

Optimal antiviral treatment strategies and the effects of resistance

Elsa Hansen^{1,*} and Troy Day^{1,2}

¹Department of Mathematics and Statistics, and ²Department of Biology, Queen's University, Jeffery Hall, Kingston, Ontario, Canada K7L 3N6

Recent pandemic planning has highlighted the importance of understanding the effect that widespread antiviral use will have on the emergence and spread of resistance. A number of recent studies have determined that if resistance to antiviral medication can evolve, then deploying treatment at a less than maximum rate often minimizes the outbreak size. This finding, however, involves the assumption that treatment levels remain constant during the entire outbreak. Using optimal control theory, we address the question of optimal antiviral use by considering a large class of time-varying treatment strategies. We prove that, contrary to previous results, it is always optimal to treat at the maximum rate provided that this treatment occurs at the right time. In general the optimal strategy is to wait some fixed amount of time and then to deploy treatment at the maximum rate for the remainder of the outbreak. We derive analytical conditions that characterize this optimal amount of delay. Our results show that it is optimal to start treatment immediately when one of the following conditions holds: (i) immediate treatment can prevent an outbreak, (ii) the initial pool of susceptibles is small, or (iii) when the maximum possible rate of treatment is low, such that there is little *de novo* emergence of resistant strains. Finally, we use numerical simulations to verify that the results also hold under more general conditions.

Keywords: optimal control; resistance; antiviral; treatment; influenza

1. INTRODUCTION

The current interest in influenza pandemics has emphasized the importance of understanding the implications of drug resistance for different types of public health interventions. Recent studies have considered the consequences of drug resistance on the effectiveness of several intervention strategies, including drug prophylaxis and treatment, vaccination, non-drug interventions, as well as combinations of interventions and multi-drug therapy [1–17]. Many of these studies have revealed interesting and unexpected behaviour. In particular, numerous studies have now demonstrated that, with respect to drug treatments, there is sometimes an intermediate optimal level of treatment that minimizes the total outbreak size [8,9,12–14].

This result is somewhat counterintuitive, and arises from a trade-off between the costs and benefits of treating infected individuals. In fact, treatment has three effects on the total outbreak size: (i) treatment is beneficial because it suppresses the spread of the sensitive strain, (ii) treatment is costly because it leads to the *de novo* appearance of resistant infections, and (iii) treatment is costly because suppression of the sensitive strain frees up susceptible hosts that can then be infected by the resistant strain [9]. The treatment level that minimizes the total outbreak size should thus strike the optimal balance among these three effects.

These findings clearly reveal that the prospect of pathogen evolution during a disease outbreak can have a significant impact on the design of optimal intervention

strategies. Nevertheless, it is important to note that most of these results make a key implicit assumption; namely that, whatever treatment level is chosen, it must remain constant during the entire course of the outbreak. An interesting exception is the studies by Moghadas *et al.* [12,13], where numerical examples are used to illustrate that an even smaller total outbreak size can sometimes be obtained by switching from one level of treatment to another once the outbreak is underway.

Given the above collection of results, it is clear that the optimal treatment strategy during an outbreak is not yet completely understood. In this paper, we address this issue through the use of optimal control theory [18–21]. In particular, we consider a model in which we allow for arbitrary, time-varying, treatment strategies during an outbreak, and we derive analytical expressions that characterize the form of the optimal strategy. We prove that, contrary to the implications of previous studies, it is always optimal to use maximum treatment levels *provided that this treatment occurs at the right time*. In general, the optimal strategy is to wait some fixed period of time once the outbreak has begun, and then to deploy treatment at a maximum level. We show that the optimal delay is one that balances the above-mentioned costs and benefits. Finally, we explore extensions of the model to more complex scenarios in order to illustrate the robustness of our conclusions.

2. MATHEMATICAL MODEL

The model, depicted in figure 1*a*, describes the basic transmission dynamics of an infectious disease and includes the effects of treatment, as well as the presence of a strain resistant to treatment. An individual can be susceptible (*S*), infected with a sensitive strain (*I*),

* Author for correspondence (elsa@mast.queensu.ca).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.1469> or via <http://rspb.royalsocietypublishing.org>.

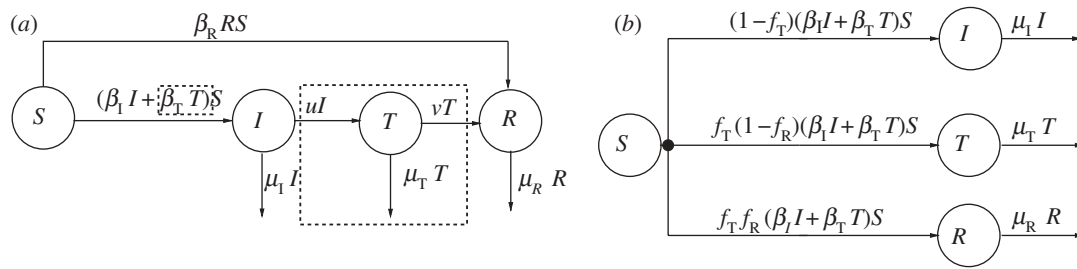


Figure 1. Model structure. (a) Schematic for the treatment model that is used to motivate model (2.1). An individual can be susceptible (S), infected with a sensitive strain (I), infected with a sensitive strain and treated (T) or infected with a resistant strain (R). The dynamics of the treated class have been enclosed in dashed boxes to emphasize the difference between this figure and the equations for model (2.1). All analytic results are derived using model (2.1), which does not include the dynamics of the treated class. (b) Schematic for the detailed model. Using numerical simulations, the analytic results for model (2.1) have been extended to this more detailed model that includes the dynamics of the treated class.

infected with a sensitive strain and treated (T) or infected with a resistant strain (R). Susceptible individuals become infected with the sensitive strain through contact with individuals in the I and T compartments (with contact parameters β_I and β_T , respectively), and become infected with the resistant strain through contact with individuals in the R compartment (with contact parameter β_R). Individuals infected with the sensitive strain can move into the treated class at some (time-varying) *per capita* rate, $u(t)$, where u is the treatment strategy of interest. Treated individuals can develop resistance at some constant *per capita* rate, ν , and finally, individuals in each of the three infected classes can leave the system, through either death or recovery with immunity, at constant *per capita* rates μ_I , μ_T and μ_R (also see electronic supplementary material, §1).

The objective of the analysis is to determine the optimal time-varying treatment strategy, u , that minimizes the total attack ratio (i.e. minimizes the attack ratio of the sensitive strain plus the attack ratio of the resistant strain), given by

$$A = \frac{S(t_0) - S(t_f)}{S(t_0)},$$

which is equivalent to minimizing the total outbreak size (and to maximizing the final number of susceptibles). The quantities t_0 and t_f are the start and end times of the outbreak, respectively. We say that the outbreak has ended as soon as the total number of all disease-transmitting individuals is equal to 1. In other words, for the model depicted in figure 1a, $t_f = \min\{t \mid I(t) + T(t) + R(t) \leq 1\}$. By defining t_f in this way, we avoid the situation where an unrealistically small number of infecteds (namely less than one infected) initiates an outbreak. The rationale for this definition of t_f is further detailed in electronic supplementary material, §2 and in Hansen & Day [22].

The treatment level, u , can vary through time, and can take on any value from 0 to some maximum value, u_{\max} . The upper bound, u_{\max} , reflects the fact that, regardless of how much we strive to increase the rate of treatment, it will typically be unavoidable that infected individuals spend some amount of time circulating in the population before they are treated. This will occur because of constraints on treatment delivery (e.g. having a finite number of health professionals available to deliver

treatment), as well as constraints imposed by the biology of the disease (e.g. if there is an asymptomatic stage).

The above model is qualitatively similar to those examined by Handel *et al.* [8], Lipsitch *et al.* [9], Moghadas *et al.* [13] and Qiu & Feng [14], but we make a further pair of simplifying assumptions to facilitate mathematical analysis: (i) that treatment is perfectly effective, meaning that essentially no transmission occurs from treated individuals, and (ii) that the time dynamics of the treated class are fast relative to those of the other classes. In this case, we can use the quasi-equilibrium value of T to simplify the model (electronic supplementary material, §1). This simplification removes the highlighted elements of the flow diagram in figure 1a, and yields the following equations:

$$\left. \begin{aligned} \dot{S} &= -\beta_I IS - \beta_R RS, \\ \dot{I} &= \beta_I IS - (\mu_I + u)I \\ \dot{R} &= \beta_R RS - \mu_R R + \kappa uI, \end{aligned} \right\}, \quad (2.1)$$

where κ denotes the probability that a treated individual develops resistance.

Although the above simplifying assumptions might not be particularly realistic for some situations, our primary aim is to develop a broad conceptual understanding of the optimal treatment strategy during an outbreak. Moreover, these assumptions do not appear to qualitatively alter the results, as evidenced by the numerical simulations presented later. In fact, these simplifications primarily accentuate the three key effects of treatment that were discussed in §1. First, model (2.1) assumes that treatment completely suppresses the spread of the sensitive strain and therefore it maximizes the first effect of treatment (i.e. the suppression of sensitive infections). Second, model (2.1) assumes that, if *de novo* resistance appears at all, it arises immediately upon treatment because the dynamics of the treated class are assumed to be fast relative to the other classes. Third, model (2.1) also removes the impact of treated individuals on the depletion of susceptible hosts, therefore maximizing the third effect of treatment (i.e. the freeing-up of susceptible hosts for infection by resistant strains).

3. RESULTS

To facilitate the understanding of our results, we first define some important quantities and briefly describe

some of the basic features of the optimal treatment strategy. We then give a complete characterization of the optimal treatment strategy in two steps. First, we consider a special case (situation 1) in which we can use standard techniques from calculus to gain an intuition for the results. Second, we characterize the optimal treatment strategy in general (situation 2). Analysis of the latter case requires techniques from optimal control theory, and these details are presented in the electronic supplementary material.

The optimal treatment strategy for model (2.1) is characterized by the following three quantities:

$$\begin{aligned}
 \text{(i)} \quad R_I &= \frac{\beta_I}{\mu_I} S(t_0) = \frac{S(t_0)}{S_I}; \\
 \text{(ii)} \quad R_R &= \frac{\beta_R}{\mu_R} S(t_0) = \frac{S(t_0)}{S_R}; \quad \text{and} \\
 \text{(iii)} \quad R_{IR} &= \left(\frac{\beta_I}{\mu_I + u_{\max}} + \frac{\beta_R}{\mu_R} \frac{\kappa u_{\max}}{\mu_I + u_{\max}} \right) S(t_0).
 \end{aligned}$$

Here, R_I and R_R are the basic reproduction numbers of the sensitive and resistant strain, respectively [23–25]. The quantities $S_I = \mu_I/\beta_I$ and $S_R = \mu_R/\beta_R$ represent threshold parameters, and denote the number of susceptibles above which a sensitive and resistant outbreak can occur, respectively. To understand the constant R_{IR} , consider the possible paths that a new sensitive infection can take. If the population is receiving treatment (i.e. $u \neq 0$), then an initially sensitive infection may develop resistance and so some initially sensitive infections will generate both sensitive and resistant infections during their lifetime. The constant R_{IR} can be interpreted as the total expected number of secondary infections (both sensitive and resistant) caused by an individual initially infected with the sensitive strain in a wholly susceptible population that is receiving maximum treatment (i.e. $u \equiv u_{\max}$). Specifically, the first term represents the number of infections generated while infected with the sensitive strain, and the second term represents the probability of developing resistance, multiplied by the number of infections generated while infected with this newly evolved resistant strain.

A few more general observations will be helpful before stating the results. First, R_{IR} can be expressed as $R_{IR} = R_I(1 - \chi) + \chi R_R$, where $\chi = u_{\max}/(\mu_I + u_{\max})$ is the probability of an infected individual receiving treatment when the treatment effort is $u \equiv u_{\max}$. This expresses R_{IR} as a mixture of the two reproduction numbers, R_I and R_R . Second, because the number of susceptibles always decreases through time, it is sometimes easier to express the optimal treatment strategy as a function of the size of the susceptible pool rather than as a function of time *per se*. This is also more useful from a health policy point of view because ‘number of susceptibles’ is a much more meaningful (and more easily measured) quantity than ‘time since start of outbreak’. Thus, although in the sequel we will discuss ‘treatment start time’ and ‘delay’, these will sometimes be measured in terms of the corresponding number of susceptible individuals. For reference all symbols have been defined in table 1.

(a) The optimal treatment strategy

The optimal treatment strategy is to delay treatment until some fixed time τ^* (possibly $\tau^* = t_0$) and then to treat maximally for the remainder of the epidemic. Thus, higher treatment levels are always better provided that the onset of treatment is delayed by the optimal amount.

(b) The optimal amount of delay

To simplify the conditions for the optimal delay, suppose that $R_I \geq R_R$ (the resistant strain suffers a fitness cost in the absence of treatment) and $\kappa < 1$ (not all treated infections develop resistance; the complete analysis is presented in the electronic supplementary material). In this case, the optimal delay can be characterized in terms of a critical number of susceptible hosts, defined by

$$S_{\min} = S_R + \frac{\kappa}{1 - \kappa} (S_R - S_I).$$

The optimal delay is achieved by balancing the effect that delay has on the sensitive and resistant outbreaks, and this is best understood by considering two separate situations: (1) the maximum treatment level is very large ($u_{\max} \rightarrow \infty$) and (2) the maximum treatment level is not very large (u_{\max} is finite).

Situation 1. If the maximum treatment level is very large, then some insight into the optimal treatment strategy can be gained using basic calculus. In particular, if the maximum level of treatment is used, then the onset of treatment effectively ends the sensitive epidemic and initiates the start of the resistant epidemic. This is because at the onset of treatment, all sensitive-infected individuals enter either the removed compartment (i.e. they are treated and no longer transmitting the infection) or the resistant compartment (i.e. treatment causes them to immediately develop resistance). Therefore, before treatment begins, model (2.1) simplifies to

$$\left. \begin{aligned}
 \dot{S} &= -\beta_I IS \\
 \dot{I} &= \beta_I IS - \mu_I I,
 \end{aligned} \right\} \tag{3.1}$$

with initial conditions $(S(t_0), I(t_0))$. Once treatment begins at time $t = \tau$, all of the sensitive-infected individuals either (i) immediately become resistant or (ii) immediately stop transmitting the infection. Thus, for $t > \tau$, model (2.1) simplifies to

$$\left. \begin{aligned}
 \dot{S} &= -\beta_R RS \\
 \dot{R} &= \beta_R RS - \mu_R R,
 \end{aligned} \right\} \tag{3.2}$$

with $(S(\tau), R(\tau)) = (S(\tau), \kappa I(\tau))$. Hence, in this special case, determining the optimal treatment strategy amounts to determining the optimal time to switch from model (3.1) to model (3.2).

Integrating and rearranging the equations for model (3.1) and model (3.2) give (see electronic supplementary material, §3, for details):

$$S_R \ln S(t_f) - S(t_f) = (1 - \kappa)(S_{\min} \ln S(\tau) - S(\tau)) + C, \tag{3.3}$$

where $C = \kappa(S_I \ln S(t_0) - S(t_0) - I(t_0)) + R(t_f)$. Now since, by definition, the outbreak ends as soon as the total number of infecteds is less than or equal to 1, there are two possibilities with respect to the optimal delay.

Table 1. Table of symbols.

symbol	brief definition
$S(t)$	number of susceptibles at time t
$I(t)$	number of individuals infected with the treatment-sensitive strain at time t
$T(t)$	number of treated individuals at time t
$R(t)$	number of individuals infected with the treatment-resistant strain at time t
β_I	contact parameter for the sensitive strain
β_T	contact parameter for the treated strain ($\beta_T = 0$ for the simple model)
β_R	contact parameter for the resistant strain
μ_I	<i>per capita</i> rate of death or recovery with immunity for the sensitive-infected class
μ_T	<i>per capita</i> rate of death or recovery with immunity for the treated class
μ_R	<i>per capita</i> rate of death or recovery with immunity for the resistant-infected class
$u(t)$	<i>per capita</i> rate of treatment for the sensitive-infected class at time t
v	<i>per capita</i> rate of developing resistance for the treated class (applies to model in figure 1a)
κ	probability that treated individual develops resistance (applies to simple model, see equation (2.1))
t_0	the outbreak start time
t_f	the outbreak end time
u_{\max}	upper bound for <i>per capita</i> treatment rate u
R_I	the basic reproduction number for the sensitive strain
R_R	the basic reproduction number for the resistant strain
R_{IR}	the expected number of secondary infections (both sensitive and resistant) caused by an individual initially infected with the sensitive strain, in a wholly susceptible population that is receiving maximum treatment (i.e. $u \equiv u_{\max}$)
S_I	the number of susceptibles above which a sensitive outbreak can occur
S_R	the number of susceptibles above which a resistant outbreak can occur
χ	the probability that an infected individual receives treatment (assuming $u \equiv u_{\max}$)
τ	an arbitrary treatment start time
τ^*	the optimal treatment start time
f_T	the fraction of infected individuals that receives treatment (applies to detailed model, see figure 1b)
S_{\min}	the number of susceptibles at the optimal treatment start time assuming that an outbreak occurs and u_{\max} is unbounded (applies to simple model)
$S_{\min,c}$	the number of susceptibles at the optimal treatment start time assuming that an outbreak occurs and $f_T = 1$ (applies to detailed model)
A	the total attack ratio

(A) *Treating immediately can prevent the outbreak:* if $\kappa I(t_0) < 1$, then starting treatment at t_0 prevents the outbreak (i.e. $R(\tau) = \kappa I(\tau) < 1$). Clearly, this is the best possible scenario, and therefore $\tau^* = t_0$. Notice that this situation is closely related to the idea of ‘containment’ discussed in Lipsitch *et al.* [9].

(B) *Treating immediately cannot prevent the outbreak:* if the outbreak cannot be prevented through treatment, then any treatment strategy will result in $S(t_f) \leq S_R$ (i.e. the outbreak will not end until the pool of susceptibles is diminished below that necessary to sustain a resistant outbreak). Therefore, the treatment start time that maximizes the left-hand side of equation (3.3) will also maximize $S(t_f)$ (and hence minimize the total attack ratio). In other words, it is optimal to delay treatment until the number of susceptible hosts has declined to

$$S(\tau^*) = S_{\min}.$$

If the initial number of susceptible hosts is smaller than this threshold, then treatment should begin immediately.

Situation 2. If the maximum treatment level is bounded, then there will be some temporal overlap in the sensitive and resistant epidemics once treatment has been initiated. In this situation, Pontryagin’s maximum principle can be used to determine the optimal (time-varying) treatment strategy (see electronic supplementary material, §4, for details). There are, again, two possibilities with respect to the optimal delay.

(A’) *Treating immediately can prevent the outbreak:* if starting treatment at t_0 prevents the outbreak from occurring, then this is clearly the best possible scenario, and so $\tau^* = t_0$. As with situation 1, this situation is closely related to the idea of ‘containment’ discussed in Lipsitch *et al.* [9] and is contingent on how t_f is defined (i.e. for model (2.1) we have that $t_f = \min\{t | I(t) + R(t) \leq 1\}$).

(B’) *Treating immediately cannot prevent the outbreak:* if scenario (A’) does not hold, then it may be optimal to delay the onset of treatment. In particular, the threshold number of susceptibles S_{\min} now provides an upper bound on the optimal delay. Thus, if the initial number of susceptible hosts is smaller than this threshold, then treatment should begin immediately. On the other hand, if $S(t_0)$ is initially larger than S_{\min} , then the optimal delay is no larger than the time it takes for the number of susceptible hosts to decline to S_{\min} . There is one other possibility, however, that overrides these conditions: if $R_{IR} > R_R$, then it is optimal to start treatment immediately.

Note that there are two major differences between the results for situation 1 and situation 2. First, when maximum treatment levels are bounded (situation 2), the critical threshold, S_{\min} , provides only an upper bound for the optimal delay. Second, in situation 2, there is an additional condition under which immediate treatment is optimal; namely, when $R_{IR} > R_R$. Intuitively, if *de novo* resistance has a negligible effect on the outbreak, then it seems reasonable that treatment should proceed

as though there is no de novo resistance. The condition that treatment should begin immediately if $R_{IR} > R_R$ actually highlights this intuition and emphasizes that the quantity that measures the effect of de novo resistance is the rate of de novo resistance (κuI) and not the probability of de novo resistance (κ). This can be seen by rearranging $R_{IR} > R_R$ and multiplying both sides by κI to produce a relationship involving the rate of de novo resistance:

$$\kappa u_{\max} I < \frac{\kappa}{1 - \kappa} \left(\frac{\mu_R}{\beta_R} - \frac{\mu_I}{\beta_I} \right) \beta_I I. \tag{3.4}$$

Hence, if the rate of de novo resistance is small, then treatment should proceed as though there is only a treatment-sensitive strain (i.e. if equation (3.4) holds then treatment should start immediately).

(c) Interpretation

We have shown that, for a simple model which includes the evolution of resistance, the optimal treatment strategy is to delay treatment for a specific amount of time (possibly no time) and then to treat with maximum effort for the remainder of the epidemic. If an outbreak is unavoidable, the optimal amount of delay is determined by balancing the benefits and costs of delay, and this trade-off can be best understood by again considering the case when the maximum treatment level is very large.

When the maximum treatment level is very large (i.e. situation 1), the marginal change in the attack ratio that comes from increasing the amount of delay before treatment is (see electronic supplementary material, §3, for details):

$$\frac{dA}{d\tau} = \frac{dA_I}{d\tau} + \frac{dA_R}{d\tau}, \tag{3.5}$$

$$= \frac{d}{d\tau} \left(\frac{S(t_0) - S(\tau)}{S(t_0)} \right) + \frac{d}{d\tau} \left(\frac{S(\tau) - S(t_f)}{S(t_0)} \right), \tag{3.6}$$

$$= \alpha[S_R - S(\tau)] - \alpha\kappa[S_I - S(\tau)], \tag{3.7}$$

where A, A_I, A_R denote the total, sensitive and resistant attack ratios, respectively, τ is the treatment start time and $\alpha = S(t_f)\beta_I I(\tau)/[S(t_0)(S_R - S(t_f))] > 0$.

Equation (3.5) emphasizes that the total attack ratio can be decomposed into a sensitive attack ratio and a resistant attack ratio. Furthermore, since for situation 1 the sensitive and resistant outbreaks are temporally separated (because the resistant strain emerges only after treatment begins), the sensitive attack ratio includes all new infections that occur before the treatment start time τ , and the resistant attack ratio includes all new infections that occur after the treatment start time τ (see equation (3.6)). The two terms in equation (3.7) represent the benefit and cost of delay, respectively. The first term represents the benefit of a delay that comes from a reduction in the size of the susceptible pool at the start of the resistant-strain epidemic. From the perspective of this benefit term only, it is best to delay treatment until $S(\tau) = S_R$, at which point there are no longer enough susceptible individuals to sustain a resistant outbreak. The second term represents the cost of delay that comes from an increase in the number of de novo resistant infections once treatment begins. From the perspective of this cost term only, it is best to treat immediately because

waiting longer results in a larger pool of infected individuals, which then have the potential to develop de novo resistance once treatment begins. The optimal amount of delay precisely balances these two effects, and this can be readily interpreted graphically. Figure 2a shows the final number of susceptibles as a function of treatment start time for different values of κ . As κ increases, the importance of the second term in equation (3.7) increases and correspondingly the number of infecteds at the optimal treatment start time decreases (figure 2b), while the number of susceptibles at the optimal treatment start time increases (figure 2c).

(d) Numerical results for more detailed models

Although the analytic expressions presented above are derived using a simple model, numerical simulations suggest that these results hold more generally. To illustrate this, we use numerical simulations to compare the relationship between treatment start time and attack ratio for our simple model and for the more detailed model in Lipsitch *et al.* [9]. Figure 1b depicts the treatment model used in Lipsitch *et al.* [9] and the corresponding system equations are provided in electronic supplementary material, §5.

To begin, numerical results show that both models exhibit two key features: (i) if the treatment level is constrained to be constant throughout the entire epidemic (as has been the case in most previously published analyses), then there is an intermediate optimal level of treatment, and (ii) if treatment is delayed by the appropriate amount, then maximum treatment is always optimal (electronic supplementary material, figures 1 and 2). The fact that delaying treatment can be better than constant treatment in both models explains why Lipsitch *et al.* [9], Moghadas *et al.* [13], Handel *et al.* [8] and Qiu & Feng [14] observed an intermediate optimal constant treatment level.

A comparison of the panels of figure 3 shows that the general relationship between the treatment level, the attack ratio and the number of susceptible hosts at the start of treatment is qualitatively similar for both models. Furthermore, figure 3a shows that for the simple model, if an outbreak occurs, then $S(\tau^*)$ approaches S_{\min} as the treatment level increases (i.e. the solid curve approaches the horizontal dashed line). In other words, the optimal solution for the case when u_{\max} is very large provides a bound for the case when u_{\max} is finite. A similar statement is true for the more detailed model. In the more detailed model, f_T denotes the fraction of infections that are treated, and so $f_T = 1$ is the analogous case to u_{\max} being unbounded (i.e. all infected individuals are treated). Using $S_{\min,c}$ to denote the number of susceptibles at the optimal treatment start time when $f_T = 1$, figure 3c shows that, for the detailed model, $S(\tau^*)$ decreases as the treatment level increases, with $S_{\min,c}$ providing a bound for all other values of f_T .

Both models also demonstrate that it is possible to avoid an outbreak by starting treatment immediately at sufficiently high treatment levels (electronic supplementary material, figure 3). Furthermore, if the initial pool of susceptibles is small enough, then it is always best to start treatment immediately (electronic supplementary

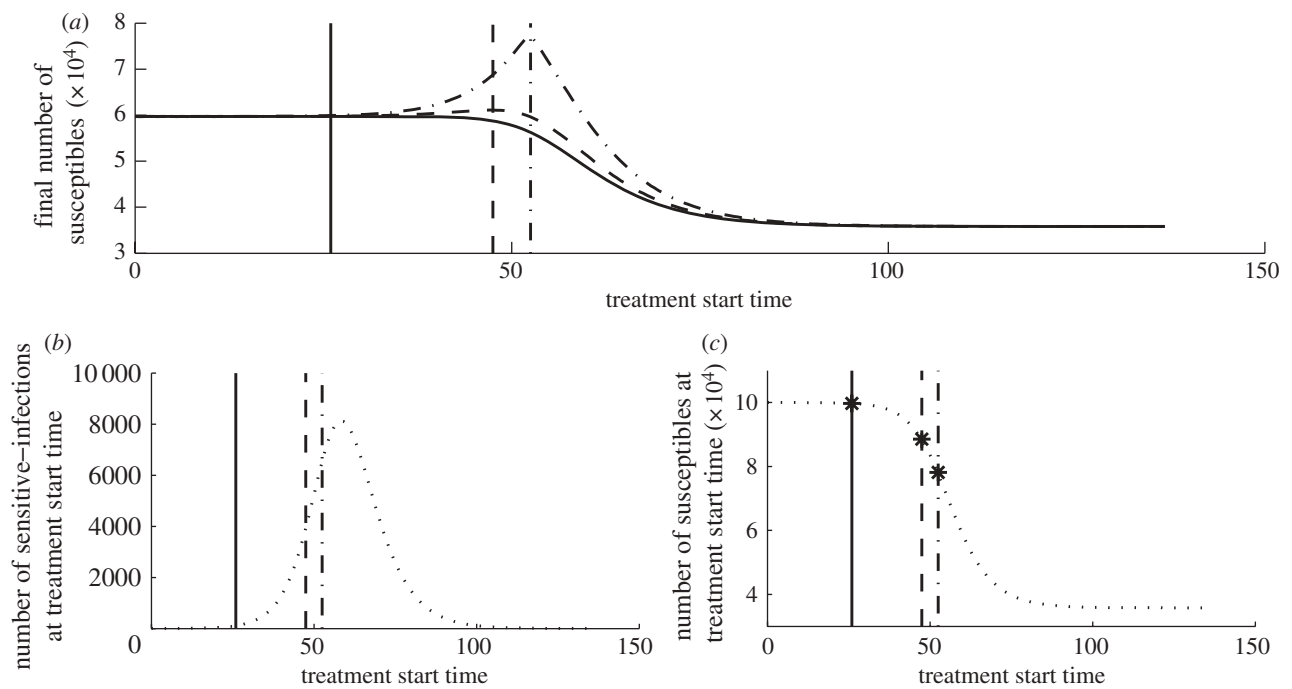


Figure 2. Effect of treatment start time on total attack rate (situation 1B). The vertical lines indicate the optimal treatment start time τ^* for a large κ (solid line), an intermediate κ (dashed line) and a small κ (dotted-dashed line). (a) The final number of susceptibles as a function of treatment start time for a large κ ($\kappa = 0.58$; solid curve), an intermediate κ ($\kappa = 0.4$; dashed curve) and a small κ ($\kappa = 0.001$; dotted-dashed curve). (b) The number of sensitive infections as a function of treatment start time, τ . The total attack rate is decreased by decreasing the number of sensitive infections at the treatment start time ($I(\tau)$) and by decreasing the number of susceptibles at the treatment start time ($S(\tau)$) (equation (3.7)). As κ increases, the effect of decreasing $I(\tau)$ becomes more important than decreasing $S(\tau)$; therefore, τ^* decreases as κ increases (from left to right, the order of the vertical lines is solid, dashed and dot-dashed). (c) The number of susceptibles as a function of treatment start time. As κ increases, $S(\tau^*)$ increases. Also, the points $S = S_{\min}$ are indicated by 'star' markers and coincide with $S(\tau^*)$. Parameters are $R_I = 1.6$, $R_R = 0.8R_I$ and $\mu_I = \mu_R = 1/3.3$.

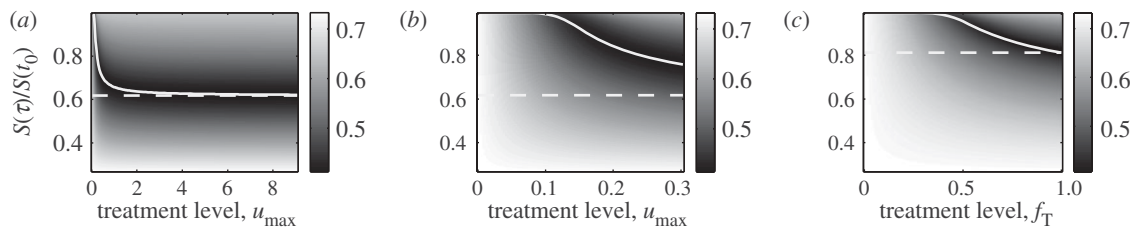


Figure 3. Total attack ratio versus treatment level and number of susceptibles at treatment start time (normalized by $S(t_0)$). The white solid curves show the number of susceptibles at the optimal treatment start time provided a resistant epidemic occurs. The white dashed horizontal lines indicate the number of susceptibles at the optimal treatment start time when the maximum possible treatment level is very large (i.e. when u_{\max} is unbounded for the simple model and when $f_T = 1$ for the detailed model). The white dashed horizontal lines in (a) and (b) were computed analytically; all other curves were computed numerically. It is important to emphasize that this figure was generated by assuming that an outbreak occurs. Indeed, for this specific choice of parameters, electronic supplementary material, figure 3, illustrates that for $u_{\max} > 0.25$ and $f_T > 0.8$ treating immediately will prevent an outbreak and so treating immediately is optimal. (a) Figure produced using simple model. (b) Magnified version of (a). (c) Figure produced using detailed model. Parameters are $R_R = 0.9R_I$, $\mu_1 = \mu_2 = \mu_3 = 1/3.3$, $\kappa = 0.0066$, $f_I = 0.002$.

material, figure 4). For the simple model, 'small enough' is measured relative to the number of susceptibles at the optimal treatment start time when u_{\max} is unbounded. Namely, if $S(t_0) < S_{\min}$, then $\tau^* = t_0$. Similarly, for the more detailed model, electronic supplementary material, figure 4b, suggests that if $S(t_0) < S_{\min,c}$ then again it is optimal to start treatment immediately.

Finally, recall that for the simple model if $R_{IR} > R_R$, then it is optimal to start treatment immediately. Using the biological interpretation of R_{IR} , namely that R_{IR} is

the expected total number of infections caused by an individual initially infected with the treatment-sensitive strain, we can define an analogous quantity for the detailed model (see electronic supplementary material, text S1, for details). Electronic supplementary material, figures 5 and 6 show that the relationship between R_R , R_{IR} and the optimal treatment start time is similar for both the simple and detailed models. This result is particularly interesting because it emphasizes that (i) using a simple model can highlight relationships in more detailed

models that may otherwise be undetected, and (ii) the quantity R_{IR} is of general importance and not just an artefact of our simple model.

4. DISCUSSION

Perhaps the most salient feature of previous analyses is that the best treatment somehow balances the effects of the sensitive and resistant epidemics. This trade-off, which is very clearly explained in the discussion found in Lipsitch *et al.* [9], is also highlighted in our model. The difference is in the way this trade-off is resolved. In Lipsitch *et al.* [9], this trade-off is used to explain why an intermediate level of constant treatment is better than a high level of constant treatment. Conversely, we have shown that the optimal treatment balances this trade-off by delaying the onset of treatment instead of capping the maximum level of treatment. By allowing the treatment strategy to vary in time, we have shown that higher levels of treatment are always better, provided they are started at the appropriate time. Furthermore, when the maximum possible treatment level is very large, our analysis shows that the optimal treatment strategy is the one that switches from a sensitive outbreak to a resistant outbreak precisely when the number of susceptible hosts has decreased to the value given by S_{\min} .

We have examined two types of model, a simple model that does not explicitly include the dynamics of the treatment class and a more detailed model that does include treatment dynamics. For the simple model, we showed, using analytic calculations, that the optimal treatment strategy is ‘off–on’ (i.e. delay treatment for a certain amount of time—possibly no time—and then treat maximally for the remainder of the outbreak). It is important to emphasize that for model (2.1) this off–on strategy is indeed optimal in the sense that any other, more complicated, strategy that involves intermediate levels of treatment will not result in a lower attack ratio. Conversely, for the more detailed model, we have only shown (via numerical simulations) that an off–on strategy can outperform a ‘constant’ strategy. Thus, although it seems likely that the off–on strategy is optimal for the more detailed model as well it remains an open problem to explicitly prove that this is indeed the case. Furthermore, although an off–on strategy can be better than a constant strategy, policy makers still face the ethical dilemma of choosing between the rights of individuals to receive treatment and the benefit to the population of delaying treatment.

Given the ethical problems associated with implementing the optimal treatment strategy when it involves a delay, it is useful to clearly delineate the conditions under which implementing treatment immediately is actually optimal. It is always optimal to start treatment immediately if one of three conditions holds: (i) immediate treatment can prevent an outbreak, (ii) the initial pool of susceptibles is small (i.e. $S(t_0) < S_{\min}$ for the simple model and $S(t_0) < S_{\min,c}$ for the detailed model), or (iii) the rate of de novo resistance is small (i.e. $R_{IR} > R_R$). Each of these three situations have relevance to public health policy.

Condition (i) underscores that our focus has been to minimize the total outbreak size, and so in this context preventing an outbreak is the best possible outcome.

This outcome, however, may not be desirable if there is a high probability of a second epidemic occurring at a later time. For example, although treating immediately at a very high level may minimize the total attack ratio, this will result in $S(t_f) > S_R > S_I$. As a result, the population will be susceptible to future sensitive and resistant outbreaks. Alternatively, if a treatment strategy is chosen to ensure that a resistant epidemic occurs, then the final number of susceptibles will be small enough to prevent future resistant outbreaks (and the severity of a possible future sensitive outbreak will be significantly decreased). Antia *et al.* [26] discusses these issues in the context of a single-strain epidemic. Interestingly, if delay is optimal, then, up to a point, increasing u_{\max} will increase the optimal amount of delay. However, if u_{\max} can be increased sufficiently, then treating immediately will effectively prevent the outbreak, and this then becomes the optimal strategy. If this can occur, then policy makers will have to balance the benefits of treating immediately (and therefore preventing the current outbreak) with the benefits of delaying treatment (and perhaps minimizing the effects of a second outbreak).

Condition (ii) highlights that the optimal treatment strategy depends on the size of the target community. Essentially, if a community is small enough, then treatment should start immediately. This observation leads to an interesting hypothesis for combined vaccination–treatment policies. Namely, since vaccination decreases the susceptible population, low vaccination levels make it more likely that treatment should be delayed. This proposed relationship is surprising since, naively, it seems reasonable to expect that if fewer individuals are vaccinated, then it would be better to start treatment early in order to reduce spread. Nevertheless, our results suggest that this intuition is, in fact, incorrect.

Condition (iii) emphasizes that if the probability of receiving treatment is low, then it is more likely that treatment should start immediately. More concretely, by expressing R_{IR} as a linear combination of the two reproduction numbers R_I and R_R , we can rewrite $R_{IR} > R_R$ as

$$\frac{R_I}{R_R} > \frac{1 - \kappa\chi}{1 - \chi}.$$

Therefore, if $R_I/R_R > (1 - \kappa\chi)/(1 - \chi)$, it is optimal to start treatment immediately and if $R_I/R_R \leq (1 - \kappa\chi)/(1 - \chi)$, then it may be best to delay treatment. For any specific values of R_I , R_R and κ , we see that as the probability of treating an infected individual, χ , decreases, it becomes more likely that starting treatment immediately is best. Intuitively, this guideline makes sense since if treatment is unlikely to reach many infected individuals, then it makes sense to start administering treatment as soon as possible in order to reach more individuals.

Since multiple intervention measures are often used to control a disease, an important next step is to consider a model that includes other intervention measures, in addition to treatment [4,5,7,10]. Some models similar to our model that have included multi-intervention strategies are the treatment–vaccination models in Ferguson *et al.* [6] and Qiu & Feng [14], and the treatment–prophylaxis model in Lipsitch *et al.* [9]. Interestingly, a number of studies that consider multi-intervention strategies also exhibit optimal constant intervention levels. This suggests

that considering multi-intervention strategies that vary in time may be very informative as well.

REFERENCES

- 1 Alexander, M. E., Bowman, C. S., Feng, Z., Gardam, M., Moghadas, S. M., Rst, G., Wu, J. & Yan, P. 2007 Emergence of drug resistance: implications for antiviral control of pandemic influenza. *Proc. R. Soc. B* **274**, 1675–1684. (doi:10.1098/rspb.2007.0422)
- 2 Bergstrom, C., Lo, M. & Lipsitch, M. 2004 Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proc. Natl Acad. Sci. USA* **101**, 13 285–13 290. (doi:10.1073/pnas.0402298101)
- 3 Bonhoeffer, S., Lipsitch, M. & Levin, B. 1997 Evaluating treatment protocols to prevent antibiotic resistance. *Proc. Natl Acad. Sci. USA* **94**, 12 106–12 111. (doi:10.1073/pnas.94.22.12106)
- 4 Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsrithaworn, S. & Burke, D. S. 2005 Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214. (doi:10.1038/nature04017)
- 5 Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C. & Burke, D. S. 2006 Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452. (doi:10.1038/nature04795)
- 6 Ferguson, N. M., Mallett, S., Jackson, H., Roberts, N. & Ward, P. 2003 A population-dynamic model for evaluating the potential spread of drug-resistant influenza virus infections during community-based use of antivirals. *J. Antimicrob. Chemother.* **51**, 977–990. (doi:10.1093/jac/dkg136)
- 7 Germann, T. C., Kadau, K., Longini, I. M. & Macken, C. A. 2006 Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103**, 5935–5940. (doi:10.1073/pnas.0601266103)
- 8 Handel, A., Longini, I. M. & Antia, R. 2009 Antiviral resistance and the control of pandemic influenza: the roles of stochasticity, evolution and model details. *J. Theor. Biol.* **256**, 117–125. (doi:10.1016/j.jtbi.2008.09.021)
- 9 Lipsitch, M., Cohen, T., Murray, M. & Levin, B. R. 2007 Antiviral resistance and the control of pandemic influenza. *PLoS Med.* **4**, 0111–0121.
- 10 Longini Jr, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaworakul, W., Cummings, D. A. T. & Halloran, M. E. 2005 Containing pandemic influenza at the source. *Science* **309**, 1083–1087. (doi:10.1126/science.1115717)
- 11 McCaw, J. M., Wood, J. G., McCaw, C. T. & McVernon, J. 2008 Impact of emerging antiviral drug resistance on influenza containment and spread: influence of subclinical infection and strategic use of a stockpile containing one or two drugs. *PLoS ONE* **3**, e2362. (doi:10.1371/journal.pone.0002362)
- 12 Moghadas, S. 2008 Management of drug resistance in the population: influenza as a case study. *Proc. R. Soc. B* **275**, 1163–1169. (doi:10.1098/rspb.2008.0016)
- 13 Moghadas, S. M., Bowman, C. S., Röst, G. & Wu, J. 2008 Population-wide emergence of antiviral resistance during pandemic influenza. *PLoS ONE* **3**, e1839. (doi:10.1371/journal.pone.0001839)
- 14 Qiu, Z. & Feng, Z. 2010 Transmission dynamics of an influenza model with vaccination and antiviral treatment. *Bull. Math. Biol.* **72**, 1–33. (doi:10.1007/s11538-009-9435-5)
- 15 Regoes, R. R. & Bonhoeffer, S. 2006 Emergence of drug-resistant influenza virus: population dynamical considerations. *Science* **312**, 389–391. (doi:10.1126/science.1122947)
- 16 Stilianakis, N., Perelson, A. & Hayden, F. 1998 Emergence of drug resistance during an influenza epidemic: insights from a mathematical model. *J. Infect. Dis.* **177**, 863–873.
- 17 Wu, J. T., Leung, G. M., Lipsitch, M., Cooper, B. S. & Riley, S. 2009 Hedging against antiviral resistance during the next influenza pandemic using small stockpiles of an alternative chemotherapy. *PLoS Med.* **6**, e1000085. (doi:10.1371/journal.pmed.1000085)
- 18 Berkovitz, L. D. 1974 *Optimal control theory, applied mathematical sciences*, vol. 12. New York, NY: Springer.
- 19 Kamien, M. I. & Schwartz, N. L. 2003 *Dynamic optimization. The calculus of variations and optimal control in economics and management*. Advanced Textbooks in Economics, no. 31. Amsterdam, The Netherlands: Elsevier.
- 20 Lenhart, S. & Workman, J. 2007 *Optimal control applied to biological models*. Boca Raton, FL: Chapman and Hall.
- 21 Pontryagin, L. S., Boltyanski, V. G., Gamkrelidze, R. V. & Mishchenko, E. F. 1964 *The mathematical theory of optimal processes*. International Series of Monographs in Pure and Applied Mathematics, no. 55. New York, NY: The Macmillan Company.
- 22 Hansen, E. & Day, T. In press. Optimal control of epidemics with limited resources. *J. Math. Biol.* (doi:10.1007/s00285-010-0341-0)
- 23 Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford Science Publications.
- 24 Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and the computation of the basic reproductive ratio r_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382. (doi:10.1007/BF00178324)
- 25 Hethcote, H. W. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
- 26 Antia, R., Handel, A. & Longini, I. 2007 What is the best control strategy for multiple infectious disease outbreaks. *Proc. R. Soc. B* **274**, 833–837. (doi:10.1098/rspb.2006.0015)