The timing of one-shot interventions for epidemic control

Francesco Di Lauro[†], István Z. Kiss[†], and Joel C. Miller[‡*]
[†] Department of Mathematics, School of Mathematical and Physical Sciences,
University of Sussex, Falmer, Brighton BN1 9QH, UK

[‡] Department of Mathematics and Statistics
School of Engineering and Mathematical Sciences
La Trobe University, Bundoora, Australia

February 2020

Abstract

The apparent early success in China's large-scale intervention to control the COVID-19 epidemic has led to interest in whether other countries can replicate it as well as concerns about a resurgence of the epidemic if or when China relaxes the interventions. In this paper we look at the impact of a single short-term intervention on an epidemic. We see that if an intervention cannot be sustained long-term, it has the greatest impact if it is imposed once infection levels have become large enough that there is an appreciable number of infections present. For minimising the total number infected it should start close to the peak so that there is no rebound once the intervention is stopped, while to minimise the peak prevalence, it should start earlier, allowing two peaks of comparable size rather than one very large peak. In populations with distinct subgroups, synchronized interventions are less effective than targeting the interventions in each sub-population separately.

We do not attempt to clearly determine what makes an intervention sustainable or not. We believe that is a policy question. If an intervention is sustainable, it should be kept in place. Our intent is to offer insight into how best to time an intervention whose impact on society is too great to maintain.

1 Introduction

The Influenza pandemic of 1918 was one of the deadliest epidemics of infectious disease the world has ever seen. In response, many cities introduced widespread interventions intended

*joel.c.miller.research@gmail.com

to reduce the spread. There is evidence [3] that those cities which implemented these interventions later had fewer deaths. This seemingly counter-intuitive observation suggests that those cities which were slow to respond were the most successful.

A similar pattern can be observed for the 2009 H1N1 influenza pandemic. When it was first recognized, there were many cases present in Mexico City. After a 2 week school closure with many additional interventions in place, the number of cases in Mexico City dropped dramatically and did not significantly increase after relaxing the controls.

As the disease was introduced into other geographic regions, many schools closed in response to the first observations of infection. Once these schools reopened, they were likely to eventually have a new introduction. The remaining susceptible population was almost as large as at the outset, so new epidemics were likely to be as large as the original epidemics would have been. The school closure provided increased time to prepare, but the overall epidemic was very similar.

We can understand these historical patterns by observing that some of the most drastic social distancing interventions are unsustainable. We will refer to these as "one-shot", meaning that the intervention cannot be maintained indefinitely or repeated. The theoretical work contained in [3] showed that if an intervention was only temporary, it was generally more effective if introduced later in the epidemic (but not too late).

The phenomenon can be explained by noting that epidemics rely on two things to spread: infected individuals and a supply of susceptible individuals. If the intervention is too early, the number infected may fall, but there will be enough susceptibles available that it can re-establish and grow again; the intervention only delays the spread. However, if the intervention occurs once the susceptible population has been noticeably depleted, then the number of infections falls quickly and it may not be able to rebound.

In the ongoing COVID-19 epidemic, China has introduced drastic control measures. These appear to have significantly reduced transmission, apparently reducing the effective reproduction number (the number of new infections per infected individual) very close or below one [6, 10], although there is still a lot of uncertainty about the effectiveness of control measures [5].

Similarly, in response to observed infections in some regions Japan has turned to widespread school closures. Other counties are implmenting or considering similar interventions. Some of these control measures are likely unsustainable. So they will be relaxed eventually. It is unclear what will happen if or when the interventions are relaxed. China, in particular, faces a challenge. It may be that in Wuhan, the epicenter of the initial spread, the susceptible fraction is sufficiently depleted that the disease is unable to return (although the apparent fraction infected thus far is quite small, so this seems unlikely). In other regions of China which have apparently been quickly brought under control with only a small number of transmissions it seems likely the the disease may easily return if the measures are relaxed fully.

Motivated in part by current decisions facing policy makers for the COVID-19 epidemic, we develop mathematical models which allow us to explore how to time short-term interventions in response to an emerging epidemic. We are particularly interested in how the timing

of the intervention might affect the total fraction infected and the peak prevalence, but we are also interested in the resulting delay of the epidemic.

We model an infection introduced into a fully susceptible population. We make an assumption that is appropriate for an established pandemic that eliminating infection soon after it is introduced to the population will not prevent future reintroductions. So we do not focus our attention on the possibility of eliminating the disease. We will investigate the impact of intervention on the timing of epidemics, the peak fraction infected, and the final fraction infected (in the absence of additional interventions).

Our results have important implications for the ongoing COVID-19 epidemic. If an intervention definitely cannot be sustained for an extended period of time, then it is best if it is "held in reserve" until depletion of susceptibles has reduced the effective reproductive number enough that the one-shot intervention will have maximal impact. However, we must exercise care in determining that an intervention cannot be sustained. The uncertainty about the case fatality rate remains high [9], and at the higher end of the plausible values, the tolerance of the population for drastic interventions may be significant. Thus what might appear to be an unsustainable intervention may in fact be sustainable if we have a better understanding of the case fatality rate. An additional consequence of our results which is applicable to sustainable interventions is that the expression "better late than never" applies quite strongly for interventions.

In this paper, we first introduce the mathematical models we use to explore the impact of a one-shot intervention against an infectious disease in a single well-mixed population and a metapopulation made up of several distinct sub-populations. Then we discuss results from those mathematical models. Finally we discuss the implications of these results. A particularly important point to raise is that for an infection like COVID-19 with a relatively high proportion of severe cases, we need to think carefully about an intervention before we decide that it is not sustainable. In the Appendix we develop some mathematical theory explaining the mechanism underlying the effect in more detail.

Our goal is not to provide specific predictions for a specific population, but rather to demonstrate the generic impact of delaying a one-off intervention and to show its robustness.

2 Methods

In this section we introduce mathematical models for an "SIR" (Susceptible–Infected–Recovered) epidemic in a single well-mixed population and in a metapopulation made up of several sub-populations. We assume that the intervention is initiated at a specific time t^* (typically once the cumulative number of infections $I + R$ reaches some threshold), and that the intervention lasts for a fixed duration D. In the metapopulation model, we compare outcomes when the intervention is implemented in all populations at the same time or in each individual population separately.

2.1 Well-mixed population

To study an intervention in a well-mixed population, we use the standard SIR model [1].

$$\begin{aligned} \dot{S} &= -\beta IS, \\ \dot{I} &= \beta IS - \gamma I, \\ \dot{R} &= \gamma I, \end{aligned} \tag{1}$$

where S, I, and R denote the susceptible, infected and recovered fractions of the population with $S + I + R = 1$. There are a few important quantities to consider.

- The *basic reproduction number* $\mathcal{R}_0$: The typical number of infections an infected individual causes early in the epidemic in the absence of intervention and the absence of any depletion of susceptibles. This is $\mathcal{R}_0 = \beta/\gamma$.
- The *effective reproduction number* $\mathcal{R}_e$: As depletion of susceptibles occurs or interventions are put into place, the number of infections an infected individual causes is reduced. When $\mathcal{R}_e < 1$, the number of infections declines.

If $\mathcal{R}_0 > 1$ the typical behavior of an epidemic without an intervention is that at $t = 0$ we have $S \approx 1$, I is very small and $R = 0$. As time increases, I and R grow and S decreases. The reduction in S reduces the effective reproduction number: $\mathcal{R}_e = \mathcal{R}_0 S$. Once $S < 1/\mathcal{R}_0$, I begins to fall as well as recoveries outweigh new infections: $I \to 0$. Some fraction remain uninfected: $S(\infty) > 0$ and $R(\infty) = 1 - S(\infty)$, see figure 1 for typical profiles of S, I and R in time. We will measure time in multiples of the typical infection duration. Making this assumption means that $\gamma = 1$ and $\beta = \mathcal{R}_0$.

We assume that at some time $t = t^*$, a social distancing intervention is introduced with duration D. The intervention reduces β by some factor c. So from time $t = t^*$ to time $t = t^* + D$ the transmission rate $\beta = \mathcal{R}_0$ is replaced by $\beta = (1 - c)\mathcal{R}_0$. During the intervention, the effective reproduction number is $\mathcal{R}_e = S(1 - c)\mathcal{R}_0$. After time $t = t^* + D$ the transmission rate returns to $\beta = \mathcal{R}_0$, and $\mathcal{R}_e = S\mathcal{R}_0$.

We will typically assume that t^* is chosen based on the cumulative number of infections $I(t) + R(t)$ crossing some threshold.

We will measure three quantities of interest:

- the *attack rate* or total fraction infected $R(\infty)$,
- the *peak prevalence* or maximum value of $I(t)$, and
- the *time of peak*, or t_p at which I is maximized.

In general the goal of our intervention is to reduce $R(\infty)$, reduce $I(t_p)$, and increase t_p.

One detail that needs attention in interpreting later results is that if an intervention is put into place before $S = 1/\mathcal{R}_0$, but after the number of infections is large, it may happen that the intervention causes an immediate decline in I, and that I never recovers to the level it

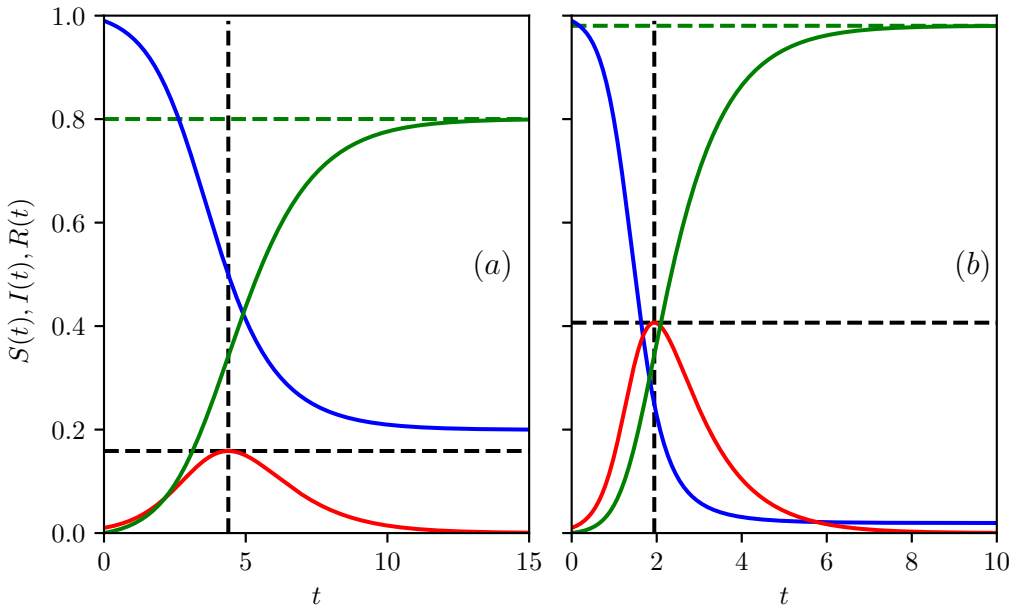

Figure 1: The time-evolution of S, I and R for epidemics with no control. (a) $\mathcal{R}_0 = \beta = 2$ and (b) $\mathcal{R}_0 = \beta = 4$ with $\gamma = 1$ in both. Horizontal and vertical dashed black lines indicate the peak prevalence $I_{\max}$ and time of the peak t_p, respectively, while green dashed horizontal lines show the attack rate $R(\infty)$ found by numerically solving $R(\infty) = 1 - S(0)e^{-\mathcal{R}_0 R(\infty)}$.

was at before the intervention. In this case, we would see that $t_p = t^*$, which corresponds to an earlier peak. This seeming acceleration in the time of the peak is actually a consequence of the intervention reducing or eliminating the later peak prevalence. We should not interpret a reduction in t_p due to this effect as a failure of the intervention.

2.2 Metapopulation model

We will also investigate the effectiveness of interventions in a metapopulation made up of distinct sub-populations. We are particularly interested in whether it is better to time the intervention to each sub-population separately or whether it is best for the intervention to be synchronized.

It is well-known that if the sub-populations have strong enough coupling, the epidemics in all sub-populations are effectively synchronised [2, 4]. In this case the time-evolution of the outbreaks in the various sub-populations are similar to that of the fully mixed population, and there is little distinction between interventions based on individual sub-population or at the global-level.

Thus to compare whether a global intervention is comparable to an individually targeted intervention and understand the magnitude of the difference, we need to explore a population with weak coupling. The model we will use is a standard meta-population model [1], allowing

most transmission to be within a sub-population and some cross interactions between the sub-populations.

$$\begin{aligned}\dot{S}_j &= -\sum \beta_{ij} I_i S_j,\\ \dot{I}_j &= \Big(\sum \beta_{ij} I_i S_j\Big) - \gamma I_j,\\ \dot{R}_j &= \gamma I_j,\end{aligned}$$

where $0 \leq S_i \leq 1$, $0 \leq I_i \leq 1$ and $0 \leq R_i \leq 1$, with $(S_i + I_i + R_i)(t) = 1$ for all $t \geq 0$, represent the fraction of susceptible, infected and infectious and recovered individuals in sub-population i, where $i = 1, 2, \ldots, N$.

To simplify the presentation, all sub-populations are of equal size. The recovery rate γ is identical for all populations. As before we measure time in multiples of the typical infectious period, so we set γ to 1. The cross-infection between sub-populations is modelled by $B = (\beta_{ij})_{i,j=1,2,\ldots N}$, where β_{ij} represents the rate at which infectious contacts are made from sub-population i towards susceptible individuals in sub-population j.

We implement a weak coupling by joining the population in a linear fashion: population i is only connected to population $(i-1)$ and $(i+1)$. The first and the last populations only connect to the second and the pen-ultimate population, respectively. The entries for the coupling/mixing matrix are generated as follows. On the main diagonal, the β_{ii} values are set to $2 + (Unif(0,1) - 0.5)$. Off-diagonal entries are set to $Unif(0,1)(\beta^*_{ii}/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) and represent a scaled and randomised version of the largest entry on the main diagonal. This yields an $\mathcal{R}_0$ above 2, comparable to current estimates for COVID-19 [6, 7].

We will use this model to explore whether it is better to implement an intervention in a synchronized fashion across all sub-populations or whether it is better to implement it in each sub-population. In particular, we will consider the following scenarios:

- track $(I_i + R_i)(t)$ in each sub-population and as soon as $(I_i + R_i)(t) > Tr_i$, a one-shot control is deployed in the corresponding sub-population,
- track $(I + R)(t) = \frac{1}{N}\sum_{i=1}^{N}(I_i + R_i)(t)$ globally and as soon as $(I + R)(t) > Tr$, a one-shot control is deployed across all sub-populations, and
- track each sub-population and deploy the one-shot control is deployed across all sub-populations as soon as threshold for $(I + R)(t)$ is breached in any of the individual epidemics.

One-shot control in a sub-population is understood to mean reduction in the internal, incoming, and outgoing rates of infection with a factor of $(1 - c)$, where $0 \leq c \leq 1$ denotes the efficiency of the intervention in sub-population i. This reduction lasts for a duration D and, as soon as the control is over, the transmission rates for that sub-population are restored to the starting levels.

3 Results

We use our mathematical models to demonstrate how the timing of an intervention can impact

- total attack rate,
- peak prevalence, and
- time of peak.

These are expected to be good measures of the total impact on the population or the burden on the health services.

We find that one-shot interventions that begin at the first sign of infection have little effect on the final outcome beyond delaying it. This is because the intervention blocks few transmissions because only a few individuals are infected when the intervention is implemented. When the restrictions are lifted, the disease eventually spreads again in an almost fully susceptible population, and its trajectory is for all practical purposes the same, just delayed. In contrast if the intervention is delayed to start closer to the peak of the epidemic it will be more effective.

For a weakly-coupled metapopulation model, the subgroups are likely to have somewhat asynchronous epidemics. In this case it is better to implement the one-shot interventions based on a local threshold rather than a global threshold. If the coupling is stronger, the epidemics are closely synchronized and there is little difference between the strategies.

3.1 Well-mixed population

For a well-mixed population we find that the timing of a one-shot intervention has an important impact on the effectiveness of an intervention the peak prevalence and total fraction infected. If the intervention is put in place early, then the impact on the peak is to delay it. If the intervention is put into place later, then the we may artificially see that the time the intervention is implemented becomes the new peak since the later peak is either dramatically reduced or eliminated altogether.

Figure 2 shows the impact of an intervention in a population with $\mathcal{R}_0 = 2.5$ and an intervention of strength $c = 4/5$ (it prevents 4 of every 5 transmissions), and duration 2 (time units measured in multiples of the typical infection duration). The figure focuses on the impact of varying the threshold value of $I + R$ at which the intervention is introduced.

Figure 3 shows how the optimal threshold changes as the parameters of the disease or intervention change.

3.1.1 Impact on attack rate

If the intervention is introduced early on, it will have an immediate impact. However, when the intervention is lifted, there are still many susceptibles around, so the epidemic can rebound. Even if the disease is eliminated locally, in a pandemic setting it will eventually

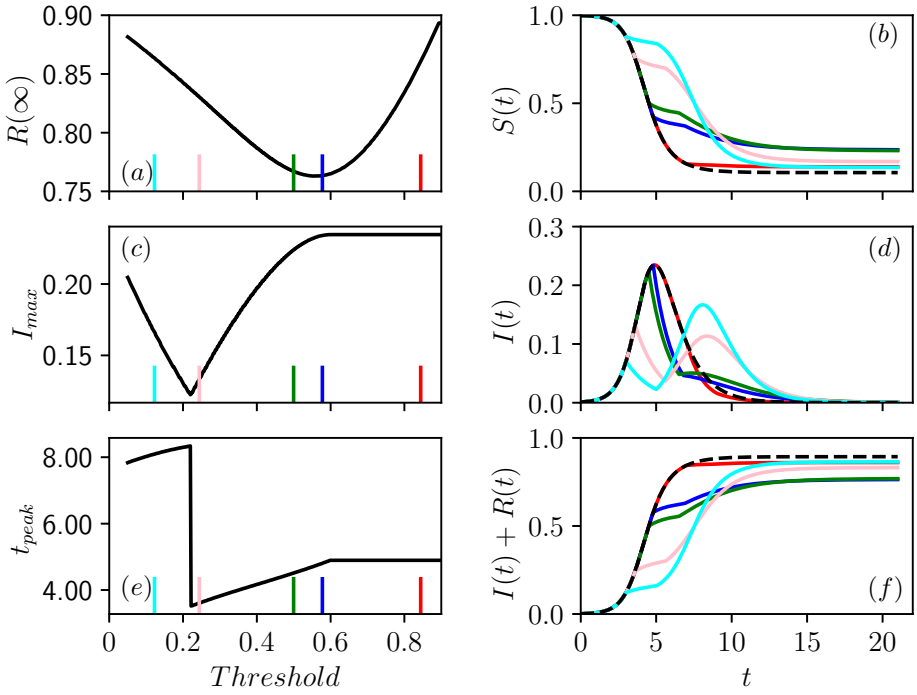

Figure 2: Illustration of the impact of one-shot intervention in a population with $\mathcal{R}_0 = 2.5$. The intervention has $c = 0.8$ for a duration of $D = 2$ time units. This intervention is introduced at different times as determined by a range of *Threshold* values. The impact of the threshold $I + R$ for implementing the intervention is shown for (a) the attack rate $R(\infty)$; (b) $S(t)$; (c) time of the peak prevalence; (d) $I(t)$; (e) time of peak; and (f) plots of $I(t) + R(t)$. In (b,d,f), the no-control case is plotted as a dashed line. The vertical lines in (a,c,e) correspond to the threshold for cumulative infections $I + R$ which yields the intervention leading to the corresponding color in (a,c,e).

return to the population. The models predict an almost identical epidemic curve once it rebounds, except with a shift to later time. So an early intervention has little impact on the attack rate $R(\infty)$.

In Figure 2(a) we see that the intervention is introduced later, we see a clear improvement in $R(\infty)$, up to a threshold of $I+R$ of 0.6, which is close to where the peak prevalence in the epidemic without intervention. As a general rule of thumb to reduce the attack rate $R(\infty)$, the intervention is most effective if it eliminates as many transmissions as possible. So we want to time the epidemic to maximize the number of infected individuals present during the intervention (Mathematically we want to maximize $\int_{t^*}^{t^*+D} I(\tau)d\tau$ given D, c, and $\mathcal{R}_0$).

Thus the ideal timing is not at the first hint of infection (when there are not enough infected individuals to cause many transmissions), but rather, close to the peak. This suggests that the more effective an intervention is, the closer we should be to the peak before

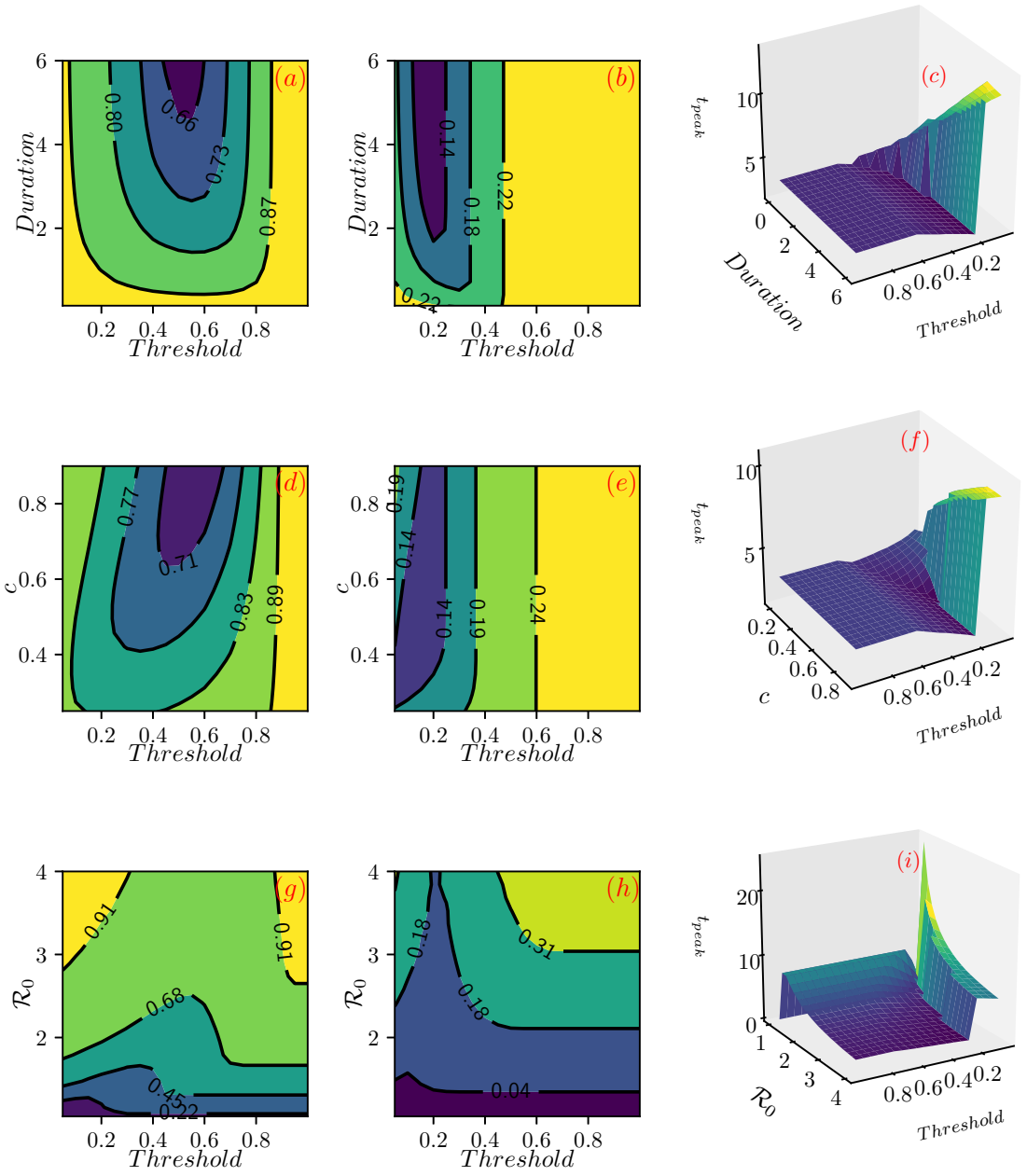

Figure 3: Contour plots for $R(\infty)$, I_{max} and surface plots for t_p as a function of parameters for the well-mixed population. We explore different threshold values of $I+R$ for the intervention to start, from a minimum of 0.05 to a max of 0.9. The first row investigates impact of duration from $D = 0.1$ to $D = 6$, holding$\beta = 2.5$ and $c = 0.8$. On the second row, intervention duration is $D = 4$ and the c ranges from 0.2 to 0.9. Finally, on the third row, $c = 0.8$ and $D = 4$, and the values of $\beta = \mathcal{R}_0$ vary from 1 to 4. In all cases $\gamma = 1$.

implementing it. It also suggests that for a longer duration intervention, we can implement it somewhat sooner, but not significantly sooner. If the infectiousness of the disease goes up, the need to begin the intervention near the peak implies that the threshold value of $I + R$ will need to be larger (though the time t^* at which it is implemented is smaller).

These predictions are borne out by observations of the first column of Figure 3 which shows how the optimal threshold value of $I + R$ for implementing the intervention changes as the the efficiency c, the duration D, or the reproductive number $\mathcal{R}_0$ change.

3.1.2 Impact on peak prevalence

As in the attack rate case, an early intervention only delays the epidemic curve. It does not significantly alter the shape. Thus the peak prevalence remains effectively the same.

If the susceptible population has been significantly depleted prior to the elimination of the intervention, then once the intervention is stopped, the epidemic rebound will be significantly muted. Moving the intervention later makes the rebound smaller. However, it means that the number of infections prior to the intervention is larger. There comes a point at which it can no longer rebound to the level of infection prior to the intervention. Delaying the intervention past this value results in a larger pre-intervention peak.

Figure 2(c) shows the optimal threshold to reduce the peak prevalence occurs sooner than to reduce the attack rate. For optimizing peak prevalence, having a moderate sized rebound is less of a concern than for optimizing the attack rate. For the purpose of reducing peak prevalence, the optimal time to introduce the intervention is when the current prevalence matches the peak prevalence that would occur once the disease rebounds.

We can crudely estimate the threshold necessary for minimizing the peak prevalence. If we know a population's reproductive number $\mathcal{R}_0$ and its initially immune fraction R^* and susceptible fraction $S^* = 1 - R^*$, we can determine the peak prevalence.[1] In the limit of a very long ($D \to \infty$) and strong intervention ($c \to 1$), at the end of the intervention $S(t^* + D) \approx S(t^*)$ and $R(t^* + D) \approx I(t^*) + R(t^*)$. This will be a reasonable estimate as long as D is longer than a typical infection, even if c is not very close to 1.

We can use this to estimate when $I(t^*)$ will approximate the rebound. As this is not strongly dependent on duration, or c, this explains why the optimal threshold for peak size does not vary much in Figures 3(b) and (e). We do not currently have an explanation for the fact that it remain similar as $\mathcal{R}_0$ changes in Figure 3(h).

3.1.3 Impact on timing of epidemic

As we noted for the attack rate and the peak prevalence discussions, an early intervention simply shifts the epidemic curve to later, which increases t_p. For very early interventions, the shift in the peak time is the regardless of the threshold[2]. However, as the threshold

[1] There is actually an analytic formula for this $1 - \frac{1}{\mathcal{R}_0} - R^* - \frac{\ln S^* \mathcal{R}_0}{\mathcal{R}_0}$ but for our purposes we just need to recognize that $\mathcal{R}_0$, R^* and S^* are sufficient to determine the peak prevalence.

[2] of course if the disease is eliminated locally which is more likely with a small threshold, then the prediction of the next peak depends on frequency of reintroduction. We do not consider this here

increases and we start to see an impact on $R(\infty)$ and $I_{\max}$, there is also an impact on t_p. As with the other targets, a later intervention tends to have an increased impact. The peak time t_p moves later as the start time increases.

There is an abrupt change that occurs for later interventions when the epidemic is no longer able to rebound to the same number of infections seen prior to the intervention. In this case, the peak becomes the time at which the intervention is implemented. This appears in Figure 2 as a an abrupt drop in t_p. It rises again and eventually settles to a constant value, corresponding to the intervention being introduced after the natural epidemic peak.

In a real-world context, we anticipate that the model may overstate the delay from a very early intervention if there is significant transmission outside the population of interest. In a setting where the disease is spreading well outside the population, the reduction of infections within the population during the intervention is immediately negated by new transmissions from outside, so the delay of the peak is effectively equal to the duration of the intervention. However, in a setting where the disease is not yet well-established outside the population (as occurred in China early in the COVID-19 epidemic), a major effort at early time may significantly delay the eventual epidemic.

3.2 Metapopulation model

We now consider a more realistic population which consists of coupled sub-populations, so effectively a metapopulation model. The most obvious reason for this setup could be location/geographic or by age, but other alternatives exist including religion, ethnicity or socio-economic status. We again consider one-shot interventions that either target the entire population at once or that target individual sub-populations at different times. A typical plot of the prevalence level in each sub-populations is shown in figure 4 in the absence of intervention. The epidemic starts in sub-population two but it then spreads to all the other.

The overall impact of these interventions is qualitatively similar to that of the single-population model. We find that if the epidemics in the sub-populations are not synchronous (as occurs if the coupling is small [2, 4]) then interventions acting at different times for each sub-population are substantially more effective than interventions responding to a global threshold. If the coupling is larger so that epidemics are synchronized, then there is no distinction between these strategies.

As before we consider the impact of intervention on attack rate, peak prevalence, and peak timing. We find that an intervention which is targeted at each sub-population individually significantly outperforms synchronized interventions that either begin when the first sub-population reaches a threshold or the global infection crosses a threshold.

3.2.1 Impact on attack rate

The smallest values of the attack rates are achieved when control acts independently in each sub-population meaning that as soon as $(I_i + R_i)(t)$ is larger than the threshold, the one-shot control is switched on in sub-population i. This is true independently of whether the efficacy or duration of control is kept fixed, while the other is varied, see figure 5(a,d).

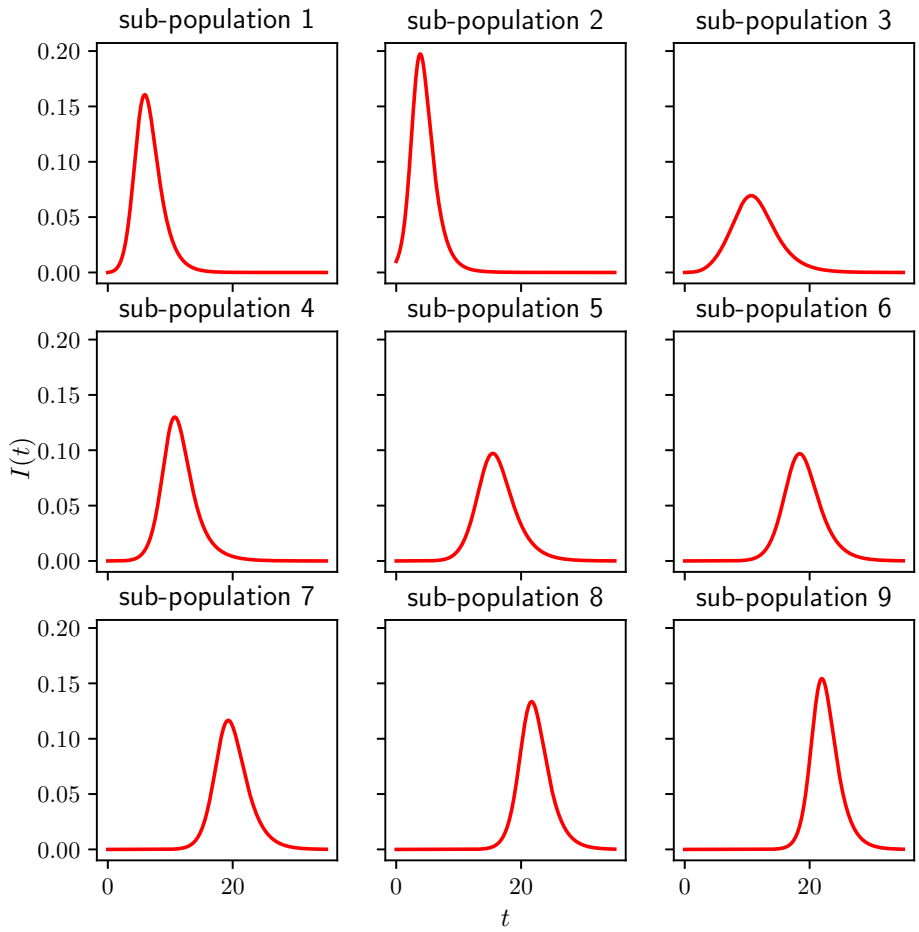

Figure 4: Example of an epidemic spreading across 9 identical populations with different contact rates, see Appendix for the precise mixing matrix B). The epidemic starts from sub-population 2 and it is run for $T = 35$ units of time. $\gamma = 1$ for all sub-populations. With no control the attack rate or final epidemic size is 0.744.

The impact of the threshold is qualitatively similar. Typically, as in the case of a single population, there seems to be a clear optimal threshold value which leads to the smallest final epidemic size. Applying the control too early or too late leads to higher attack rates. Fixing the threshold value and increasing the duration of control, see Figure 5(a), or the efficacy of control [Figure 5(d)]leads to smaller attack rates. Both these effects are nullified if the intervention is too early or too late.

The impact of the intervention based on the global level of $(I + R)(t)$ [see figure 5(b,e)] or on the first sub-population to reach a threshold [see figure 5(c,f)] are similar, but qualitatively different compared to the individually-targeted intervention. First of all, there a is a monotonic but slow decrease in the attack rate with increasing values of the threshold, duration of control and efficacy. However, the decrease is modest and this is due to the fact that when control is deployed the epidemics in some sub-populations may have effectively

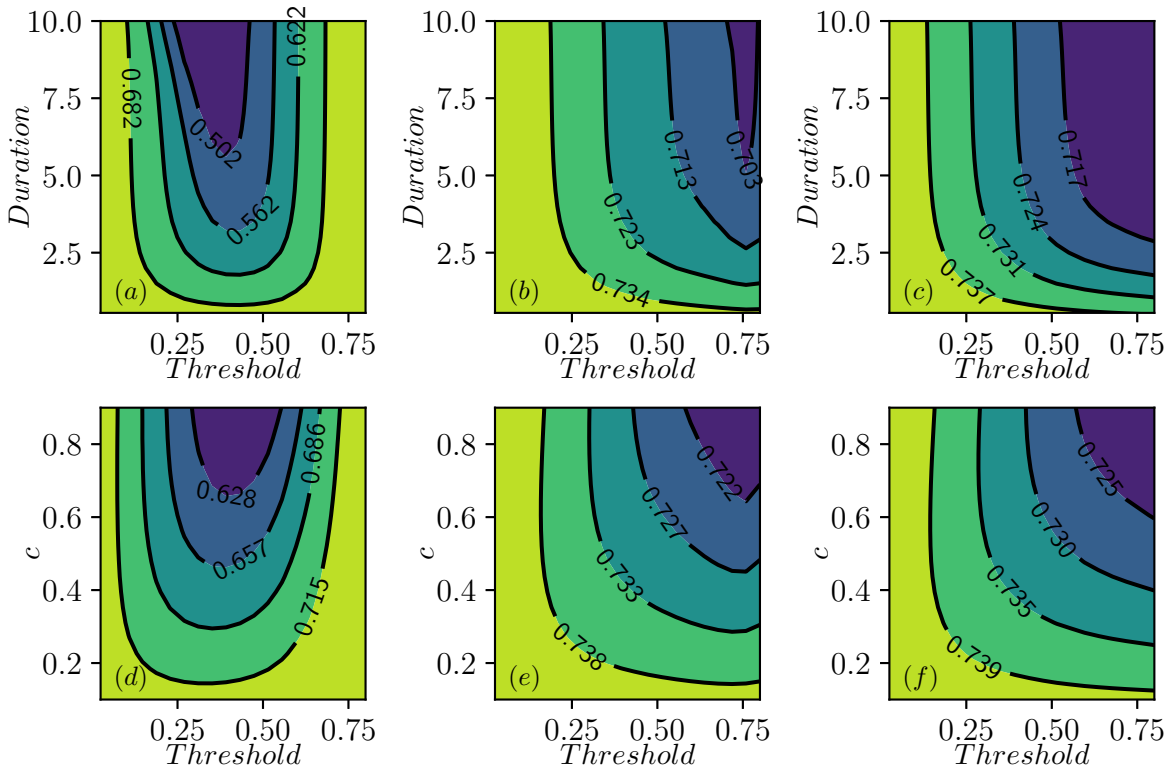

Figure 5: Contour plots showing the attack rate (final epidemic size) under different conditions. In the first row c is fixed and the duration of control varies on the vertical axis, while in the second row duration is fixed and c varies. Each column corresponds to one of the three strategies: (a,d) intervention in each sub-population, (b,e) global intervention at global threshold, and (c,f) global intervention when the first sub-population breaches the threshold) for a population consisting of 9 sub-populations. In each plot, the $x-axis$ shows the values that the threshold for intervention can take (from a minimum of 0.05 to a maximum of 0.8). In the first row $c = 0.8$ is constant, while the duration of control varies from a minimum of $T = 1$ to a maximum of $T = 10$. On the second row instead, the duration of control is kept fixed at $T = 2$, and the values of c varies from $c = 0.1$ to $c = 0.9$. The mixing matrix B is given in the Appendix, and $\gamma = 1$ for all sub-populations.

finished or not even started (see the asynchrony in Figure 4). The situation is very similar when control is based on the first population to reach a threshold, see figure 5(c,f). Finally, it is clear that the biggest impact on the attack rate still comes from the value of the threshold. Larger values of the threshold mean that control is more likely to hit more of the individual epidemics as they become established at different times. The duration of control and its efficacy only really matter at lower values.

Finally, in figure 6 we show how the best one-shot control works when the optimal threshold for fixed control efficacy and duration is implemented. As expected, this plot confirms that intervention happens close to the peak of the epidemic in each sub-population

and secondary waves of infection do not occur.

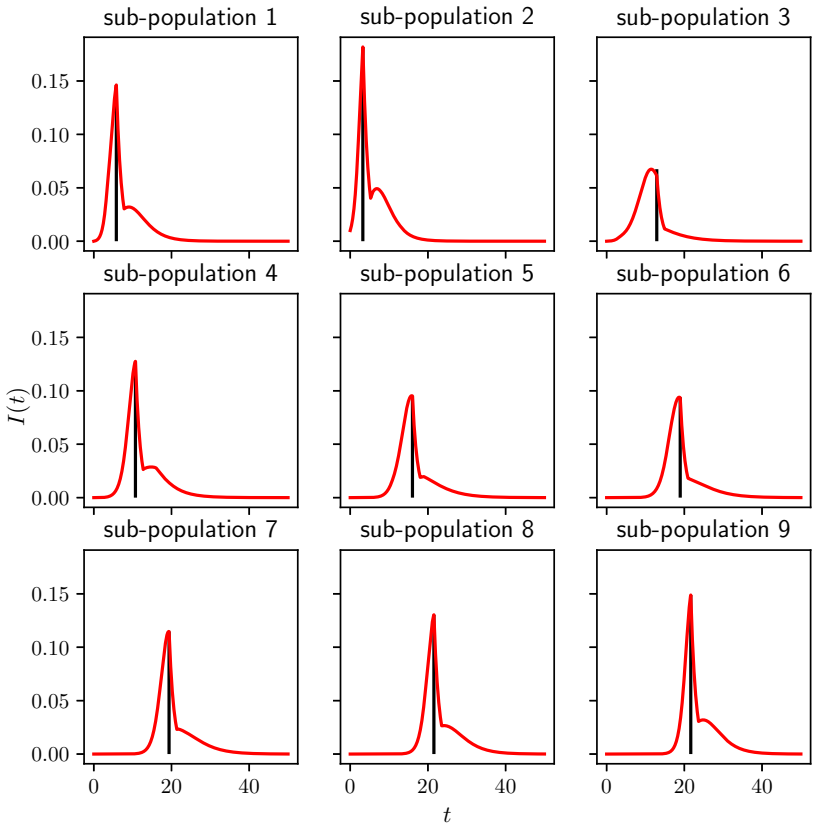

Figure 6: Illustration of best control strategy (i.e. smallest attack rate) (controlling sub-populations individually but using the same threshold for each) when efficacy and duration of control are fixed at $c = 0.8$ and $D = 2$, respectively. It turns out that the optimal threshold is close to (0.4). This combination represents the point $(0.4, 2)$ in Fig 5 panel (a), or equivalently the point $(0.4, 0.8)$ in panel (d). With this strategy, we find that $R(\infty)$ goes from $R(\infty) = 0.75$ to $R(\infty) = 0.63$. If we increase control duration from 2 to 10 we would achieve a further reduction to $R(\infty) = 0.44$. The vertical black line show the onset of control.

3.2.2 Impact on peak size

Here we look at the effect of the intervention in terms of the peak prevalence, that is the maximum value achieved by $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time course of the epidemics. Perhaps not surprisingly, figure 7 is qualitatively similar to figure 5. In figure 7 (a,d) the optimal threshold for intervention is evident and is shifted to the left, this is in line with the trend observed in figure 2 for the one single population case. Here we notice that different strategies lead to different optimal choices for intervention, as we observed when we studied the attack rate. If each sub-population is followed locally and the reduction in peak is the

main target then it is best to start the intervention with some delay, rather than immediately enforce it as soon as a few cases emerge. This is contrary to our other two strategies where following the global epidemic or the first local outbreak means that in many other sub-populations the intervention will be too early to have a significant impact. Thus, in line with our previous observation when looking at reduction in the attack rate, targeting each sub-population locally appears to be the overall best strategy to reduce the peak prevalence. We also notice that, if interventions are global, the duration of control has a limited impact on the overall peak, unless it is started at a later stage of the epidemic. Finally, in all the cases it appears that the strength of control is not as significant as its duration.

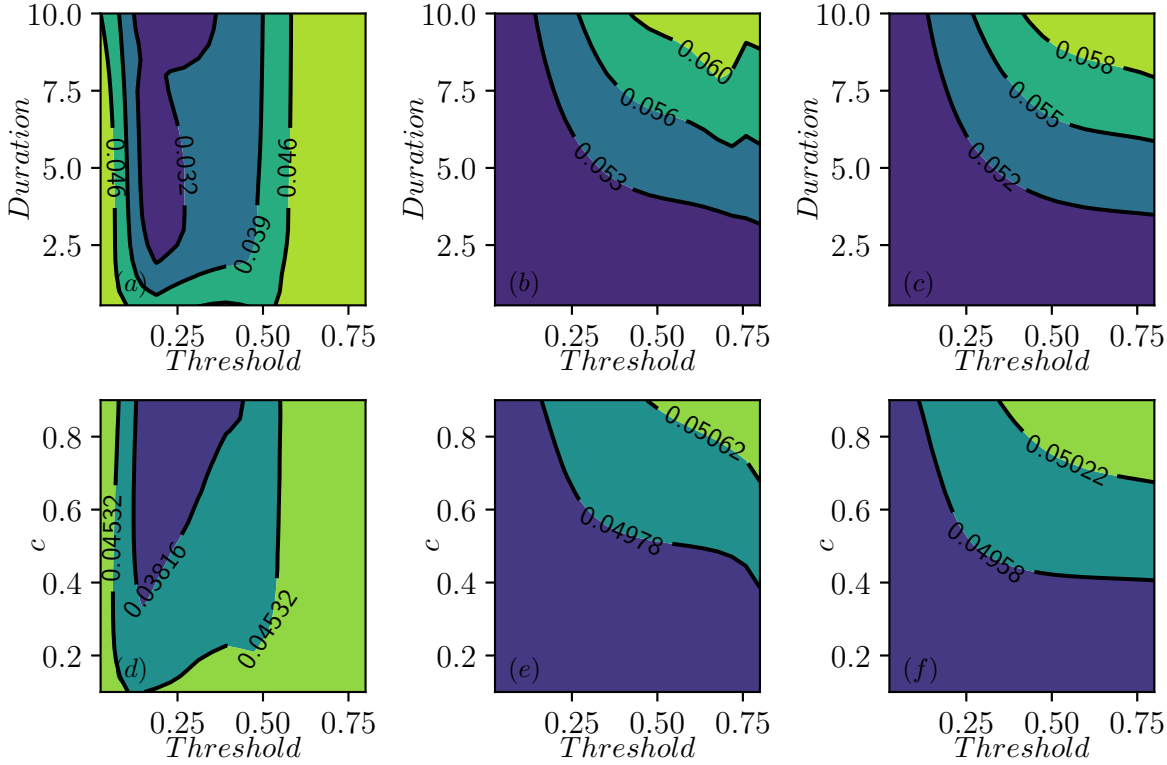

Figure 7: Contour plots of the peak prevalence, I_{peak}, that is the maximum value achieved by $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time-course of the epidemic. Control strategies and setup the same as in figure 5.

3.2.3 Impact on timing of epidemic

In the spirit of last subsection, we can also notice how different control strategies affect the peak prevalence time. The most striking feature emerging from simulations is that a control strategy tailored to the needs of each sub-population, figure 8(a)-(d) can delay the time of the peak by quite a margin, in accordance to what is observed in the single population case. When the intervention acts on a global level, the effect of control on peak-time appears to be

more regular, being proportional to the duration of the intervention being mildly affected by the threshold at which intervention happens. This is due to the fact that if the intervention is not perfectly timed on each epidemic curve, the peaks do not change by much, exactly as discussed in the single population case. The lower peak observed in panel (d) happens for $c \to 1$ and a threshold lower than 0.2 in a sub-population based approach, and it means that as soon as control is triggered, the epidemic is completely stopped for a duration D and it cannot return to values higher than the ones that triggered intervention.

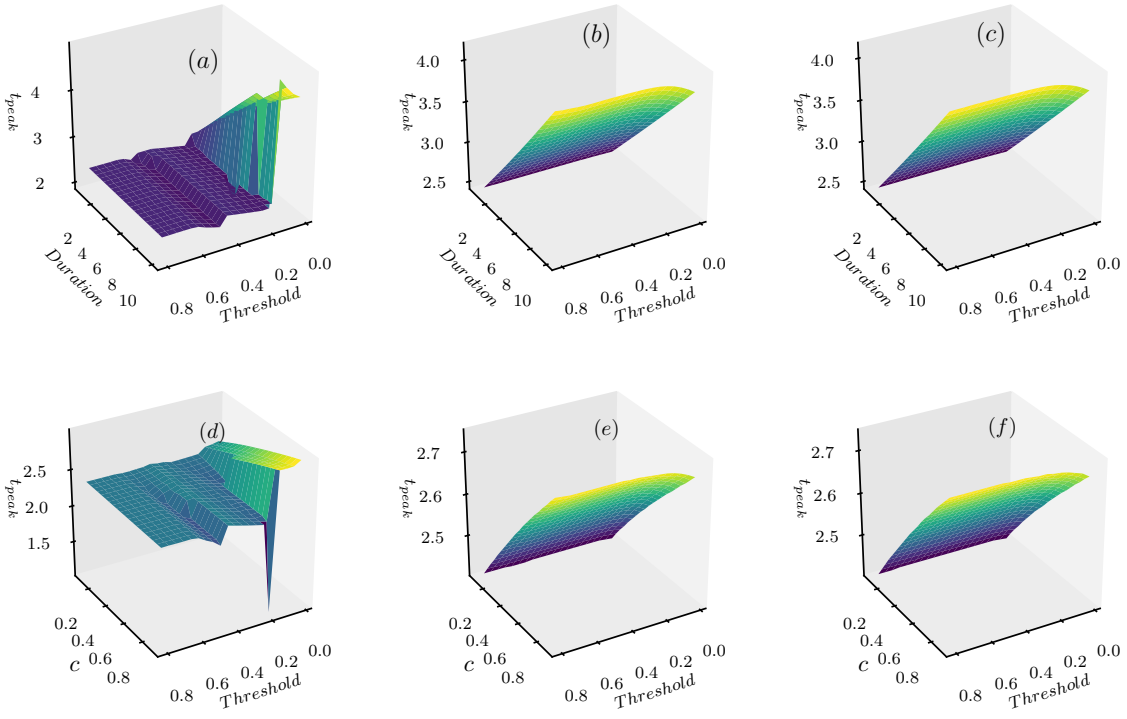

Figure 8: Surface plot of the time at which peak prevalence in the global epidemic happens, i.e. the maximum point in $\sum_i I_i$. In terms of control strategies and parameter values the same setup as in figures 7 and 5 are used.

4 Discussion

Our analysis shows clearly that a one-shot intervention is much more effective if implemented once the number of infections is reasonably large. This pattern is consistent for well-mixed populations and for metapopulations. In the metapopulation model, the intervention is significantly more effective when the timing in each subgroup is based on the infection levels in that group as opposed to synchronized across the population. Our analysis also shows that if a sustainable transmission-reducing intervention has not yet been implemented once the

disease is already established, then there is still significant benefit even if the introduction is delayed.

There is some difference in the optimal threshold to use for the intervention based on whether the goal is to control peak prevalence (thus reducing the maximal load on the health care system) or reducing the total number infected. However in both cases the optimal threshold is well after the disease is established.

So for a one-shot intervention, it is clear that delaying the intervention until infection levels have become relatively large is optimal. However, we must think critically about what constitutes a one-shot intervention. Obviously an intervention that prevents the population from getting food cannot be maintained. But what about other interventions?

We find that in the metapopulation model, the targeted one-shot intervention achieves significant improvements over a synchronized one-shot intervention. This is because the inherent asynchrony of the epidemics means that many communities have an epidemic either before or after the intervention. Our results offer strong support for targeting the interventions if they cannot be maintained for a long period. We have ignored the logistical challenges that might be associated with implementing the intervention separately for each sub-population. On a large scale (states within a country) we anticipate that this is logistically feasible, while on a small scale (suburbs in a city) it is more likely that the epidemics will be synchronized.

Current estimates of case fatality rate (not to be confused with infection fatality rate) range from 0.7% in China outside of Hubei province to around 2% in much of the world, to around 5.8% in Wuhan [8]. These estimates may depend on what proportion of cases are identified, and whether the health system is over capacity. With such high fatality rates, our tolerance for drastic interventions should increase. Thus an intervention that would be considered one-shot for the 2009 H1N1 pandemic which had a low fatality rate might be considered sustainable for the COVID-19 epidemic.

In deciding whether an intervention is sustainable policy maker could formulate an answer to this question: "if infection rates are the same or higher at some future time, and would increase if they were dropped, would you be willing to maintain the intervention in place?" If so, then the intervention is sustainable. If the answer is "no", and the intervention will be abandoned at some future time regardless of the new infection profile, then this is a one-shot intervention, and it should be held in reserve until it will have more impact.

A Mathematical Analysis

In this section we provide mathematical analyses of the single population model to support our results for reducing attack rate and peak prevalence.

A.1 A Phase-plane based analysis

Because $S + I + R = 1$, we can fully specify the current state and the future dynamics by knowing S and R, in which case $I = 1 - S - R$. It will be useful to use this to explore the

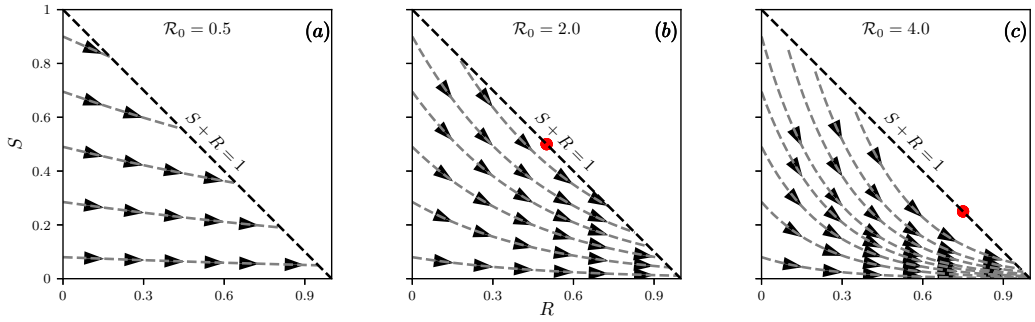

Figure 9: We plot $S(t)$ versus $R(t)$ for $\mathcal{R}_0 = 0.5$ (a), 2 (b), and 4 (c). For given $S(t)$ and $R(t)$, the proportion infected is $I(t) = 1 - S(t) - R(t)$, which equals the vertical or horizontal distance from the point $(R(t), S(t))$ to the line $S + R = 1$. The curves and arrows show how a solution to System (1) evolves in time. At points $S > 1/\mathcal{R}_0$ (which occurs only for $\mathcal{R}_0 > 1$) orbits move farther from the diagonal, representing an increase in I. Note that the velocity an orbit is traversed varies depending on location, and goes to zero close to $S + R = 1$. Red dots in panel (b)-(c) indicate the point $(S = \frac{1}{R_0}, I = 0, R = 1 - S)$.

dynamics of an epidemic and the impact of an intervention.

In Figure 9 we show how $S(t)$ and $R(t)$ evolve together in time for three values of $R(0)$ (0.5, 2, and 4) and for several different initial conditions. For a given point (S_0, R_0), the value of I at that time is given by the horizontal (or vertical) distance to the diagonal line $S + R = 1$.

In the figure, we can see that if $S > 1/\mathcal{R}_0$ (which is only possible if $\mathcal{R}_0 > 1$), then the horizontal distance from the orbit to the $S + R = 1$ line is increasing as the orbit moves forward. In other words, I is increasing. Once $S < 1/\mathcal{R}_0$, the distance decreases and eventually goes to 0.

Using these orbits, we can investigate the impact of an intervention, as shown in Figure 10. We follow S and R along an orbit. When we turn on the one-shot intervention at time t^*, it no longer follows the original curve. Instead the change in S and R follows the paths we would find for $(1 - c)\mathcal{R}_0$, starting from the point $(R(t^*), S(t^*))$. It follows this curve until reaching $(R(t^* + D), S(t^* + D))$ when the intervention is halted. It then follows the curves for the original $\mathcal{R}_0$, but starting from this new point.

Note that there is a point $(R, S) = (1 - 1/\mathcal{R}_0, 1/\mathcal{R}_0)$ at which separates the points on the line $R + S = 1$ from which an epidemic could start from the points at which epidemics finish. The closer an orbit is to this point, the smaller the final size.

So for $\mathcal{R}_0 = 2$, a temporary intervention gives us a way to move from one of these curves in the $\mathcal{R}_0 = 2$ plot to another. We see this in figure 10. The timing of the intervention determines which of the orbits the system lands on.

In this context the goal of reducing the attack rate is equivalent to ensuring that the intervention shifts the curve to an orbit as close as possible to $(R, S) = (1 - 1/\mathcal{R}_0, S = 1/\mathcal{R}_0)$. Reducing the peak prevalence is equivalent to ensuring that the orbit remains as close as

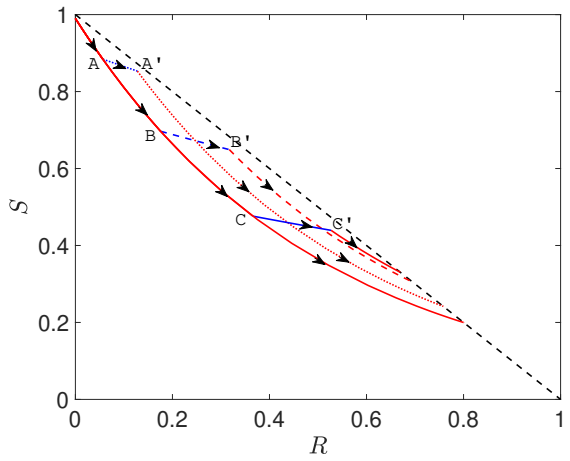

Figure 10: (S,R) phase portrait (arrows indicate growing time) based on an SIR model in a single population with $\beta = 2$, $\gamma = 1$ (giving $\mathcal{R}_0 = 2$) and initial condition $I(0) = 0.01$. The plot shows a trajectory with no control (continuous red line) as well as three other trajectories where $\beta = 0.5$ for a time period of length $D = 2$ but with the intervention setting in only once $(I + R)(t)$ goes past 0.1 (partially dotted line), 0.3 (partially dashed line) and 0.5 (continuous broken line), respectively. Control for the three different scenarios sets in at the points denoted by A, B and C and control ends at A', B' and C', respectively.

possible to the line $S + R = 1$.

A.1.1 Attack rate

If our goal is to minimize the number of infections, we accomplish this by having the curve $(R(t), S(t))$ land on an orbit that is as close to $(R, S) = (1 - 1/\mathcal{R}_0, 1/\mathcal{R}_0)$ as possible given the constraints on the intervention.

Typically we have to wait until the curve has moved closer to the desirable orbits before implementing the intervention. Implementing the intervention early, see the dotted line in figure 10 means that at the end of the intervention there is still a large pool of susceptibles which are at risk of becoming infected. Crossing from A to A' simply puts the epidemic on a slightly different trajectory but the final size is very close to the case with not control. An intervention at a later stage, see dashed line, improves the final outcome resulting in a final size that is smaller when compared to the case on no-control. Finally, the continuous broken line shows an almost optimal intervention with a further small reduction in the final fraction infected.

In general, the intervention that will get us closest to the optimal value occurs when the original curve is close to, but has not yet reached, the largest value of I, which occurs when $S = 1/\mathcal{R}_0$. As the effectiveness of the intervention increases, the orbits it follows during the intervention become more horizontal. For very effective interventions, this suggests we should wait until very close to the epidemic peak, while for less effective interventions (which

will slope downwards more), we will want to implement them somewhat sooner.

A.1.2 Peak prevalence

For peak prevalence, the goal is to keep the curve as close as possible to $S + R = 1$. The longer we wait to implement the intervention, the closer the final orbit is to $S + R = 1$, but the farther the original orbit moves from the line. With this in mind it becomes clear that the optimal t^* to reduce peak prevalence is smaller than the optimal value to reduce attack rate.

A.2 The mixing matrix

The cross-infection between sub-populations is modelled by $B = (\beta_{ij})_{i,j=1,2,\ldots N}$, where β_{ij} represents the rate at which infectious contacts are made from sub-population i towards susceptible individuals in sub-population j.

We implement a weak coupling by joining the population in a linear fashion: population i is only connected to population $(i-1)$ and $(i+1)$. The first and the last populations only connect to the second and the pen-ultimate population, respectively. The entries for the coupling/mixing matrix are generated as follows. On the main diagonal, the β_{ii} values are set to $2 + (Unif(0,1) - 0.5)$. Off-diagonal entries are set to $Unif(0,1)(\beta^*_{ii}/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) and represent a scaled and randomised version of the largest entry on the main diagonal. This yields an $\mathcal{R}_0$ above 2, comparable to current estimates for COVID-19.

We only use a single realisation of the mixing matrix and this is given below,

$$B = \begin{pmatrix} 1.917 & 0.059817 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.062024 & 2.2203 & 0.03117 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0094413 & 1.5001 & 0.0043357 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.070055 & 1.8023 & 0.076213 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.01146 & 1.6468 & 0.049723 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.02948 & 1.5923 & 0.054573 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.07709 & 1.6863 & 0.045981 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.015262 & 1.8456 & 0.015462 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.098061 & 1.8968 \end{pmatrix}. \quad (2)$$

References

[1] Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans*. Oxford University Press, Oxford, 1991.

[2] Frank Ball, Tom Britton, Thomas House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015.

[3] Martin CJ Bootsma and Neil M Ferguson. The effect of public health measures on the 1918 influenza pandemic in US cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.

[4] Mark Dickison, Shlomo Havlin, and H Eugene Stanley. Epidemics on interconnected networks. *Physical Review E*, 85(6):066109, 2012.

[5] Natsuko Imai et al. Report 3: Transmissibility of 2019-ncov. *Imperial College London*.

[6] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, , John Edmunds, Sebastian Funk, and Rosalind M Eggo. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *medRxiv*, 2020.

[7] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 02 2020. taaa021.

[8] WHO. Report of the who-china joint mission on coronavirus disease 2019 (covid-19), 2020.

[9] Li et al Xu. Full spectrum of covid-19 severity still being depicted. *Lancet. 2020*, 2020.

[10] Chong You, Yuhao Deng, Wenjie Hu, Jiarui Sun, Qiushi Lin, Feng Zhou, Cheng Heng Pang, Yuan Zhang, Zhengchao Chen, and Xiao-Hua Zhou. Estimation of the time-varying reproduction number of covid-19 outbreak in china. *medRxiv*, 2020.