Optimal 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 interventions and outcomes in the ongoing SARS-CoV-2 pandemic are highly varied. The disease and the interventions both impose costs and harm on society. Some interventions with particularly high costs may only be implemented briefly. The design of optimal policy requires consideration of many intervention scenarios. In this paper we investigate the optimal timing of interventions that are not sustainable for a long period. Specifically, we look at at the impact of a single short-term non-repeated intervention (a "one-shot intervention") on an epidemic and consider the impact of the intervention's timing. To minimize the total number infected, the intervention should start close to the peak so that there is minimal rebound once the intervention is stopped. To minimise the peak prevalence, it should start earlier, leading to initial reduction and then having a rebound to the same prevalence as the pre-intervention peak rather than one very large peak. To delay infections as much as possible (as might be appropriate if we expect improved interventions or treatments to be developed), earlier interventions have clear benefit. In populations with distinct subgroups, synchronized interventions are less effective than targeting the interventions in each subcommunity separately.

Author Summary

Some interventions which help control a spreading epidemic have significant adverse effects on the population, and cannot be maintained long-term. The optimal timing of such an intervention will depend on the ultimate goal.

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

- Interventions to delay the epidemic while new treatments or interventions are developed are best implemented as soon as possible.
- Interventions to minimize the peak prevalence are best implemented partway through the growth phase allowing immunity to build up so that the eventual rebound is not larger than the initial peak.
- Interventions to minimize the total number of infections are best implemented late in the growth phase to minimize the amount of rebound.

For a population with subcommunities which would have asynchronous outbreaks, similar results hold. Additionally, we find that it is best to target the intervention asynchronously to each subcommunity rather than synchronously across the population.

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 to reduce the spread. There is evidence [6] that some cities which implemented these interventions later had fewer deaths. This seemingly counter-intuitive observation suggests that they were more successful by being slow to respond.

When the 2009 influenza pandemic first arrived outside of Mexico, many schools shut after the first observed infection. Once these schools reopened, and received a new introduction, the remaining susceptible population was almost as large as at the outset, so the resulting epidemic was likely to be nearly as large as the original epidemics would have been. The closure provided increased time to prepare a response and learn more about the disease, but the overall epidemic was very similar to what would have happened without the closure. In contrast, evidence suggests that summer holidays altered the final outcome of the pandemic (at least in the UK), significantly reducing the total number of infections by splitting the epidemic into two smaller peaks [13].

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. When it returns to the original size, the remaining susceptible population will be effectively the same size as it was the first time. Thus, the intervention primarily delays the spread; the resulting epidemic is comparable to what would have been seen before. However, if the intervention occurs once the susceptible population has been noticeably depleted, then the number of infections falls and when the intervention is relaxed, the depleted susceptible population makes the rebound smaller or even nonexistent.

To make this explanation more robust, we note that is well-known that after an unmitigated epidemic, the total number of infections exceeds the number of infections required to achieve the "herd immunity threshold" (the level of immunity required to reduce the effective reproduction number below one) [8]. We refer to this extra level of infection as the

"overshoot". It is a consequence of the fact that when the effective reproduction number (in absence of intervention) finally falls to 1, the population reaches the "herd immunity threshold" and incidence no longer increases. However, because the epidemic is at its peak, this is the time at which those who have escaped infection so far face the highest force of infection. As the number infected falls, significant transmission still happens and the epidemic overshoots the herd immunity threshold. In the absence of further intervention, the size of the overshoot is determined by the number infected when the herd immunity threshold is reached.

When we think about this in terms of a temporary intervention, the option to minimize the total number of infections becomes clearer. A short intervention that ends with the effective reproduction above one would see a rebound and would see a larger overshoot than a slightly later intervention that ends with the effective reproduction number equal to 1.

This underlies the explanation of [6] for why temporary interventions are generally more effective if introduced later in the epidemic (but not too late). Similar, more detailed theoretical results have been found by [19, 1]. Most studies of these effects are focused on a single population, and they do not carefully consider the tradeoffs between competing goals of delaying infections, reducing the peak prevalence, or reducing the total size.

In the ongoing COVID-19 epidemic, China introduced drastic control measures very early. These significantly reduced transmission, apparently reducing the effective reproduction number (the number of new infections per infected individual) below one [23, 39], although it took a very long time for new cases to stop. Despite quite significant interventions in Italy in force for a long period of time, the rate of new cases was slow to fall [9].

Many other places have turned to aggressive control of infection in an attempt to keep transmission suppressed [20, 24, 18, 33, 37, 3, 22].

In places which have have nearly eliminated the disease, the threat of re-emergence requires constant vigilance. In places which have failed to contain transmission, the pervasive interventions that would be required to get transmission low would impose significant costs through the entire population, and such extensive interventions are unlikely to be maintained long term. Thus policy-makers face challenges about whether or when to implement such restrictive interventions.

Motivated by ongoing decisions facing policy makers for the COVID-19 pandemic, we develop mathematical models which allow us to explore how to time short-term interventions in response to an emerging epidemic. We will refer to these temporary interventions as "one-shot" interventions, meaning that the intervention cannot be maintained indefinitely or repeated. We are particularly interested in how the timing might affect the total fraction infected and the peak prevalence, but we are also interested in the resulting delay of infections.

We must exercise care in determining that a given intervention cannot be sustained. In the initial phase of the pandemic there was significant uncertainty in the fatality rate. With this in mind [38], the tolerance of the population for drastic interventions could be significant. What might appear to be an unsustainable intervention given one set of assumptions about severity may in fact be sustainable under another set of assumptions. We assume perfect

information and focus on choosing the time at which a given strategy will be more effective. A separate, but related line of research focuses on whether (and how long) we should hold an intervention in reserve while we learn more about the disease: sometimes the greatest expected benefit comes from learning more before choosing the intervention [5, 26, 32]. For an epidemic that grows quickly (like the early stage of the SARS-CoV-2 pandemic), there is effectively no time to learn about the disease before a decision is needed, and so these strategies would not be relevant until strong enough interventions are in place to suppress transmission.

We model an infection spreading in an initially fully susceptible population. We will model the spread within a single well-mixed population and a population made up of several weakly-coupled subcommunities (a *metapopulation*). We will investigate the impact of intervention on the attack rate (the final fraction infected), the peak prevalence, and the timing of infections, and in the metapopulation model we will additionally consider whether it is better to have a synchronized intervention or to have the intervention timed separately for each subcommunity. The important question of whether disease can be eliminated locally is beyond our scope.

Our goal is not to provide predictions for a specific population, but rather to demonstrate the generic impact of delaying a one-shot intervention, to show its robustness, and to provide intuition and some guiding principles which will apply to more complex scenarios.

Our results have important implications for the ongoing COVID-19 epidemic. If an intervention cannot be sustained for an extended period of time but new interventions or treatments are being developed, it is likely to be best to perform the intervention sooner to delay potential infections until other methods are available to treat or further delay infection (e.g., masks distributed, contact tracing implemented, healthcare capacity increased, therapeutic treatments identified, or even vaccine produced). However, if no other intervention or treatment improvement is likely to emerge, then it is best if the intervention is "held in reserve" until depletion of susceptibles has reduced the effective reproductive number enough that the intervention will have maximal impact on the total number of infections (by preventing the overshoot) or the peak prevalence.

We completed and released this research early in the initial stages of the pandemic (at the end of February 2020) [11], but did not immediately pursue it further due to other pressing questions. In the interim, a number of other papers have emerged studying related questions, including [17, 14, 10, 15, 30]. It is clear from much of this work that nonpharmaceutical interventions are an important part of epidemic control. In particular, the timing of an intervention, be it in a single population or over different communities, has a major impact on its effectiveness and overall outcome.

Our results provide insights into ongoing discussions of "circuit-breaker" interventions: in particular, such an intervention particularly valuable because it can delay infections while other interventions are brought into place, and it can keep the infection count low enough that interventions that cannot scale well can remain effective. However, if there is no significant effort to increase other interventions, then a repeated sequence of such "circuit-breakers" may be needed or the circuit-breakers should be delayed.

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 in a metapopulation made up of several distinct subcommunities. Then we discuss results from those mathematical models. Finally we discuss the implications of these results. In the Appendix we develop some mathematical theory explaining the mechanism underlying the effect in more detail.

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 subcommunities. 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. It reduces the transmission rate by a "strength factor" c. We explore the impact of the threshold, duration and strength of intervention. 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. In both models we measure time in multiples of the typical infection duration.

We will measure the impact of interventions on three quantities of interest:

- the *attack rate* or *final size*: the total fraction infected $R(\infty)$,
- the *peak prevalence* or maximum value of $I(t)$, and
- the *average time of infection*, $\bar{t}$, the average time (or date) at which individuals become infected. It is given by $\int_0^\infty -t\dot{S}\,dt / \int_0^\infty \dot{S}\,dt = \int_0^\infty \beta I S\,dt / R(\infty)$.

In general the (often conflicting) goals of our intervention are to reduce $R(\infty)$, reduce $I_{\max}$, and increase $\bar{t}$.

The value of minimizing the attack rate is clear as it minimizes the number of infections. The value of minimizing the peak prevalence is highlighted by the struggles that many health systems have faced during the early phase of the COVID-19 pandemic. The importance of increasing the average date at which infections occur is somewhat less clear. However, early in an epidemic, medical knowledge about the disease, health care capacity, and testing/contact tracing capacities are likely to be limited. Important knowledge about the transmission mechanisms may be missing. In this early stage where knowledge is increasing, any intervention that delays the bulk of infections until later is likely to help increase both the quality of medical care provided and the effectiveness of interventions that may prevent those infections altogether in the future. This may be particularly important for interventions such as contact tracing which take time to put in place and lose effectiveness when there are many infections.

2.1 Well-mixed population

To study an intervention in a well-mixed population, we use the standard SIR model [2].

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

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 average 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.

By measuring time in multiples of the typical infection duration, we impose that $\gamma = 1$, and so $\beta = \mathcal{R}_0$.

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 because recoveries outweigh new infections: $I \to 0$. Some fraction remains uninfected: $S(\infty) > 0$ and $R(\infty) = 1 - S(\infty)$ [2, 27, 29]. See Figure 1 for typical profiles of S, I, and R in time.

We assume that at some time $t = t^*$, an intervention that reduces the transmission rate is introduced for a 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 choose a monotonically increasing measure $I + R$ because this lets us choose an arbitrary t^*, which would not be possible if we focused on prevalence (I) or instantaneous rate of infection ($-\dot{S}$).

2.2 Weakly-coupled Metapopulation model

We will also investigate the effectiveness of interventions in a metapopulation made up of distinct subcommunities that do not have synchronized epidemics. The most obvious reason for this setup would be geographically separated populations. However there could be stratification by age, religion, ethnicity or socio-economic status. We are particularly interested in whether it is better to time interventions to the dynamics within each subcommunity

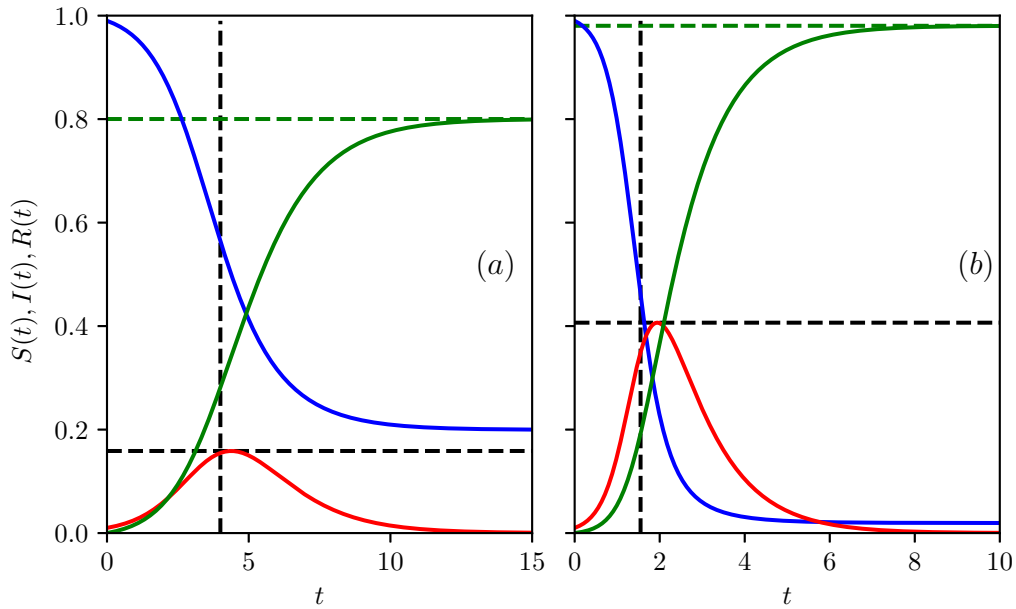

Figure 1: The time-evolution of S, I and R for epidemics with no control. (a) $\mathcal{R}_0 = \beta = 2$ and (b) $\mathcal{R}_0 = \beta = 4$ with $\gamma = 1$ in both. Horizontal and vertical dashed black lines indicate the peak prevalence $I_{\max}$ and average time of infection $\bar{t}$ respectively,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

separately or for the intervention to be synchronized even through the respective epidemics are not.

It is well-known that if the subcommunities have strong enough coupling, the epidemics in all subcommunities are effectively synchronised [4, 12]. In this case there is little distinction between asynchronous interventions for each subcommunity or interventions synchronized across all subcommunities. Thus to compare the results from synchronized interventions with asynchronous interventions targeted to each subcommunity, we need to explore a population with weak coupling. We use a standard meta-population model [2], allowing most transmission to be within a subcommunity and some cross interactions between the subcommunities.

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

where $0 \leq S_i \leq 1$, $0 \leq I_i \leq 1$ and $0 \leq R_i \leq 1$, with $S_i + I_i + R_i = 1$ for all t, represent the

fraction of susceptible, infected and infectious and recovered individuals in subcommunity i, where $i = 1, 2, \ldots, N$.

To simplify the presentation, all subcommunities are of equal size. The recovery rate γ is identical for all subcommunities. As before we measure time in multiples of the typical infectious period, so we set γ to 1. The cross-infection between subcommunities 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 subcommunity i towards susceptible individuals in subcommunity 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 $2 + (\text{Unif}(0,1) - 0.5)$ where $\text{Unif}(0,1)$ produces a random number chosen uniformly between 0 and 1. Off-diagonal entries are set to $\text{Unif}(0,1)(\beta^*/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) 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 [23, 25].

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

- track $I_i + R_i$ in each subcommunity and as soon as $I_i + R_i > T$ for some threshold T, a one-shot control is deployed in the corresponding subcommunity,
- track $I + R = \frac{1}{N}\sum_{i=1}^{N} I_i + R_i$ globally and as soon as $I + R > T$, a one-shot control across all subcommunities, and
- track each subcommunity and deploy the one-shot control is deployed across *all* subcommunities as soon as $I_i + R_i > T$ for the first subcommunity.

One-shot control in a subcommunity 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 intervention strength (we assume that the strength is the same in each subcommunity, and if two communities are both acting, then the movement between them is scaled by $(1-c)^2$). This reduction lasts for a duration D and, as soon as the control is over, the transmission rates for that subcommunity are restored to the starting levels.

In our results, we will present the average outcome of simulations across 100 distinct populations whose mixing matrices are chosen stochastically based on the rules described above.

3 Results

We use our mathematical models to explore how the timing of a one-shot intervention can impact

- total attack rate,

- peak prevalence, and
- average time of infection.

These are expected to be good proxies 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 the most impact on delaying the epidemic, but they have little impact on the attack rate or the peak prevalence. This is because only a few individuals are infected when the intervention is implemented so not many transmissions are blocked. When the restrictions are lifted, almost as many transmissions end up happening: the disease spreads in an almost fully susceptible population, and its trajectory is very nearly the same, just delayed. In contrast if the intervention is delayed until a non-negligible fraction of the population has been infected it will have more impact on the epidemic's shape.

For the 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

We can think of a strong, but temporary, intervention as dividing the overall epidemic into two phases. We allow an epidemic to spread until the intervention is started. The intervention resets I to a small value (that is, the intervention shifts the epidemic to a new trajectory with a similar S, but a smaller I). Depending on how long the epidemic was allowed to spread prior to the intervention, we have some new value of $S(t^*+D)$. Then a new epidemic happens starting from the new initial state, spreading as if a fraction $1-S(t^*+D)$ were vaccinated. The longer we allow the first phase epidemic to spread, the smaller the value of $S(t^* + D)$, and so the smaller the second phase will be. An early intervention truncates the first phase, but a later intervention reduces the second phase.

For a well-mixed population we find that the timing of a one-shot intervention has an important impact on the total epidemic. If the intervention is put in place very early, then the impact is to simply delay the epidemic. Because $S(t^* + D) \approx S(0)$ the second phase is effectively like the first phase without an intervention, but delayed. The delay is somewhat larger than D because it takes some time for I to grow back to $I(t^*)$.

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.

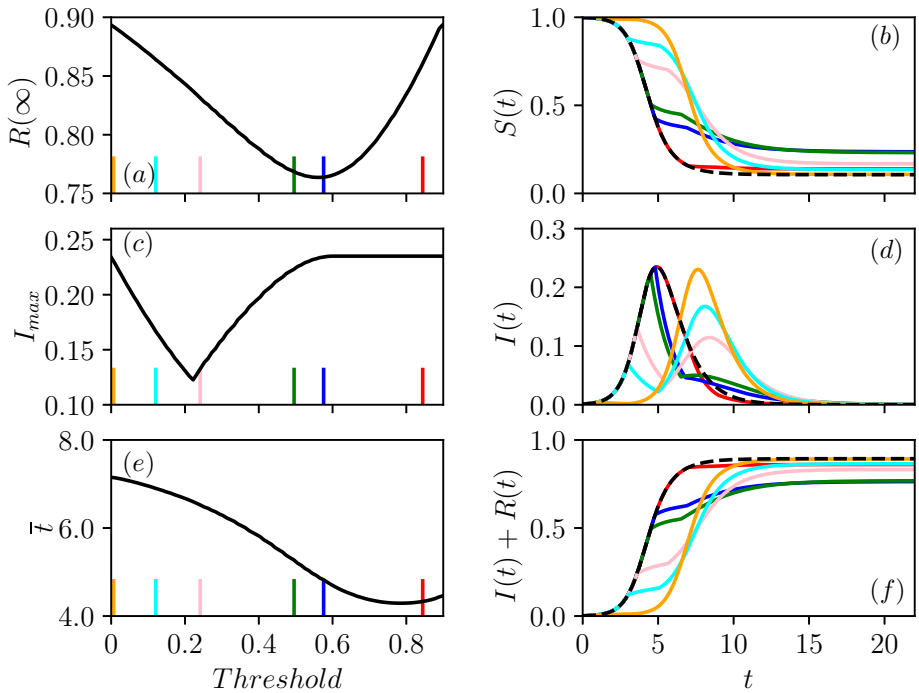

Figure 2: Illustration of the impact of one-shot intervention in a population with $\mathcal{R}_0 = 2.5$. The intervention has $c = 0.8$ for a duration of $D = 2$ time units. This intervention is introduced at different times as determined by a range of *Threshold* values. The impact of the threshold ($I+R > Tr$) for implementing the intervention is shown for (a) the attack rate $R(\infty)$; (b) $S(t)$; (c) peak prevalence $I_{\max}$; (d) $I(t)$; (e) average time of infection $\bar{t}$; 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).

3.1.1 Impact on attack rate

The impact on the attack rate (the total number infected) can be understood by a mental model of the intervention as a way to shift from the current epidemic trajectory to a new epidemic trajectory with a similar number of susceptibles, but fewer infected (this is made more rigorous in Appendix A.1).

If the intervention is introduced early on, it will have an immediate impact. However, when the intervention is lifted, the epidemic rebounds until the number of infections is the same as the original value. The number susceptible is relatively unchanged, and so a similar epidemic happens with 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 if the intervention is introduced later, there is clear improve-

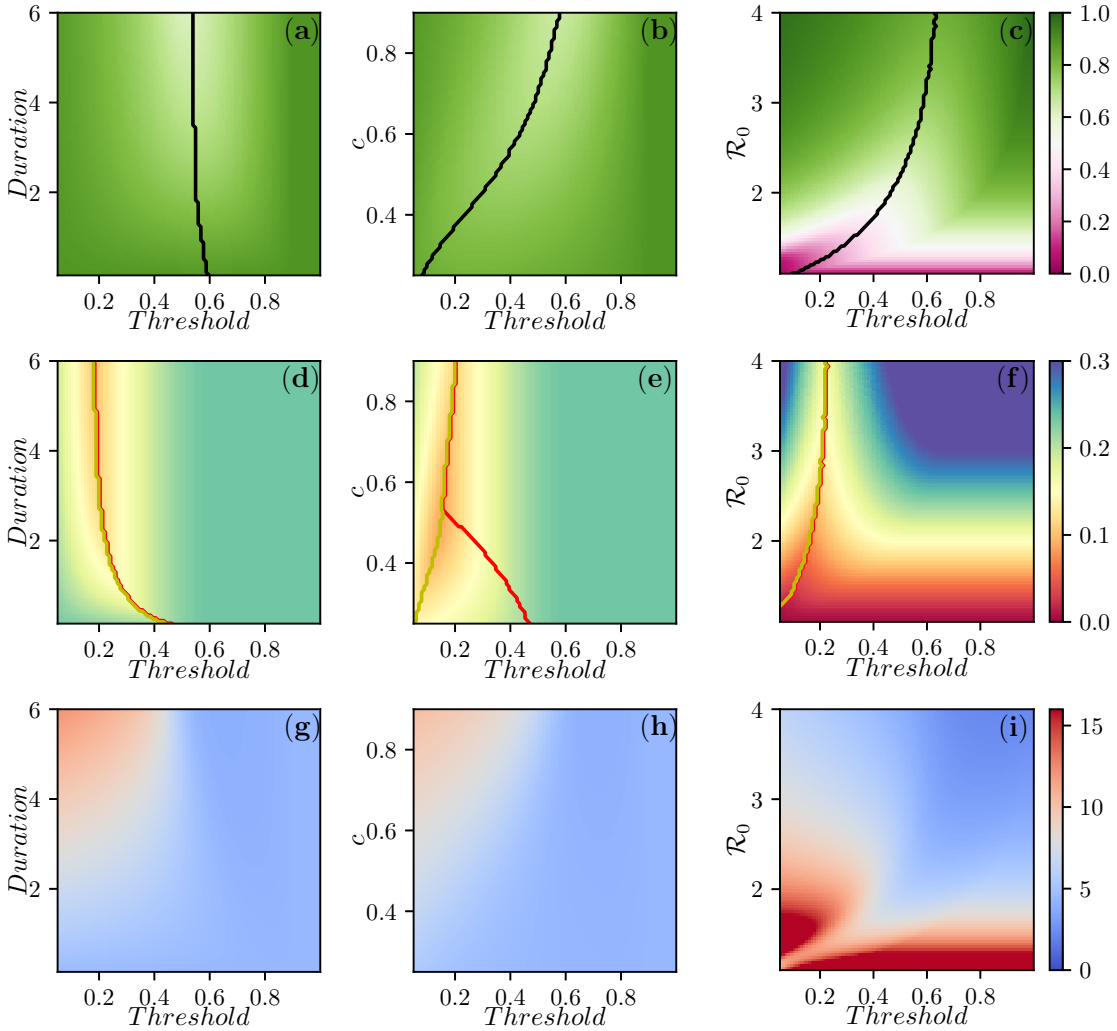

Figure 3: Contour plots for $R(\infty)$ (top), $I_{\max}$ (middle) and the mean time of infection $\bar{t}$ (bottom) as a function of parameters for the well-mixed population. We explore different threshold values of $I + R$ for the intervention to start, from a minimum of 0.05 to a max of 0.9. The first column investigates the impact of duration from $D = 0.1$ to $D = 6$, holding $\beta = 2.5$ and $c = 0.8$. On the second column, intervention duration is $D = 4$ and c ranges from 0.2 to 0.9. Finally, on the third column, $c = 0.8$ and $D = 4$, and the values of $\beta = \mathcal{R}_0$ vary from 1 to 4. In all cases $\gamma = 1$. In the first row, the black curve denotes the threshold for which $\mathcal{R}_e = 1$ when the intervention completes. In the three regions defined by the two lines in the panels of the second row, the peak prevalence is observed after the intervention has ended (from left to yellow curve), during intervention (area between the curves), or before intervention (from red curve to the end of the figure). Where the two curves align, the prevalence decays as soon as the intervention is implemented and then recovers to the pre-intervention peak.

ment in $R(\infty)$, up to a threshold of $I + R$ of 0.6, which is close to where the peak prevalence occurs in the epidemic without intervention. This is because when the epidemic peaks, $\mathcal{R}_e = 1$, and so if we immediately and dramatically reduce the number infected at this point the epidemic quickly dies out.

As new infections do happen during the intervention, this mental model is only an approximation. It can be made more precise by recognizing that to reduce the attack rate $R(\infty)$, the intervention is most effective if it is timed to directly block as many transmissions as possible. So we want to time the intervention to maximize the number of infected individuals present while it is in place (mathematically we want to maximize $\int_{t^*}^{t^*+D} I(\tau)d\tau$ given D, c, and $\mathcal{R}_0$).

Thus the ideal timing to reduce the total number of infections is not at the first hint of infection (when there are not enough infected individuals to cause many transmissions), but rather, a little before the peak, and if the intervention is perfect ($c = 1$), then at the peak. This suggests that the more effective an intervention is, the closer we should be to the peak before implementing it. It also suggests that for an intervention of a longer duration, we can implement it somewhat sooner, but not significantly sooner. For a more infectious disease, 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 strength c, the duration D, or the reproductive number $\mathcal{R}_0$ change. The earliest interventions have the most impact on the average time of infection, while somewhat delayed interventions affect the peak prevalence the most, and later interventions (near the epidemic peak) affect the final attack rate.

3.1.2 Impact on peak prevalence

As in the attack rate case, an early intervention primarily delays the epidemic curve. It does not significantly alter the shape. Thus the peak prevalence remains effectively the same unless the intervention is delayed until S is noticeably depleted.

If the susceptible population has been sufficiently depleted prior to the elimination of the intervention, then once the intervention is stopped, the epidemic rebound will be muted. Moving the intervention later makes the rebound smaller still. However, it means that the number of infections prior to the intervention is larger. There comes a threshold at which the phase before and the phase after the intervention have the same maximum. This is the time that minimizes the peak prevalence. Delaying the intervention past this value results in a larger pre-intervention peak, while doing it sooner results in a larger post-intervention peak.

Figure 2(c) shows the optimal threshold to reduce the peak prevalence occurs sooner than to reduce the attack rate. We can understand this intuitively, because for optimizing peak prevalence a moderate rebound is less of a concern than for optimizing the attack rate. For the purpose of reducing peak prevalence, Figure 2(d) shows that 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 $R(t^* + D) \approx I(t^*) + R(t^*)$. If D is long, but c is small enough that we cannot ignore transmissions occurring during the intervention, then we need to correct for the fact that $S(t^* + D)$ may be somewhat smaller than $S(t^*)$. Accounting for these transmissions further reduces the size of the second peak.

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 prevalence does not vary much in Figures 3(b) and (e).

It is worth highlighting that in related recent work [30] showed that the penalty for making a small error in the *timing* of the intervention is larger if it is too late compared to too early. As we see in our figure, the error as a function of the threshold $I + R$ appears roughly symmetric, but because the optimal intervention time often occurs while the epidemic growth is increasing, this means that being a little too late means a larger error in $I + R$ than being a little too early.

3.1.3 Impact on timing of infections

The impact on the timing of an emerging epidemic is an additional factor that plays an important role. If we anticipate rapid development of new treatments or interventions, then this may be more important than reducing the anticipated peak prevalence or total fraction infected.

As we noted for the attack rate and the peak prevalence discussions, for very early interventions (for which $I + R$ is very small), the entire epidemic curve shifts in time.[2]. However, as the threshold increases and we start to see an impact on the final attack rate $R(\infty)$ and peak prevalence $I_{\max}$, we also see an additional impact on the average time of infection. Unlike the other targets, a later intervention tends to have an decreased impact because more of the infections have occurred earlier.

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 outside the population, the reduction of infections within the population during the intervention may be immediately negated by new transmissions from outside, which are likely to be increasing. So the effect to delay the epidemic is largest if accompanied by reduction in transmissions from outside. However, in a setting where the disease is not well-established outside the population (as occurred in China early

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

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

in the COVID-19 pandemic), or travel from outside can be restricted, a major effort at early time may significantly delay the eventual epidemic.

3.2 Weakly-coupled metapopulation model

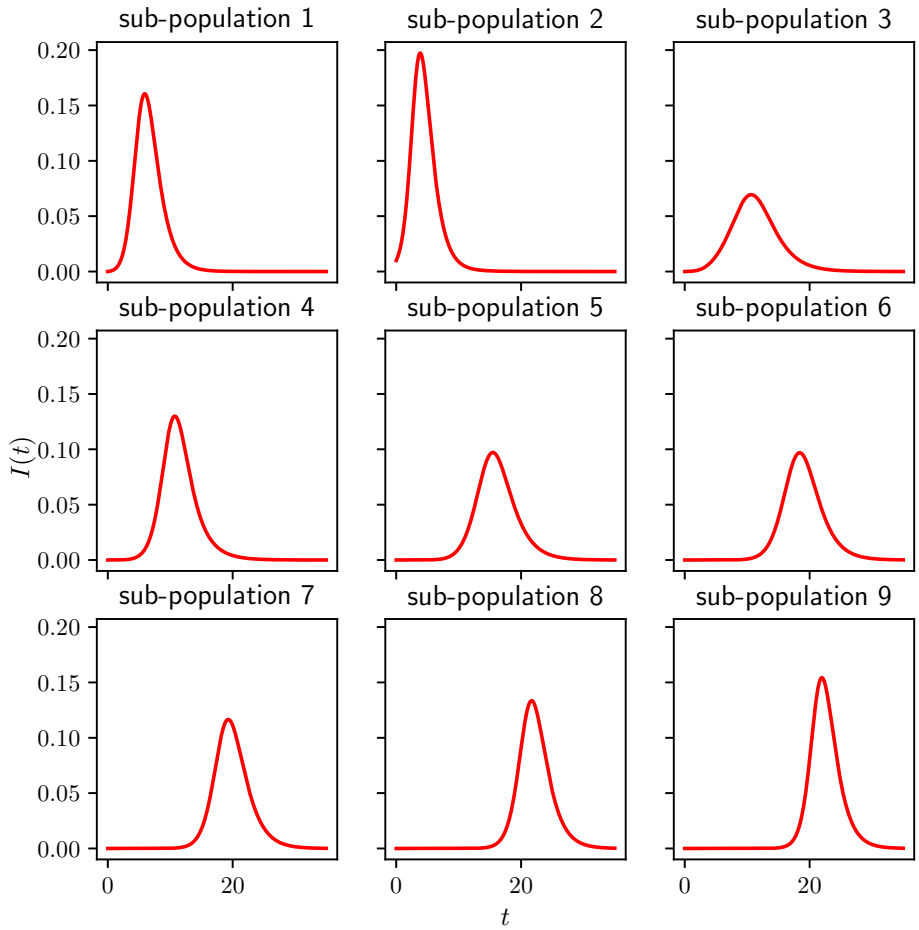

Figure 4: Example of an epidemic spreading across 9 subcommunities with different contact rates (see the Appendix A.2 for the precise mixing matrix B). The epidemic starts from subcommunity 2 and it is run for $T = 35$ units of time. $\gamma = 1$ for all subcommunities. With no control the attack rate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

We now consider a more realistic population which consists of coupled subcommunities, effectively a metapopulation model. We again consider one-shot interventions that either target the entire population at once (synchronous interventions) or that target individual subcommunities at different times (asynchronous interventions). If they were strongly coupled, the epidemics would be synchronous [4, 12]. So the single well-mixed population results would carry over. Our focus is on weakly-coupled subcommunities.

A typical plot of the prevalence level in each subcommunity is shown in Figure 4 in the absence of intervention. The epidemic starts in subcommunity two but it then spreads to the others. The entries of the cross-infection/mixing matrix are generated following the description in Section 2.2, and the specific mixing parameters are given in the Appendix.

As before we consider the impact of intervention on attack rate, peak prevalence, and peak timing. The overall effect of interventions is qualitatively similar to that of the single-population model. However, we find that asynchronous interventions that target each subcommunity significantly outperform synchronized interventions that begin when either the first subcommunity reaches a threshold or the global infection crosses a threshold.

For synchronized interventions, the overall impact is smaller, and the best outcomes are not driven by the actual threshold value. Rather they result from the intervention being timed to have significant impact on multiple communities, or optimally delaying the spread between communities. Consequently, the ideal times for this will depend on the parameters for between-community transmission, and are likely to be population-dependent.

Because the epidemics subcommunity may not be synchronized across subcommunities, this means that when a synchronous intervention is applied some subcommunities may have already completed their epidemic, or others have not yet begun. Interventions that are based on the first population to reach a threshold may not be valuable if the particular intervention is most effective if it disrupts patterns that do not appear until the first subcommunity has effectively completed its epidemic.

In our results, we consider 100 simulated populations consisting of 9 subcommunities, whose contact structure is generated from the random process described in Section 2.2. For most results we present only the average behavior. We note that this aggregation may hide important behavior from individual simulations [21]. However, except where noted, our averaged results are qualitatively similar to individual populations. Where we look at an individual simulation, we use the specific population of Figure 4.

3.2.1 Impact on attack rate

Our primary observation about the attack rate is that interventions acting at different times for each subcommunity are substantially more effective than synchronized interventions.

The smallest values of the attack rates are achieved when control acts independently in each subcommunity meaning that as soon as $I_i + R_i$ crosses a threshold, the one-shot control is switched on in subcommunity i. This is done independently of whether the efficacy or duration of control is kept fixed, while the other is varied, see Figure 5(a,d). Typically, as in the case of a single population, there seems to be a clear optimal threshold value which leads to the smallest attack rate. 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 strength of control, Figure 5(d), leads to smaller attack rates. Both strength and duration of control have no significant impact on the attack rate if the intervention is too early or too late.

Figure 6 shows 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

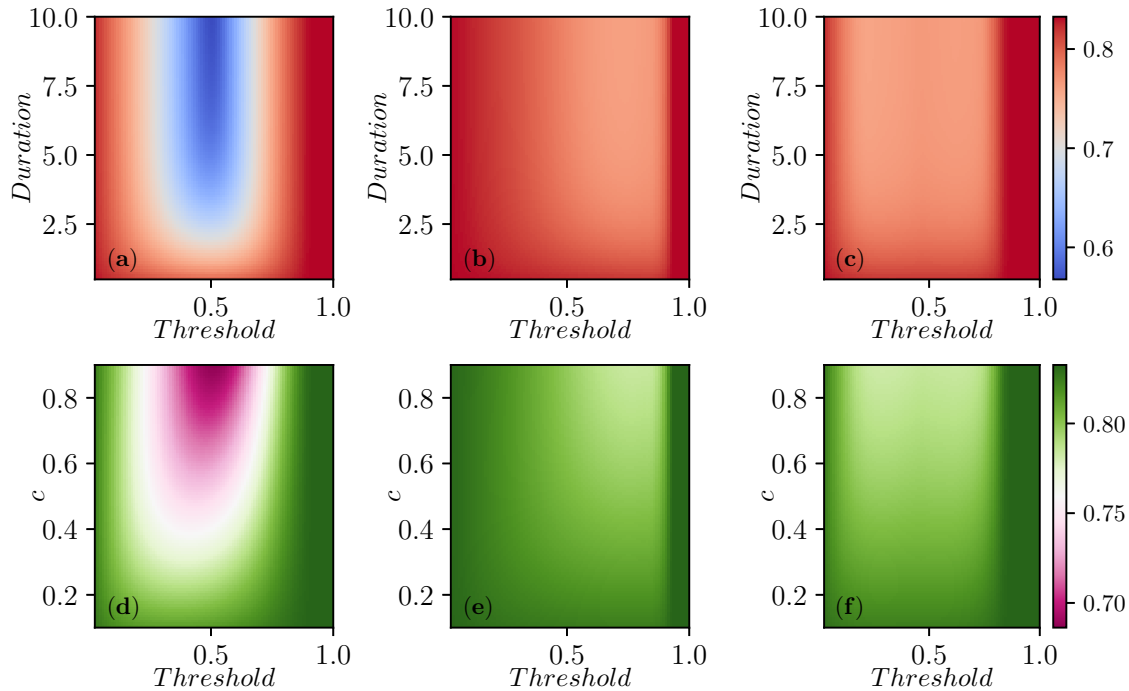

Figure 5: Contour plots showing the average attack rate (final epidemic size) over 100 simulated populations for each set of parameter values. In the first row c is fixed and the duration of control varies on the vertical axis, while in the second row duration is fixed and c varies. Each column corresponds to one of the three strategies: (a,d) intervention in each subcommunity, (b,e) global intervention when the first subcommunity breaches the threshold, and (c,f) global intervention at global threshold for a population consisting of 9 subcommunities. 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 recovery rate is $\gamma = 1$ for all subcommunities. In all cases, if the threshold is set too large the intervention is never implemented. The two synchronized interventions can be approximately mapped to one another by noting the largest $I_i + R_i$ at the time the global $I + R$ reaches a given threshold. The subcommunity threshold gives more resolution at small values while the global threshold gives more resolution at large values.

intervention happens close to the peak of the epidemic in each subcommunity so secondary waves of infection are heavily suppressed.

The impact of the synchronized intervention based on the global level of $I + R$, see Figure 5(c,f), or on the first subcommunity to reach a threshold, see Figure 5(b,e), are much smaller than asynchronous interventions. This is because when it is implemented in the synchronous case, some communities have already completed their epidemic while others

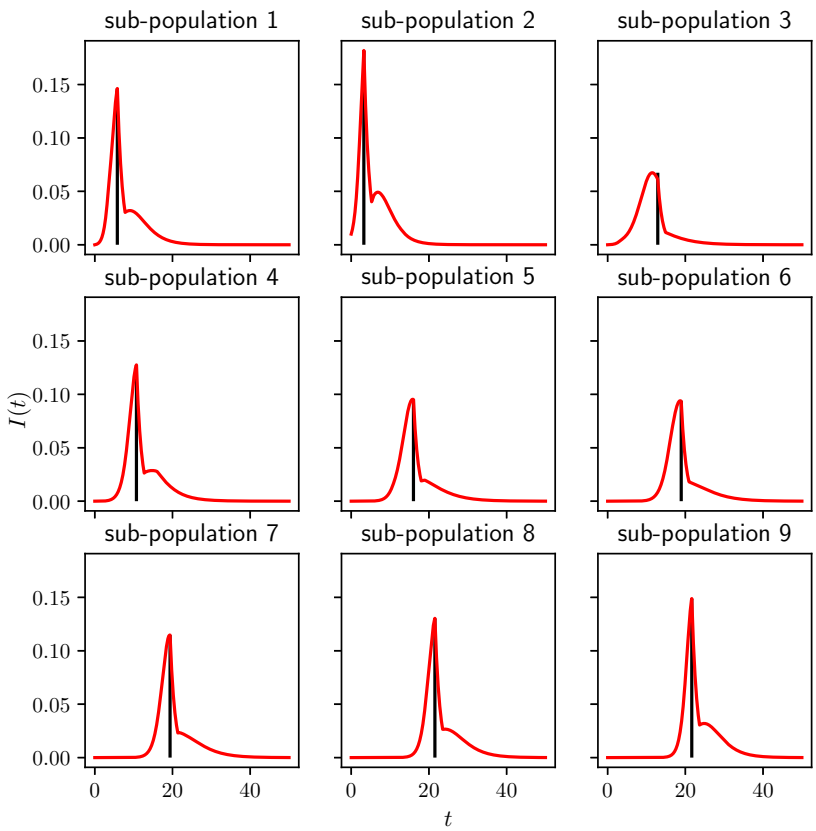

Figure 6: Illustration of best control strategy (i.e. smallest attack rate) (controlling subcommunities individually but using the same threshold for each) when efficacy and duration of control are fixed at $c = 0.8$ and $D = 2$, respectively. It turns out that the optimal threshold is close to (0.4). 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 lines show the onset of control.

have not yet begun. So there is less overall impact (see the asynchrony in Figure 4).

When the intervention is based on the first time a subcommunity crosses a threshold, we find that the optimal thresholds are at relatively large values. This suggests that the value of the synchronized intervention comes from disrupting transmission when the disease is spreading in multiple subcommunities.

Under the synchronous intervention scenario, we also see some surprising behavior where there are multiple local maxima for the specific metapopulation used in Figure 4 (not shown in Figure 5). This effect is because the timing aligns with different outbreaks. If we intervene at one time, we may have a big impact on one subcommunity, and if we miss that window, it is best to wait until another subcommunity begins to have an outbreak. This effect disappears in the aggregated data of Figure 5 because the specific ideal timing is a consequence of the

randomly chosen parameters of each population.

3.2.2 Impact on peak prevalence

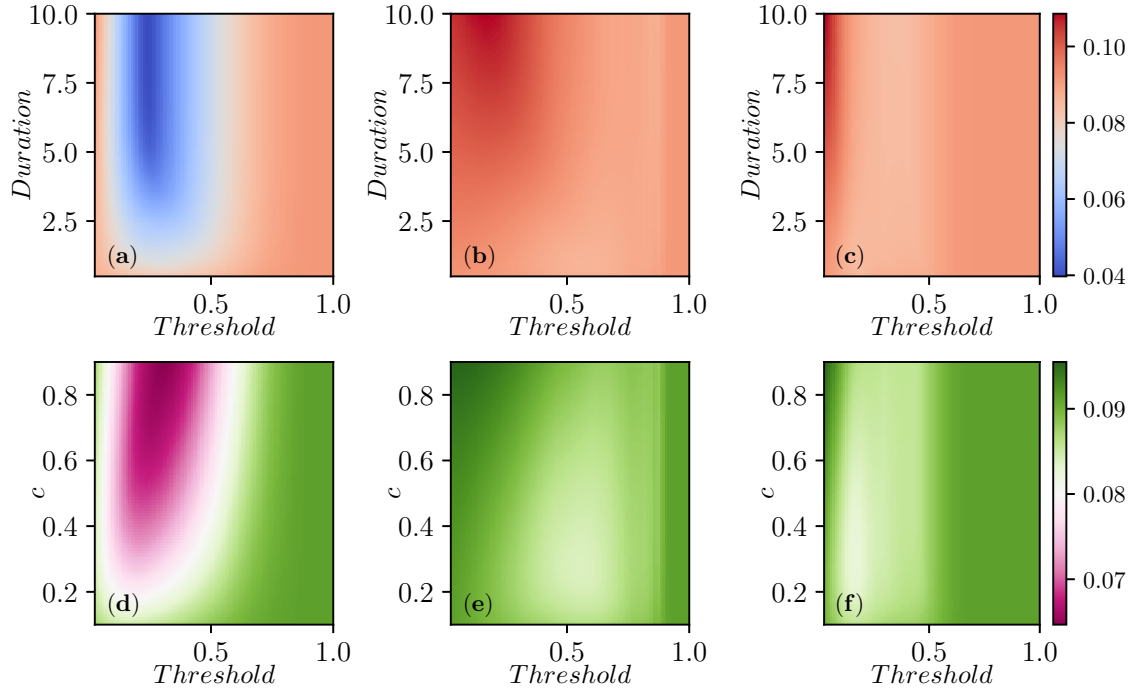

Figure 7: Contour plots of the peak prevalence $I_{peak} = \max_t \left(\frac{1}{9}\sum_i I_i(t)\right)$, averaged across 100 simulated populations each with 9 subcommunities. Control strategies and setup the same as in Figure 5.

Here we look at the effect of the intervention on the peak prevalence, that is the maximum value of $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time course of the epidemics. As with the attack rate, our primary observation for the peak prevalence is that it is significantly reduced by targeting based on the individual subcommunity.

Figure 7 shows the average of the peak prevalence across the same 100 populations as Figure 5. Perhaps not surprisingly, Figure 7 is qualitatively similar to Figure 5. The most impact is through having interventions occurring when the individual populations reach a threshold. The optimal choices for intervention come earlier in the epidemic. We still observe that if the intervention is too soon or too late then there is no significant reduction in peak prevalence.

In Figure 7 (a,d) the optimal threshold for intervention is relatively early, this is in line with the trend observed in Figure 2 for the single population case. We should wait until some immunity builds up before intervening, so that the rebound in each population is muted.

For our two synchronized strategies, the effectiveness is much less, because the overall peak prevalence is related to how the individual subcommunities' peaks align, and different

details of intervention timing, combined with the random parameters of the simulation, can make the individual peaks align or not.

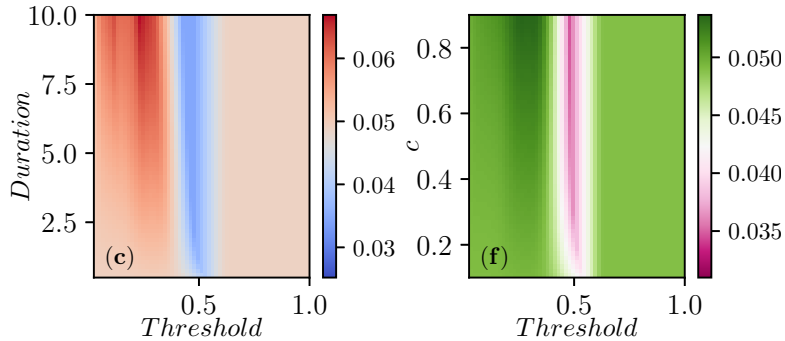

Figure 8: Contour plots of the peak prevalence, I_{peak}, that is the maximum value achieved by $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time-course of the epidemic. Control strategies and setup the same as in Figure 5.

Interestingly, if we look at an individual simulations, there are thresholds which yield significantly larger improvements in the peak prevalence than we see in the aggregated data. This is because in each simulated population, the relative timing of the epidemics subpopulations depends on the system parameters. In a weakly-coupled metapopulation model with relatively few subcommunities, the global peak prevalence is likely to occur when multiple subcommunities happen to be aligned. This is highly sensitive to parameters, and so the optimal intervention time will vary.

3.2.3 Impact on epidemic timing

When we investigate the average time of infection, we see that targeting the intervention at each subcommunity is again the most effective. In general the interventions need to be implemented very early in the epidemic.

Most of the impact comes from slowing the epidemic in the initial subcommunity. A delay in the initial place of introduction results in a delay in all subcommunities. Once the intervention is no longer in place in the initial subcommunity it begins to grow and spill over into other communities. If other subcommunities wait until then to begin their response, they gain some benefit. However, once they stop, they face rapid reseeding from the initial subcommunity. So the main benefit comes from the initial subcommunity's actions. When we use a synchronized intervention, the effect is somewhat smaller, but it is not significantly smaller.

In fact, most of the benefit comes from the initial subcommunity engaging in preventative measures. There is relatively little impact on the average time of infection to be gained from the other subcommunities acting early, unless they can maintain a very small spillover rate for a long period through extensive travel restrictions or similar interventions. This suggests that significant benefit may come from a hybrid strategy which focuses on delaying infections

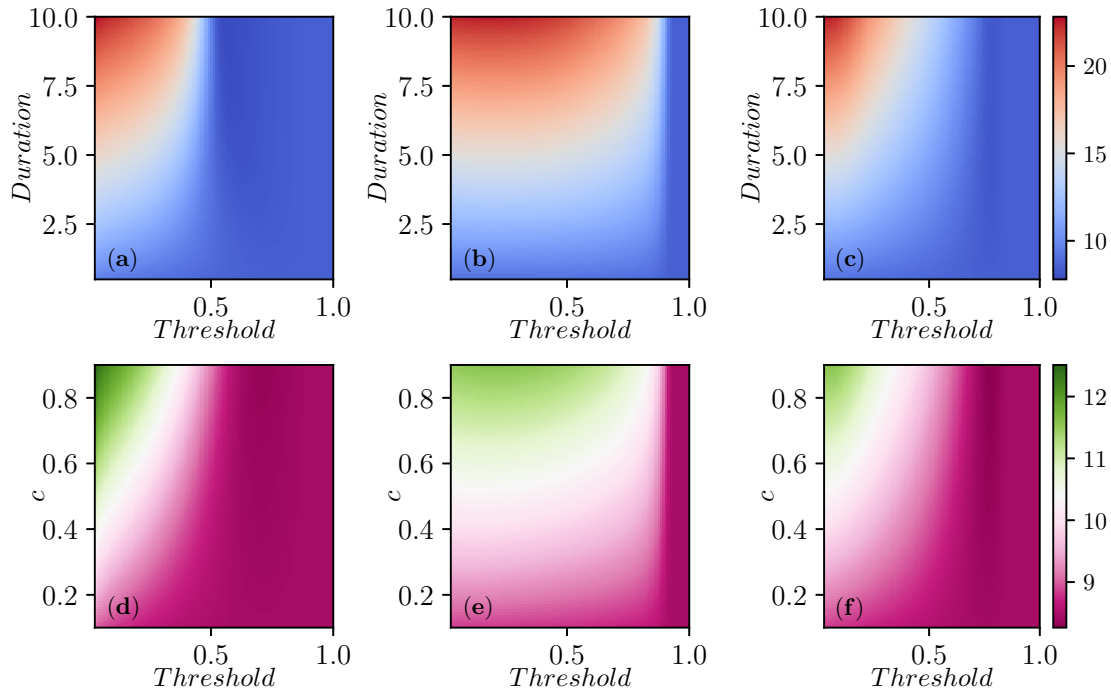

Figure 9: Countour plots of the global mean infection time, defined as $T = -\frac{1}{R(\infty)}\int_0^\infty t\sum_i \frac{dS_i}{dt}dt$, averaged over 100 simulations. In terms of control strategies and parameter values the same setup as in figures 7 and 5 are used.

out of the initial subcommunity while other locations focus their interventions on optimizing peak prevalence or attack rate.

4 Discussion

We have considered the impact of a single one-shot limited duration impact on the spread of an infectious SIR disease, in both a single well-mixed community and in a weakly-coupled metapopulation model.

We have found that in a single well-mixed population, an intervention at the first hint of infection is best for delaying infections. An intervention that waits until the epidemic is well-established but still well short of the peak is best to reduce the peak prevalence of the epidemic. An intervention (whose duration would be D) that starts a little less than D units of time before the peak would otherwise be reached is best to reduce the total number of infections.

In a weakly-coupled metapopulation model, we find qualitatively similar results. The best strategy to reduce either the peak prevalence or the total number of infections, a strategy that times the interventions asynchronously to when each subcommunity reaches a threshold rather than being synchronized to when the global average crosses a threshold or when the first subcommunity crosses a threshold.

For delaying the average time of infection, the most important detail is that the subcommunity with the first introduction responds as quickly as possible. Whether the other subcommunities respond immediately or delay their response until more infections are present within the subcommunities has a smaller effect.

4.1 Limitations

Our results are somewhat limited by the assumptions we have made to produce a tractable problem.

We assume that behavior responds immediately to changes in interventions. In reality, behavior may change prior to an intervention being implemented. Additionally adherence may drop as the intervention continues, and some adherence to the intervention may remain even once the intervention is removed.

The assumption that the individuals are largely homogeneous may lead to pessimistic predictions of when the herd immunity threshold is reached. In the presence of significant population heterogeneity, there is evidence suggesting that the herd immunity threshold would be reached earlier, and the epidemic could proceed significantly faster [16, 7]. Our qualitative predictions remain robust, but the timings would need to move sooner.

We must think critically about what constitutes a one-shot intervention. Whether an intervention can be maintained may depend on context. Early estimates of case fatality rate (not to be confused with infection fatality rate) of COVID-19 ranged from 0.7% in China outside of Hubei province to around 2% in much of the world, to around 5.8% in Wuhan [35]. These estimates were affected by the proportion of cases identified (leading to uncertainty in the denominator), and whether the health system was over capacity (which would increase the death rate leading to uncertainty in the numerator). True infection fatality rates appear to lie between 0.5% and 1%, with many estimates closer to 1% than 0.5% [28, 34, 31]. 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 significantly lower fatality rate [36] might be considered sustainable for the COVID-19 pandemic.

In deciding whether an intervention is sustainable policy makers could formulate an answer to this question: "Assume you impose the intervention now, and infection rates remain the same or higher in the future, 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 maximal impact.

We have ignored logistical challenges that might be associated with implementing the intervention separately for each subcommunity. On a large scale (e.g., states within a country or cities within a state) we anticipate that this is logistically feasible, while on a small scale (e.g., suburbs in a city) it is more likely that the epidemics will be synchronized and this benefit is small compared to the logistical challenges.

4.2 Policy Implications

Our observations have a number of important policy implications for an epidemic which is sufficiently established that elimination is not a goal. Primarily:

- To reduce the attack rate of an epidemic, a one-shot intervention should be introduced shortly before the epidemic peak.
- To reduce the peak size of an epidemic, the intervention should be late enough to allow significant immunity to develop, but early enough to allow a substantial rebound after the intervention.
- To delay infections as much as possible, the intervention should be implemented early on.
- In a population made up of many weakly-coupled subcommunities, interventions should be asynchronous.

Because they require different timings, the three goals we have considered are in conflict. The benefits of reducing the total number of infections are clear, and if health care capacity is threatened, the benefits of minimizing the peak prevalence are also clear. It is less obvious that delaying infections may help. However when there is an expectation that improved treatments and improved inteventions may be developed, a delay is likely to be the best available option.

An additional benefit of delaying the epidemic is observed when we have an intervention, such as testing & tracing, which does not scale well. As the number of infections grows ebyond the testing capacity, the effectiveness per infection goes down. At low infection levels these interventions may be enough to suppress transmission. However, if levels get too high, then a quick intervention that causes infections to drop may be more effective than an intervention that waits until the optimal time to minimize the size.

Although we have focused on three distinct goals which lead to different optimal strategies, the ideal goal is likely to be a combination of these effects. So the timing will need to respond to different pressures. If, for example, the goal is to keep the peak prevalence below a certain value while minimizing the maximum prevalence, then the ideal strategy would let as much immunity develop in the population as possible before the prevalence limit is reached and then intervene (this is of course only feasible if there is a time that can keep prevalence below that level). Delaying the intervention until this point would mean that the second peak would be lower. If the goal is instead to keep the peak prevalence below some level while maximizing the delay of infections, then the intervention would be sooner and timed so that the second peak would reach the target prevalence.

We finally note that in a population made up of weakly-coupled subcommunities whose epidemics will not be synchronized, the ideal intervention might be to react strongly and immediately in the first subcommunity where the infection begins to spread. This can provide protection to the other communities and significantly delay the spread. Once it spreads beyond that initial subcommunity then the focus may turn to minimizing the peak prevalence or the attack rate.

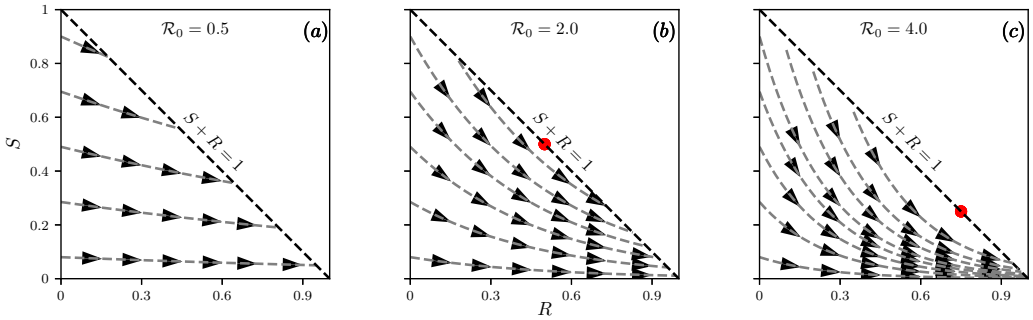

Figure 10: We plot $S(t)$ versus $R(t)$ for $\mathcal{R}_0 = 0.5$ (a), 2 (b), and 4 (c). For given $S(t)$ and $R(t)$, the proportion infected is $I(t) = 1 - S(t) - R(t)$, which equals the vertical or horizontal distance from the point $(R(t), S(t))$ to the line $S + R = 1$. The curves and arrows show 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 $(S = \frac{1}{R_0},\ \ I = 0,\ \ R = 1 - S)$.

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 dynamics of an epidemic and the impact of an intervention.

In Figure 10 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 11. 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.

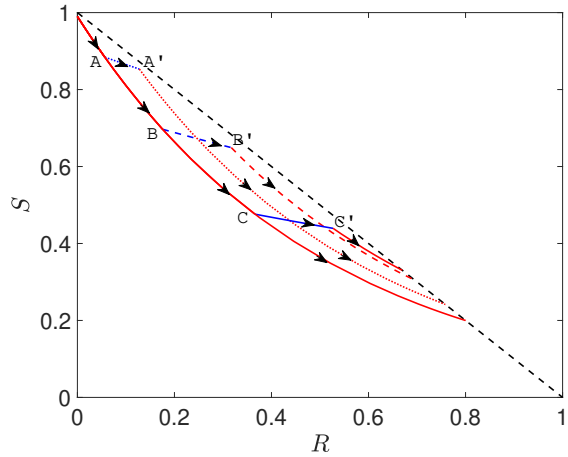

Figure 11: (S,R) phase portrait (arrows indicate growing time) based on an SIR model in a single population with $\beta = 2$, $\gamma = 1$ (giving $\mathcal{R}_0 = 2$) and initial condition $I(0) = 0.01$. The plot shows a trajectory with no control (continuous red line) as well as three other trajectories where $\beta = 0.5$ for a time period of length $D = 2$ but with the intervention setting in only once $I + R$ 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.

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 attack rate.

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 11. 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 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 11 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 attack rate is very close to the case with not control.

An intervention at a later stage, see dashed line, improves the final outcome resulting in an attack rate that is smaller when compared to the case of no control. Finally, the continuous broken line shows an almost optimal intervention with a further small reduction in the attack rate.

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 subcommunities 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 subcommunity i towards susceptible individuals in subcommunity 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 $2 + (\mathrm{Unif}(0,1) - 0.5)$. Off-diagonal entries are set to $\mathrm{Unif}(0,1)(\beta^*_{ii}/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) 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.

For most of our results we aggregate over 100 distinct simulated populations. However, in figures 4, 6, and 8 we use a single realization of the metapopulation model. For this we take the mixing matrix

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

References

[1] Roy M Anderson, Hans Heesterbeek, Don Klinkenberg, and T Déirdre Hollingsworth. How will country-based mitigation measures influence the course of the covid-19 epidemic? *The Lancet*, 395(10228):931–934, 2020.

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

[3] Michael G Baker, Nick Wilson, and Andrew Anglemyer. Successful elimination of covid-19 transmission in new zealand. *New England Journal of Medicine*, 383(8):e56, 2020.

[4] 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.

[5] Peter WJ Baxter and Hugh P Possingham. Optimizing search strategies for invasive pests: learn before you leap. *Journal of applied ecology*, 48(1):86–95, 2011.

[6] 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.

[7] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2. *Science*, 369(6505):846–849, 2020.

[8] Sarah Cobey. Modeling infectious disease dynamics. *Science*, 368(6492):713–714, 2020.

[9] S. De Flora and S. La Maestra. Growth and decline of the covid-19 epidemic wave in italy from march to june 2020. *Journal of Medical Virology*, 2020.

[10] Fabio Della Rossa, Davide Salzano, Anna Di Meglio, Francesco De Lellis, Marco Coraggio, Carmela Calabrese, Agostino Guarino, Ricardo Cardona, Pietro DeLellis, Davide Liuzza, et al. Intermittent yet coordinated regional strategies can alleviate the covid-19 epidemic: a network model of the italian case. *arXiv preprint arXiv:2005.07594*, 2020.

[11] Francesco Di Lauro, István Z Kiss, and Joel Miller. The timing of one-shot interventions for epidemic control. *medRxiv*, 2020.

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

[13] KTD Eames. The influence of school holiday timing on epidemic impact. *Epidemiology & Infection*, 142(9):1963–1971, 2014.

[14] Jana Gevertz, James Greene, Cynthia Hixahuary Sanchez Tapia, and Eduardo D Sontag. A novel covid-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. *medRxiv*, 2020.

[15] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6, 2020.

[16] M Gabriela M Gomes, Ricardo Aguas, Rodrigo M Corder, Jessica G King, Kate E Langwig, Caetano Souto-Maior, Jorge Carneiro, Marcelo U Ferreira, and Carlos Penha-Goncalves. Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold. *medRxiv*, 2020.

[17] Johannes Haushofer and C Jessica E Metcalf. Which interventions work best in a pandemic? *Science*, 368(6495):1063–1065, 2020.

[18] Van Minh Hoang, Hong Hanh Hoang, Quynh Long Khuong, Ngoc Quang La, and Thi Tuyet Hanh Tran. Describing the pattern of the covid-19 epidemic in vietnam. *Global health action*, 13(1):1776526, 2020.

[19] T Déirdre Hollingsworth, Don Klinkenberg, Hans Heesterbeek, and Roy M Anderson. Mitigation strategies for pandemic influenza a: balancing conflicting policy objectives. *PLoS computational biology*, 7(2), 2011.

[20] Akiko Iwasaki and Nathan D Grubaugh. Why does japan have so few cases of covid-19? *EMBO Molecular Medicine*, 12(5):e12481, 2020.

[21] Jonas L Juul, Kaare Græsbøll, Lasse Engbo Christiansen, and Sune Lehmann. Fixed-time descriptive statistics underestimate extremes of epidemic curve ensembles. *arXiv preprint arXiv:2007.05035*, 2020.

[22] June-Ho Kim, Julia Ah-Reum An, Pok-kee Min, Asaf Bitton, and Atul A Gawande. How south korea responded to the covid-19 outbreak in daegu. *NEJM Catalyst Innovations in Care Delivery*, 1(4), 2020.

[23] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, and the LSHTM CMMID COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.

[24] Viet-Phuong La, Thanh-Hang Pham, Manh-Toan Ho, Minh-Hoang Nguyen, Khanh-Linh P Nguyen, Thu-Trang Vuong, Trung Tran, Quy Khuc, Manh-Tung Ho, Quan-Hoang Vuong, et al. Policy response, social media and science journalism for the sustainability of the public health system amid the covid-19 outbreak: The vietnam lessons. *Sustainability*, 12(7):2931, 2020.

[25] 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.

[26] Michael Ludkovski and Jarad Niemi. Optimal dynamic policies for influenza management. *Statistical Communications in Infectious Diseases*, 2(1), 2010.

[27] Junling J. Ma and David J. D. Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68(3):679–702, 2006.

[28] Gideon Meyerowitz-Katz and Lea Merone. A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates. *medRxiv*, 2020.

[29] Joel C. Miller. A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74(9):2125–2141, 2012.

[30] Dylan H. Morris, Fernando W. Rossine, Joshua B. Plotkin, , and Simon A. Levin. Optimal, near-optimal, and robust epidemic control. *arXiv preprint arXiv:2004.02209*, 2020.

[31] Roberto Pastor-Barriuso, Beatriz Perez-Gomez, Miguel A Hernan, Mayte Perez-Olmeda, Raquel Yotti, Jesus Oteo, Jose Luis Sanmartin, Inmaculada Leon-Gomez, Aurora Fernandez-Garcia, Pablo Fernandez-Navarro, et al. Sars-cov-2 infection fatality risk in a nationwide seroepidemiological study. *medRxiv*, 2020.

[32] Robin N Thompson, Christopher A Gilligan, and Nik J Cunniffe. Control fast or control smart: When should invading pathogens be controlled? *PLoS computational biology*, 14(2):e1006014, 2018.

[33] C Jason Wang, Chun Y Ng, and Robert H Brook. Response to covid-19 in taiwan: big data analytics, new technology, and proactive testing. *Jama*, 323(14):1341–1342, 2020.

[34] Helen Ward, Christina J Atchison, Matthew Whitaker, Kylie EC Ainslie, Joshua Elliot, Lucy C Okell, Rozlyn Redd, Deborah Ashby, Christl A Donnelly, Wendy Barclay, et al. Antibody prevalence for sars-cov-2 in england following first peak of the pandemic: React2 study in 100,000 adults. *MedRxiv*, 2020.

[35] WHO. Report of the WHO–China joint mission on Coronavirus Disease 2019 (COVID-19), 2020.

[36] Jessica Y Wong, Peng Wu, Hiroshi Nishiura, Edward Goldstein, Eric HY Lau, Lin Yang, SK Chuang, Thomas Tsang, JS Malik Peiris, Joseph T Wu, et al. Infection fatality risk of the pandemic a (h1n1) 2009 virus in hong kong. *American journal of epidemiology*, 177(8):834–840, 2013.

[37] Yi-Chi Wu, Ching-Sung Chen, and Yu-Jiun Chan. The outbreak of covid-19: An overview. *Journal of the Chinese Medical Association*, 83(3):217, 2020.

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

[39] 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.