{J}} \beta_{jk} \omega_k i_k - \gamma i_j \quad (2)$$

$$r'_j = r_j + (1 - \delta_j) \gamma i_j \quad (3)$$

$$d'_j = d_j + \delta_j \gamma i_j, \quad (4)$$

for $j \in \mathcal{J}$, where a prime denotes the value next period. Note that these equations describe the dynamics for a population with any number J of groups. However, for simplicity we only consider the case with two groups. By analogy with the standard SIR model without heterogeneity (see Appendix A), the reproduction ratio for group j can be defined as $\mathcal{R}_{0,j}(t) = \frac{s_j \sum_{k \in \mathcal{J}} \beta_{jk} \omega_k i_k}{\gamma i_j}$ and for the population overall as $\mathcal{R}_0(t) = \frac{\sum_j \omega_j s_j \sum_{k \in \mathcal{J}} \beta_{jk} \omega_k i_k}{\gamma i}$. If this reproduction ratio exceeds 1 in group j or the population overall, the pandemic will spread in that group or the population, respectively.

Note that the assumptions about mortality δ_j in each group do not affect the dynamics of the epidemic in terms of s_j and i_j , as equations (1) and (2) are independent of equations (3) and (4) and do not involve the parameters δ_j . The assumptions about δ_j do of course affect the fraction of the population that eventually recovers and the implications of the pandemic for mortality.

We model the demand on the health system by simply assuming that a fraction $\phi_j^h \in (0, 1)$ of the infected in group j require hospitalization and a fraction $\phi_j^c \in (0, 1)$ require critical care. Therefore, the fraction hospitalized in group j is $h_j = \phi_j^h i_j$ and the fraction requiring critical care is $c_j = \phi_j^c i_j$. The fraction hospitalized in the overall population is $h = \sum_{j \in \mathcal{J}} \phi_j^h \omega_j i_j$ and the fraction requiring critical care $c = \sum_{j \in \mathcal{J}} \phi_j^c \omega_j i_j$.

We model the economy simply as output being produced by individuals who participate in the labor force using a linear technology requiring only labor inputs. Labor force participation l_j depends on the group j . We assume all individuals who work are equally productive and normalize the productivity to one. We assume that the infected population cannot participate in the labor force, while the susceptible and recovered population participate such that output of group j is $y_j = l_j (s_j + r_j)$. Aggregate output in the economy is hence $y = \sum_{j \in \mathcal{J}} l_j \omega_j (s_j + r_j)$. Finally, we assume that if a group is subject to the intervention due to the pandemic, only a fraction of the population of the group can participate in the labor force from home. This is the main way in which intervention reduces economic activity.

2 Evaluation of strategies to lift interventions

This section evaluates a sequential strategy to lift interventions as well as two alternative strategies which treat the entire population symmetrically, one lifting the interventions early and the other lifting the interventions late. We find that the sequential strategy reduces mortality and the demands on the health system substantially compared to the two alternative strategies considered. In terms of economic consequences, lifting interventions early for the entire population is associated with the lowest output loss, but the sequential strategy has drastically lower output losses than lifting interventions late for the entire population.

2.1 Calibration

We calibrate the baseline model using data for the COVID-19 pandemic mostly from Ferguson et al. (2020). Panel A of Table 1 displays the significant age-dependence of morbidity and mortality due to infection. Hospitalization rates increase substantially with age; they are low for individuals below 50 years of age and rise considerably with age above 50. The pattern in the need for critical care is more drastic, with very few cases requiring critical care for individuals below 50, limited need for individuals between 50 and 59, and a substantial increase for older individuals. The pattern in mortality is the most dramatic; mortality is low for individuals below 60 but increases very substantially for the groups in their 60s, 70s and 80 and above. Notice that each ten-year age bracket contains roughly equal population, except for the top two: the population age 70 or older is roughly equal in population to each of the ten-year age brackets below.

Labor force participation displays the opposite pattern: labor force participation is roughly constant at above 80% for the age groups 25-34 to 45-54 and falls to 65% and 27% for individuals aged 55-64 and 65-74, respectively. Labor force participation below 25 is also lower, although one could argue that schooling is investment in human capital and thus should be considered as producing output as well.

Based on this data we calibrate a model with two groups, Groups 1 and 2; see Panel B of Table 1. Group 1 comprises the population 0 to 54 years old, which is about 71% of the population overall. Group 2 comprises the population 55 years old and up, about 29% of the population. Since the hospitalization and critical care rates from Ferguson et al. (2020) are for symptomatic cases only, we adjust these downward by 50% to account for asymptomatic cases. The hospitalization and critical care rates are 1.22% and 0.08% in the younger group and 9.56% and 3.75% in the older group. We take the mortality rate from Ferguson et al. (2020), but adjust the overall rate, so that the average mortality rate

replicates their predicted number of deaths in the unmitigated scenario.¹ The mortality in Group 1 is 0.06% compared to 2.67% in Group 2, a dramatic difference.

As a fraction of the working age population 16 years old and up, Group 1 makes up about 64% and Group 2 about 36%. The labor force participation rate (among the working age population) is about 76% in Group 1 and 40% in Group 2. This implies that Group 1 constitutes 77% of the labor force while Group 2 constitutes 23%. Given the assumption that all individuals are equally productive, these fractions are also the fraction of output generated by each group. Income data by age could be used to calibrate the economic importance of the two groups more precisely.

We choose 55 years of age as the cutoff so that the younger group contains a large fraction of the labor force and the older group a large fraction of the more vulnerable population, but a different cutoff could be considered. We choose to split the population in just two groups for simplicity and to facilitate the interpretation, but one could consider a more stratified approach where interventions are lifted for an even younger group first, and then progressively for older and older groups.

We calibrate the epidemiological parameters largely based on some of the baseline values considered by Ferguson et al. (2020). The basic reproduction number for COVID-19 we consider is 2.4, but there is considerable uncertainty about this value.² Following Atkeson (2020) we calibrate the recovery rate to 7/18, as the expected duration of the disease is 18 days and a period in our model corresponds to a week. The implied baseline contact rate without heterogeneity is thus $\beta \equiv \beta_{jk} = \mathcal{R}_0\gamma = 0.93$, which we assume is identical across groups without intervention. We assume that intervention reduces the contact rate by 75% to 0.25β in our baseline intervention strategy, with details discussed below.

In terms of output, we calibrate the labor force participation rate to the one observed in the data. Moreover, we assume that all individuals are equally productive, that infected individuals cannot work, and that individuals subject to intervention have to work from home, if they can, and calibrate the fraction of work which can be done from home to 46% (which is a wage-weighted average) following Dingel and Neiman (2020).

¹Ferguson et al. (2020) report that in their unmitigated scenario 81% of the U.S. population would be eventually infected and there would be 2.2 million deaths; given a U.S. population of 330 million, this implies a death rate of 0.82%. We adjust the population-weighted average mortality rate from Panel A of Figure 1 of 1.12% to match that number and adjust the mortality rate in each group accordingly.

²A lower number implies the pandemic spreads more slowly, which reduces peak health system demands but would mean the lifting of interventions for Group 2 would need to be further delayed.

2.2 Alternative strategies

In our baseline calibration, we consider the following three intervention lifting strategies. The first two treat the entire population homogenously, lifting the general intervention for both groups at the same time, either early or late. The third strategy involves sequential lifting of interventions for the younger group (Group 1) first and for the older group (Group 2) later on.

The specific timing is as follows. We start the population in week 1 (week of January 5, 2020) with 0.01% of the population infected. The pandemic spreads with contact rates assumed to be constant at β for all groups. In week 11 (week of March 15, 2020), a general intervention is implemented which reduces all contact rates to 0.25β . The strategies we consider differ in when and how these interventions are lifted. The two homogenous strategies lift the interventions for both groups at the same time, either in week 16 (week of April 19, 2020) (we call this strategy “Early Lifting”) or in week 45 (week of November 8, 2020) (“Late Lifting”), when the contact rates go back to their pre-intervention values. The sequential strategy lifts the intervention for Group 1 (young) in week 16, but delays lifting the intervention for Group 2 (old) until week 45; we call this strategy “Sequential Lifting.” During the time where the intervention is lifted for Group 1 but not Group 2, we assume that the contact rate among Group 2 (old) and between Group 1 and Group 2 is 0.25β , while the contact rate among Group 1 (young) returns to its pre-intervention level. The contact rates for these interventions are summarized in Panel A of Table 2.

2.3 Quantitative evaluation

Table 3 evaluates the three strategies quantitatively and Figures 1 and 2 illustrate the dynamics. We first describe the quantitative results for the Early Lifting strategy, our first homogenous strategy, which we use as our benchmark to assess the gains from the sequential strategy. We note that, except for the economic costs, the results from our second homogenous strategy, the Late Lifting strategy, are rather similar, so this description applies to both homogenous strategies we consider. The overall mortality rate is high: 0.74% of the population eventually die from the disease.³ The mortality rate is low for Group 1, namely 0.05%, but rather high for Group 2 at 2.41%. The demands on the health system are also massive, with the peak hospitalization rate at 0.86% of the population and the peak critical care rate at 0.27%; these vastly exceed hospital bed capacity of 0.29% (according to the WHO) and ICU capacity of around 0.02% (according

³This is because our calibration follows that of Ferguson et al. (2020); calibrating mortality based on Lourenço et al. (2020) results in considerably lower numbers as we show in the sensitivity analysis in Section 2.4.

to IHME). The reason is that with a homogenous lifting strategy there is a massive spike in infections (and hence hospitalizations and demand for critical care) when the interventions are lifted (see Column A in Figure 2). When interventions are lifted across the board, the reproduction ratio goes back to almost the basic reproduction number \mathcal{R}_0 (as the middle panel in Figure 1, Column A, shows) and the pandemic spreads very quickly, resulting in a spike in deaths. Economic output drops during the intervention, then rebounds, but drops again when infections spike. The output losses are large, roughly 10% of output, but the downturn is quite transitory (see the bottom panel in Column A of the figure). We should caution that our model of the economy is simplistic, and does not have its own dynamics or consider financial factors. In this early lifting strategy, about 90% are eventually infected, or in other words only about 10% of the population are never infected.

Consider now the Sequential Lifting strategy. Mortality is still high, at 0.46%, but this is a reduction of about 38% relative to the homogenous Early Lifting strategy, which is quite substantial. The mortality of both groups drops, to 0.05% and 1.47%, respectively. Most of the reduction in mortality comes from Group 2 (old). Peak demand for hospitalizations and critical care drop to 0.22% and 0.05%, respectively. These are reductions of 75% to 80% compared to the Early Lifting strategy, which is massive. In fact, peak demands are now more in line with capacity, although critical care needs still exceed normal capacity by a factor 2 or more.

How is this reduction accomplished? From Figures 1 and 2 we see that the fraction infected is considerably flatter in this strategy and shows two, much flatter peaks. This is because the interventions are lifted sequentially. The reproduction ratio rises substantially when the interventions for Group 1 (young) are lifted, but not nearly to the basic reproduction number \mathcal{R}_0 . Over time, it falls below one, and while it jumps when the interventions for Group 2 (old) are lifted, it stays below one for the population overall. The result is a double hump shape in infections, hospitalizations, demand for critical care, and deaths, but with lower peaks. While Group 2 (old) remains subject to the interventions during the first peak, most of the deaths even in this phase are from this older group. This is because there is contact between the two groups even during this sequential phase. The large reduction in mortality is possible, because by the time the interventions are lifted for the older group, a sizable fraction of the population has already recovered, which reduces the reproduction ratio and results in a much lower fraction of the older group ever getting infected: under the homogenous Early Lifting strategy, about 10% are never infected, whereas under the sequential strategy 45% of Group 2 (old) are never infected (see Table 3). This is the mechanism that allows a substantial reduction

in overall mortality.

The output loss associated with the sequential strategy is larger than that when all interventions are lifted early, a loss of about 16% of annual output compared to 10%, as Group 2 (old) is subject to the interventions for longer and only a fraction of individuals can work from home. However, since the older group comprises only 23% of the labor force, the economy recovers considerably when the interventions are lifted on the younger group. This initial recovery may be enough to stave off an economic disaster. The implications of a homogenous Late Lifting strategy are very similar to those associated with the homogenous Early Lifting strategy except for the economic consequences: because of the extended period of interventions, the output loss associated with Late Lifting are about 40%, clearly an economic disaster.

We conclude that the Sequential Lifting strategy vastly improves on lifting interventions later for the entire population in terms of economic consequences, while at the same time substantially reducing mortality and the demands on the health system relative to homogenous policies that do not take into account the heterogeneity in the population in terms of vulnerability and economic activity. To state this provocatively: when considering the options of lifting interventions earlier for the less vulnerable population vs. later for the entire population, there is *no trade off* between economic activity and mortality.

2.4 Sensitivity analysis

We now consider the sensitivity of our quantitative results to the assumptions about the effectiveness of the intervention and about the mortality rates. We find that a sequential policy offers sizable benefits even if interventions are less effective during the sequential phase or less effective overall, and even if the mortality rate of the disease itself is calibrated to a much lower level. These results are reported in Table 4 and the alternative assumptions about the effectiveness of the policy are described in Panels B and C in Table 2.

Consider again the Sequential Lifting strategy, but suppose that shielding the vulnerable population from infection during the phase in which the interventions are lifted for Group 1 (young) but not Group 2 (old) is difficult. Specifically, suppose the contact rate between the two groups only drops by 50% (to 0.5β) instead of by 75% (to 0.25β) as assumed before (see Panel B of Table 2). This results in a more pronounced peak in infections (and hence demands on the health system and mortality) when the interventions are lifted for Group 1 (young). However, sizable gains from the sequential policy remain (see Panel A of Table 4). Compared to our baseline calibration, overall mortality increases to 0.50% (and 1.60% for the older group) and the peak hospitalization and crit-

ical care rates increase to 0.37% and 0.11%, but these still represent reductions of 32% in mortality and around 60% in demands on the health system relative to the homogenous strategies. This suggests that a sequential approach is desirable even in this case. The basic mechanism remains the same: by the time the interventions are lifted on the more vulnerable population, a larger fraction of the population has already recovered, which in turn increases the fraction of the more vulnerable population that is never infected.

Next suppose that all interventions are less effective in reducing contact rates, and that all contact rates are reduced by 50% instead of by 75% as in the baseline scenario (see Panel C of Table 2). As Panel B of Table 4 shows, the performance of the Early Lifting strategy is very similar to the baseline case. The Sequential Lifting strategy does not do quite as well as in our baseline case. Mortality is 0.51% for the population overall (and 1.64% for the more vulnerable group) and peak demand on the health system is 0.41% in terms of hospitalizations and 0.12% in terms of critical care. These are still sizable reductions of over 30% in mortality and more than 50% in terms of demand on the health system relative to the homogenous Early Lifting strategy. Lifting the interventions sequentially still offers benefits as a larger fraction of the more vulnerable population is never infected. Notice that in this parametrization the Late Lifting strategy also offers some benefits in terms of reduced mortality and demands on the health system, but it does still come at an enormous economic cost.

Finally, there is considerable debate about the mortality rate of COVID-19. Lourenço et al. (2020) use a much lower mortality rate of 0.14%. Adjusting the mortality rates of the two groups accordingly, we find an overall mortality rate in the homogenous Early Lifting and Late Lifting strategies of 0.13% and 0.41% in the more vulnerable group (see Panel C in Table 4). In contrast, the Sequential Lifting strategy results in mortality rates of 0.08% in the population overall and 0.25% in the more vulnerable group, reductions in mortality of about 40%. In sum, these results suggest that the advantages of a sequential approach to lifting interventions are reasonably robust to alternative calibrations of the parameters.

3 Extensions and Discussion

The analysis in this paper is kept deliberately simple to emphasize the key insight that lifting interventions on the less vulnerable population earlier can increase economic output while reducing mortality and health system demands at the same time. Here we discuss several additional advantages of the sequential approach as well as some limitations of the analysis.

3.1 Growth in testing, hospital, and critical care capacity

Much of the recent debate has focused on the desirability of an expansion of testing, hospital, and critical care capacity. A sequential approach to lifting interventions reduces peak demand on the health system, but it also allows additional time to increase such capabilities significantly by the time the interventions are lifted for the more vulnerable population. This is an additional advantage of the sequential strategy in our view.

Moreover, there is considerable uncertainty about many of the basic parameters of the disease, including mortality and the extent to which the disease has already spread in the population. The sequential approach, which shields the more vulnerable population for longer, raises the possibility that this uncertainty can be reduced considerably before the interventions are lifted for that part of the population. Indeed, if it turns out that a much larger fraction of the population has already been exposed and has recovered from the disease, it may be possible to lift the restrictions for the vulnerable part of the population much earlier than the analysis here suggests.

3.2 Technological progress in treatment and vaccine

We have not considered the possibility of technological progress in terms of the development of treatments and a vaccine. Our basic insight does not depend on such treatments or vaccines becoming available, but if they do, the delayed lifting of interventions for the more vulnerable population makes it more likely that they will be available before that population is exposed to the disease. That said, if such pharmaceutical progress were likely to be made very soon, an argument for delaying the lifting of interventions until then could be made. Similarly, the possibility of significant progress in diagnostic technologies could make delaying the lifting of interventions optimal, and speed up the lifting of interventions once such progress has been made. That said, since the approach considered here involves achieving herd immunity, exposing a large fraction of the less vulnerable population to the disease is paradoxically beneficial, especially for the more vulnerable population.

3.3 Discussion

We consider the effect of policies in a simple epidemiological model. In an economic model, individuals make choices about whether to work, shop, or stay home taking the risks and consequences of various actions into account. Recent work on economic models considers the endogenous response of individuals to a pandemic (see, e.g., Eichenbaum et al. (2020)). In such models, individuals may choose to modify their behavior to reduce

contact, and this should be especially true for more vulnerable individuals. Thus, some of the benefits of the policies analyzed here may arise in equilibrium even in the absence of policy, although individuals of course do not take into account how their own actions affect the equilibrium dynamics.

From a policy perspective, the analysis here suggests that employment protections and financial assistance might optimally be targeted at the more vulnerable population, reducing the demands on fiscal policy. But the spirit of the analysis is to allow a large fraction of economic activity by the less vulnerable population to resume, reducing the financing needs for individuals, firms, and governments.

Several limitations should be noted. First, we stratify the population by age only, but given that comorbidities significantly affect vulnerability, these should probably also be taken into account. Second, we stratify the population by age of the individual, but the households structure should be taken into account. Households should be stratified by the age of the oldest member; [data on the household age structure](#) from the U.S. Census Bureau suggests that 55% of households have a householder who is less than 55 years old, which suggests that taking this into account would not dramatically alter our conclusions.

4 Conclusion

Sequentially lifting interventions in a pandemic, first for the less vulnerable population and subsequently for the more vulnerable one, can substantially reduce mortality and the demand on the health care system. In our baseline calibration mortality is reduced by close to 40% and the peak demand for hospitalization and critical care are reduced by 75% and 80%. Mortality is reduced because by the time restrictions are lifted for the more vulnerable population a sizable part of the population has recovered reducing the infectiousness of the pandemic. This means that a larger fraction of the more vulnerable population is never infected with the virus. Peak demand on the health care system is reduced for two reasons: on the one hand, there is a direct effect as the restrictions are lifted for different parts of the population at different times; on the other hand, there is an indirect effect, namely the fact that a smaller part of the more vulnerable population is ever infected. Lifting restrictions earlier for the less vulnerable part of the population substantially reduces the economic cost compared to keeping the restrictions in place for the entire population for a longer period of time. An additional advantage of a sequential strategy is that, before the restrictions on the more vulnerable part of the population are lifted, more is learned about the morbidity and mortality risk of the disease and there is potential to increase health system capacity and develop testing and possibly treatments

and a vaccine, none of which are modeled here.

Several limitations of the analysis should be emphasized. First, there remains substantial parameter uncertainty about morbidity and mortality of COVID-19 and this paper provides only limited sensitivity analysis to parameters. Second, the cutoff for the two groups, the number of groups, and the timing of when to lift interventions for the more vulnerable group has not been optimized. Third, vulnerability should be determined by the oldest member of the household, not individual-by-individual. Finally, and most importantly, whether a strategy to achieve herd immunity is inevitable, and hence called for, much depends on the expectations about the feasibility and timing of potential treatments and vaccines, if any. But if lifting restrictions is advised, the age-dependence of the vulnerability to the disease should be a primary consideration.

References

- Allen, L. J. S. (1994). Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences* 124, 83–105.
- Atkeson, A. (2020). What will be the economic impact of COVID-19 in the us? rough estimates of disease scenarios. *NBER Working Paper No. 26867*.
- Chikina, M. and W. Pegdeny (2020). Modeling strict age-targeted mitigation strategies for COVID-19. *Working paper*.
- Dingel, J. I. and B. Neiman (2020). How many jobs can be done at home? *Working Paper*.
- Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020). The macroeconomics of epidemics. *NBER Working Paper No. 26882*.
- Ferguson, N. M., D. Laydon, G. Nedjati-Gilani, N. Imai, and et al. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Report of the Imperial College COVID-19 Response Team*.
- Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 115, 700–721.
- Lourenço, J., R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klennerman, and S. Gupta (2020). Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *Working Paper*.

Table 1: **Data and Calibration**

This table presents the data and assumptions for the calibration. Data sources: Data for the COVID-19 pandemic are from Ferguson et al. (2020). [Labor force participation data](#) for 2018 is from the U.S. Bureau of Labor Statistics. [Population data](#) for 2018 is from the U.S. Census Bureau.

Panel A: Data									
<i>Effects of COVID-19 Infection by Age Group</i> (Ferguson et al. (2020), Table 1)									
Age group (years)	0 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80+
Hospitalization rate	0.10%	0.30%	1.20%	3.20%	4.90%	10.20%	16.60%	24.30%	27.30%
Critical care rate	0.01%	0.02%	0.06%	0.16%	0.31%	1.24%	4.55%	10.50%	19.36%
Mortality rate	0.00%	0.01%	0.03%	0.08%	0.15%	0.60%	2.20%	5.10%	9.30%
Population	12.22%	12.83%	13.89%	13.38%	12.37%	13.11%	11.43%	6.93%	3.86%
<i>Labor Force Participation by Age Group</i>									
Age group (years)	0 to 15	16 to 24	25 to 34	35 to 44	45 to 54	55 to 64	65 to 74	75+	
Participation rate	na	55.2%	82.5%	82.9%	80.8%	65.0%	27.0%	8.7%	
Population	19.87%	11.87%	13.98%	12.62%	12.73%	12.93%	9.34%	6.68%	
Panel B: Calibration									
				Group 1	Group 2				
				(young)	(old)	Overall	Source		
Age group (years)				0 to 54	55+				
Population (ω_j)				71.06%	28.94%		U.S. Census Bureau		
<i>Effects of COVID-19 Infection</i>									
Hospitalization rate (symptomatic cases)				2.44%	19.13%	7.27%	Ferguson et al. (2020)		
Adj. for 50% asymptomatic cases (ϕ_j^h)				1.22%	9.56%	3.64%			
Critical care rate				0.16%	7.51%	2.29%	Ferguson et al. (2020)		
Adj. for 50% asymptomatic cases (ϕ_j^c)				0.08%	3.75%	1.14%			
Mortality rate				0.08%	3.65%	1.12%	Ferguson et al. (2020)		
Adj. to match Ferguson et al. (2020) (δ_j)				0.06%	2.67%	0.82%			
<i>COVID-19 Epidemiological Parameters</i>									
\mathcal{R}_0						2.4	Ferguson et al. (2020)		
Recovery rate γ (weekly)						7/18	Atkeson (2020)		
Baseline contact rate $\beta_{jk} = \mathcal{R}_0\gamma = \beta$						0.93			
Contact rate during intervention						0.25 β	Ferguson et al. (2020)		(see Table 2 for details)
<i>Labor Force Participation</i>									
Fraction of population 16 and older				63.88%	36.12%	100%	U.S. Census Bureau		
Participation rate (l_j)				75.85%	39.75%	62.81%	U.S. Bureau of Labor Statistics		
Fraction able to work at home (wage-weighted)				46%	46%	46%	Dingel and Neiman (2020)		

Table 2: **Policy Interventions**

This table describes the assumptions about the contact rates matrix

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

for the various lifting strategies. The analysis evaluates three strategies to lifting interventions. Two strategies treat the entire population homogenously, lifting the general intervention for both groups either early or late. The third strategy involves sequential lifting of interventions for group 1 (young) first and group 2 (old) later on. Each panel shows the contact rates before the general intervention (first column), during the general intervention (second column), during the sequential phase (if applicable) (third column), and after interventions are lifted for the entire population (fourth column). In the sequential phase (if any), contact rates differ across group 1 (young) and group 2 (old). Panel A provides the baseline calibration. Panels B and C consider sensitivity analysis. Panel B assumes the contact rate between group 1 and 2 is higher during the sequential phase. Panel C assumes interventions are less effective at reducing all contact rates. The effects of interventions on contact rates are comparable to those considered in Ferguson et al. (2020). The caption of Figure 1 describes the specific timing assumptions for each strategy.

Panel A: Baseline Calibration

	Before intervention	General intervention	Sequential lifting (if applicable)	After intervention
Homogenous lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.25\beta & 0.25\beta \\ 0.25\beta & 0.25\beta \end{bmatrix}$		$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$
Sequential lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.25\beta & 0.25\beta \\ 0.25\beta & 0.25\beta \end{bmatrix}$	$\begin{bmatrix} \beta & 0.25\beta \\ 0.25\beta & 0.25\beta \end{bmatrix}$	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$

Panel B: Sensitivity Analysis – Less Effective Sequential Intervention Phase

	Before intervention	General intervention	Sequential lifting (if applicable)	After intervention
Homogenous lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.25\beta & 0.25\beta \\ 0.25\beta & 0.25\beta \end{bmatrix}$		$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$
Sequential lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.25\beta & 0.25\beta \\ 0.25\beta & 0.25\beta \end{bmatrix}$	$\begin{bmatrix} \beta & 0.5\beta \\ 0.5\beta & 0.25\beta \end{bmatrix}$	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$

Panel C: Sensitivity Analysis – Overall Less Effective Intervention

	Before intervention	General intervention	Sequential lifting (if applicable)	After intervention
Homogenous lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.5\beta & 0.5\beta \\ 0.5\beta & 0.5\beta \end{bmatrix}$		$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$
Sequential lifting	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$	$\begin{bmatrix} 0.5\beta & 0.5\beta \\ 0.5\beta & 0.5\beta \end{bmatrix}$	$\begin{bmatrix} \beta & 0.5\beta \\ 0.5\beta & 0.5\beta \end{bmatrix}$	$\begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$

Table 3: Evaluation of Lifting of Interventions Strategies: Baseline Calibration

This table presents the quantitative results for the three strategies under the baseline calibration. For each strategy (Early Lifting; Sequential Lifting; and Late Lifting), the table reports the implications for mortality (deaths overall and for each group) (as percent of population), health system demand (peak hospitalization and critical care rate) (as percent of population), economic output loss (as percent of steady state output), as well as the fraction of the population that was never infected and that has recovered after two years (overall and for each group). The interventions are described in Panel A of Table 2. The caption of Figure 1 describes the specific timing assumptions for each strategy.

Strategy	Early Lifting	Sequential Lifting	Late Lifting
<i>Mortality</i>			
Overall deaths (%)	0.74%	0.46%	0.73%
Group 1 (young)	0.05%	0.05%	0.05%
Group 2 (old)	2.41%	1.47%	2.40%
<i>Health system demands</i>			
Peak hospitalization rate	0.86%	0.22%	0.83%
Peak critical care rate	0.27%	0.05%	0.26%
<i>Economic consequences</i>			
Output loss (%)	10.47%	16.01%	40.09%
<i>Never infected</i>			
Overall	9.81%	26.67%	10.10%
Group 1 (young)	9.81%	19.20%	10.10%
Group 2 (old)	9.81%	45.02%	10.10%
<i>Recovered</i>			
Overall	89.45%	72.87%	89.16%
Group 1 (young)	90.13%	80.75%	89.84%
Group 2 (old)	87.78%	53.51%	87.50%

Table 4: **Evaluation of Lifting of Interventions Strategies: Sensitivity Analysis**

This table reports sensitivity analysis of the quantitative results in terms of the effectiveness of the intervention (Panels A and B) and the assumption about the mortality rates. Panel A considers the sensitivity to a less effective sequential intervention, in which the contact rate between Group 1 (young) and Group 2 (old) is reduced by 50% only (see Panel B of Table 2 for details). Panel B considers the sensitivity to a less effective intervention overall, in which all contact rates are reduced by 50% only (see Panel C of Table 2 for details). Panel C considers the sensitivity to an alternative calibration of the mortality rate based on Lourenço et al. (2020): mortality rates δ_j in Table 1 are scaled to match their baseline infection-to-fatality ratio (IFR) of 0.14% (instead of 0.82% as in our baseline calibration). In all panels we report the mortality rates (as percent of population) for the three strategies (early, sequential, and late lifting of interventions) overall and for each group. In Panels A and B we also report the demands on the health system and the fraction never infected for the three strategies.

Panel A: Sensitivity Analysis – Less Effective Sequential Intervention Phase			
Strategy	Early Lifting	Sequential Lifting	Late Lifting
<i>Mortality</i>			
Overall deaths (%)	0.74%	0.50%	0.73%
Group 1 (young)	0.05%	0.05%	0.05%
Group 2 (old)	2.41%	1.60%	2.40%
<i>Health system demands</i>			
Peak hospitalization rate	0.86%	0.37%	0.83%
Peak critical care rate	0.27%	0.11%	0.26%
<i>Never infected</i>			
Overall	9.81%	25.02%	10.10%
Group 1 (young)	9.81%	18.85%	10.10%
Group 2 (old)	9.81%	40.14%	10.10%
Panel B: Sensitivity Analysis – Overall Less Effective Intervention			
Strategy	Early Lifting	Sequential Lifting	Late Lifting
<i>Mortality</i>			
Overall deaths (%)	0.74%	0.51%	0.68%
Group 1 (young)	0.05%	0.05%	0.05%
Group 2 (old)	2.41%	1.64%	2.23%
<i>Health system demands</i>			
Peak hospitalization rate	0.85%	0.41%	0.44%
Peak critical care rate	0.27%	0.12%	0.14%
<i>Never infected</i>			
Overall	9.84%	24.83%	16.35%
Group 1 (young)	9.84%	19.22%	16.35%
Group 2 (old)	9.84%	38.58%	16.35%
Panel C: Sensitivity Analysis – Mortality Rate based on Lourenço et al. (2020)			
Strategy	Early Lifting	Sequential Lifting	Late Lifting
<i>Mortality</i>			
Overall deaths (%)	0.13%	0.08%	0.13%
Group 1 (young)	0.01%	0.01%	0.01%
Group 2 (old)	0.41%	0.25%	0.41%

Figure 1: **Evaluation of Lifting of Interventions Strategies**

This figure displays the dynamics of mortality, the reproduction ratio, and economic output for three strategies over time. The population starts in week of January 5, 2020 (week 1) with 0.01% infected and the rest susceptible. The general intervention is implemented the week of March 15 (week 11). Column A considers early lifting of interventions for both groups the week of April 19 (week 16). Column B considers sequential lifting of interventions for group 1 (young) first (week of April 19; week 16) and group 2 (old) later (week of November 8; week 45). Column C considers a delayed lifting of interventions for both groups (week of November 8; week 45). The first row shows death per week (red; right axis) and cumulative deaths (blue; left axis), both as a fraction of the total population; the values in the overall population are solid, the values for group 1 (young) are dash-dotted and the values for group 2 (old) are dashed. The second row shows the reproduction ratio for the population overall $\mathcal{R}_0(t)$ (solid), for group 1 (young) $\mathcal{R}_{0,1}(t)$ (dash-dotted), and for group 2 (old) $\mathcal{R}_{0,2}(t)$ (old). The third row shows economic output of the population overall (solid), of group 1 (young) (dash-dotted), and of group 2 (old) (dashed).

Column A: Early Lifting

Column B: Sequential Lifting

Column C: Late Lifting

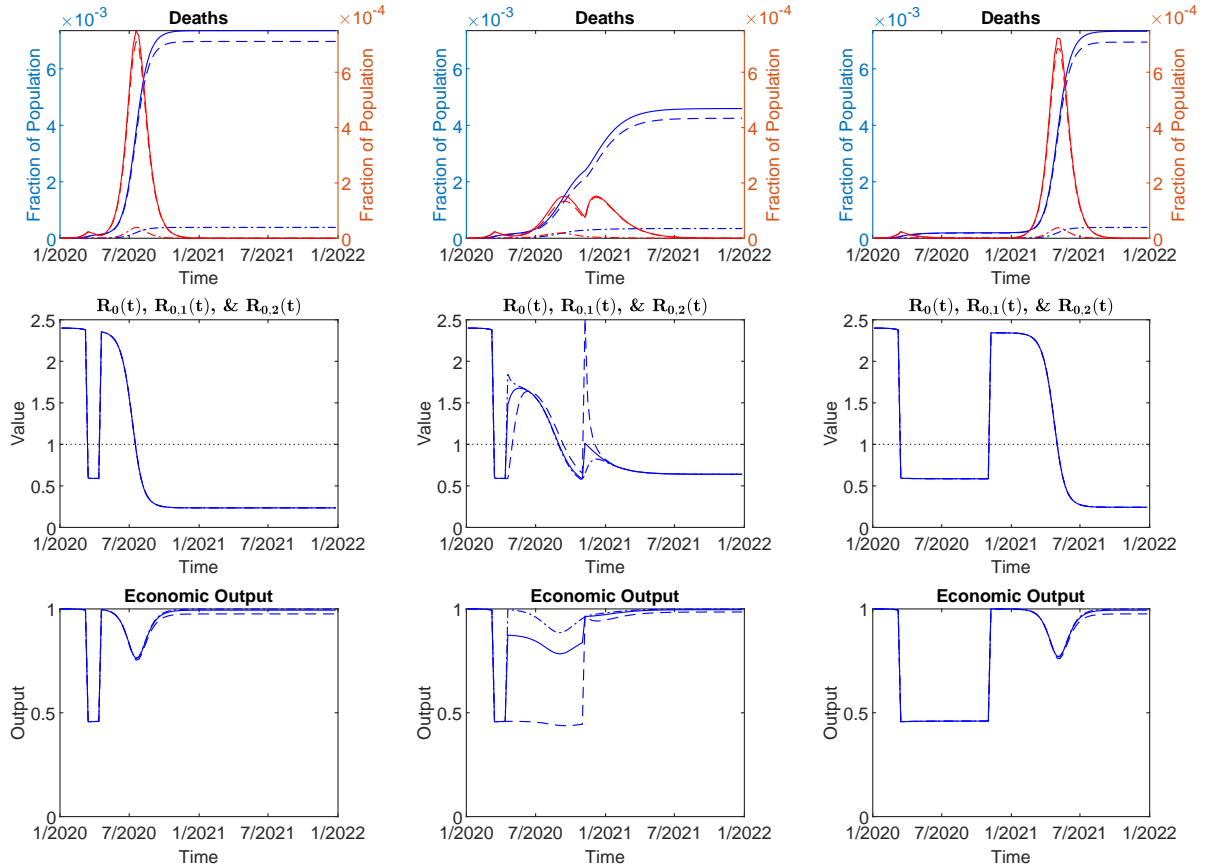
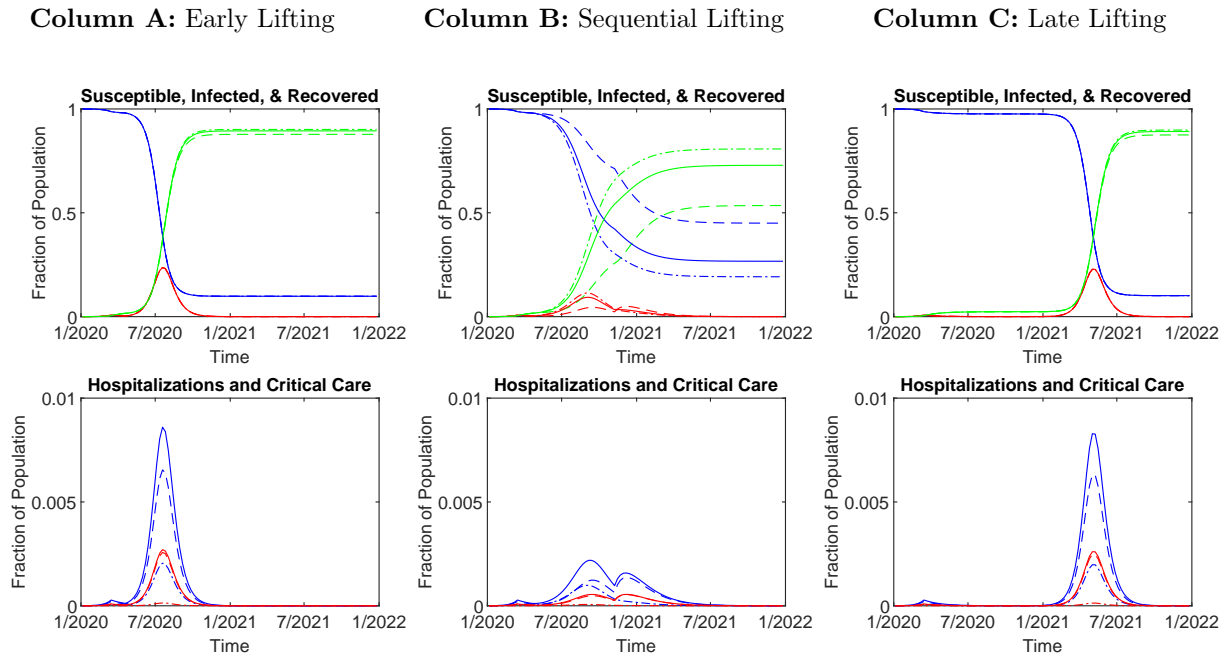


Figure 2: **Evaluation of Lifting of Interventions Strategies**

This figure displays the dynamics of the fraction of the population that is susceptible, infected, and recovered and the demands on the health system for the three strategies described in Figure 1 over time. Column A considers early lifting of interventions for both groups; Column B considers sequential lifting of interventions for group 1 (young) first and group 2 (old) later. Column C considers a delayed lifting of interventions for both groups. The first row shows the fraction susceptible (blue), infected (red) and recovered (green) as a fraction of the total population; the values in the overall population are solid, the values for group 1 (young) are dash-dotted and the values for group 2 (old) are dashed. The second row shows the hospitalization rate (blue) and the demand for critical care (red) for the population overall (solid), for group 1 (young) (dash-dotted), and for group 2 (old) (dashed).



Appendix A Basic SIR model without heterogeneity

This appendix describes the basic discrete-time SIR model without heterogeneity. Consider a population of size normalized to one. At each point in time, the population is comprised of fraction s of susceptible individuals, fraction i of infected individuals, and fraction r of recovered individuals. Susceptible individuals contact other individuals in the population at rate $\beta > 0$. Contact with an infected person, which happens with probability i , results in infection. Thus the fraction s' of the population susceptible next period is

$$s' = s - \beta si. \quad (\text{A1})$$

The current infected population i increases at a rate βs due to new infections while fraction γ of the infected recover; therefore, the infected population i' next period is

$$i' = i + \beta si - \gamma i. \quad (\text{A2})$$

Finally, the recovered population r this period grows due to the infected that recovered last period; the recovered population r' next period is therefore

$$r' = r + \gamma i. \quad (\text{A3})$$

Note that the infected population i grows at rate

$$\frac{i'}{i} - 1 = \beta s - \gamma = \left(\frac{\beta s}{\gamma} - 1 \right) \gamma.$$

Define the reproduction ratio $\mathcal{R}_0(t) = \frac{\beta s(t)}{\gamma}$, where $s(t)$ is the fraction susceptible at time t . For a novel disease to which the entire population is susceptible, $s(0)$ is approximately one, and so $\mathcal{R}_0 \equiv \mathcal{R}_0(0) \approx \frac{\beta}{\gamma}$, often referred to as the basic reproduction number.⁴ If $\mathcal{R}_0(0) > 1$, the infection will initially spread, but the growth rate of the infected population $\mathcal{R}_0(t) = \frac{\beta s(t)}{\gamma}$ decreases as the susceptible population $s(t)$ decreases.

To account for deaths separately, we can easily modify the model by assuming that γ is the fraction of infected that recovers or dies per period, of which fraction δ dies (while fraction $1 - \delta$ recovers); the modified law of motion for the fraction recovered (A3) is

$$r' = r + (1 - \delta)\gamma i. \quad (\text{A4})$$

The modified law of motion for the fraction of the population that has died is

$$d' = d + \delta\gamma i. \quad (\text{A5})$$

The laws of motion for the fraction susceptible (A1) and fraction infected (A2) and the definition of reproduction ratio $\mathcal{R}_0(t)$ are unchanged as they do not depend on δ .

⁴If the entire population is susceptible, the basic reproduction number is the expected number of individuals an infected person infects before recovering, as $\beta(1 + (1 - \gamma) + (1 - \gamma)^2 + \dots) = \frac{\beta}{\gamma} = \mathcal{R}_0$.