CrossMark

The optimal treatment of an infectious disease with two strains

Robert Rowthorn[1] · **Selma Walther**[2]

Abstract This paper explores the optimal treatment of an infectious disease in a Susceptible-Infected-Susceptible model, where there are two strains of the disease and one strain is more infectious than the other. The strains are perfectly distinguishable, instantly diagnosed and equally costly in terms of social welfare. Treatment is equally costly and effective for both strains. Eradication is not possible, and there is no superinfection. In this model, we characterise two types of fixed points: coexistence equilibria, where both strains prevail, and boundary equilibria, where one strain is asymptotically eradicated and the other prevails at a positive level. We derive regimes of feasibility that determine which equilibria are feasible for which parameter combinations. Numerically, we show that optimal policy exhibits switch points over time, and that the paths to coexistence equilibria exhibit spirals, suggesting that coexistence equilibria are never the end points of optimal paths.

Keywords Multiple strain infection · Economic epidemiology · Treatment · Optimal control

Mathematics Subject Classification 91B55 · 91B76 · 93C15

✉ Selma Walther
s.walther@warwick.ac.uk

Robert Rowthorn
rer3@econ.cam.ac.uk

[1] Faculty of Economics, University of Cambridge, Cambridge, UK

[2] Department of Economics, University of Warwick, Coventry CV4 7AL, UK

1 Introduction

Several infectious diseases that pose a pressing problem in developing countries exist in multiple strains, including tuberculosis, HIV, Lyme disease, Human Papillomavirus, Helicobacter pylori and influenza (Balmer and Tanner 2011). Where resources are limited, a uniform treatment policy at the national level may be the most efficient way to tackle these diseases. This paper addresses the following question: what is the optimal treatment policy for an infectious disease that has two strains, which differ in their infectiousness? We take a standard model in epidemiology, the Susceptible-Infected-Susceptible (SIS) model, and extend it to include two strains of infection with different transmission probabilities. In the standard SIS model, individuals move between two states, susceptible and infected, based on exogenous probabilities. The probability of an individual catching a disease when he encounters an infected person depends on an infectivity or transmission parameter. This parameter is usually assumed to be homogeneous, which does not allow for policy differentiation if the disease exists in several strains.

The SIS model is often used to describe sexually transmitted diseases such as chlamydia. *Chlamydia trachomatis* is an organism that has three human strains: one strain causes *trachoma*, which is an infection of the eyes; a second strain causes sexually transmitted urethritis, often referred to simply as *chlamydia*, while a third strain causes the sexually transmitted *lymphogranuloma venereum* (*LGV*) (Byrne 2010). Chlamydia, caused by serovars D-K, is also the most common sexually transmitted disease in the United States and the United Kingdom, with over 1.2 million infections detected in the former and over 200 thousand in the latter.[1] It can result in serious complications, such as pelvic inflammatory disease, which can lead to infertility if left untreated (Shaw et al. 2011). There are currently no vaccines for chlamydia, and treatment with antibiotics, usually with azithromycin or doxycycline, is necessary (Shaw et al. 2011). LGV is also treated with doxycycline (McLean et al. 2007), while trachoma is treated with azithromycin (Tabbara et al. 1996). Therefore, the optimal treatment mix for the various strains of chlamydia trachomatis is an important public health concern. It has been established in the literature that LGV strains are more infectious than trachoma strains, and that there may even be variation in virulence among trachoma strains (Pearce and Prain 1986; Kari et al. 2008). Reinfection with chlamydia is possible and superinfection does not appear to occur in practise (Heijne et al. 2011; Heiligenberg et al. 2014). Therefore, chlamydia trachomatis satisfies the assumptions of the model analysed in this paper, and as such, the results can provide guidance on the optimal treatment of the diseases that it causes.

We analyse the optimal treatment policy of a two-strain SIS disease, both analytically and numerically. We assume that the two strains differ in their infectivities, and neither superinfection nor eradication of the disease is possible. The policymaker can perfectly observe and target both strains with two separate treatment instruments. First, we provide theoretical results on the possible equilibria of this model. There are two sets of equilibria: non-boundary equilibria, where the two strains co-exist, and

---

[1] Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report (April 1 2011) and Public Health England, Infection Report, Volume 9 Number 22.

boundary equilibria, where one of the strains is endemic while the other is eliminated asymptotically. The feasibility of these equilibria depends on parameter conditions. Broadly speaking, when the infectivities of the two strains are very different, only the boundary equilibria can be achieved, and in particular those where the more infective strain prevails and the less infective strain is asymptotically eradicated. When the strains are quite similar, all equilibria are feasible. We demonstrate that in the coexistence equilibria, the more infective strain must receive more treatment than the less infective strain. These results lie in contrast to the equilibrium with no intervention, when the less infectious strain dies out.

We provide numerical examples where we simulate the model and look for optimal treatment policy along the path to the fixed points. We find that policy exhibits switch points along the path, implying that policy flexibility is beneficial, and that a static model would be too restrictive. We are unable to find a path towards a coexistence equilibrium that is optimal, in the sense of not being a spiral. We tentatively conclude that the coexistence equilibria can never be the end points of optimal paths, and that the boundary equilibria are always the optimal ones.

The ideas in this paper relate closely to the theory of competitive exclusion in ecology, which postulates that the coexistence of two or more species limited by one resource (in our case, two strains limited by one host population) is not possible (Armstrong and McGehee 1980). For example, Iggidr et al. (2006) demonstrate in the context of multiple-strain malaria that an equilibrium is either disease-free or has at most one surviving strain.

However, this is not always the case, provided some assumptions are relaxed, such as allowing for cycles in the persistence of species (Armstrong and McGehee 1980). In a model of renewable natural resources, coexistence of species is possible, with the stocks of the two resources higher in a mutualistic than a competitive equilibrium (Ströbele and Wacker 1991). In the case of an SIS model, coexistence of two strains is possible if one allows for superinfection and age structure (Li et al. 2010). Elbasha and Galvani (2005), in a model of two-strain Human Papillomavirus vaccines, show that when individuals are vaccinated for only one of the two strains, the equilibrium depends on whether the strains are synergistic (in which case vaccination also reduces the prevalence of the non-targeted strain) or antagonistic (where vaccination increases the prevalence of the other strain). Castillo-Chavez et al. (1999) demonstrate in a two-strain two-sex SIS model of a sexually transmitted disease that a unique coexistence equilibrium can occur under certain conditions. We show that in our two-strain SIS model, it is sometimes feasible for both strains to coexist, especially if they are not too different in their infectiousness. If one strain is much more infectious than the other, then it is only feasible to strive for the eradication of the less infectious strain, while allowing the more infectious strain to be endemic.

This paper builds on the economic epidemiology literature, which examines optimal intervention in tackling infectious diseases. Rowthorn et al. (2009) examine the optimal way to treat a disease with spatial dynamics, while Rowthorn (2006) considers the optimal treatment of a disease when policymakers have limited funds. There are important tradeoffs between treatment and vaccination as two instruments available to a policymaker (Rowthorn and Toxvaerd 2015). In general, a lack of intervention yields to inefficiently low uptake of vaccination (Chen and Toxvaerd 2014). While

there is a rich literature that describes disease dynamics in the absence of intervention, the study of *optimal intervention* is still in its early stages, and this paper contributes to that literature.

The biology literature has shown the pervasiveness of diseases that exist in multiple strains, for which treatment and vaccination may have to vary. Lyme disease has at least 14 known strains.[2] Seinost et al. (1999) present strong evidence that strains of Lyme disease have different infectivities. Other multiple-strain infections include Human Papillomavirus, which includes high-risk strains that are a leading cause of cervical cancer and low-risk strains that do not cause cancer, as well as tuberculosis, where some strains are particularly drug resistant (Castillo-Chavez and Feng 1997).[3] Influenza is a multiple-strain disease: Truscott et al. (2012) argue that accurate modelling of influenza should not neglect the presence of several strains. HIV also exists in two forms: HIV-1 and HIV-2. Studies show that the less common HIV-2 strain is also less infectious than its counterpart for most of its infectious period.[4] Accurate measurement of the transmission rate is the key to realistic simulations of HIV/AIDS prevalence rates for the USA and Africa (Oster 2005), which highlights both the importance of the transmission rate in describing an infection and the particular relevance of this paper to developing countries.

The paper proceeds as follows. Section 2 provides an overview of some theoretical concepts while Sect. 3 outlines the SIS model with two strains. Section 4 discusses the equilibria of the model and Sect. 5 provides examples of simulations. Section 6 concludes.

2 Relevant concepts

It is useful to review some relevant concepts before the central analysis of the paper. The SIS model belongs to the category of transmission system models, which are built on differential equations that describe the evolution of disease prevalence over time as a function of parameters. It has become standard to assume random mixing in models of this type: agents are equally likely to encounter any other agent in the population.[5] There are two possible states: individuals are either susceptible or infected. They can move between the two states an unlimited number of times. Agents are homogeneous and the population is closed. If a susceptible agent encounters an infected agent, the susceptible agent becomes infected with a certain probability, termed the transmission probability. Typically, these models assume a homogeneous transmission parameter for all agents. The standard SIS model is in continuous time. At every time increment, a proportion of infected individuals is treated with a certain success rate of

---

[2] Lyme Disease Association of Australia (http://lymedisease.org.au).

[3] Centers for Disease Control and Prevention, http://www.cdc.gov.

[4] More information on this can be found at the Centers for Disease Control and Prevention (http://www.cdc.gov/hiv/topics/basic) as well as the charity AVERT (http://www.avert.org/hiv-types.htm).

[5] See Keeling and Eames (2005) and Bansal et al. (2007) for a discussion of how mixing networks can be applied to the study of epidemiology.

treatment (which can be interpreted as a rate of recovery). There is also the possibility of spontaneous or natural recovery.

Our model differs from the standard SIS model in two key ways: the disease has two strains with distinct transmission rates and the policymaker has two policy instruments. We make three further assumptions about the nature of the disease: transmission is through human-to-human contact rather than vector-borne, superinfection is not possible and agents can recover fully from the disease. SIS models are often used to describe sexually transmitted diseases (Keeling and Eames 2005), and chlamydia is an appropriate example of an SIS disease with multiple strains (Byrne 2010).

Optimal policy determines what proportion of individuals to treat at each point in time subject to the cost of treatment, the value of susceptible and infected individuals to social welfare, as well as the transmission dynamics. Studies focus on deriving equilibria and looking for policy switches over time. Policy generally takes the form of a Most Rapid Approach Path (MRAP)—the policy that takes the system the fastest way to the desired equilibrium. Another point of interest is Skiba points or curves, which delineate points where the policymaker is indifferent between policies.[6]

An analysis of optimal treatment in the one-strain SIS model can be found in Rowthorn (2006) and Goldman and Lightwood (2002). We summarise their results here as a brief indication of what one might expect from a two-strain model. Two types of equilibria of the model exist: the policy instrument is either at a boundary, where no one or everyone is treated, or policy is interior, with partial treatment of the population. It is never optimal to treat partially; it can be shown that optimal policy will always take on one of the two boundary values. Whether it is optimal to treat everyone or no one depends on the parameters at hand.

3 The SIS model with two strains

The model has the following characteristics. The two variants of the infection are High (H, high infectivity) and Low (L, low infectivity). The more infectious variant H has transmission rate β_H while the less infectious variant L is characterised by transmission rate β_L.[7] The two policy instruments are f_H, $f_L \in [0, 1]$, which measure the proportion of the infectees in each strain that are treated. The success rate of treatment is $\alpha \in (0, 1)$. We assume that the policymaker has perfect information about the prevalence of each strain in the population and can therefore target treatment of each strain perfectly.[8]

Individuals can catch either strain at the outset. When infected, they transmit the strain that they themselves are infected with. There is a possibility of exogenous recovery at rate $\tau \in (0, 1]$. If individuals recover, they are again susceptible to either infection strain. The proportion of the total population infected with the High strain

---

[6] For more detail on these concepts, the reader is referred to Skiba (1978) and Spence and Starrett (1975).

[7] An index of notation and acronyms used in the paper can be found in Appendix A.

[8] Other assumptions are possible; for example, a policymaker may know an individual is ill but not which strain of the disease he has. In this case, the policymaker would effectively have only one policy instrument at hand. Manski (2013) discusses treatment under uncertainty.

is I_H. The proportion infected with the Low strain is I_L. To ease calculations, the total population is normalised to size 1: $I_H + I_L + S = 1$, where S denotes the population of healthy individuals (susceptibles). All parameters are strictly positive. The policymaker places value p on healthy individuals, value 0 on infected individuals and discounts the future at rate δ. She also faces a constant marginal cost c of treatment.

In order to choose the optimal policy mix of f_H and f_L, the policymaker maximises the social welfare function

$$V(I_H^0, I_L^0) = \int_0^\infty e^{-\delta t} (p * (1 - I_H(t) - I_L(t)) - c * (f_H(t)I_H(t) + f_L(t)I_L(t)))dt, \tag{1}$$

subject to the equations of motion for the two infection types:

$$\dot{I}_H = I_H(t) [\beta_H (1 - I_H(t) - I_L(t)) - \tau - \alpha f_H(t)], \tag{2}$$
$$\dot{I}_L = I_L(t) [\beta_L (1 - I_H(t) - I_L(t)) - \tau - \alpha f_L(t)]. \tag{3}$$

Thus, the population of individuals infected with H increases at rate $\beta I_H S$, which is a direct consequence of the assumption of random mixing, and decreases at rate $I_H(\tau + \alpha f_H)$. The population of agents infected with L changes in a symmetric way. We further assume that

$$I_H(0) = I_H^0 > 0 \text{ given},$$
$$I_L(0) = I_L^0 > 0 \text{ given},$$
$$\beta_H > \beta_L > \tau + \alpha. \tag{4}$$

Thus, we assume that the population begins with a positive level of prevalence of each strain. Assumption (4) states that the infection probability in an encounter between a healthy individual and a sick individual is higher than the probability of a treated sick individual recovering. This ensures that neither variant of the disease can be eliminated, even asymptotically, by treating all infected people (see Sect. 4.4 for a proof of this). Thus, $I_H(t), I_L(t) > 0$ for all t.

In order to provide a benchmark for the optimal policy analysis, we discuss two simplifications of the model. First, we consider what happens when the two strains are independent, so that agents can be infected with both strains i.e. superinfection is possible, and infection with one strain has no impact on infection with the other strain. In this case, the disease dynamics are governed by the following equations:

$$\dot{I}_H = \beta_H I_H(t)(1 - I_H(t)) - I_H(t)\tau - I_H(t)\alpha f_H(t),$$
$$\dot{I}_L = \beta_L I_L(t)(1 - I_L(t)) - I_L(t)\tau - I_L(t)\alpha f_L(t).$$

The equilibrium prevalence of the two strains is

$$I_H^* = 1 - \frac{\alpha f_H + \tau}{\beta_H},$$
$$I_L^* = 1 - \frac{\alpha f_L + \tau}{\beta_L}.$$

Thus, prevalence depends only on the infectiousness of that strain and the extent to which it is treated. For the same level of treatment, the total number of infections (where a person infected with both strains counts as two infections) is higher than in the case when superinfection is not possible.

Second, we consider what happens to the infection when no intervention is possible; that is, $f_H = f_L = 0$ at all points in time (but the infections are no longer independent). In this case, the two Eqs. (2) and (3) become

$$\dot{I}_H = \beta_H I_H(t)(1 - I_H(t) - I_L(t)) - I_H(t)\tau,$$
$$\dot{I}_L = \beta_L I_L(t)(1 - I_H(t) - I_L(t)) - I_L(t)\tau.$$

For each strain to be in equilibrium, we require $\dot{I}_H = \dot{I}_L = 0$. As long as $\beta_H \neq \beta_L$, this can never hold with strictly positive I_H and I_L when there is no intervention. This implies that the two strains can never co-exist. Instead, there are two possibilities. Either the Low strain will die out and the High strain will prevail, or both strains will die out. The Low strain dies out because the probability of infection in each encounter of a healthy individual with one infected with the Low strain is lower than in the healthy individual's encounter with one infected with the High strain. As a result, the High strain spreads more quickly and eventually causes the Low strain to disappear. The High strain will prevail if $\beta_H > \tau$ and die out if $\beta_H < \tau$. Assumption (4) ensures that $\beta_H > \tau$ and so the High strain will prevail in equilibrium when there is no treatment, with $I_H^* = 1 - \frac{\tau}{\beta_H}$.

To determine the optimal treatment policy in the general model with treatment, the policymaker solves the current value Hamiltonian:

$$
\begin{aligned}
Q = {} & p(1 - I_H - I_L) - c(f_H I_H + f_L I_L) \\
& + \lambda_H(\beta_H I_H(1 - I_H - I_L) - I_H(\tau + f_H \alpha) \\
& + \lambda_L(\beta_L I_L(1 - I_H - I_L) - I_L(\tau + f_L \alpha).
\end{aligned}
$$

As the control variables f_H and f_L enter the Hamiltonian in a linear fashion, there is a bang-bang solution. The first-order conditions describing the optimal choice of treatment at a given point in time are

$$
f_H \begin{bmatrix} = 0 \\ \in [0, 1] \\ = 1 \end{bmatrix} \quad \text{if } \lambda_H \begin{bmatrix} < \\ = \\ > \end{bmatrix} -\frac{c}{\alpha}, \tag{5}
$$

$$
f_L \begin{bmatrix} = 0 \\ \in [0, 1] \\ = 1 \end{bmatrix} \quad \text{if } \lambda_L \begin{bmatrix} < \\ = \\ > \end{bmatrix} -\frac{c}{\alpha}, \tag{6}
$$

where we drop time subscripts in these and many subsequent expressions for ease of notation. These conditions imply that if $\lambda_H < -\frac{c}{\alpha}$ for example, then it is optimal to treat no one infected with the High strain. On the other hand, if $\lambda_L = -\frac{c}{\alpha}$,

then the policymaker is indifferent between any level of treatment of the Low strain. These conditions can also be interpreted intuitively. Optimal treatment levels depend on the shadow price (λ) of having an additional infected individual in the population. This shadow price will differ for the two strains. Considering the H type as an example, the optimal value of f_H will depend on how $|\lambda_H^*|$ compares to $\frac{c}{\alpha}$.[9] The latter is the price of treating an individual divided by the probability that the treatment will be successful: it is the effective cost of treating the individual. It is constant for all infectees in the population. If the shadow price exceeds the effective cost of treatment, then all individuals are treated because the price of having them infected is greater than the price of treating them. On the other hand, if the shadow price is less than the effective cost, the price of treating an individual is greater than the price to social welfare of allowing this individual to remain infected and no one is treated. Finally, where the shadow price and effective cost are equal, the policymaker is indifferent between treating and not treating the individual and any level of treatment is optimal.

To complete the solution, the equations of motion for the two co-state variables are given below:

$$
\begin{aligned}
\dot{\lambda}_H &= \delta\lambda_H - \frac{\partial Q}{\partial I_H} \\
&= -p + cf_H - \lambda_H \left[-\delta + \beta_H (1 - I_H - I_L) - \tau - \alpha f_H \right] \\
&\quad + (\lambda_H \beta_H I_H + \lambda_L \beta_L I_L)
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
\dot{\lambda}_L &= \delta\lambda_L - \frac{\partial Q}{\partial I_L} \\
&= -p + cf_L - \lambda_L \left[-\delta + \beta_L (1 - I_H - I_L) - \tau - \alpha f_L \right] \\
&\quad + (\lambda_H \beta_H I_H + \lambda_L \beta_L I_L)
\end{aligned}
\tag{8}
$$

These demonstrate how the shadow prices change with prevalence levels. The reproduction ratio describes the number of new infections arising from one infection (Anderson and May 1991). If it is greater than one, the infection persists; otherwise, it dies out. In this model, the reproduction ratios of the High and Low strains are

$$
R^H = \frac{\beta_H}{\tau + \alpha f_H},
$$

$$
R^L = \frac{\beta_L}{\tau + \alpha f_L}.
$$

As a result of Assumption (4), these reproduction ratios are greater than one even if everyone infected with either strain is treated, so that neither strain can be eradicated (except in the case with no treatment, when the High strain will eradicate the Low strain).

---

[9] Note that λ_H and λ_L are always negative, as an additional infected person in the population lowers social welfare. I write the condition here in absolute values to ease interpretation.

4 Fixed points

In the analysis of dynamic models, of primary interest are the equilibria or fixed points of the model. In this section, we discuss the fixed points of the system and under which conditions they are attainable.

4.1 Analysis of non-boundary fixed points

We propose that there are two types of equilibria: non-boundary fixed points and asymptotic fixed points. They are defined as follows.

Definition 1 A non-boundary fixed point (FP) is a solution $(f_H^*, f_L^*, I_H^*, I_L^*, \lambda_H^*, \lambda_L^*)$ satisfying Eqs. (2), (3), (5), (6), (7) and (8), as well as $\dot{I}_H = \dot{I}_L = \dot{\lambda}_H = \dot{\lambda}_L = \dot{f}_H = \dot{f}_L = 0$.

Definition 2 An asymptotic fixed point (AFP) is a fixed point of the degenerate system in which one of (I_H, I_L) equals zero and towards which there are paths with $I_H > 0, I_L > 0$ that converge asymptotically to this fixed point. We look for the treatment values (f_H^*, f_L^*) in the vicinity of this fixed point that ensure this asymptotic convergence. On the boundary where one of the strains reaches zero prevalence (e.g. I_H), the co-state variable for that strain (e.g. λ_H) will be infinite, while the co-state variable of the other strain (e.g. λ_L) will be finite. In this sense, it is not a fixed point of the four variable system $(I_H, I_L, \lambda_H, \lambda_L)$.

The key difference between AFPs and FPs is that some variables at an AFP are not constant: they move towards a constant but only reach it in the limit. The following Lemma reduces the set of feasible non-boundary fixed points to three:

Lemma 1 At any fixed point, the proportion of I_H treated must be greater than the proportion of I_L treated: $f_H^* > f_L^*.$[10] Therefore, only the following combinations of treatment at the fixed point are feasible: $(f_H^* = 1, f_L^* = 0)$; $(f_H^* = 1, f_L^* \in (0, 1))$; $(f_H^* \in (0, 1), f_L^* = 0)$. This implies that there are three potential non-boundary fixed points.

Proof At a non-boundary fixed point, it must be the case that

$$I_H^* > 0,$$
$$I_L^* > 0,$$
$$\dot{I}_H = \dot{I}_L = 0.$$

---

[10] This result depends on the assumption that the success rate of treatment α and the natural rate of recovery τ are the same between the two strains. If α or τ were substantially higher for the High strain than the Low strain, this result could be reversed, with more treatment mandated for the Low strain. See Sect. 4.5.

This implies

$$0 = \frac{\dot{I}_H}{I_H^*} = \beta_H(1 - I_H^* - I_L^*) - \tau - \alpha f_H^*,$$

$$0 = \frac{\dot{I}_L}{I_L^*} = \beta_L(1 - I_H^* - I_L^*) - \tau - \alpha f_L^*.$$

Eliminating $(1 - I_H^* - I_L^*)$ yields

$$\beta_L \left(\tau + \alpha f_H^* \right) = \beta_H \left(\tau + \alpha f_L^* \right),$$

and hence

$$1 \geq f_H^* = \frac{\beta_H}{\beta_L} f_L^* + \frac{\tau}{\alpha} \left(\frac{\beta_H - \beta_L}{\beta_L} \right) > f_L^* \geq 0. \tag{9}$$

$\square$

Since $f_H^* > 0$ it follows that $\lambda_H^* \leq -\frac{c}{\alpha}$. Alternatively,

$$f_L^* = \frac{\beta_L}{\beta_H} \left[f_H^* - \left(\frac{\beta_H - \beta_L}{\beta_L} \right) \frac{\tau}{\alpha} \right]. \tag{10}$$

At a non-boundary fixed point $\dot{I}_H = 0$ and $I_H > 0$. From (2) it follows that the total prevalence of the disease is given by

$$I_L^* + I_H^* = 1 - \frac{\tau + \alpha f_H^*}{\beta_H}. \tag{11}$$

Since $\dot{\lambda}_H = 0$ it follows from (7) that

$$0 = -p + c f_H^* + \delta \lambda_H^* + \lambda_H^* \left[\beta_H(1 - I_H^* - I_L^*) - \tau - \alpha f_H^* \right] + \lambda_H^* \beta_H I_H^* + \lambda_L^* \beta_L I_L^*.$$

Eliminating I_L^* yields

$$0 = -p + c f_H^* + \delta \lambda_H^* + \lambda_H^* \beta_H I_H^* + \lambda_L^* \beta_L \left(1 - \frac{\tau + \alpha f_H^*}{\beta_H} - I_H^* \right),$$

which can be written as

$$0 = -p + c f_H^* + \delta \lambda_H^* + \lambda_L^* \beta_L \left(1 - \frac{\tau + \alpha f_H^*}{\beta_H} \right) - \left(\lambda_H^* \beta_H - \lambda_L^* \beta_L \right) I_H^*. \tag{12}$$

The above equation determines I_H^* uniquely as a function of f_H^*, λ_H^* and λ_L^*:

$$I_H^* = \frac{p + c f_H^* + \delta \lambda_H^* + \lambda_L^* \beta_L \left(1 - \frac{\tau + \alpha f_H^*}{\beta_H} \right)}{\left(\lambda_L^* \beta_L - \lambda_H^* \beta_H \right)}. \tag{13}$$

At a non-boundary fixed point $\dot{\lambda}_H = \dot{\lambda}_L = 0$. Subtracting (8) from (7), using the fact that $\dot{I}_H = \dot{I}_L = 0$, and rearranging yields

$$f_H^* - f_L^* = -\frac{\delta}{c}\left(\lambda_H^* - \lambda_L^*\right). \tag{14}$$

Lemma 1 shows that $f_H^* - f_L^* > 0$ at any non-boundary fixed point, so the above equation implies that $\lambda_H^* < \lambda_L^*$.

Following Wagener (2003), the parameter space can be split into different regimes that mandate which fixed points are feasible for which parameter constellations. We define the constant K, which will be used to create three regimes:

$$K = \frac{\beta_H - \beta_L}{\beta_L}\frac{\tau}{\alpha}.$$

The value of this constant is increasing in β_H and τ but decreasing in β_L and α. Intuitively, this implies that K will be high when the difference between the two infectivities is high and when the success rate of treatment is low but the natural rate of recovery is high. We will show that there are three regimes of feasibility: $K < 1$, $K = 1$ and $K > 1$. We distinguish between two cases of fixed points, $f_H^* \in (0, 1)$ and $f_H^* = 1$.

4.1.1 Case 1: $f_H^* \in (0, 1)$

We shall show that this case requires that $K = \left(\frac{\beta_H - \beta_L}{\beta_L}\right)\frac{\tau}{\alpha} < 1$ and $K < \frac{\delta}{\alpha}$. In this case $\lambda_H^* = -\frac{c}{\alpha}$ because f_H^* takes an interior value. Hence from (14), $\lambda_L^* > -\frac{c}{\alpha}$ and $f_L^* = 0$. From (9) and (14) it follows that this fixed point is characterised by the following values of the control variables and co-state variables:

$$f_H^* = \frac{\beta_H}{\beta_L}f_L^* + \frac{\tau}{\alpha}\left(\frac{\beta_H - \beta_L}{\beta_L}\right) = K,$$

$$f_L^* = 0,$$

$$\lambda_H^* = -\frac{c}{\alpha},$$

$$\lambda_L^* = \frac{c}{\delta}\left(f_H^* - f_L^*\right) + \lambda_H^* = \frac{c}{\delta}\left[K - \frac{\delta}{\alpha}\right].$$

$$I_H^* = \frac{p + cK - \frac{c}{\alpha}\delta + \frac{c}{\delta}\left[K - \frac{\delta}{\alpha}\right]\beta_L\left(1 - \frac{\tau + \alpha K}{\beta_H}\right)}{\left(\frac{c}{\delta}\left[K - \frac{\delta}{\alpha}\right]\beta_L + \frac{c}{\alpha}\beta_H\right)} > 0$$

Thus, f_H^* is decreasing in α but increasing in τ. It is also increasing in β_H because a more infectious strain requires more treatment but decreasing in β_L, as this implies a lower overall spread of the disease. Since $f_H^* < 1$ the parameters must satisfy the following condition:

$$K < 1.$$

Economic logic requires that $\lambda_L^* < 0$, which yields the additional condition:

$$K < \frac{\delta}{\alpha}.$$

Logic also requires that $I_H^* < 1 - \frac{\tau}{\beta_L}$, which yields the condition

$$\frac{p + cK - \frac{c}{\alpha}\delta + \frac{c}{\delta}\left[K - \frac{\delta}{\alpha}\right]\beta_L\left(1 - \frac{\tau + \alpha K}{\beta_H}\right)}{\left(\frac{c}{\delta}\left[K - \frac{\delta}{\alpha}\right]\beta_L + \frac{c}{\alpha}\beta_H\right)} < 1 - \frac{\tau}{\beta_L}.$$

In practice, this places constraints on c and p, which are otherwise free parameters. The value of I_L^* can be determined from the equation for total prevalence:

$$I_H^* + I_L^* = 1 - \frac{\tau}{\beta_L}.$$

Total prevalence depends positively on β_L and negatively on τ; there is no dependence on β_H as this enters through the choice of f_H^*. The success rate of treatment, α, does not affect total prevalence because no one infected with the Low strain is treated; the dependence of the High strain on α enters, again, through the choice of f_H^*.

4.1.2 Case 2: $f_H^* = 1$

Using (10), we find that in this case,

$$f_L^* = \frac{\beta_L}{\beta_H}\left[f_H^* - \left(\frac{\beta_H - \beta_L}{\beta_L}\right)\frac{\tau}{\alpha}\right] = \frac{\beta_L}{\beta_H}[1 - K] \tag{15}$$

For this solution to be feasible, we must have $f_L^* \geq 0$ and hence

$$K = \left(\frac{\beta_H - \beta_L}{\beta_L}\right)\frac{\tau}{\alpha} \leq 1.$$

Therefore, two subcases of this fixed point can be distinguished, depending on whether $K = 1$ or $K < 1$.

Case 2a $f_H^* = 1$, $K < 1$. Since $K < 1$, this implies that $f_L^* = \frac{\beta_L}{\beta_H}[1 - K] > 0$. In particular, f_L^* is increasing in β_L but decreasing in β_H. There is interdependence between the two strains: the more infectious is the High strain, the less the Low strain is treated. However, the more infectious is the Low strain, the more it is treated. The optimal choice of f_L^* also depends intuitively on α and τ: the faster the rate of recovery, whether through treatment or naturally, the lower is the optimal treatment

level. Since $f_L^* \in (0, 1)$, it follows that $\lambda_L^* = -\frac{c}{\alpha}$ and the value of λ_H^* is determined as follows:

$$
\begin{aligned}
\lambda_H^* &= \lambda_L^* - \frac{c}{\delta} \left(f_H^* - f_L^* \right) \\
&= -\frac{c}{\alpha} - \frac{c}{\delta} \left(1 - \frac{\beta_L}{\beta_H} [1 - K] \right) \\
&< -\frac{c}{\alpha}.
\end{aligned}
$$

Thus, this fixed point is characterised by the following values of the control and co-state variables:

$$
\begin{aligned}
f_H^* &= 1, \\
f_L^* &= \frac{\beta_L}{\beta_H} [1 - K] \in (0, 1), \\
\lambda_L^* &= -\frac{c}{\alpha}, \\
\lambda_H^* &= -\frac{c}{\alpha} - \frac{c}{\delta} \left(1 - \frac{\beta_L}{\beta_H} [1 - K] \right) < -\frac{c}{\alpha}.
\end{aligned}
$$

Inserting the above values of f_H^*, λ_H^* and λ_L^* into (12) determines I_H^* uniquely:

$$
I_H^* = \frac{p - \frac{\delta c}{\alpha} + \frac{c\beta_L}{\beta_H}(1 - K) - \frac{c\beta_L}{\alpha} \left(1 - \frac{\tau+\alpha}{\beta_H} \right)}{\frac{c}{\alpha}(\beta_H - \beta_H) + \frac{c}{\delta}\beta_H \left(1 - \frac{\beta_L}{\beta_H}(1 - K) \right)} > 0.
$$

The value of I_L^* can then be determined from (11):

$$
I_L^* = 1 - \frac{\tau + \alpha}{\beta_H} - I_H^*,
$$

which is increasing in β_H but decreasing in τ and α. This is intuitive: the more infectious is the H type, the higher the prevalence of the disease in equilibrium; on the other hand, the more successful the treatment or the higher the rate of natural recovery, the lower the total prevalence in equilibrium. A point of interest is that β_L does not affect total prevalence; this is because the level of f_L^* is fixed and chosen based on the value of β_L. The dependence of f_L^* on β_L ensures that β_L does not affect total prevalence. Note that we require I_H^* and I_L^* to be bounded by zero and one, so that an additional condition for the feasibility of this fixed point is

$$
\frac{p - \frac{\delta c}{\alpha} + \frac{c\beta_L}{\beta_H}(1 - K) - \frac{c\beta_L}{\alpha} \left(1 - \frac{\tau+\alpha}{\beta_H} \right)}{\frac{c}{\alpha}(\beta_H - \beta_H) + \frac{c}{\delta}\beta_H \left(1 - \frac{\beta_L}{\beta_H}(1 - K) \right)} < 1 - \frac{\tau + \alpha}{\beta_H},
$$

which will place constraints on p and c.

Case 2b $f_H^* = 1$, $K = 1$. Equation (15) implies that in this case $f_L^* = 0$. Hence $\lambda_H^* \leq -\frac{c}{\alpha} \leq \lambda_L^*$ and

$$
\lambda_L^* = \lambda_H^* + \frac{c}{\delta} \left(f_H^* - f_L^* \right)
$$
$$
= \lambda_H^* + \frac{c}{\delta}.
$$

To ensure that $\lambda_H^* < -\frac{c}{\alpha}$ and $\lambda_L^* < 0$ we require that

$$
-\left(\frac{c}{\alpha} + \frac{c}{\delta} \right) \leq \lambda_H^* < -\max\left(\frac{c}{\alpha}, \frac{c}{\delta} \right).
$$

Thus, this fixed point is described as follows:

$$
f_H^* = 1,
$$
$$
f_L^* = 0,
$$
$$
\lambda_H^* : -\left(\frac{c}{\alpha} + \frac{c}{\delta} \right) \leq \lambda_H^* < -\max\left(\frac{c}{\alpha}, \frac{c}{\delta} \right),
$$
$$
\lambda_L^* = \lambda_H^* + \frac{c}{\delta}.
$$

There is one degree of freedom. The shadow price λ_H^* is indeterminate within the range shown. Eliminating f_H^* and λ_L^* from (12) determines I_H^* as a function of λ_H^*. The value of I_L^* can then be determined from (11). There is a continuum of feasible fixed points with total prevalence

$$
I_H^* + I_L^* = 1 - \frac{\tau + \alpha}{\beta_H}.
$$

4.2 Summary of non-boundary fixed points

This summary uses multiple asterisks (*), (**), and (***) to denote distinct types of fixed point. Note that feasibility requires that $\dot{I}_H = \dot{I}_L = 0$ and $\dot{\lambda}_H = \dot{\lambda}_L = 0$.

(a) If $K < 1$ and $K \geq \frac{\delta}{\alpha}$ there is a unique non-boundary feasible fixed point (I_H^*, I_L^*). The control variables are $f_H^* = 1$ and $f_L^* = \frac{\beta_L}{\beta_H} [1 - K]$.

(b) If $K < 1$ and $K < \frac{\delta}{\alpha}$ there are two feasible non-boundary fixed points: (I_H^*, I_L^*) with the control variables $f_H^* = 1$ and $f_L^* = \frac{\beta_L}{\beta_H} [1 - K]$; and (I_H^{**}, I_L^{**}) with control variables $f_H^{**} = K$ and $f_L^{**} = 0$.

(c) If $K = 1$ the control variables satisfy $f_H^{***} = 1$, $f_L^{***} = 0$. In this case, there is a continuum of feasible fixed points (I_H^{***}, I_L^{***}) which lie on the line $I_H^{***} + I_L^{***} = 1 - \frac{\tau + \alpha}{\beta_H} = 1 - \frac{\tau}{\beta_L}$.

The above are the only feasible non-boundary fixed points. Note there are no feasible non-boundary fixed points if $K > 1$. These findings suggest the possibility of a Skiba case when $K < 1$ and $K < \frac{\delta}{\alpha}$. Inserting control variables in (11) shows that fixed

point (**) always has higher total infection than fixed point (*). This is intuitive as in the former, both treatment levels are lower. For ease of notation, we label fixed point (*) as F^{low} (low total prevalence), fixed point (**) as F^{high} (high total prevalence) and fixed point (***) as F^{int} (intermediate total prevalence).

4.3 Analysis of asymptotic fixed points

The asymptotic fixed points are defined such that one strain approaches eradication asymptotically, while the other prevails at a positive level. As a result, the behaviour of the system in the neighbourhood of the fixed point can be approximated by the behaviour of a one-infection system. This is because the behaviour of the system with small I_L is very similar to the behaviour when $I_L = 0$. Naturally, this holds for I_H close to zero as well.

The behaviour of a one-infection system has been studied by Rowthorn (2006) and discussed in Sect. 2. Only boundary values for policy are optimal. This is because the co-state variable is a function of prevalence only in this type of problem. This implies that there must be a one-to-one mapping between prevalence and the co-state variable. However, the co-state variable changes with time. Therefore, a path cannot go back on itself in order to achieve two different values of the co-state variable for one level of prevalence. Interior values of policy require a path that goes back on itself (Rowthorn 2006). Therefore, interior values of policy cannot be optimal because they violate the assumption of the co-state variable being a function of prevalence.

To analyse asymptotic fixed points, we analyse a one-infection system and perturb it around this equilibrium point slightly. There are two possible categories of fixed points: those were I_H tends asymptotically towards zero, and those where I_L tends asymptotically towards zero.

4.3.1 $I_H = 0$ boundary

First, suppose that $I_H = 0$. The system then reduces to the following maximisation problem:

$$\max V(I_L^0) = \int_0^\infty e^{-\delta t} \left[p(1 - I_L(t)) - c f_L(t) I_L(t) \right] dt$$

subject to

$$\dot{I}_L = I_L(t) \left[\beta_L (1 - I_L(t)) - \tau - \alpha f_L(t) \right],$$
$$f_L(t) \in [0, 1],$$
$$I_L(t) \in [0, 1],$$
$$I_L(0) = I_L^0 > 0,$$
$$\beta_L > \tau + \alpha. \tag{16}$$

Optimal treatment is governed by the following bang-bang condition:

$$f_L \begin{Bmatrix} = 0 \\ \in [0, 1] \\ = 1 \end{Bmatrix} \quad \text{if } \lambda_L \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} -\frac{c}{\alpha}. \tag{17}$$

Finally, the co-state variable evolves according to the following function:

$$\begin{aligned}
\dot{\lambda}_L &= \delta\lambda_L - \frac{\partial Q}{\partial I_L} \\
&= \delta\lambda_L + p + cf_L - \lambda_L \left[\beta_L(1 - I_L) - \tau - \alpha f_L\right] \\
&\quad + \lambda_L \beta_L I_L.
\end{aligned} \tag{18}$$

At a Hamiltonian fixed point, $\dot{\lambda}_L = 0$ and $\dot{I}_L = 0$:

$$0 = (\delta + \beta_L I_L)\lambda_L + p + cf_L$$
$$0 = \beta_L(1 - I_L) - \tau - \alpha f_L$$

If the fixed point is interior then $\lambda_L = -\frac{c}{\alpha}$ and we can solve the resulting equations to yield a unique interior fixed point I_L^{int}. There are also fixed points with $f_L = 0$ and $f_L = 1$. Let I_L^{high}, I_L^{low} be the corresponding levels of infection. It has been shown in the literature (Rowthorn 2006) that it is never optimal to go to I_L^{int}. Thus, the possible candidates for an optimum are I_L^{high} and I_L^{low}. Let $A_L^{high} = (0, I_L^{high})$, $A_L^{low} = (0, I_L^{low})$.

First, note that the point A_L^{low} cannot be reached even asymptotically from the interior. Consider any point $(x, I_L^{low} + y)$, where $x > 0$:

$$\begin{aligned}
\dot{I}_H &= I_H(t)\left[\beta_H(1 - I_H(t) - I_L(t)) - \tau - \alpha f_H(t)\right] \\
&= x\left[\beta_H(1 - x - I_L^{low} - y) - \tau - \alpha f_H(t)\right] \\
&= x\left[-\beta_H(x + y) + \beta_H(1 - I_L^{low}) - \tau - \alpha f_H(t)\right]
\end{aligned} \tag{19}$$

Recall that $\dot{I}_L = 0$ implies that $\beta_L(1 - I_L^{low}) - \tau - \alpha = 0$. Hence,

$$\dot{I}_H = x\left[-\beta_H(x + y) + \left(\frac{\beta_H}{\beta_L} - 1\right)(\tau + \alpha) + \alpha(1 - f_H(t))\right]. \tag{20}$$

For sufficiently small values of x and y, and for any value $f_H(t) \in [0, 1]$, the right hand side of this equation is positive. Hence, the prevalence of the High strain is increasing away from zero rather than decreasing towards zero, and A_L^{low} cannot be reached even asymptotically from the interior.

What about the point A_L^{high}? Consider any point $(x, I_L^{high} + y)$ where $x > 0$:

$$\dot{I}_H = I_H(t) \left[\beta_H (1 - I_H(t) - I_L(t)) - \tau - \alpha f_H(t) \right]$$
$$= x \left[\beta_{LH}(1 - x - I_L^{high} - y) - \tau - \alpha f_H(t) \right]$$
$$= x \left[-\beta_H (x + y) + \beta_H (1 - I_L^{high}) - \tau - \alpha f_H(t) \right] \qquad (21)$$

As before, $\beta_L (1 - I_L^{high}) - \tau = 0$, so that

$$\dot{I}_H = x \left[-\beta_H (x + y) + \left(\frac{\beta_H}{\beta_L} - 1 \right) \tau - \alpha f_H(t) \right]. \qquad (22)$$

For sufficiently small values of x and y, the right hand side is positive. In particular, for small x and y the right hand side is positive for all $f_H(t) \in [0, 1]$ if

$$K = \left(\frac{\beta_H - \beta_L}{\beta_L} \right) \frac{\alpha}{\tau} > 1.$$

If this condition is satisfied, there is no path that converges to A_L^{high} from the interior. On the other hand, for $f_H(t) = 1$, the asymptotic fixed point A_L^{high} is attainable if $K < 1$. Thus, there is a feasible asymptotic fixed point with

$$f_H^* = 1,$$
$$f_L^* = 0,$$
$$I_H \to 0,$$
$$I_L^* = 1 - \frac{\tau}{\beta_L}.$$

4.3.2 $I_L = 0$ boundary

Suppose next that $I_L = 0$. By symmetry, there are two potential categories of fixed points, with $f_H^* = 0$ or $= 1$. Let I_H^{high}, I_H^{low} be the corresponding levels of infection. When are these feasible?

Let $A_H^{high} = (0, I_H^{high})$, and $A_H^{low} = (0, I_H^{low})$. Note that the point A_L^{low} can always be reached asymptotically from the interior. Consider any point $(I_H(t), I_L(t)) = (I_H^{low} + x, y)$ where $y > 0$. Suppose that $f_H^* = 1$ and $I_H^{low} = 1 - \frac{\tau + \alpha}{\beta_H}$. In this case, the evolution of I_L follows

$$\dot{I}_L = y \left[-\beta_L(x + y) + \left(\frac{\beta_L}{\beta_H} - 1 \right) \tau + \alpha \frac{\beta_L}{\beta_H} - \alpha f_L(t) \right]. \qquad (23)$$

The right-hand side of this equation is negative for small x and y when $f_L > \frac{\beta_L}{\beta_H} > 0$. Since an interior value of f_L is suboptimal, this implies there is one final asymptotic fixed point, which is feasible for all K:

$$f_L^* = 1,$$
$$f_H^* = 1,$$
$$I_L \to 0,$$
$$I_H^* = 1 - \frac{\tau + \alpha}{\beta_H}.$$

What about A_H^{high} with $f_H^* = 0$? Consider any point $(I_H(t), I_L(t)) = (I_H^{high} + x, y)$ where $y > 0$. The evolution of I_L is governed by the following equation:

$$\dot{I}_L = y \left[-\beta_L(x + y) + \left(\frac{\beta_L}{\beta_H} - 1 \right) \tau - \alpha f_L(t) \right]. \tag{24}$$

The right-hand side is negative for all values of $f_L(t) \in [0, 1]$. Since interior values of f_L can be ruled out by Rowthorn (2006), this asymptotic fixed points can be reached for all values of K by setting $f_L = 0$ or $f_L = 1$:

$$f_L^* \in \{0, 1\},$$
$$f_H^* = 0,$$
$$I_L \to 0,$$
$$I_H^* = 1 - \frac{\tau}{\beta_H}.$$

Thus, when $K > 1$, the only feasible candidates for asymptotic equilibria are those with $I_L \to 0$.

4.4 Summary of asymptotic fixed points

If $K < 1$, there are three asymptotic fixed points:

(a) A_L^{high}, with $f_H^* = 1$, $f_L^* = 0$ and $I_H \to 0$,
(b) A_H^{high} with $f_L^* \in \{0, 1\}$, $f_H^* = 0$ and $I_L \to 0$,
(c) A_H^{low} with $f_L^* = 1$, $f_H^* = 1$ and $I_L \to 0$.

If $K > 1$ there are no interior points that satisfy the Hamiltonian conditions. Moreover, the boundary points with $I_H = 0$ cannot be reached. In this case, the only feasible candidates for an optimum are boundary points with $I_L = 0$. In particular, there are three possible cases with the following control variables:

(a) A_H^{high} with $f_L^* \in \{0, 1\}$, $f_H^* = 0$ and $I_L \to 0$,
(b) A_H^{low} with $f_L^* = 1$, $f_H^* = 1$ and $I_L \to 0$,
(c) Skiba case. The optimum path goes to either A_H^{low} or A_H^{high} depending on the starting point.

The central conclusion when $K > 1$ is that optimal policy involves asymptotic elimination of the low infectivity strain. The possibility of the Skiba case indicates

that it is not enough simply to compare fixed points—the initial prevalence of the disease will affect which fixed point is optimal.

However, the following Proposition shows that asymptotic eradication of both strains of the disease is never possible:

Proposition 1 *Both variants of the disease cannot be simultaneously eradicated even asymptotically in equilibrium if*

$$\frac{\tau + \alpha}{\beta_H} < 1,$$
$$\frac{\tau + \alpha}{\beta_L} < 1,$$

which hold as a consequence of Assumption (4).

Proof In Appendix B. □

This proposition shows that even if both AFPs are feasible, asymptotic eradication of the disease is not possible and at least one strain always prevails.

4.5 Extensions to the model

In reality, strains do vary in factors such as drug resistance and the cost of treatment. In the present model, this could be modelled by varying the parameters α, τ and c by strain.

Suppose that the natural recovery rate of strain H is τ_H and that of strain L is τ_L. Similarly, the success rate of treatment of the High strain is α_H and of the Low strain is α_L. We can make some tentative claims about what would happen to the equilibria of the model. First, it should be noticed that coexistence of the two strains without intervention is now possible. The two strains evolve according to the following differential equations:

$$\frac{\dot{I}_H}{I_H} = (1 - I_H - I_L)\beta_H - \tau_H - \alpha_H f_H,$$
$$\frac{\dot{I}_L}{I_L} = (1 - I_H - I_L)\beta_L - \tau_L - \alpha_L f_L.$$

When there is no intervention, the strains can co-exist if

$$\frac{\tau_H}{\beta_H} = \frac{\tau_L}{\beta_L}.$$

Intuitively, this condition requires that the rate of infection and rate of recovery of the two strains are proportional. Thus, the High strain spreads its infection more quickly but also returns healthy individuals back into the population more quickly, so that the Low strain can survive. Next, we can analyse what would happen to the non-boundary fixed points. Lemma 1 states that only those fixed points where the High

strain is treated more are feasible. In fact, if α_H is higher than α_L, and τ_H is higher than τ_L, this result could be reversed. A non-boundary fixed point requires

$$\frac{\tau_H + \alpha_H f_H}{\beta_H} = \frac{\tau_L + \alpha_L f_L}{\beta_L},$$

so that the relationship between the two treatment levels is

$$f_L = \frac{\alpha_H \beta_L}{\alpha_L \beta_H} f_H + \frac{\beta_L \tau_H - \beta_H \tau_L}{\alpha_L \beta_H}.$$

Thus, depending on the values of α_H, α_L, τ_H and τ_L, f_L could be higher than f_H. This makes intuitive sense, as higher values of α_H and τ_H make the High strain comparatively easier to contain, so that this may outweigh the fact that a higher transmission probability makes the High strain more difficult to contain. Finally, we can consider the implications for the asymptotic fixed points. Here, also, the main result may be reversed: namely that the only feasible equilibrium is one where the Low strain is asymptotically eradicated. This is easy to see by rewriting Eqs. (19) and (21) allowing for differences in α and τ by strain, and performing a similar exercise for asymptotic eradication of the Low strain. When α_H is much higher than α_L, and similarly τ_H is much higher than τ_L, then it may be that the only feasible equilibrium is asymptotic eradication of the High strain. This is because the low success rate of treatment of the Low strain, and its low natural rate of recovery, make it impossible to eradicate, even asymptotically.

Varying cost by strain would not affect prevalence or the fixed points that can be attained by the policymaker, but would affect the choice of equilibrium to converge to. This is ultimately a numerical question. It seems intuitive to assume that $c_H > c_L$: it is more expensive to treat those infected with the High strain. The effect on the optimal equilibrium would depend how these values compare to the homogeneous cost parameter c. It could be that the overall cost goes down for a given level of treatment (such as if c_H is slightly higher than c, but c_L is much lower than c), in which case the policymaker would be more likely to pick an equilibrium where more people are treated, such as F^{low}.

4.6 Feasibility

4.6.1 Summary of regimes of feasible equilibria

In this section we plot the feasible equilibria for each regime of K. Recall that if $K < 1$, the set of feasible equilibria is $Z_{K<1} = \{F^{low}, F^{high}, A_L^{high}, A_H^{high}, A_H^{low}\}$. If $K = 1$, the set of feasible equilibria is $Z_{K=1} = \{F^{int}, A_H^{high}, A_H^{low}\}$. If $K > 1$, the set of feasible equilibria is $Z_{K>1} = \{A_H^{high}, A_H^{low}\}$. As K increases, the set of feasible equilibria falls: there are fewer ways in which the disease can be contained. This is because a higher K occurs when treatment is ineffective or when the strains differ substantially in their infectivities, making it difficult to control both of them.
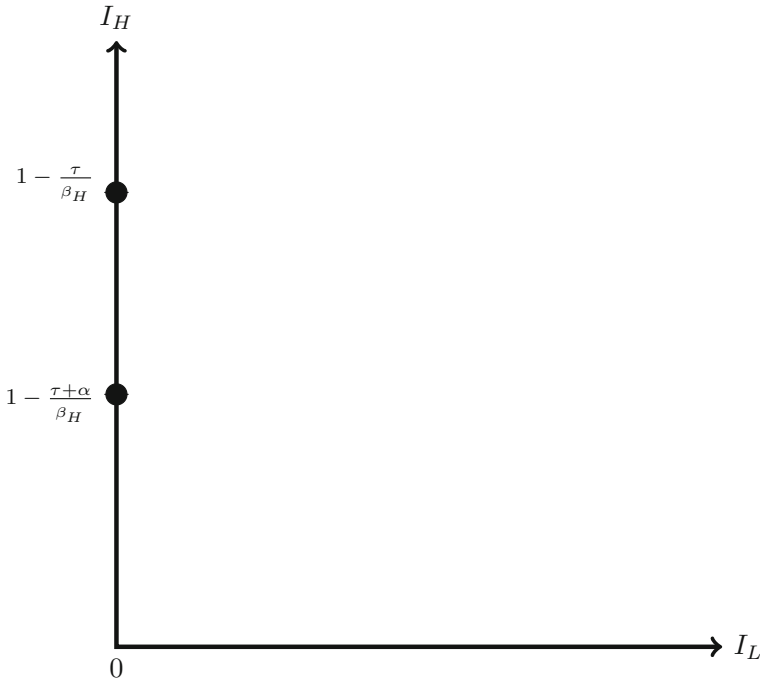
Fig. 1 The set of feasible equilibria when $K > 1$. The *full circles* denote the two equilibria A_H^{high} and A_H^{low}

As K increases, the minimum attainable equilibrium prevalence of I_H rises. This makes intuitive sense, as K is increasing in β_H and in the difference between β_H and β_L. Where this difference is very large, $K > 1$ and the only feasible equilibrium is the AFP where I_L tends asymptotically towards zero and I_H prevails at a positive level. In contrast, when β_H is low, $K < 1$ and one can attain equilibria such as F^{low}, where the equilibrium prevalence of I_H can be close to zero.

The set of feasible equilibria can be depicted graphically. Figure 1 shows the set of feasible equilibria when $K > 1$. The feasible set when $K = 1$ is shown in Fig. 2. Figure 3 shows the set of feasible equilibria when $K < 1$.[11] The graphs clearly illustrate the rising number of feasible equilibria as K falls.

4.6.2 An intuitive explanation and some examples

It is useful to introduce some parameter values in order to better understand the policy implications of these results. There are, in total, five individual fixed points: three non-boundary fixed points (one of which is in fact a continuum of feasible fixed points with the same total prevalence across the two strains) and two asymptotic fixed points.

---

[11] These figures are illustrated for the following parameter values: $\beta_L = 0.4$, $\tau = 0.15$ and $\alpha = 0.2$, with $\beta_H = 0.95$ for $K > 1$, $\beta_H = 0.9\dot{3}$ for $K = 1$, and $\beta_H = 0.8$ for $K < 1$.
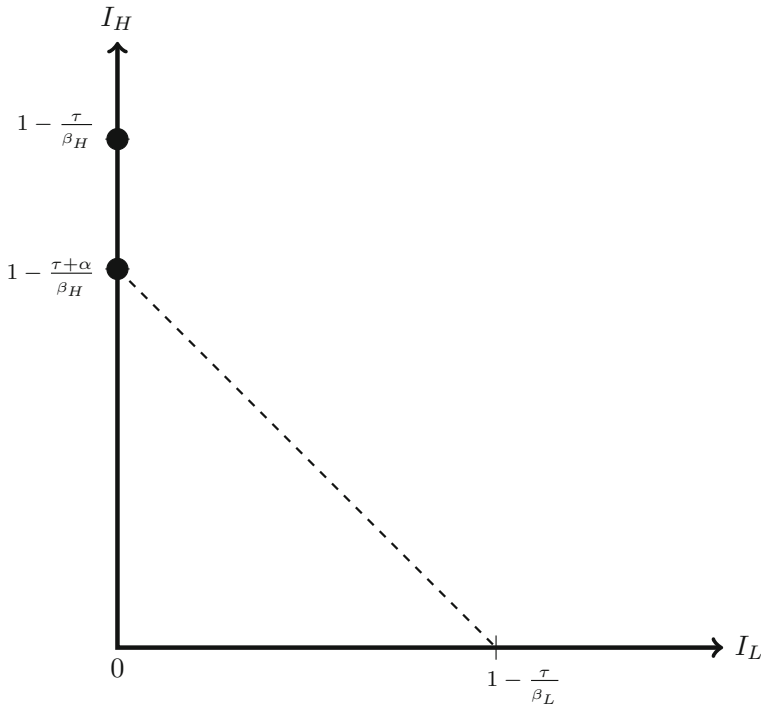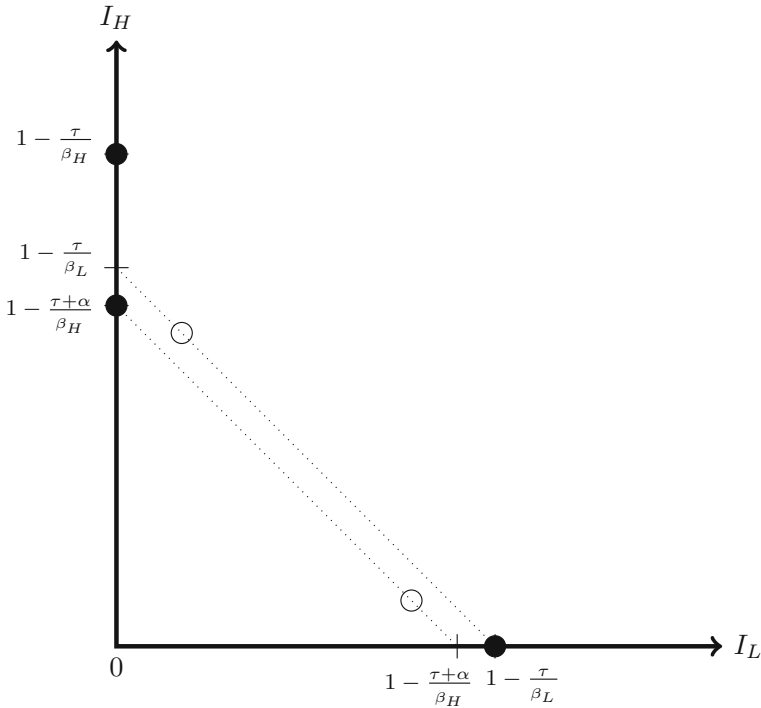
Fig. 2 The set of feasible equilibria when $K = 1$. The *full circles* show the equilibria A_H^{high} and A_H^{low}. The *dashed line* denotes the continuum of equilibria of type F^{int}

Feasibility depends on the value of $K = \left(\frac{\beta_H - \beta_L}{\beta_L}\right) \frac{\tau}{\alpha}$. In order for the interior fixed points to exist, we require $K \leq 1$. If $K < 1$ and $K < \frac{\delta}{\alpha}$, then two non-boundary fixed points exist, which suggest the existence of a Skiba line. A further consideration is whether the assumption $\beta_H > \beta_L > \tau + \alpha$ is satisfied, which ensures that the disease cannot be eradicated, even asymptotically.

Generally, the existence of a non-boundary fixed point with positive equilibrium prevalence of the disease requires the following conditions, where (a) to (c) ensure the existence of the non-boundary fixed points and (d) ensures that at this non-boundary fixed point, the disease is not eradicated.

(a) The relative difference between the infectivities, $\frac{\beta_H - \beta_L}{\beta_L}$, is not too high.
(b) The natural recovery rate is not too high relative to the success rate of treatment: $\frac{\tau}{\alpha}$ is not too high.
(c) The future is not discounted too highly: δ is high enough.
(d) The infectivity probabilities, β_H and β_L, are high relative to the recovery probabilities, τ and α

We can interpret these conditions in an intuitive way. In order for both strains of the disease to co-exist in equilibrium, we require that their infectivities are not too different. This is because if the High strain is much more infectious, then individuals will be much more susceptible to it relative to the Low strain, so that the Low strain will

Fig. 3 The set of feasible equilibria when $K < 1$. The *full circles* denote the equilibria A_L^{high} (on the x-axis), A_H^{high} and A_H^{low} (on the y-axis). The *empty circles* denote the feasible equilibria F^{low} and F^{high} for this set of parameters, with the latter having higher total prevalence

eventually die out (asymptotically). On the other hand, if the infection probabilities are similar but low, relative to the recovery probabilities, then both strains will die out, because the rate of reinfection is not high enough to ensure the survival of the disease. The magnitude of the natural recovery rate relative to the success rate of treatment is important because a high natural recovery rate will likely lead to the elimination of both strains of the disease, while a low value of α will imply that treatment is ineffective (and a particular level of treatment is required to ensure that the disease survives at a positive, constant level in equilibrium). Finally, the condition on the discount rate arises because economic logic requires $\lambda_L^* < 0$ in the non-boundary fixed point F^{high}; otherwise, the marginal social value of an additional infected person would be positive. Intuitively, not discounting the future too highly ensures that the impact of an additional infected person is felt for a longer period of time, hence ensuring that the marginal social impact of this infected person is negative at the fixed point.

Let us consider two examples that illustrate these points. First, suppose that $\beta_H = 0.9$, $\beta_L = 0.1$: the High strain is very infectious, relative to the Low strain. In order for $K < 1$, we require $\tau < \frac{\alpha}{8}$: the treatment needs to be very effective, relative to the natural rate of recovery. In addition, we require $\tau + \alpha < 0.1$, in order for the disease to not be eradicated. One possible set of values for (τ, α) is $(0.004, 0.04)$.

Thus, in the example of a disease where the High strain is very infectious, the only way that both strains can co-exist in a non-boundary fixed point is when both the natural rate of recovery and the success rate of treatment are extremely low. Note that this example also requires a value of $\delta > 0.8$, which is reasonable, given that the economics literature usually assumes a discount rate of 0.95 or 0.99 (the latter is implied by an annual interest rate of 4%). If natural recovery is somewhat higher, e.g. if $\tau = \alpha = 0.04$, then we will end up in the situation where the Low strain is asymptotically eradicated, and the High strain prevails in equilibrium. If either the natural recovery rate or the treatment success are high, e.g. if $\alpha = 0.2$, even if $\tau = 0.004$, then both strains will be (asymptotically) eradicated.

Second, suppose that we have a very effective form of treatment with $\alpha = 0.9$, but individuals naturally recover from the disease only rarely, so that $\tau = 0.1$. In this case, the disease will be eradicated, because the condition $\beta_H > \beta_L > \tau + \alpha$ can never be satisfied. Suppose that instead we have a disease where $\alpha = 0.8$ and $\tau = 0.1$. Here, we require $\beta_H > \beta_L > 0.9$, so that infection is almost guaranteed when a susceptible individual meets an infected individual of either strain. One combination of values of (β_H, β_L) that ensures coexistence of both strains in equilibrium is (0.99, 0.95). This set of parameters requires $\delta > 0.004$, which is easily satisfied.

The general lesson from these examples is that where treatment is very successful, the disease needs to be highly infectious in order to prevail. An interesting finding is that the rate of natural recovery relative to the success of treatment is also important: coexistence of both strains is more likely when natural recovery is low, but the treatment is very successful. Finally, coexistence of both strains is more likely when they are similar in their infectivities. This makes sense intuitively, as the bigger the difference, the more likely that the infectivity of the High strain will cause the Low strain to die out.

### 4.6.3 Feasible policies along the path to the fixed points

Another area where restrictions on feasibility may apply is the choice of treatment policy along the path to the non-boundary fixed points. Policies along the path towards the fixed points will always be boundary policies (i.e. 0 or 1), as these are Most Rapid Approach Paths. We take each of the boundary policies in turn and examine the feasibility conditions required for I_H and I_L to converge to their equilibrium values if this is the chosen policy. Different feasibility constraints will exist depending on whether the system begins *below* or *above* each equilibrium; below means beginning at an initial level $I_H^0 + I_L^0 < I_H^* + I_L^*$, which requires I_H and I_L to *increase* to their equilibrium values, while beginning above implies an initial prevalence $I_H^0 + I_L^0 > I_H^* + I_L^*$, which requires I_H and I_L to *decrease* to their equilibrium values. Letting P^{ab} denote the policy $f_H = a$, $f_L = b$, the feasibility restrictions are summarised in Table 1. Derivations can be found in Appendix C.

The table is interpreted as follows. If the system begins below $I_H^* + I_L^*$, I_H and I_L need to increase. Not all policies can achieve this at all times: the range where the policy $f_H = 0$, $f_L = 0$ will allow I_H and I_L to increase is given by the top left entry in the table. There is an *upper bound* on $I_H + I_L$; if this condition is not satisfied, then

Table 1 Feasible policies along the path

		$\dot{i}_H, \dot{i}_L > 0$	$\dot{i}_H, \dot{i}_L < 0$
P^{00}		$1 - \frac{\tau}{\beta_L} > I_H + I_L$	$1 - \frac{\tau}{\beta_H} < I_H + I_L$
P^{10}		$1 - \frac{\tau}{\beta_L} > I_H + I_L$	$1 - \frac{\tau+\alpha}{\beta_H} < I_H + I_L$
P^{11}		$1 - \frac{\tau+\alpha}{\beta_L} > I_H + I_L$	$1 - \frac{\tau+\alpha}{\beta_H} < I_H + I_L$
P^{01}		$1 - \frac{\tau+\alpha}{\beta_L} > I_H + I_L$	$1 - \frac{\tau}{\beta_H} < I_H + I_L$
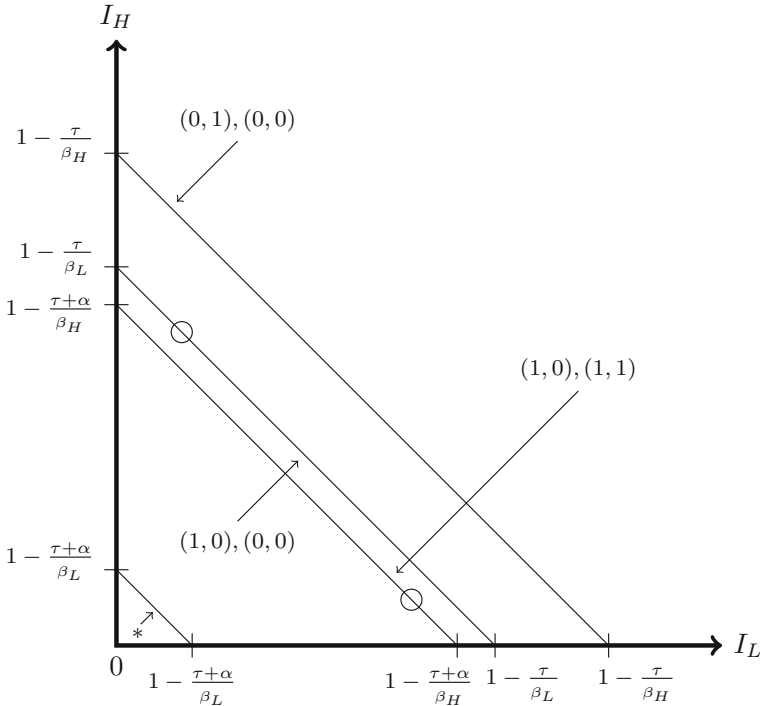
Fig. 4 Feasible policies depicted for the case when $K < 1$, where the *arrows* denote the range where the policy (f_H, f_L) is feasible. At ∗, the policies are $(0, 1)$ and $(1, 1)$. The *empty circles* denote the feasible equilibria F^{high} and F^{low} for this set of parameters

setting $f_H = 0$, $f_L = 0$ will cause the system to move away from an equilibrium. The conditions in the table are more clearly shown in graphical form in Fig. 4.

The message of this figure is that when one starts below an equilibrium, feasible policies to reach the equilibrium usually involve treating no one or just one strain. This allows total prevalence to increase. On the other hand, when one starts above an equilibrium, feasible policies that ensure the equilibrium is reached treat at least one group, in order to ensure that prevalence falls. Thus, F^{high} and F^{low} can be reached from above with the policies $(f_H, f_L) = (1, 0)$ and $(1, 1)$, while they can be reached from below with the policies $(1, 0)$ and $(0, 0)$.

Table 2 Summary of analytical results

Equilibrium	f_H^*	f_L^*	Feasible
F^{high}	$\frac{(\beta_H - \beta_L)}{\beta_L} \frac{\tau}{\alpha}$	0	$K < 1, K < \frac{\delta}{\alpha}$
F^{low}	1	$1 - \frac{(\beta_H - \beta_L)}{\beta_H} \frac{\tau + \alpha}{\alpha}$	$K < 1$
F^{int}	1	0	$K = 1$
A_L^{high}	1	0	$K < 1$
A_H^{high}	0	$\in \{0, 1\}$	All K
A_H^{low}	1	1	All K

4.7 Summary of analytical results

Table 2 summarises the analytical results—the feasible fixed points, their associated optimal policies and for which values of K they are attainable.

5 Simulations

In this section, we provide examples of optimal policy under various parameter combinations. When $K > 1$, the candidate equilibria are A_H^{high} and A_H^{low}. Simulations can tell us, under particular parameter values, which equilibrium it is optimal to go to, and what is the best way to get there. We explore the case when $K > 1$ in the first part of this section. Next, we consider the situation when $K < 1$. In this case, there is a wide set of feasible equilibria, and it is not possible to simulate an optimal choice of equilibrium because the derivations of optimal policy are only correct in the neighbourhood of each equilibrium. Instead, we suppose that it is optimal to go to the non-boundary fixed point F^{low}, and simulate optimal policy towards this point. We find that, in all the simulations that we estimate, optimal policy involves a spiral. We argue that this suggests that the interior fixed points are unlikely to ever be an optimal target for the policymaker.[12]

5.1 Optimal policy when $K > 1$

The values used for the parameters in the simulations are given in Table 3.

Some relevant concepts ought to be explained at this stage. First, we examine both optimal fixed and optimal variable policy. Optimal fixed policy compares the social welfare from choosing various policies at the beginning and not changing them until equilibrium. Optimal variable policy uses the Hamiltonian optimality conditions (5) and (6) to evaluate optimal policy at each time increment, using a fourth-order Runge-

---

[12] In our simulations, we also explored the properties of "singular solutions". A singular solution is a path along which one of the control variables takes an interior value (i.e. strictly less than one and greater than zero). Such paths were not optimal in the cases we examined.

Table 3 Parameter values for $K > 1$

Parameter	Value
β_H	0.95
β_L	0.4
τ	0.15
α	0.2
p	1
δ	0.111

Kutta procedure. Paths with optimal variable policy are described as Hamiltonian paths because they satisfy these Hamiltonian optimality conditions. In the case of variable policy, we examine both forward and backward paths. This means that, starting from an initial prevalence, we simulate the system both towards and away from the fixed point. This is carried out by setting time increments equal to $+1$ and -1 respectively. As a result, it is possible to describe a longer optimal path than if we were to only simulate the system towards equilibrium.

5.1.1 Fixed policy

In this first set of simulations, we assume that policy is fixed along the entire path to equilibrium. This is a special case that can guide intuition, as policy may not always have the benefit of full flexibility. We assume that the policymaker chooses $f_L^* = 1$, and compare social welfare V when $f_H^* = 1$ or $f_H^* = 0$. When the system begins, the policymaker chooses whether to treat everyone or no one infected with the High strain. The optimal policy is the choice of f_H^* that gives the highest value of social welfare.[13] The simulations set $t = 90$. Each time increment can be interpreted as one day, which implies that the results simulate an infection evolving over 90 days. The initial prevalence of the Low strain is set to 0.1.

Results are shown in Table 4. Optimal policy depends on costs. Where costs are low, it is optimal to treat everyone infected with the High strain. As costs rise, it becomes optimal to treat no one with the High strain. There is a unique optimal policy in regions I and III; in region II, optimal policy depends on the initial prevalence of the High strain. If the High strain is very prevalent, it is optimal to treat no one. If it is not, it is optimal to treat everyone. This strongly suggests that there exists a value of I_H^0 that is a Skiba point in Region II. By varying I_H^0 by small increments in region II and evaluating social welfare at each initial value for the two candidate policies, we find that a Skiba point exists when $I_H^0 = 0.7125$. At this point, the social welfare from selecting $f_H^* = 1$ or $f_H^* = 0$ is equal.[14]

---

[13] Note that when $f_H^* = 1$, $I_H^* = 1 - \frac{\tau + \alpha}{\beta_H} = 0.6316$. When $f_H^* = 0$, $I_H^* = 1 - \frac{\tau}{\beta_H} = 0.8421$.

[14] All of these simulations show the same qualitative results when the initial prevalence of the Low strain is set to a smaller value: $I_L^0 = 0.01$, $I_L^0 = 0.001$ and $I_L^0 = 0.0001$.

Table 4 Optimal fixed policy as cost varies

Region	c	f_H^*	f_L^*
I (low costs)	$c < 0.2875$	1	1
II (intermediate costs)	$0.2875 \leq c \leq 0.3006$	0 or 1, depending on I_H^0	1
III (high costs)	$c > 0.3006$	0	1

5.1.2 Variable policy

In this set of simulations we allow policy to vary optimally over time. We look for strictly Hamiltonian paths; that is, we look for paths that are optimal because they satisfy the first-order conditions on λ_H and λ_L. For example, the policy ($f_H = 1$, $f_L = 1$) is Hamiltonian at a particular point in time $\bar{t}$ if $|\lambda_H|, |\lambda_L| > \frac{c}{\alpha}$ at $(I_H(\bar{t}), I_L(\bar{t}))$. In order to evaluate the Hamiltonian conditions, the co-state variables are needed at each point in time. These can be calculated using the facts that

$$\lambda_H(t) = \frac{\partial V(I_H(t), I_L(t))}{\partial I_H(t)},$$

$$\lambda_L(t) = \frac{\partial V(I_H(t), I_L(t))}{\partial I_L(t)}.$$

These partial derivatives can be approximated by perturbing the infection levels slightly. Thus, for initial infection levels I_H^0 and I_L^0, the initial values of the co-state variables are

$$\lambda_H^0 \approx \frac{V(I_H^0 + \Delta, I_L^0) - V(I_H^0, I_L^0)}{\Delta}, \tag{25}$$

$$\lambda_L^0 \approx \frac{V(I_H^0, I_L^0 + \Delta) - V(I_H^0, I_L^0)}{\Delta}, \tag{26}$$

for small Δ. In these simulations we set $\Delta = 0.001$. For a given initial infection level, we find the initial values of the co-state variables λ_H^0 and λ_L^0 using (25) and (26), and find the corresponding optimal initial policy. We then allow the system to evolve, checking for optimal policy at each increment of time by re-evaluating $\lambda_H(t)$ and $\lambda_L(t)$ using the equations for $\dot{\lambda}_H$ and $\dot{\lambda}_L$ that were derived in Sect. 3. We also simulate the paths backwards, starting at the initial point (I_H^0, I_L^0) but moving away from equilibrium.[15] Together, the forward and backward paths give a fairly complete picture of optimal policy over a period of length $2t$, with (I_H^0, I_L^0) in the middle. In this set of simulations we set $t = 30$, so that overall we examine the behaviour of the infection over 60 days. We assume costs are $c = 0.4$ and set initial prevalence at $(I_H^0, I_L^0) = (0.6316, 2 * 10^{-10})$.

---

[15] The backward paths are necessarily Hamiltonian paths as long as the Hamiltonian conditions for the forward paths are satisfied. This is because the backward paths are merely a continuation, albeit in the opposite direction, of the Hamiltonian paths.
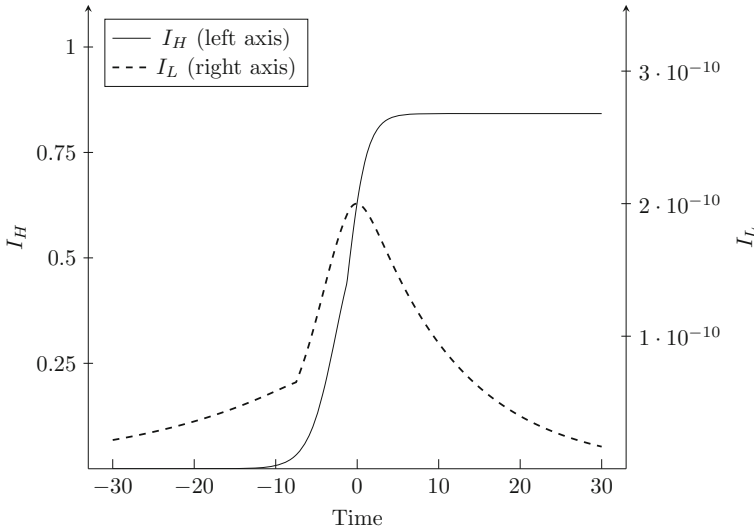
Fig. 5 Evolution of prevalence along the path towards equilibrium A_L^{high}, with $I_H^0 = 0.6316$, $I_L^0 = 2 * 10^{-10}$ and $c = 0.4$. Time >0 is the forward path, while time <0 is the backward path

We find that the system tends towards the equilibrium where $f_H^* = 0$, so that $I_H^* = 0.8421 = 1 - \frac{\tau}{\beta_H}$ and $I_L^* \to 0$. This is consistent with the findings in Table 4, as $c = 0.4$ falls into Region III of costs. The evolution of the two strains is depicted in Fig. 5. This figure shows both the forward and backward paths. We find that prevalence of I_H rises from a low initial point at first exponentially, and then roughly logarithmically until reaching the fixed point. Attainment of equilibrium takes approximately 15 days. The evolution of I_L takes a slightly different path. Prevalence first increases, and then decreases towards zero. This is because as I_H increases, there are fewer susceptibles that can become infected with the Low strain, so that the Low strain can be asymptotically eradicated.

When examining the sequence of optimal policy, we find that optimal policy exhibits switch points. This is depicted in Fig. 6. The treatment of the High strain switches from one to zero just before the start of the forward path, while the treatment of the Low strain switches from one to zero earlier.

Thus, optimal policy in this particular example has three phases:

$$f_H^* = 1, f_L^* = 1,$$
$$f_H^* = 1, f_L^* = 0,$$
$$f_H^* = 0, f_L^* = 0.$$

Thus, when policy is variable, it is always optimal to set $f_H^* = 0$ in equilibrium when $c = 0.4$. However, policy does not begin at this level: it switches from one to zero when the marginal cost of an additional infected individual in the population is low, relative to the cost of treating him. In addition, the asymptotic eradication of the
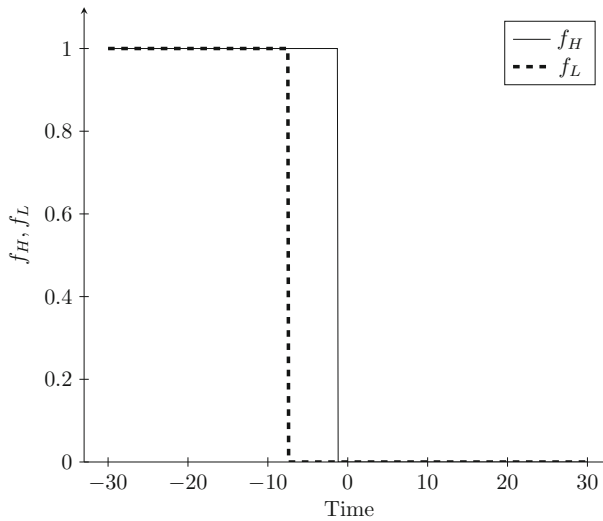
Fig. 6 Optimal policy along the path towards A_L^{high}, with $I_H^0 = 0.6316$, $I_L^0 = 2 * 10^{-10}$ and $c = 0.4$. Time > 0 is the forward path, while time < 0 is the backward path

Table 5 Parameter values for K>1

| Parameter	Value
β_H	0.8
β_L	0.4
τ	0.15
α	0.2
p	1
δ	0.222

Low strain is possible in the closing part of the path even when no one in the Low strain is treated. Since this is feasible, it is also optimal, as it is cheaper than treating unnecessarily.

5.2 Optimal policy when $K < 1$

In this section we address the case when $K < 1$. In this regime, it is possible to have equilibria where both strains of the disease prevail. The parameter values are given in Table 5.

We simulate optimal policy in the region of one of the non-boundary fixed points. In particular, we assume that it is optimal to go to the fixed point F^{low}, and simulate optimal policy backwards, starting from that fixed point. The evolution of prevalence is shown in Fig. 7.

Fig. 7 Evolution of prevalence along the path towards equilibrium F^{low}, with $I_H^0 = 0.0755$, $I_L^0 = 0.4870$ and $c = 0.4$. This path is simulated only backwards
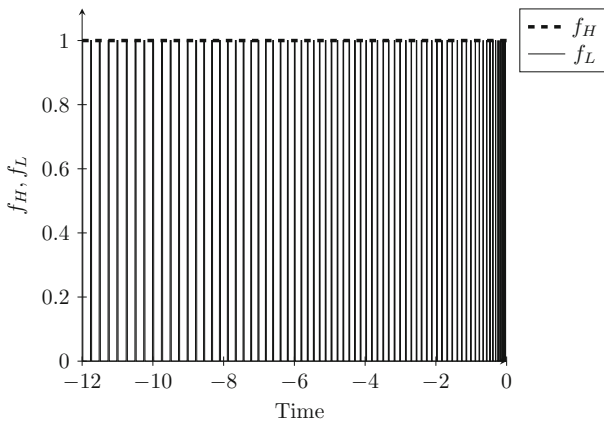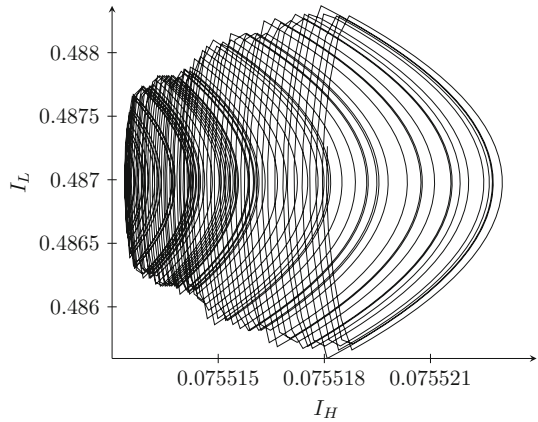
Fig. 8 Evolution of optimal policy along the path towards equilibrium F^{low}, with $I_H^0 = 0.0755$, $I_L^0 = 0.4870$ and $c = 0.4$. This path is simulated only backwards

This figure clearly shows a spiral. These simulations suggest that non-boundary fixed points are never the end points of optimal paths. In every case examined, perturbing the fixed point solution and running the path backwards generates a high frequency cycle. Reversing the path from its end point generates an integral which is exceedingly large. Figure 8 shows the evolution of optimal policy along this same path. The treatment of the Low strain switches between zero and one many times. Therefore, we conclude that even in the case when $K < 1$, it is optimal to move to one of the boundary fixed points. It appears that coexistence of the two strains is never optimal.

5.3 Chlamydia trachomatis

We began this paper with the example of chlamydia trachomatis, which is a disease with three strains that satisfies the behaviour of an SIS infection. As accurate modelling

Table 6 Parameters characterising chlamydia trachomatis

Parameter	Value	Source
β_H (LGV)	0.60	N/A
β_L (Chlamydia)	0.55	Althaus et al. (2012)
τ	0.18	Geisler et al. (2008)
α	0.97	Lau and Qureshi (2002)
K	0.02	
Eradicate High strain?	Yes	
Eradicate Low strain?	Yes	

of the evolution of chlamydia trachomatis has been done in detail in the literature (e.g. Heijne et al. 2011; Sharomi and Gumel 2009; De Vries et al. 2006), we focus here on the distinction between the various strains, and the calculation of the cut-off parameter K, which is the novel contribution of this paper. In Table 6, we present the parameters that characterise chlamydia (serovars D-K) and LGV (serovars L1–L3) in the context of our SIS model with two strains.

There are challenges to estimating the right parameter values for diseases such as chlamydia and LGV. Due to a lack of studies on LGV, we use the recovery rate and treatment success rate of chlamydia. The transmission parameter depends on the number of contacts between the infected partner and the non-infected partner, and we use the per-partnership parameter estimated by Althaus et al. (2012). We were unable to find an estimate of the transmission parameter for LGV, but we know that it is higher than that for chlamydia, so we set it marginally higher at 0.6. Next, the spontaneous recovery parameter is challenging to measure because of the ethical considerations of not treating individuals diagnosed with chlamydia (Geisler 2010). Geisler et al. (2008) screened individuals for chlamydia and then screened them again when they came for a treatment visit, for which the median was 13 days later, where they found spontaneous resolution in 18% of patients. Finally, the treatment success probability faces the challenge of non-compliance (Bachmann et al. 1999). We use the measured success rate of treatment with azithromycin in a meta-analysis of randomised control trials, which is found to be 97% (Lau and Qureshi 2002).

These parameters imply two observations about chlamydia trachomatis in our model: first, the value of K is less than one, so that the full range of fixed points should be attainable—those that involve coexistence of chlamydia and LGV, and those that involve asymptotic eradication of chlamydia and positive prevalence of LGV, although our simulations suggest that coexistence is not the end of an optimal path. Second, and more importantly, both diseases should be possible to eradicate, if everyone who is infected is treated. The question then is why the prevalence of chlamydia, for example, is around 6% among people younger than 25 in the UK (Macleod et al. 2005). The likely reason is that chlamydia is often an asymptomatic disease: 75-85% women are asymptomatic, while over 50% of men are asymptomatic (Risser et al. 2005). Thus, the treatment parameter is far from equal to one. In fact, a rough back-of-the envelope calculation, assuming values of $I_H^* = 0.001$, $I_L^* = 0.06$ and the fact that total preva-

lence in equilibrium is given by Eq. (3) implies a treatment parameter of $f_L = 0.35$.[16] This suggests that only 35% of individuals infected with chlamydia at any given time are treated.

6 Conclusion

We have derived analytically the fixed points in an SIS model with two strains, where neither superinfection nor eradication of the strains are possible, and the policymaker has access to two separate treatment instruments. We have characterised two types of fixed points: non-boundary fixed points, where the two strains co-exist, and boundary fixed points, which involve the asymptotic eradication of one strain. The feasibility of these fixed points depends on parameter values: we have defined the parameter K, which delineates three regimes of feasibility. For $K > 1$, the only attainable equilibria are those where the Low strain is asymptotically eradicated and the High strain prevails. For $K \leq 1$, the coexistence of the two strains is a feasible equilibrium.

The simulations present two interesting results. First, when $K > 1$, optimal policy along the path to equilibrium exhibits switch points. This implies that a fixed policy is suboptimal. Further, in the neighbourhood of the equilibrium in our example, it is optimal to treat no one. This still allows the eradication of the Low strain. Second, when $K < 1$, we simulate optimal policy in the neighbourhood of one of the non-boundary fixed points. We find that optimal policy exhibits a spiral. This suggests that this equilibrium can never be the end point of an optimal path, and hence that coexistence, although feasible, is unlikely to be optimal. This suggests that for all values of K, the welfare-maximising policy is one that leads to one of the asymptotic equilibria.

Further research should consider extending this model to include protection via vaccination as another instrument available to the policymaker.

Appendices

A Index of notation and acronyms

Table 7 defines the notation and acronyms used in this paper.

---

[16] This is implied by an equilibrium prevalence of $I_H + I_L = 1 - \frac{\tau + \alpha f_L}{\beta_L}$. The prevalence of LGV is very low: Ward et al. (2007) estimate that there were only 327 cases in the UK between 2004 and 2006. We assume a prevalence of 0.001 but setting this number closer to zero would not have a substantive impact on our estimate of f_L.

Table 7 Index of notation and acronyms

Notation	Definition
$j = H, L$	High, Low infectivity strain
$I_j(t)$	The prevalence of strain j in the population at time t
I_j^0	The initial prevalence of strain j
f_j	The proportion of individuals infected with strain j that are treated
β_j	The probability that a healthy person encountering a strain j infectee will become infected
α	The success rate of treatment
τ	The natural rate of recovery from the infection
c	The marginal cost of treatment
p	The value of a healthy individual in social welfare
δ	The discount rate
λ_j	The multiplier in the Hamiltonian on the equation of motion for infection j
F^{high}	Fixed point where $f_H^* = \frac{(\beta_H - \beta_L)}{\beta_L} \frac{\tau}{\alpha}$ and $f_L^* = 0$
F^{low}	Fixed point where $f_H^* = 1$ and $f_L^* = 1 - \frac{(\beta_H - \beta_L)}{\beta_H} \frac{\tau + \alpha}{\alpha}$
F^{int}	Continuum of fixed points where $f_H^* = 1$ and $f_L^* = 0$
A_L^{high}	The asymptotic fixed point where H is asymptotically eradicated, $f_H^* = 1$ and $f_L^* = 0$
A_H^{high}	The asymptotic fixed point where L is asymptotically eradicated, $f_H^* = 0$ and $f_L^* = \in \{0, 1\}$
A_H^{low}	The asymptotic fixed point where L is asymptotically eradicated, $f_H^* = 1$ and $f_L^* = 1$
Z	The set of feasible equilibria
K	Equal to $\frac{\beta_H - \beta_L}{\beta_L} \frac{\tau}{\alpha}$, the parameter that characterises regimes of feasibility
P^{ab}	The treatment policy $f_H = a$, $f_L = b$
AFP	Asymptotic fixed point
FP	Fixed point
MRAP	Most Rapid Approach Path

B Asymptotic eradication

Proposition 1 *Both variants of the disease cannot be simultaneously eradicated even asymptotically in equilibrium if*

$$\frac{\tau + \alpha}{\beta_H} < 1,$$

$$\frac{\tau + \alpha}{\beta_L} < 1,$$

which hold as a consequence of Assumption (4).

Proof To see this, first consider the case of the interior fixed points. Here, it is trivial. In the case of F^{low}, $I_H^* + I_L^* = 1 - \frac{\alpha + \tau}{\beta_H}$. It is not possible for $I_H^* + I_L^* = 0$. Similarly

for F^{high} where $I_H^* + I_L^* = 1 - \frac{\tau}{\beta_L}$, it is not possible to have $I_H^* + I_L^* = 0$. Next, consider the asymptotic fixed points. In the case of A_L, it is known that $I_L^* \to 0$ so the question is what happens to I_H^*. For I_L to tend towards zero asymptotically, the necessary condition is $I_H > 1 - \frac{\alpha f_H + \tau}{\beta_L}$. Clearly $I_H \neq 0$ is necessary for this to be satisfied. Similarly, fixed point A_H implies that $I_H^* \to 0$. The necessary condition for this is $I_L > 1 - \frac{\alpha f_L + \tau}{\beta_H}$, which can only be satisfied if $I_L \neq 0$. Thus, both variants of the disease cannot be eradicated in equilibrium, even asymptotically. $\qquad\square$

C Feasible policy along the path to equilibrium

In order to derive which policies are feasible along the path towards equilibrium, it is necessary to examine each MRAP policy in turn and derive the conditions that are required for I_H and I_L to either increase or decrease.

First, consider P^{00}. This policy implies

$$\frac{\dot{I}_H}{I_H} = \beta_H(1 - I_H - I_L) - \tau,$$

$$\frac{\dot{I}_L}{I_L} = \beta_L(1 - I_H - I_L) - \tau.$$

There are two ways of approaching a fixed point with this policy. First, it is possible that $\dot{I}_H, \dot{I}_L > 0$ (i.e. I_H and I_L are increasing towards I_H^* and I_L^*). The required conditions for this are

$$1 - \frac{\tau}{\beta_L} > I_H + I_L,$$

$$1 - \frac{\tau}{\beta_H} > I_H + I_L,$$

which collapse to

$$1 - \frac{\tau}{\beta_L} > I_H + I_L.$$

Similarly, it is possible that $\dot{I}_H, \dot{I}_L < 0$ (i.e. I_H and I_L are decreasing towards I_H^* and I_L^*). The required conditions for this are

$$1 - \frac{\tau}{\beta_L} < I_H + I_L,$$

$$1 - \frac{\tau}{\beta_H} < I_H + I_L,$$

which collapse to

$$1 - \frac{\tau}{\beta_H} < I_H + I_L.$$

Next, consider P^{10}. For $\dot{I}_H, \dot{I}_L > 0$, the required conditions are

$$1 - \frac{\tau}{\beta_L} > I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_H} > I_H + I_L,$$

where the overriding condition is

$$1 - \frac{\tau}{\beta_L} > I_H + I_L.$$

For $\dot{I}_H, \dot{I}_L < 0$, the required conditions are,

$$1 - \frac{\tau}{\beta_L} < I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_H} < I_H + I_L,$$

both of which are satisfied when

$$1 - \frac{\tau + \alpha}{\beta_H} < I_H + I_L.$$

Third, take P^{11}. For $\dot{I}_H, \dot{I}_L > 0$, it needs to be the case that

$$1 - \frac{\tau + \alpha}{\beta_H} > I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_L} > I_H + I_L,$$

where the overriding condition is

$$1 - \frac{\tau + \alpha}{\beta_L} > I_H + I_L.$$

For $\dot{I}_H, \dot{I}_L < 0$, it is required that

$$1 - \frac{\tau + \alpha}{\beta_H} < I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_L} < I_H + I_L,$$

both of which are satisfied when

$$1 - \frac{\tau + \alpha}{\beta_H} < I_H + I_L.$$

Last, consider P^{01}. For $\dot{I}_H, \dot{I}_L > 0$, it needs to be the case that

$$1 - \frac{\tau}{\beta_H} > I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_L} > I_H + I_L.$$

where the overriding condition is

$$1 - \frac{\tau + \alpha}{\beta_L} > I_H + I_L.$$

For $\dot{I}_H, \dot{I}_L < 0$, it is required that

$$1 - \frac{\tau}{\beta_H} < I_H + I_L,$$

$$1 - \frac{\tau + \alpha}{\beta_L} < I_H + I_L,$$

both of which are satisfied when

$$1 - \frac{\tau}{\beta_H} < I_H + I_L.$$

References

Althaus CL, Heijne JC, Low N (2012) Towards more robust estimates of the transmissibility of Chlamydia trachomatis. Sex Transm Dis 39(5):402–404

Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford

Armstrong RA, McGehee R (1980) Competitive exclusion. Am Nat 115(2):151–170

Bachmann LH, Stephens J, Richey CM, Hook EW 3rd (1999) Measured versus self-reported compliance with doxycycline therapy for Chlamydia-associated syndromes: high therapeutic success rates despite poor compliance. Sex Transm Dis 26(5):272–278

Balmer O, Tanner M (2011) Prevalence and implications of multiple-strain infections. Lancet Infect Dis 11(11):868–878

Bansal S, Grenfell BT, Meyers LA (2007) When individual behaviour matters: homogeneous and network models in epidemiology. J R Soc Interface 4:16

Byrne GI (2010) Chlamydia trachomatis strains and virulence: rethinking links to infection prevalence and disease severity. J Infect Dis 201(Suppl 2):S126–S133

Castillo-Chavez C, Huang W, Li J (1999) Competitive exclusion and coexistence of multiple strains in an SIS STD model. SIAM J Appl Math 59(5):1790–1811

Castillo-Chavez C, Feng Z (1997) To treat or not to treat: the case of tuberculosis. J Math Biol 35:629–656

Chen FH, Toxvaerd F (2014) The economics of vaccination. J Theor Biol 363:105–117

De Vries R, Van Bergen JEAM, Jong-van De, den Berg LTW, Postma MJ (2006) Systematic screening for Chlamydia trachomatis: estimating cost-effectiveness using dynamic modeling and Dutch data. Value Health 9(1):1–11

Elbasha EH, Galvani AP (2005) Vaccination against multiple HPV types. Math Biosci 197(1):88–117

Geisler WM (2010) Duration of untreated, uncomplicated Chlamydia trachomatis genital infection and factors associated with chlamydia resolution: a review of human studies. J Infect Dis 201(Suppl 2):S104–S13

Geisler WM, Wang C, Morrison SG, Black CM, Bandea CI, Hook EW 3rd (2008) The natural history of untreated Chlamydia trachomatis infection in the interval between screening and returning for treatment. Sex Transm Dis 35:119–123

Goldman SM, Lightwood J (2002) Cost optimisation in the SIS model of infectious disease with treatment. B. E. J Econ Anal Policy (Topics) 2(1):1–22

Heiligenberg M, Verweij SP, Speksnijder AGCL, Morre SA, De Vries HJC, Schim van der Loeff MF (2014) No evidence for LGV transmission among heterosexuals in Amsterdam, the Netherlands. BMC Res Notes 7

Heijne JCM, Althaus CL, Herzog SA, Kretzschmar M, Low N (2011) The role of reinfection and partner notification in the efficacy of Chlamydia screening programs. J Infect Dis 203(3):372–377

Iggidr A, Kamgang J-C, Sallet G, Tewa J-J (2006) Global analysis of new malaria intrahost models with a competitive exclusion principle. SIAM J Appl Math 67(1):260–278

Kari L, Whitmire WM, Carlson JH, Crane DD, Reveneau N, Nelson DE, Mabey DCW, Bailey RL, Holland MJ, McClarty G, Caldwell HD (2008) Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis 197(3):449–456

Keeling MJ, Eames KTD (2005) Networks and epidemic models. J R Soc Interface 2:295–307

Lau C-Y, Qureshi AK (2002) Azithromycin versus doxycycline for genital chlamydial infections: a meta-analysis of randomized clinical trials. Sex Transm Dis 29(9):497–502

Li X-Z, Liu J-X, Martcheva M (2010) An age-structured two-strain epidemic model with super-infection. Math Biosci Eng 7:123–147

Macleod J, Salisbury C, Low N et al (2005) Coverage and uptake of systematic postal screening for genital Chlamydia trachomatis and prevalence of infection in the United Kingdom general population: cross sectional study. Brit Med J 330:940–942

Manski CF (2013) Diagnostic testing and treatment under ambiguity: using decision analysis to inform clinical practice. Proc Natl Acad Sci 110(6):2064–2069

McLean CA, Stoner BP, Workowski KA (2007) Treatment of lymphogranuloma venereum. Clin Infect Dis 44(Suppl 3):S147–S152

Oster E (2005) Sexually transmitted infections, sexual behaviour and the HIV/AIDS epidemic. Q J Econ 120(2):467–515

Pearce JH, Prain CJ (1986) Intracellular association of Chlamydia psittaci with lysosomes during infection of McCoy cells. In: Oriel D, Ridgeway D, Schachter J, Taylor-Robinson D, Ward M (eds) Chlamydial infections. Cambridge University Press, London

Risser W, Bortot A, Benjamins L et al (2005) The epidemiology of sexually transmitted infections in adolescents. Semin Pediatr Infect Dis 16:160–167

Rowthorn RE, Laxminarayan R, Gilligan CA (2009) Optimal control of epidemics in metapopulations. J R Soc Interface 6:1135–1144

Rowthorn RE (2006) Optimal treatment of disease under a budget constraint. In: Halvorsen R, Layton DF (eds) Explorations in environmental and natural resource economics: essays in honor of Gardner M. Brown, Jr. Edward Elgar Publishing, Cheltenham, pp 20–35

Rowthorn RE, Toxvaerd F (2015) The optimal control of infectious diseases via prevention and treatment. Mimeo

Seinost G, Dykhuizen DE, Dattwyler RJ, Golde WT, Dunn JJ, Wang I, Wormser GP, Schriefer ME, Luft BJ (1999) Four clones of Borrelia burgdorferi sensu stricto cause invasive infection in humans. Infect Immun 67(7):3518–3524

Sharomi O, Gumel AB (2009) Re-infection-induced backward bifurcation in the transmission dynamics of Chlamydia trachomatis. J Math Anal Appl 356(1):96–118

Shaw K, Coleman D, O'Sullivan M, Stephens N (2011) Public health policies and management strategies for genital Chlamydia trachomatis infection. Risk Manag Healthc Policy 4:57–65

Skiba AK (1978) Optimal growth with a convex-concave production function. Econometrica 46(3):527–539

Spence M, Starrett D (1975) Most rapid approach paths in accumulation problems. Int Econ Rev 16(2):88–403

Ströbele WJ, Wacker H (1991) The concept of sustainable yield in multi-species fisheries. Ecol Model 53:61–74

Tabbara KF, Abu-el-Asrar A, Al-Omar O, Choudhury AH, Al-Faisal Z (1996) Single-dose azithromycin in the treatment of trachoma: a randomized, controlled study. Ophthalmology 103(5):842–846

Truscott J, Fraser C, Cauchemez S, Meeyai A, Hinsley W, Donnelly CA, Ghani A, Ferguson N (2012) Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. J R Soc Interface 67:304–312

Wagener FOO (2003) Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. J Econ Dyn Control 27(9):1533–1561

Ward H, Martin I, Macdonalds N, Alexander S, Simms I, Fenton K, French P, Dean G, Ison C (2007) Lymphogranuloma venereum in the United Kingdom. Clin Infect Dis 44(1):26–32