June 12, 2003.

The Economical Control of Infectious Diseases*

by

Mark Gersovitz
Department of Economics
The Johns Hopkins University

and

Jeffrey S. Hammer
The World Bank

Abstract

The structure of representative agents and decentralization of the social planner's problem provide a framework for the economics of infection and associated externalities. Optimal implementation of prevention and therapy depends on: (1) biology including whether infection is person to person or by vectors; (2) whether the infected progress to recovery and susceptibility, immunity, or death; (3) costs of interventions; (4) whether interventions target everyone, the uninfected, the infected, or contacts between the two; (5) individual behaviour leading to two types of externalities. By way of example, if people recover to be susceptible, government subsidies should equally favour prevention and therapy.

JEL Classification: H41, I12, O15.

Key Words: Health, Infections, Prevention, Therapies, Epidemiology.

*We thank Edward F. Buffie, Daniel Leonard, Steve Salant and Carl Simon for comments. The findings, interpretations and conclusions in this paper are entirely those of the authors and do not necessarily represent the views of the World Bank, its executive directors or the countries they represent.

The economic approach to infectious diseases is in its infancy, somewhat oddly because many economists have long had the intuition that epidemics and infectious diseases are quintessential manifestations of the principle of an externality, itself a central concept in economics (Gersovitz and Hammer, forthcoming a). Furthermore, epidemiology provides ready-made dynamic models of disease transmission and economics provides methods of valuing the costs and benefits of health interventions and methods of dynamic optimization to guide policy. Policy toward infections is of great importance. Yet only recently have economists begun to look at these questions in a formal way.

This paper has two main goals. The overarching goal is to dissect two important externalities involving infectious diseases when there are the options of both prevention and therapy, so that the relative phasing of these two types of interventions is important. To achieve this goal, we need to state the social planner's or first-best problem and to compare it to a representative agent's problem. To avoid compounding the identification of any externalities with problems of myopia, imperfect ability to insure health outcomes, or a disregard for the welfare of future generations, we make assumptions about the behaviour of the representative agent such that these problems do not arise, allowing us to focus exclusively on the externalities that we identify. These assumptions may not always be adequate approximations and a relaxation of these assumptions would lead to further considerations in the design of interventions, but we do not deal with these issues in this paper.

The economic literature on infectious diseases has taken as its starting point some special although important concerns. This starting point has influenced modelling strategies in ways that we believe obscures the general structure of private choice and consequent externalities in the process of disease transmission.

One focus of the previous literature has been on vaccinations (Brito et al, 1991, Francis,

1996, Geoffard and Philipson, 1996 and Kremer, 1996). It is certainly an important intervention but one that is only possible for a limited number of diseases. Importantly from a modelling strategy, moreover, vaccination is plausibly a discrete decision, to vaccinate or not, and whether it is or not, that is how it has been modelled. Indeed, Francis models vaccination as a discrete choice by identical individuals. With neither an intensive margin (choice of intensity of vaccination) nor an extensive margin (heterogenous individuals), he concludes that vaccination does not exhibit externalities, a result that does not, however, generalize (Gersovitz, 2003).

Philipson (2000 and cited references) and associates and Kremer (1996) have focussed on HIV/AIDS. For this disease prevention is naturally the almost exclusive focus, none of this work analyses therapeutic behaviour, and the transition from infection is to death, rather than to recovered-and-susceptible or immune. As we will show, the state that follows infection is important for the analytical tractability of the model, and a fatal disease is not the easiest to analyse. In fact, Geoffard and Philipson (1996) analyse HIV under the assumption that infection raises an individual's discount rate but does not lead to a diminution in the population; while tractable, such a formulation hardly incorporates the salient feature of this fatal infection. Furthermore, the work of Philipson and his associates formulates prevention as an entirely discrete choice (safe or risky sex). Whether this formulation is realistic for HIV is open to question; regardless, it is clear to us that it ignores the important scope for varying preventive effort that arises for many other diseases which present a virtual continuum of degrees of prevention. Oral-fecal diseases provide good examples of such a continuum: boil water progressively longer, wash hands on more and more occasions. Kremer (1996) stresses heterogeneous behaviour in the population which leads him to a wide range of interesting conclusions about HIV. For our purposes, however, his assumption that people are not future-

oriented in their risk-taking behaviour limits the applicability of his model to our question because it confounds myopia with externalities.

In the economics literature on infections, Wiemer(1987) is an exception to the study of one type of intervention in isolation. She models the use of two interventions in the case of schistosomiasis (bilharzia), but under assumptions about costs and benefits that imply that one of the two interventions is used maximally or not at all. Furthermore, she does not model decisions by individuals and the associated externalities of this type of vector-spread disease. Epidemiologists have also analysed the problem of optimal interventions to control infections, but have not surprisingly ignored many of the issues we investigate.[1]

Once we have adopted a framework that can achieve our first goal of analysing externalities and infectious diseases, our second goal is to examine a typology of infectious diseases. This typology allows for people to progress from being susceptible to infected (and infectious) to: (1) recovered but again susceptible, or (2) immune, or (3) dead. Furthermore, the typology includes diseases that spread from person to person and those that are spread by intermediate vectors, such as mosquitos. Finally, the typology allows for the targeting of different interventions at different groups. We show how these different characteristics of the epidemic affect conclusions about how to offset externalities and how the preventive and therapeutic interventions are phased relative to each other.

---

[1]Wickwire (1977) surveys early work and Sethi and Staats (1978), Greenhalgh (1988) and Hocking (1991) are more recent. This literature, however, has several shortcomings. First, the objective function is often poorly specified, with no discounting and a finite horizon. More importantly, marginal costs and benefits are typically assumed constant leading to unrealistic bang-bang solutions. Usually, there is no discussion of the co-ordination of multiple interventions. Finally, these models do not incorporate health choices by individuals, and therefore do not discuss externalities and the decentralization of the optimal policies, a topic we stress.

In fact, we identify two types of externalities: First, infectious people can infect other people who in turn infect others and so on, the source of what we call the pure infection externality. This externality arises if, in choosing their own levels of preventive and therapeutic effort, people do not fully take into account the costs to others who will become infected as a consequence of their being infectious. Second, there is a pure prevention externality that arises because the preventive actions of one individual may directly affect the probability that other people become infected, whether or not the preventive action prevents infection of the individual undertaking it. Typically, the pure prevention externality arises only for diseases that involve a vector and therefore appears only in the last model of the paper. An example is the use of insecticides to kill mosquitos that carry disease. The person using the insecticide may or may not be bitten and infected but the killing of mosquitos lessens the probability that others will be bitten and infected, something that the user of the insecticide may disregard. A third type of externality involves pathogen resistance to drugs, both preventive and therapeutic, but we do not examine this potentially important problem.

As is conventional in modelling externalities, the paper begins with the problem of a hypothetical social planner who can directly control all preventive and therapeutic actions, initially in a model of infection from person to person. First, we lay out the accounting for the people who are in different disease statuses and the dynamics that move people from one status to another, the constraints on the optimization problem. Next, we introduce the objective of decision makers: the maximization of utility of income net of the costs of the disease in terms of discomfort, fear and economic loss and net of the costs of preventive and therapeutic measures. We then maximize the objective function subject to the constraints, and look at the optimal solution to the social planner's problem and how it depends on what happens to people who are

infected. The next section looks at the decentralized decisions, their deviation from the social planner's choices, and hence the existence of externalities and the role for public interventions. A penultimate section looks at some of these issues when vectors transmit disease. The paper ends with some concluding remarks.

1. The Social Planner's Problem

1.1 The Dynamic Constraints:

The starting point for the study of optimal policy toward infectious diseases is the classic literature on mathematical epidemiology. It provides the dynamic constraints that condition decisions about infectious diseases. This literature models many diseases, ones transmitted directly from person to person and ones transmitted by vectors (Anderson and May, 1991).

In the most general model of diseases transmitted from person to person that we consider, the total number of people (N) is the sum of the number who are: (1) susceptible (S); (2) infected and infectious (I); and (3) recovered and immune, i.e. uninfectible (U):

$$N = S+I+U \quad . \tag{1}$$

The proportions of these groups in the population are denoted by s, i, and u, with $s + i + u = 1$. The birth rate of the population is ϵ while deaths only occur as a proportion, δ, of the infections at any time, so that the net change in the population is :

$$\dot{N} = \epsilon N - \delta I \quad . \tag{2}$$

Note if people do not die of the disease, they never die; this assumption is inconsequential for the qualitative results that we derive and saves on a nuisance parameter, the baseline level of mortality for people who do not die of the disease.

The number of susceptibles changes according to:

(3) $\dot{S} = \epsilon N - \alpha Si + \beta I$.

The first part of the right-hand side embodies the assumption that all newborns are susceptible. The second part reduces the number of susceptibles by those people who become infected. Under the assumption of random contacts, the probability per contact of a susceptible person's meeting an infected (and infectious) person is the proportion of infected people in the population, i = I/N.[2] The product, Si, is the number of susceptibles who do so. The factor α is an adjustment incorporating both the rate of contact and the inherent infectiousness of an infected (or susceptibility of a susceptible). The third part is the addition to the susceptible pool resulting from the recovery of a fraction, β, of the infecteds.

The number of infecteds evolves according to:

(4) $\dot{I} = \alpha Si - \beta I - \delta I - \gamma I$.

The first three terms on the right-hand side have been discussed in connection with equations (2) and (3). The last term accounts for the transition of the fraction γ of the infecteds to the status of immunes. Correspondingly, the number of immunes evolves according to:

(5) $\dot{U} = \gamma I$.

These equations can be solved for the change in the three proportions:

(6) $\dot{s} = (1-s)\epsilon + (\delta - \alpha)si + \beta i$,

(7) $\dot{i} = \alpha si + \delta i^2 - (\epsilon + \delta + \beta + \gamma)i$,

and

(8) $\dot{u} = \delta iu + \gamma i - \epsilon u$.

---

[2] We assume that each person has the same rate of contacts. Kremer (1996) examines some behavioural aspects of models in which different people have different rates of contact.

In the subsequent discussion, our main purpose is to consider how preventive and therapeutic actions by governments and individuals affect the parameters of the model: α, β, δ, and γ. Without any such interventions, however, these parameters are fixed, and the model of equations (6)-(8) evolves to a steady state. In particular, the steady state may be one in which s = 1 and i = u = 0 and the disease disappears, rather than one in which s, i and u lie strictly between 0 and 1. In the former case, optimal policy would be to approach optimally the steady state in which the disease disappears. In other cases, it may be desirable to adopt policies that eradicate the disease even though in the absence of interventions the steady-state proportion of infecteds would be positive. With few exceptions, however, we do not believe that eradication, optimal or otherwise, is feasible, and we will be studying situations in which optimal policy involves optimally moving to and sustaining a steady state with a positive level of infection.[3]

1.2 The Costs and Benefits of Interventions and the Social Planner's Optimization:

We are now ready to specify the social planner's objective function. In the total absence of the disease, each person in the society would have a level of income of V_0. Because we

[3]While this paper was in the last stage of review, a paper by Goldman and Lightwood (2002) appeared that is the only mathematical discussion of when eradication dominates chronic control of an infection (and vice versa) that we know. Their analysis shows that there can be more than one optimal steady state in a model of the optimal control of infections that is similar to the ones we discuss here, and that there can be both stable and unstable optimal steady states. For instance, they provide an example where for high rates of infection the system moves to a steady state with endemic infection while for low levels of infection the system moves toward the (asymptotic) eradication of the infection. They model a somewhat special case where therapy is the only intervention and people recover to be again susceptible. Their analysis should be of pivotal importance to the investigation of this important question in the more complex models we discuss here where we are basically restricting the analysis to the discussion of the properties of a stable endemic optimal steady state and the approach to it on either the assumption that other steady states do not exist or that the system is in the region of attraction to an endemic steady state.

assume that utility is linear in income, V_0 would also be their level of utility.

When the disease exists, however, the social planner likely will be spending resources on prevention and on therapies. Either intervention may be targeted in the sense that only a proportion of the population generates costs associated with the intervention. Let θ^j, j = a, b be the proportions of the population that generate either preventive or therapeutic costs associated with an infectious disease. We refer to the θ^j as targeting functions; in general, they depend on s and i. The most natural formulation would be for prevention to be targeted at the susceptible ($\theta^a = s$) and for therapies to be targeted at the infected ($\theta^b = i$). Other formulations may, however, be plausible depending on the ability to identify and reach different groups and what makes sense in terms of the disease and the balance of costs and benefits, something considered in more detail in the following sections that deal with special cases. The type of targeting may be a choice variable, but in this paper we will assume that it is a technical given and compare the behaviour of the model under different targeting assumptions.

The government uses two policy interventions, preventive effort of $a \geq 0$ units per person at whom prevention is targeted and therapeutic effort of $b \geq 0$ units per person at whom therapeutic effort is targeted. The total number of units of these interventions are therefore $a\theta^a N$ and $b\theta^b N$. The level of these health inputs per targeted person affects the parameters of the model, $\alpha(a)$ and $\beta(b)$, $\gamma(b)$ and $\delta(b)$, and thereby determine respectively the rate of new infections and the rates of transition to recovered-but-susceptible, immune, and dead. The controls exhibit diminishing marginal products so that: $\alpha'<0$, $\alpha''>0$, $\beta'>0$, $\beta''<0$, $\gamma'>0$, $\gamma''<0$, and $\delta'<0$, $\delta''>0$. This property of our formulation distinguishes it from all the preceding work that we know, and it opens the scope for internal solutions to the optimal policy problem and for the analysis of the co-ordination of multiple interventions phased in a smooth way over the course of

an epidemic. We believe that for many if not all diseases there is scope for undertaking additional preventive and therapeutic interventions although they are marginally less and less productive. Interventions are costly; preventive effort costs p_a per unit and therapeutic effort costs p_b per unit; total costs of the interventions are therefore $p_a a\theta^a N$ and $p_b b\theta^b N$.

The objective of government policy is to maximize the present discounted value of social welfare as given by the present discounted value of total income net of the total costs of the disease and the total costs of the interventions:

$$(9) \qquad W = \int_0^\infty \{ V_0 N - [p_I iN + p_a a\theta^a N + p_b b\theta^b N] \} e^{-rt} dt$$

in which r is the discount rate, V_0 is income in the absence of the disease (received by everyone who is alive whether well or sick), p_I is the current money cost of being infected (and sick) such as foregone wages while ill and including the monetary equivalent of pain and suffering and iN are the total number of the sick, and the remaining two terms are the total costs of the preventive and therapeutic interventions. The integrand of equation (9) is therefore the sum of the current incomes of the infected, the uninfected, and the uninfectible, less the total current costs of illness and health interventions. Future costs of a current illness arising from the failure to be cured instantly or the subsequent infection of others are accounted for when they happen, either by the continuation of the infected status in future periods (again at a cost of p_I) or the accretion of new infections as they occur; these future costs are not included in the current cost of being infected.

As mentioned, the objective function embodies an assumption of linearity in the value of income net of the costs of illness and expenditures on health interventions. This assumption

simplifies many aspects of the subsequent calculations; we introduce concavity into the model via the diminishing returns of health expenditures on the parameters of the dynamics of the epidemic rather than through diminishing marginal utility of income net of all the costs associated with the disease. Without the assumption of linearity, it would be important to specify who pays which costs of the interventions. For instance, the costs of therapies may be deducted only from the incomes of the infected or health insurance may spread these costs across everyone. If each person has a concave utility of consumption net of the total health costs (the costs of prevention, therapies and illness) that each pays, it would be necessary to pay attention to different constraints on the way the social planner can distribute these costs. The social planner's objective function would be the sum of the concave utilities of the people in different disease statuses; these utilities would depend on the difference between income V_0 and the total health costs people pay. Among other simplifications, linearity of the objective function allows us to sidestep the question of the interaction between the epidemic and the health insurance regime, the latter itself a complex topic and one worthy of analysis in future research. Gersovitz and Hammer (forthcoming b) provide numerical results when utility is concave in income net of health expenditures and all the costs of the infection are shared equally in a model of vector-borne infection.

Equation (9) therefore provides the objective function while equations (2) and (6)-(8) provide the dynamic equations that constrain the optimization problem. The current-value Hamiltonian, H, is:

$$
\begin{aligned}
(10) \qquad H \;=\; & N\,\{\, V_0 - [p_I i + p_a a\theta^a + p_b b\theta^b] \,\} \\
& + (\lambda_s\, N)[(1-s)\epsilon + (\delta-\alpha)si + \beta i] \\
& + \lambda_N\, [N(\epsilon - \delta i)] \\
& + (\lambda_i\, N)[\alpha si + \delta i^2 - (\epsilon + \gamma + \delta + \beta)i] \quad ,
\end{aligned}
$$

in which ($\lambda_s N$), λ_N and ($\lambda_i N$) are the current value multipliers. We can now provide necessary conditions for a maximization of this objective function with respect to expenditures on preventive and therapeutic actions

The first derivatives of H with respect to the controls, a and b, set equal to zero imply:

$$(11a) \qquad p_a\theta^a = (\lambda_i - \lambda_s)\alpha' s i$$

and

$$(11b) \qquad p_b\theta^b = \lambda_s(\delta' s + \beta')i - \lambda_N \delta' i + \lambda_i[\delta' i^2 - (\delta' + \beta' + \gamma')i] \quad .$$

Under the assumptions on the θ^j and on α' and β', the expression (λ_s - λ_i) must be positive if the first-order conditions are to hold. In addition, the dynamic equations for the multipliers imply:

$$(11c) \qquad \dot{\lambda}_s = r\lambda_s + p_a a\theta_s^a + p_b b\theta_s^b + (\lambda_s - \lambda_i)\alpha i \quad ,$$

$$(11d) \qquad \dot{\lambda}_N = \lambda_N(r + \delta i - \epsilon) - [V_0 - (p_I i + p_a a\theta^a + p_b b\theta^b)] \quad ,$$

and

$$(11e) \qquad \dot{\lambda}_i = [r + \gamma + \delta(1-i) + \beta - \alpha s]\lambda_i + p_I + p_a a\theta_i^a + p_b b\theta_i^b - \lambda_s[(\delta - \alpha)s + \beta] + \lambda_N \delta \quad .$$

Inspection of equations (6), (7) and (11a-e) shows that the dynamic equations for s, i, and the λ_j, j = s, i, N do not involve N so that the dynamic system is independent of N although not of ϵ or λ_N.

To develop the analysis further we turn to some special cases.

1.3 The Special Case of Susceptible-Infected-Susceptible (SIS):

The model of a disease in which people recover only to become susceptible rather than immune or die is the simplest case of the preceding model of any relevance. Many of the classic sexually transmitted diseases fall in this category. In this case, $\gamma = \delta = U = u = 0$; the controls remain the variables a and b while the states are s and i. With the substitution of i = (1-s) in the current-value Hamiltonian, equation (10), λ_i can be dropped as can be λ_N because the model is expressible without reference to the total size of the population, N, once no one dies. The birth rate of the population is still a parameter because it determines part of the growth of the susceptibles relative to other categories of the population. The model therefore only has one state variable, s, with dynamic equation:

$$\dot{s} = -\alpha s(1-s)+(\beta+\epsilon)(1-s) \quad . \tag{12}$$

Equations (11a-b) simplify in this case to:

$$p_a\theta^a = -\lambda_s\alpha' s(1-s) \tag{13a}$$

and

$$p_b\theta^b = \lambda_s\beta'(1-s) \quad . \tag{13b}$$

Equation (13a) equates the marginal cost of an increase in the preventive intervention as determined by the product of its price and targeting function to the marginal benefit of the increase in the proportion of the population that is uninfected achieved by the increase in prevention. Equation (13b) similarly equates the marginal cost of an increase in therapeutic intervention as determined by the product of its price and targeting function to the marginal benefit of the increase in the proportion of the population that is uninfected achieved by the

increase in therapeutic effort. All marginal costs and benefits are expressed in terms of the welfare of the average member of the economy measured in dollars.

The variable λ_s equals the shadow benefit in dollars to the average member of the economy of an increase in the proportion of the population that is uninfected. Under the assumptions on the θ^j and on α' and β', the λ_s must be positive if the first-order conditions are to hold. In addition, the dynamic equation for the multiplier implies:

$$(13c) \qquad \dot{\lambda}_s = r\lambda_s - [p_I - p_a a\theta_s^a - p_b b\theta_s^b] + [\alpha(1-2s)+\beta]\lambda_s \quad .$$

As before, the θ^j, j = a, b in equation (12)-(13c) are targeting functions that specify the proportion of the population affected by an intervention, and depend only on s because u = 0 and therefore i = (1-s). For example, if the disease is sexually transmitted and the preventive policy is condom distribution, then θ^a could plausibly take values of 1, s, 1-s, and s(1-s). In the first case, condoms are made available to everyone, in the second only to the uninfected, in the third only to the infected and in the fourth only to matchings involving an uninfected and an infected person. For therapeutic interventions, the simplest case is targeting exclusively at the sick so that $\theta^b = (1-s)$, but other situations are possible. For instance, without any ability to diagnose the disease, it may be that $\theta^b = 1$, while if it is difficult to distinguish the disease under consideration from another, a value of θ^b between 1-s and 1 is possible as determined by the prevalence of the other disease.

Equations (13a) and (13b) in combination imply that:

$$(14) \qquad \Phi \equiv -\frac{\beta'/p_b}{\alpha'/p_a} = \frac{\theta^b s}{\theta^a} \quad .$$

Equation (14), in turn, reveals a very rich set of possibilities for the pattern of relative dependence on preventive and therapeutic interventions over the course of an epidemic. The left-hand side of this equation gives the absolute value of the ratio of the marginal products of the two interventions relative to their prices (which are fixed exogenously), a measure of the relative dependence of the policy package on prevention, a, relative to therapy, b. The right-hand side depends only on s, the state of the epidemic; in an SIS disease there is an explicit solution for the dependence of Φ on s.

Although the relative emphasis on the two types of policies in terms of their marginal products depends only on the state of the epidemic, it is extremely sensitive to the form of the targeting functions. Table 1 records the eight possible relationships between Φ and s for the combinations of the four formulations of the targeting function for α and the two for β. In case IIA, Φ is relatively high when susceptibles and infecteds are nearly equal and is low for s near either of its extremes, 0 or 1. In this case, prevention efforts are most important when transmission rates are at their maximum because the total (and marginal) costs of prevention do not rise with the level of susceptibles, but the marginal effect of prevention is highest when transmission is fastest. In case IIB, Φ falls monotonically as s rises because costs of prevention rise and those of therapy fall with an increase in susceptibles. In case IIC, Φ rises monotonically as s rises and in case IID, Φ is constant regardless of the value of s, two cases that are less easy to interpret intuitively. The ratios of the marginal products do not translate directly into the ratios of the inputs themselves, perhaps the most direct measure of the relative size of the two efforts. But, there are special cases when the two ratios do move together (see Appendix A for an example) and so Φ helps to understand the possibilities for the co-movements of the physical inputs during the evolution of an epidemic.

To simplify what follows, we assume that $\theta^b = 1\text{-}s$, so that targeting of therapeutic interventions is restricted to the infected and therefore $\theta^b_s = -1$ (cases IIA-D of Table 1). Total differentiation of equations (13a) and (13b), the first-order conditions, implies:

$$(15a) \qquad a_\lambda \equiv \frac{\partial a}{\partial \lambda_s} = -\frac{\alpha'}{\lambda_s \alpha''} > 0 \ ,$$

$$(15b) \qquad a_s \equiv \frac{\partial a}{\partial s} = \frac{\alpha'}{\theta^a}\left[\frac{\theta^a_s s(1-s) - \theta^a(1-2s)}{\alpha'' s(1-s)}\right] \lessgtr 0 \ ,$$

$$(15c) \qquad a_p \equiv \frac{\partial a}{\partial p_a} = \frac{-\theta^a}{\lambda_s \alpha'' s(1-s)} < 0 \ ,$$

$$(15d) \qquad b_\lambda \equiv \frac{\partial b}{\partial \lambda_s} = -\frac{\beta'}{\lambda_s \beta''} > 0 \ ,$$

$$(15e) \qquad b_s \equiv \frac{\partial b}{\partial s} = 0 \ ,$$

and

$$(15f) \qquad b_p \equiv \frac{\partial b}{\partial p_b} = \frac{1}{\lambda_s \beta''} < 0 \ .$$

These expressions simplify the following discussion. They all have straightforward interpretations, except perhaps the indeterminate sign of the expression in (15b) for the partial effect of s on a. This ambiguity arises because s influences both the marginal cost of an increase in a, via its role in the targeting function, and the marginal benefit of an increase in a, via the

effect of s on the dynamics of the epidemic. The variable a_s: (1) has the same sign as (1-2s) if $\theta^a = 1$; (2) is negative if $\theta^a = s$; (3) is positive if $\theta^a = (1-s)$; and (4) is zero if $\theta^a = s(1-s)$.

So far we have proceeded on the presumption that the first-order conditions determine a maximum. In fact, the most generally-used sufficiency conditions for a maximum do not obtain in this model, but we believe that the way we have characterized the maximization is, in fact, correct for a very large class of these models . While the problem is concave in the controls because they are subject to diminishing marginal returns, it is not concave in the state because the dynamic equation exhibits increasing marginal returns in the state, a fundamental property of contagion as posited by epidemiologists. Correspondingly, the failure of the sufficiency conditions is not a reflection of the linearity of the instantaneous utility function [the integrand of equation (9)]. Appendix A discusses these issues in more detail.

Setting equations (12) and (13c) to zero produces the phase diagram in s-λ_s space. The slope of the locus from setting equation (12) to zero is:

$$(16a) \qquad \left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{s}=0} = \frac{\alpha + \alpha' s a_s}{\beta' b_\lambda - \alpha' s a_\lambda} = \frac{?}{+} ,$$

and the slope of the locus from setting equation (13c) to zero is:

$$(16b) \qquad \left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{\lambda}_s=0} = \frac{-[p_a a_s \theta_s^a + p_a a \theta_{ss}^a - 2\alpha\lambda_s + (1-2s)\alpha' a_s \lambda_s]}{[r - \alpha s + \beta + (1-s)(\alpha + \alpha' s a_s)]} = \frac{+}{?} .$$

The signs of both slopes are ambiguous, partially for the same reason that the sign of a_s is ambiguous. Further progress requires the separate consideration of Cases IIA-D.

In two cases, IIB with $\theta^a = s$ and IID with $\theta^a = s(1-s)$, both slopes are positive when the

equations of motion are linearized about the steady-state so long as a variant of the conventional condition that the interest rate at least equals the population growth rate (in the absence of disease) holds, that is $r \geq \epsilon$.[4] If the slope of the $\dot{\lambda}_s = 0$ locus is flatter than that of the $\dot{s} = 0$ locus in s-λ_s space, there is a unique stable path to the steady state (Fig. 1a) because the characteristic equation of the linearized dynamic system has one positive and one negative real root. The variables s and λ_s move together toward the steady state, and b and β move with them so that therapeutic effort increases with more susceptibles. Preventive effort, a, and α move as determined by the relation between a and b as given by Φ, decreasing with the number of susceptibles in IIB and varying with it in IID. If the slope of the $\dot{\lambda}_s = 0$ locus is steeper than that of the $\dot{s} = 0$ locus in s-λ_s space, however, there is no stable path to the steady state (Fig. 1b). In these cases, the model could evolve either to the eradication of the disease, with s = 1, or to the equilibrium without intervention, a = b = 0, or to an intermediate stable equilibrium of the type illustrated in Fig. 1a if such exists.[5] That such unstable cases exist is consistent with the structure of the model, which should allow the possibility of optimally eradicating the disease or optimally doing nothing in the steady state for some configurations of the phase diagram. We do not pursue these divergent cases here; instead, we restrict the discussion to diseases that are (optimally) neither eradicated nor ignored and in which policy takes the system to an optimal steady state.

The two remaining cases, IIA with $\theta^a = 1$ and IIC with $\theta^a = 1\text{-}s$, are more complex and

---

[4] Note that in cases IIB and D the term $(\alpha+\alpha' s a_s)$ is always positive.

[5] Goldman and Lightwood (2002) demonstrate exactly this possibility of an unstable optimal steady state co-existing with a stable optimal steady state for their model which only allows for a therapeutic intervention. In this case, if the unstable steady state is perturbed, the system evolves either to the stable steady state of endemic infection or to the eradication of the disease.

there are several sub-cases.[6] When the model is linearized about the steady state both slopes may be positive, as in cases IIB and D, and the foregoing analysis obtains. The slope of the $\dot{\lambda}_s = 0$ locus may be positive while that of the $\dot{s} = 0$ locus is negative; this case is unstable and we do not discuss it further. Both slopes may be negative; if the $\dot{\lambda}_s = 0$ locus is less negatively sloped there is a stable saddlepoint (Fig. 1c) while the reverse situation is unstable.[7]

The parameters of the model are the four prices: p_I, p_a, p_b and r. The parameters p_I and r enter equation (13c) for $\dot{\lambda}_s$ but not equation (12) for $\dot{s}$, nor do they enter the first-order conditions for a and b. They therefore shift the $\dot{\lambda}_s = 0$ locus but not the $\dot{s} = 0$ locus. Consider the effect of an increase in p_I on s^*, the steady-state number of susceptibles in the two saddlepoint cases. In Fig. 1a the $\dot{\lambda}_s$ locus shifts up (from $\lambda_s\lambda_s$ to $\lambda_s'\lambda_s'$) and s^* rises; in Fig. 1c the $\dot{\lambda}_s$ locus shifts down and s^* also rises. Thus in all cases an increase in the direct cost of being infected in terms of pain and suffering increases the steady-state proportion of the population that is uninfected. The effect on s^* of an increase in r is opposite to that of p_I. The costs of prevention or therapy are borne immediately while their benefits are received over time. Because an increase in r leads to a diminished weight of the future in decisions, an increase in r leads to an increase in the optimal steady-state proportion of the population that is infected.

The effects of the other two parameters are more complicated, however, because both loci shift. The impact effect (s and λ_s fixed) of an increase in the price of either preventive or

---

[6] In these cases the term $(\alpha+\alpha' s a_s)$ may be positive or negative and therefore the numerator of equation (16a) and the denominator of equation (16b) may be positive or negative. Note, however, that if the denominator of (16b) is negative, so must be the numerator of equation (16a); it is not possible to have a positively sloped $\dot{s} = 0$ locus and a negatively sloped $\dot{\lambda}_s = 0$ locus.

[7] The results in this sentence follow from the application of such standard references as Kamien and Schwartz (1981, Appendix B). The saddlepoint case is their Case IC.

therapeutic interventions is to decrease the amount used via equations (15c) and (15f) and therefore either a and α or b and β are affected in both equations.

In the case of an increase in p_b, the $\dot{s} = 0$ locus always shifts up regardless of the sign of its slope. The $\dot{\lambda}_s = 0$ locus shifts up if its slope is positive and down if its slope is negative. In the case of an equilibrium of the type illustrated in Fig. 1a, therefore, the shift in the $\dot{s} = 0$ locus tends to lower s^* while the shift in the $\dot{\lambda}_s = 0$ locus tends to raise s^* and the net outcome is ambiguous even when the algebraic magnitudes of these shifts are taken into account. The rationale for this ambiguity is as follows: The price of a therapeutic intervention, p_b, enters the dynamic equation for the co-state variable in the same way as the cost of being infected, p_I. One of the effects of an increase in p_b is therefore to raise s, just as an increase in p_I does; in effect an increase in the cost of being cured is like an increase in the cost of being infected because every infection induces expenditures on therapeutic inputs. But there is also the fact that it is more expensive to be cured so that it may be desirable to spend less on b and be cured less quickly. That the first effect can dominate is easily seen from the special case when b is fixed at some positive value (perhaps for technological reasons) so that therapeutic effort is not adjusted in response to its price increase. The preventive intervention can still respond, however, as it would to a change in p_I and the steady state proportion of the uninfected, s^*, is thereby increased. Recall that cases IIB and D of Table 1 must conform to this latter pattern of an ambiguous impact of p_b. In contrast, starting from an equilibrium of the type illustrated in Fig. 1c, the upward shift of the $\dot{s} = 0$ locus and the downward shift of the $\dot{\lambda}_s = 0$ locus work to raise s^*.

In the case of an increase in p_a, the $\dot{s} = 0$ locus also always shifts up regardless of the sign of its slope. When the system is linearized about the steady state, the $\dot{\lambda}_s = 0$ locus shifts according to the sign of

$$\frac{-a\theta_s^a \ -\theta^a a_s}{r-\epsilon+(1-s)(\alpha+\alpha' s a_s)},$$

rising with an increase in p_a if this expression is positive and falling if it is negative. The denominator is unambiguously positive in cases IIB and D and ambiguous in the other two cases of Table 1. The sign of the numerator is ambiguous in all four cases; in cases II A and D, this numerator has the sign of (2s-1). Once again, these ambiguities stem from the role of s in affecting both the costs and benefits of an increase in a (see the discussion of a_s). Consequently, little can be said about the effect of p_a on s^*.

1.4 The Special Case of Susceptible-Infected-Dead (SID):

If all people who become infected die, the general model of equation (10) can be specialized to: U = 0, $\beta = \gamma = 0$, $\lambda_i = 0$. The states are s and i = (1-s) and the controls are a and b. On the further assumption that being sick is not per se costly, $p_I = 0$. The value of therapeutic measures is to reduce the death rate and thereby gain utility from prolonging a life; formally, the negative valuation of death is embodied in the fact that the dead are not part of N and do not get the utility associated with being alive, V_0, net of the expenditures on a and b. To save space, we do not repeat the versions of equations (11a-d) specialized for this case, but merely comment on the properties of this case that follow from these equations. Furthermore, we discuss only the results for the case in which $\theta^b = (1-s)$.

The specialized versions of equations (6) and (11a-d) are independent of N but not of λ_N and therefore so are the solutions for the optimal interventions, a and b, and for the state of the epidemic, s. The first result implies that the interventions are independent of the scale of the

economy. The second result implies that the dimension of the system is three (s, λ_s and λ_N) rather than the two of the SIS model. This important difference between the two models arises because the infected die in the SID model which is valued by λ_N, rather than returning to the susceptible state which is valued at λ_s, as in the SIS model. Consequently, the SID model is significantly less tractable than the SIS model. The solutions for the SID case are not independent of ϵ which appears in the specialized versions of equations (6) and (11c and d). The multipliers, λ_s and λ_N, must be positive under the assumptions about θ^a, α' and δ'.

The SID first-order conditions imply that:

$$\frac{\delta'/p_b}{\alpha'/p_a} = \frac{(1-s)[p_b+\lambda_N\delta']}{p_b\theta^a} .$$

In contrast to the SIS model, there is therefore no closed-form solution corresponding to the relationship between s and Φ in the SIS case, so there is the potential for very much more complicated relationships between the price adjusted marginal products of the interventions than those reported in Table 1.

1.5 The Special Case of Susceptible-Infected-Uninfectible (SIU):

This case corresponds to $\beta = \delta = 0$; individuals who are susceptible become infected and then immune. The states are s, i and u = (1- s - i) and the controls are a and b. The only substantive simplification of the social planner's problem that we have been able to identify in this case is that the five-dimensional system of section 1.2 can be reduced to four because λ_N does not appear in the dynamic equations for the other variables because people do not die of the disease. We therefore do not present any results for this case except in section 2.4 on

decentralization.

2. Decentralization

2.1 The General Problem of Decentralization:

To this point we have discussed the problem of the social planner who directly controls the values of a and b in a model without people who make decisions that affect their own health. The next step is to consider private decisions and their implications for government policy. If people do not take into account the effect on the infection of the general population caused by their ability to infect others if they become infected, they generate a pure infection externality. In our formulation of diseases in which one person directly infects another, the preventive activity of one individual does not, however, affect the probability that other people become infected independently of whether the first person becomes infected, so there is no preventive externality. After identifying the externality, we examine how government interventions with subsidies or taxes can decentralize the social planner's first-best solution.

In our abstract formulation, governments can subsidize preventive and therapeutic activities, the privately chosen values of a and b. In reality, for some diseases, there will be some inputs that are marketed and some inputs that involve individuals' non-marketed and unobservable actions such as avoiding crowded places to varying degrees in the case of tuberculosis prevention or adhering meticulously to drug regimens. When a and b involve such non-marketed and unobservable actions the subsidy/tax interventions we propose may be infeasible or may have to be targeted only on the marketed components of preventive and therapeutic activities with second-best implications.

The simplest way to illustrate the pure infection externality and its implications for policy is to assume that private decisions are made by a group of people that we call a household, a construct that we use as the representative decision-making agent. This construct provides a logically consistent and analytically tractable model to contrast with the model of the social planner: First, the household's objective function is fully congruent with the social planner's. Furthermore, the household understands and anticipates how the epidemic will evolve and is fully forward-looking with regard to its possible future statuses as well as its present situation. Unlike Kremer's (1996) modelling of individuals' behaviour which is only oriented to the conditions in the current period, our household takes account in its current decisions of the evolution of the epidemic, its implications for the future risk of infection, and its implications for all the household's descendants. For instance, if the future probability of infection is high it affects the current incentive of the household to make therapeutic expenditures. It is therefore the case that our rationale for government interventions does not depend either on myopia or on a discrepancy between the social planner's and the representative agents' valuation of outcomes over the path of the epidemic. Instead, our assumptions isolate the pure externality motivation for government intervention. To the extent that there are deviations from the preceding assumptions on the behaviour of the household, there may be other important reasons for government interventions but they are not the subject of this paper.

As is conventional in the public-economics treatment of externalities, the only distinction between the social planner and the representative agent is that the household is assumed to be small relative to the population as a whole, in this case so that the proportion of the household in any disease status does not affect the proportion of the population as a whole that is in that status. In particular, this household takes as given the proportion of the population that is infected,

which equals the probability, π, that any random contact is with an infected person. Second, the household is assumed to be sufficiently large that it can fulfill the role of a representative agent and therefore that the proportion of the household in each disease status is identical to the corresponding population proportion. Finally, it is this household that takes decisions about the interventions, a and b. Because the instantaneous utility function is linear, there is no sense in which the household is performing any implicit insurance function for its members. A perhaps more realistic but only perhaps (because people do indeed live in households) and less tractable approach would build the private economy up from representative individuals each of whom is in one or another diseases statuses at any one time and taking decisions about either prevention or therapy (or neither if already immune), with regard to their possible future statuses as well as their present situation. For diseases in which people transit from susceptible to infected to dead, the individual's problem seems tractable, because death is obviously an absorbing state. But for individuals who cycle between susceptible and infected, we have not been able to manage a formulation that does not adversely affect the tractability of the model and we leave this task for the future, but see no reason why such a formulation should fundamentally alter the nature of the externalities that we identify.

The dynamic equations of this version of the model are the same as in section 1.1, except that in equations (3) and (4) the term αSi is replaced by $\alpha S\pi$ to denote the exogeneity from the household's viewpoint of the proportion, π, of the population (in contrast to the proportion, i, of the household) that is infected. There is a consequent change in equations (6) and (7).

A further change has to be made to the objective function to reflect the possibility of government interventions. If there is an externality, the government may find it optimal to subsidize or tax preventive and/or therapeutic inputs. To allow for these possibilities, we assume

that the representative household faces prices of $q_j = (1+t_j)p_j$, j = a, b. As is standard in public economics, so that any interventions are revenue neutral in a way that does not have any incentive effects beyond the t_j, we assume that the household receives a lump sum payment (possibly negative) of T that it takes as exogenous to its own actions but that in fact equals $t_a p_a a^h \theta^a + t_b p_b b^h \theta^b$; a superscript "h" indicates that the variables are evaluated at the household's values rather than the social planner's. If this lump sum offset were not part of the package, the household's welfare would be affected by its experiencing a net loss or gain of income as the government intervenes with taxes or subsidies to offset the externality. The decentralization results that follow would not obtain as can be seen by following the steps of the proofs without the assumption of revenue neutrality.

With these modifications, the household's current-value Hamiltonian is:

$$
\begin{aligned}
(17) \qquad H^h = N^h \{ V_0^h - [p_I i^h + q_a a^h \theta^a + q_b b^h \theta^b] + T \} \\
+ (\lambda_s^h N^h)[(1 - s^h)\epsilon + \delta s^h i^h - \alpha s^h \pi + \beta i^h] \\
+ \lambda_N^h [N^h(\epsilon - \delta i^h)] \\
+ (\lambda_i^h N^h)[\alpha s^h \pi + \delta i^{h\,2} - (\epsilon + \gamma + \delta + \beta) i^h] \quad .
\end{aligned}
$$

All functions of variables (θ, α, β, δ and γ) are evaluated at the household values of their arguments while ϵ is a constant common to both the social planner's and the household's models. We now proceed to the special cases.

2.2 Decentralization in the SIS Model:

The household uses the version of equation (17) specialized to this case. We also assume that only the infected are targeted by therapies, so that $\theta^b = (1\text{-}s^h)$. The first-order conditions imply:

$$(18a) \qquad q_a\theta^a = -\lambda_s^h \alpha' s^h (1-s)$$

and

$$(18b) \qquad q_b = \lambda_s^h \beta' \quad ,$$

and the co-state equation is:

$$(18c) \qquad \dot{\lambda}_s^h = r\lambda_s^h - [p_I - q_a a^h \theta_s^a + q_b b^h] + [\alpha(1-s)+\beta]\lambda_s^h \quad .$$

Because the group is representative of society, s must equal s^h. Once this substitution is made, the only differences between equations (13a-c), the planner's problem, and equations (18a-c), the private problem, are the q_j and the (1-s) term at the end of equation (18c) rather than the (1-2s) term at the end of equation (13c). This latter difference reflects precisely the fact that the household takes the general rate of infection as exogenous in making its decisions and this difference determines whether the government's optimal intervention is a tax or a subsidy as is shown below.

The government can induce private decision makers to make decisions that coincide with the planner's problem by instituting equiproportionate changes in p_a and p_b; comparison of the two sets of first-order conditions for a and b shows that $t_a = t_b = t$. In other words, the government compensates for any differences between λ_s and λ_s^h in equations (13a-b) and (18a-b). It does so with a lump-sum offset, T, so that any revenues or expenditures from the price interventions also appear in the household's budget. Because the intervention is only to λ_s^h and because of the way λ_s^h enters equations (18a-b), a and b activities are affected to the same degree and Φ is unaffected. At the steady state, the intervention is a subsidy (negative tax) at rate t^*:

$$(19) \qquad t^* = -\frac{\lambda_s^* \alpha s^*}{p_I + \lambda_s^* \alpha s^*} < 0 \quad ,$$

in which λ_s^* and s^* are the values from the planner's steady state.[8] Furthermore, for any non-steady-state s, the government must intervene with a subsidy.[9] This finding that the intervention is a subsidy coincides with the intuition that private decisions ignore the benefits to society as a whole from taking preventive and therapeutic measures. Subsidization is at equal rates because it is equally beneficial in preventing further infection to get a person out of the infected pool as to have prevented the person from getting into it in the first place. These benefits are equally overlooked by the private decision makers. This result contradicts what may be an often-held presumption that preventive rather than therapeutic efforts are associated with externalities and are the proper domain of public health. In this model in contrast to the model of vector-borne diseases in a subsequent section of the paper, preventive activities are pure private goods in that one person's preventive effort does not affect another person's risk of infection if the first person does not become infected. Furthermore, the government does not have at its disposal any technology of intervention that private people do not have at theirs. The externality therefore arises only through an individual's being in the infected pool, either through getting into it

[8]This result follows from multiplying equation (13c) by (1+t) and setting it equal to equation (18c) because both equations equal zero at the steady state.

[9]At s^* there is a subsidy, so that $\lambda_s^h < \lambda_s$. Now if there were ever a tax corresponding to some lower level of s, $\lambda_s^h > \lambda_s$. Between this point and the steady state, there would therefore have to be some intermediate value of $s < s^*$ at which $\lambda_s^h = \lambda_s$. But at such a point, λ_s^h is increasing faster than λ_s, compare equations (18c) with (13c) evaluated at the common values of all variables because the state and costate are equal for the private and public problems at this point where the tax/subsidy equals zero. Consequently, λ_s^h could never fall relative to λ_s which would contradict the existence of a steady-state subsidy. A similar argument holds for $s > s^*$ once it is noted that in these situations s is falling (rather than rising) toward the steady state.

without consideration of the risks posed to others or through not getting out of it fast enough for the same reason.

2.3 Decentralization in the SID Model:

In this case, private decisions are made by a household as given by the Hamiltonian of equation (17) specialized to the SID model as defined in section 1.4. As in the SIS model, if the household is to be representative of society as a whole, s must equal s^h. Once this substitution is made, the only differences between the equations of the planner's problem and of the private problem are the q_j and an additional term of $-\alpha s\lambda_s$ in the equation for $\dot{\lambda}_s$ as opposed to the equation for $\dot{\lambda}_s^h$.

In contrast to an SIS disease, however, equiproportionate interventions directed at p_a and p_b do not induce private decision makers to make the decisions that coincide with the social planner's. Instead, the planner's path for s, a and b can be achieved by price interventions t_a and t_b such that:

$$(20a) \qquad 1+t_a = \frac{\lambda_s^h}{\lambda_s}$$

and

$$(20b) \qquad t_b = t_a \frac{\delta' s\lambda_s}{p_b} \quad .$$

This result follows from the equation of the α' and δ' prevailing under the private economy to those under the social plan as given by the two sets of first-order conditions and because the values of the multipliers on N are the same in the private economy and the social plan.[10] Because $\delta' < 0$ and the multiplier is positive (section 1.4) the right-hand expression in equation (20b) is negative and implies that the interventions are of opposite sign; if a is subsidized ($t_a < 0$), then b is taxed.

In fact, at the steady state, a is subsidized and b is taxed. The steady-state subsidy is:[11]

$$(21) \qquad t_a^* = -\left[\frac{\lambda_s^* \alpha s^*}{\lambda_s^* \alpha s^* + \lambda_N^* (\delta - b\delta')}\right] ,$$

in which the (positive) multipliers are evaluated at the social planner's values. The reason that therapeutic expenditures are taxed here but subsidized in the SIS model is that here therapeutic expenditures keep people in the pool of infectious people whereas in the SIS case therapeutic expenditures moved them out of the pool. In this model, interventions are only designed to offset the pure infection externality; there is neither a concern with insurance nor with therapies as merit goods, either of which may change this result. Away from the steady state, the preventive intervention is also a subsidy (and correspondingly, the therapeutic intervention is a tax). The

[10]The multipliers for the state N are equal because when there is revenue neutrality, the values of these multipliers are the same in the steady state for the private and social optimizations, as can be seen by setting the equations for the changes in these multipliers (not shown) to zero. Furthermore, the form of these equations for the changes in these multipliers is the same and therefore the entire path of these two multipliers must be the same.

[11]The result follows from multiplying the equation for the change in λ_s by $(1+t_a)$ and equating it to the equation for the change in λ_s^h because both changes are zero in the steady state.

reason is the same as for the SIS model: at a point where $\lambda_s = \lambda_s^h$, comparison of the expressions for the change in these two multipliers shows that the change in the former is smaller than the change in the latter.

2.4 Decentralization in the SIU Model:

Decentralization is harder to analyse in the SIU model than in either the SIS or the SID models. In the SIS and SID models there is only one multiplier equation that differs between the social planner's and the household's problems, that associated with the change in s, whereas in the SIU model there are two, associated with the changes in s and in i. We have therefore only been able to characterize the government's optimal interventions at the steady state.

Comparison of the results from the first-order conditions of the two problems shows that:

$$(22a) \qquad (1+t_a) = \frac{\lambda_s^h - \lambda_i^h}{\lambda_s - \lambda_i}$$

and

$$(22b) \qquad (1+t_b) = \frac{\lambda_i^h}{\lambda_i} \quad .$$

These two equations plus the property that in the steady state $(1+t_a)\,\dot{\lambda}_i^h = \dot{\lambda}_i = 0$ imply that the steady-state values of the two interventions are equal. Thus, in the steady state, there is an echo of the SIS result that the government's optimal interventions are at the same rate. And a similar

intuition applies: At the steady state, either going into the infected pool or not getting out of it imposes a similar externality because either is a permanent (i.e. steady-state) increase in the size of the infected pool. Furthermore, the fact that $(1+t_b)\,\dot{\lambda}_s^h = \dot{\lambda}_s = 0$ and $(1+t_a)\,\dot{\lambda}_i^h = \dot{\lambda}_i = 0$ in steady state implies that:

$$(23) \qquad t_a^* = t_b^* = \frac{-(\lambda_s^* - \lambda_i^*)\alpha s^*}{p_I + (\lambda_s^* - \lambda_i^*)\alpha s^*} < 0 \ ,$$

with an inequality because the first-order condition with respect to a implies $\lambda_s - \lambda_i > 0$. At the steady state the intervention is therefore a subsidy. We have not been able to derive any results outside the steady state that parallel those for the SIS and SID cases. There seems to us to be no reason to expect the SIS result on equal subsidies to carry over to situations away from the steady state where converting someone from an infected to an uninfectible is a permanent increase in people outside the infected pool whereas keeping someone among the susceptibles may only be a temporary decrease in the infected pool.

3. Control of a Vector-Borne SIS Disease

Many infectious diseases are transmitted through intermediate hosts, such as mosquitos, flies, ticks and snails. These hosts must be infected to play their part in the cycle of infection. For instance, in the case of malaria, infected mosquitos inject people with one stage of the parasite thereby infecting them. Uninfected mosquitos that bite infected people are in turn infected by a later developmental stage of the parasite in an infected person's blood, thereby continuing the cycle. Interventions to affect the prevalence of these diseases can take many

forms and the package is more complicated than the distinction between prevention and therapy in sections 1 and 2.

A model of malaria transmission must, therefore, comprise equations for the number of people who are infected (I) and the number of mosquitos that are infected (Y). For simplicity, we will assume that the total population (of people) is fixed at N, so that there is neither natural growth in the population nor are there deaths associated with malaria. The sum of the number of susceptibles (S) and infecteds equals the whole population, S+I = N, and s, as before, is the proportion of the population that is susceptible. Because we are interested in interventions that affect the total number of mosquitos (M), we also specify the dynamics of the total mosquito population.[12]

The number of infected people evolves according to:

$$\dot{I} = \alpha_1\alpha_2 my(N-I) - \beta I \quad . \tag{24}$$

The first part of this equation is the product of the number of people who are susceptible (N -I), the proportion of mosquitos that are infected (y = Y/M), the ratio of mosquitos to people (m = M/N) and two parameters α_1, the number of bites that the average mosquito manages per unit time, and α_2, the proportion of bites by infected mosquitos that lead to a human infection. The second part of the equation is the rate of recovery of infecteds.

The number of infected mosquitos evolves according to:

[12]The dynamic structure of infection in our model, equations (24) and (25), is an adaptation of the discussion in Anderson and May (1991). As does their basic model, we ignore deaths from malaria and the existence of an animal reservoir of infection on which mosquitos feed in addition to the human population.

$$(25) \qquad \dot{Y} = \alpha_1\alpha_3(1-s)(M-Y) - \delta_M Y \ .$$

The first part of this equation is the product of the number of mosquitos (M -Y) that are susceptible to infection by the chance (1-s) that a person whom they bite is infected adjusted by the rate of biting (α_1) and the chance that such a bite leads to infection of the mosquito (α_3). The second part of the equation subtracts the infected mosquitos that die; δ_M is the death rate of mosquitos regardless of whether they are infected or not. Finally, the change in the total number of mosquitos is:

$$(26) \qquad \dot{M} = F(M) - \delta_M M \ ,$$

in which F'>0 and F''<0 so that there is a steady state population of mosquitos for a given value of δ_M.

The objective of government policy is to maximize the present discounted value of social welfare:

$$(27) \qquad W = \int_0^\infty N\{V_0 - [p_I(1-s) + p_{a_1}a_1 + p_{a_2}a_2 s + p_d d + p_b b(1-s)]\}\ e^{-rt}dt \ .$$

Equation (27) incorporates a plausible targeting scenario for each of four interventions: (1) some interventions affect both the probability that an infected person infects an uninfected vector as well as the probability that an infected vector infects an uninfected person, for example, promotion of the wearing of clothes and use of bed nets that lower the chance of bites, summarized by a_1, so that $\alpha_1(a_1)$ with $\alpha_1' < 0$; (2) other interventions affect the probability that an uninfected person is infected by an infectious vector but do not prevent uninfected vectors from

becoming infected by an infectious person, for example, provision of prophylactic drugs, a_2, so that $\alpha_2(a_2)$ with $\alpha_2' < 0$; (3) spraying mosquitos with insecticides with intensity d, so that $\delta(d)$ with $\delta_M' > 0$; (4) provision of drugs that promote recovery, of amount b, so that $\beta(b)$ with $\beta' > 0$. Interventions to change the value of α_3 affect the probability that an infectious person infects an uninfected vector but do not affect the probability that an infectious vector infects an uninfected person. These interventions are perhaps the least implementable. They might include deterring people from voiding parasites in the case of schistosomiasis, but this type of action has little direct benefit to the infected person and would therefore require monitoring by the government and could not easily be decentralized through taxes and subsidies. Scientific advances provide genuine prospects for genetic engineering of the vector to resist infection, but we leave this type of intervention aside. In the subsequent discussion, α_3 is therefore treated as a fixed parameter. We also ignore possibilities for changing the form of the relation F(M), for instance through the draining of swamps.

The current value Hamiltonian, H, is

$$(28) \quad \begin{aligned} H \;=\; & N\{\, V_0 - [p_I(1-s) + p_{a_1}a_1 + p_{a_2}a_2 s + p_d d + p_b b(1-s)] \,\} \\ & + \lambda_s[-\alpha_1\alpha_2 mys + \beta(1-s)] \\ & + \lambda_y[\alpha_1\alpha_3(1-s)(1-y) \;-yF(M)M^{-1}] \\ & + \lambda_M[F(M) \;-\delta_M M] \;, \end{aligned}$$

in which the λ_j are the current value multipliers associated with the states j = s, y and M. The first-order conditions for this problem are:

$$(29a) \quad \frac{\partial H}{\partial a_1} \;=\; -Np_{a_1} \;-\lambda_s\alpha_1'\alpha_2 mys \;+\lambda_y\alpha_1'\alpha_3(1-s)(1-y) \;=\; 0 \;,$$

$$(29b) \qquad \frac{\partial H}{\partial a_2} = -Np_{a_2}s - \lambda_s\alpha_1\alpha_2' mys = 0 \ ,$$

$$(29c) \qquad \frac{\partial H}{\partial b} = -Np_b(1-s) + \lambda_s\beta'(1-s) = 0 \ ,$$

and

$$(29d) \qquad \frac{\partial H}{\partial d} = -Np_d - \lambda_M M\delta_M' = 0 \ .$$

The associated equations for the multipliers are:

$$(29e) \qquad \dot{\lambda}_s = r\lambda_s - [p_I - p_{a_2}a_2 + p_b b] + \lambda_s(\alpha_1\alpha_2 my + \beta) + \lambda_y\alpha_1\alpha_3(1-y) \ ,$$

$$(29f) \qquad \dot{\lambda}_y = r\lambda_y + \lambda_s\alpha_1\alpha_2 ms + \lambda_y[\alpha_1\alpha_3(1-s) + F(M)M^{-1}] \ ,$$

and

$$(29g) \qquad \dot{\lambda}_M = r\lambda_M + \lambda_s\alpha_1\alpha_2 syN^{-1} + \lambda_y yM^{-1}(F' - FM^{-1}) + \lambda_M(F' - \delta_M) \ .$$

Equations (29b) or (29c) imply that the multiplier on the number of susceptibles, λ_s is positive. The right-hand side of equation (29f) set to zero consequently implies that the steady-state value of the multiplier on the proportion of vectors that are infected, λ_y, is negative. Furthermore, the value of this multiplier must always be negative. If it were not, there would be a contradiction because equation (29f) would then imply that the derivative of this multiplier would be positive if the multiplier itself is positive. Therefore this multiplier could never get to its negative steady-state value starting from a positive initial value.

Inspection of equations (29a-d) shows that there is, in general, no simple closed-form solution for the relationship between the price-adjusted marginal products of the different interventions. There is, however, one special case of interest. If people either always take the maximal precautions possible or do not take any at all, α_1 is not determined endogenously in the model. In this case, equation (29a) is no longer operative and is instead replaced by an equation that gives the exogenous value of α_1. Equations (29b-c) then show that the price-adjusted marginal products of the α_2 and b interventions depend only on the ratio of infected mosquitos to people, and not on the proportion of people who are infected:

$$(30) \qquad \frac{\beta'/p_b}{\alpha_2'/p_{a_2}} = \alpha_1 \frac{Y}{N} .$$

As in the other models, we introduce private decision making by assuming that certain variables are taken as given by a household that is susceptible in the proportion s^h. In this case, the total mosquito population (M) and the extent of its infection (y) are exogenous to private decision makers. These assumptions are extreme representations of any actual situation. We are treating insecticidal spraying as a pure public good, an extreme case of a pure prevention externality in which no individual perceives any personal benefit from spraying but society as a whole does. In reality, even at the household level, there is scope for diminishing the population of mosquitos through insecticidal spraying, but the infiltration of mosquitos from outside the household's area of control is much more rapid than if the government is doing co-ordinated spraying over a large area.

The current value Hamiltonian, H, is

$$(31) \qquad H = N\{V_0 - [p_I(1-s^h) + (1+t_{a_1})p_{a_1}a_1^h + (1+t_{a_2})p_{a_2}a_2^h s^h + (1+t_b)p_b b^h(1-s^h)] + T\} + \lambda_s^f[-\alpha_1\alpha_2 mys^h + \beta(1-s^h)]$$

in which λ_s^h is the only multiplier and it is associated with the state s^h; y and M are exogenous functions of time from the household's perspective and the government's expenditure on d is subsumed in T. The first order conditions for this problem are:

$$(32a) \qquad \frac{\partial H}{\partial a_1} = -N(1+t_{a_1})p_{a_1} - \lambda_s^h \alpha_1' \alpha_2 mys^h = 0 ,$$

$$(32b) \qquad \frac{\partial H}{\partial a_2} = -N(1+t_{a_2})p_{a_2}s^h - \lambda_s^h \alpha_1 \alpha_2' mys^h = 0 ,$$

and

$$(32c) \qquad \frac{\partial H}{\partial b} = -N(1+t_b)p_b(1-s^h) + \lambda_s^h \beta'(1-s^h) = 0 .$$

There is no equation for d because it is not a control, the assumption of a pure public good. The associated equation for the multiplier is:

$$(32d) \qquad \dot{\lambda}_s^h = r\lambda_s^h - [p_I - (1+t_{a_2})p_{a_2}a_2 + (1+t_b)p_b b] + \lambda_s^h(\alpha_1\alpha_2 my + \beta) .$$

The complete dynamics of the system with private decision making is given by the addition of $s^h = s$ and equations (25) and (26).

Government intervention to transform the decisions of the household into the dynamics that maximize social welfare requires two types of interventions: (1) a program of spraying (d)

paid for by the government consistent with equation (29d); and (2) a set of price interventions to transform equations (32a-c) into equations (29a-c). Once this complete package is implemented, it is clear from comparison of equations (29b-c) and (32b-c) that the effect of price interventions for a_2 and b is only through the transformation of λ_s^h into λ_s. In other words, whatever the price intervention it is applied at equal rates to the prices of these two instruments; this result parallels the equal-rate result of the SIS model and has the same intuition. Because $\lambda_y < 0$, comparison of equation (29a) with equation (32a) shows that a_1 is subsidized more (or taxed less) than a_2 and b.

5. Conclusions

Our starting point for this paper has been a commonsensical one: Mathematical epidemiology provides a parsimonious representation of how infectious diseases spread. Economics suggests an objective function to evaluate the costs of infection and its associated offsetting interventions, and especially a role for diminishing returns in the interventions that affect the evolution of the epidemic. All these elements taken together specify the objectives and constraints that determine optimal interventions for public health in a hypothetical socially planned economy. The discrepancy between the social planner's solution and decentralized decision making defines externalities in the economy, and the scope for subsidy/tax interventions by government to maximize private welfare in the absence of central control of all decisions.

Although we set out a general formulation of this problem for infections that proceed either to recovery and further susceptibility, immunity or death as well as a variant of an infection that depends on vectors, by far the most tractable case is the first, the SIS model. For this type of disease, we looked at the effects of different targeting schemes on the phasing of preventive

inputs relative to therapeutic ones, the response of the steady-state level of susceptibility to the different prices in the model, and the role of subsidies in decentralizing the social planner's problem. In this model, the optimal intervention is a subsidy at the same rate to both preventive and therapeutic activities, both in the steady state and during the approach to it, and the result does not depend on the form of targeting. By contrast, the phasing of the two interventions is qualitatively sensitive to the form of targeting. Even so, this model poses difficult technical problems, most especially as regards the sufficient conditions for a welfare-maximizing dynamic policy, something we largely addressed through numerical examples.

With regard to the other models, we provided some results, on the phasing of interventions and on the subsidies and taxes necessary to achieve decentralization. Targeting is again central to the phasing of interventions. What happens to the infected is critical to the qualitative properties of the tax/subsidy interventions. The SIS model provides an anchor for the discussion of these models, and some of the SIS results re-appear in one form or another underlining the usefulness of the SIS model as a starting point for the analysis. But these models are all inherently of higher dimension than the SIS model and their properties as a whole are generally more difficult to analyse. Undoubtedly, further progress will depend on more numerical work, but again the insights of the SIS model should provide guidance in interpretation; Gersovitz and Hammer (forthcoming b) examine numerically a model of a vector-borne infection.

Despite the variety of assumptions that we have incorporated into the analysis, there are many alternatives that we have not considered. The basic structure of the model provides scope for analysing many of these alternatives. The assumptions that the representative agent is fully forward looking and cares fully about present and future household members could be modified;

deviations from these assumptions would generate additional rationales for government interventions. So would the acquisition of drug resistance by the pathogens. Finally, there may be further economic implications of large-scale epidemics such as HIV that necessitate a more detailed modelling of the costs of infections as they work their way through the economy as a whole.

Appendix A

The Second-Order Conditions for the SIS Model

While the current-value Hamiltonian for the SIS model, H, is jointly concave in the controls and the multiplier λ_s is positive, the maximized Hamiltonian, H^*, is not concave in the state variable, s. Instead, the maximized Hamiltonian is unambiguously convex in s because

$$(A.1) \qquad \frac{\partial^2 H^*}{\partial s^2} = -N\{p_a(a_s\theta_s^a + a\theta_{ss}^a) + \lambda_s[\alpha'(1-2s)a_s - 2\alpha]\} > 0 \ ,$$

as can be shown using equations (13a) and (15b) and any of the four definitions of θ^a. The commonly-used Arrow sufficiency condition (e.g. Kamien and Schwartz, 1981) for a maximum is therefore not fulfilled.

Despite this failure of the commonly-used sufficiency condition, we believe that the problem of section 1.3 is well posed and its solution is characterized by the first-order conditions and the phase diagrams. There are three reasons for our belief.

First, at an intuitive level, the α and β functions may exhibit steeply diminishing returns. In this case, it is hard to see how the solution could be characterized by anything but a

conventional optimizing approach to a steady-state level of infection.

Second, consistent with this intuition, the Arrow conditions seem much stronger than are necessary. Zeiden (1984) presents weaker conditions but they are not easy to apply.

Third, and to us most convincing, we next present the results of a numerical implementation of the discrete version of the SIS model of section 1.3 using dynamic programming. This implementation seems well-behaved and consistent with the optimal-control treatment of the SIS model in section 1.3. The parameters we use are: $V_0 = 1600$; $p_I = 1600$; $p_a = 0.25$; $p_b = 2$; $\epsilon = 0$; $1+ r = 1/.9$; $\theta^a = s$; $\theta^b = (1-s)$. The functional forms for α and β are:

$$(A.2) \qquad \alpha = 0.2 - \frac{a^{0.5}}{800} \quad ; \quad \beta = 0.08 + \frac{b^{0.5}}{800} \quad .$$

In principle, sufficiently high values of a or b would violate $\alpha > 0$ and, in the discrete model, $\beta < 1$; in practice, the simulations do not produce such values. Because both exponents in equation (A.2) are the same, equation (14) for Φ implies a closed-form relation between the ratios of the inputs, b/a, and (1 -s).

The dynamic equation is the discrete version of equation (12). In this case, if nothing is done, the equilibrium is an internal one with the steady- state fraction of the population being $s^* = .08/.20 = 0.40$ and the corresponding value of $W = 6400$. We found the policy rule by iterating on the value function.[13] The computation seemed well-behaved and straightforward,

[13]We used a grid of 417 points, with s ranging from 0.005 to 0.985. The density of points was not uniform and was chosen to provide increased accuracy at the limits of the range and around the steady state . We interpolated the value function for intermediate values of s using a cubic spline. We accepted that the value function had converged when the maximum proportional

except that there was some slight instability in the values of the controls for the last few values of s at either end of its range. This instability seemed to be associated with the interpolation because it seemed always to be pushed to the end when we added additional points. To save space, Table A.1 shows values of the (endogenous) parameters of the model (α and β) and the value function (W) for ten of the 417 values of the state variable , s, used in the computation. The steady-state value of s is (approximately) 0.443 and the value of moving there from the no-intervention steady-state value of s is the difference between 6400 and 6568.

We increased the four price parameters to indicate the responsiveness of the steady-state value of s, s*. A ten percent increase in p_I increases s* by 0.9 percent. A ten percent increase in p_a decreases s* by 0.7 percent. A ten percent increase in p_b decreases s* by 0.6 percent. A ten percent increase in (1+r) decreases s* by 3.2 percent. All the simulations for these comparative steady-state results as well as other similar sensitivity analysis suggest a well-behaved problem.

6. References

Anderson, R. M. and May, R. M. (1991). Infectious Diseases of Humans .Oxford: Oxford University Press.

Brito, D. L., Sheshinski, E. and Intriligator, M. D, (1991). "Externalities and Compulsory Vaccinations," Journal of Public Economics, vol. 45, no. 1 (June), pp. 69-90.

Francis, P. J. (1997). "Dynamic Epidemiology and the Market for Vaccinations," Journal of Public Economics, vol. 63, no. 3 (February) pp. 383-406.

Geoffard, P.-Y. and Philipson, T. (1996). "Rational Epidemics and their Public Control,"

---

change in the value function at every s in an iteration was less than .000001. All calculations were done in double-precision FORTRAN.

International Economic Review, vol. 37, no. 3 (August) pp. 603-624.

Gersovitz, M. (2003). “Births, Recoveries, Vaccinations and Externalities,” in Essays in Honor of Joseph E. Stiglitz eds., R. J. Arnott, et al. Cambridge: MIT, pp. 469-483.

Gersovitz, M. and Hammer, J. S.(forthcoming a). “Infectious Diseases, Public Policy, and the Marriage of Economics and Epidemiology,” World Bank Research Observer.

Gersovitz, M. and Hammer, J. S. (forthcoming b). “Tax/Subsidy Policies toward Vector-Borne Infectious Diseases,” Journal of Public Economics.

Goldman, S. M. and Lightwood, J. (2002). “Cost Optimization in the SIS Model of Infectious Disease with Treatment,” Topics in Economic Analysis & Policy, vol. 2, no. 1, pp. 1-22.

Greenhalgh, D. (1988). "Some Results on Optimal Control Applied to Epidemics," Mathematical Biosciences, vol. 88, no. 2 (April) pp. 125-158.

Hocking, L. M. (1991). Optimal Control: An Introduction to the Theory with Applications. Oxford: Clarendon Press.

Kamien, M. I. and Schwartz, N. L. (1981). Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management. New York: North Holland.

Kremer, M. (1996). "Integrating Behavioral Choice into Epidemiological Models of AIDS," Quarterly Journal of Economics, vol. 111, no. 2 (May), pp. 549-573.

Philipson, T. (2000). “Economic Epidemiology and Infectious Diseases,”, ch. 33 in A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics. Amsterdam: Elsevier.

Sethi, S. P. and Staats, P. W. (1978). “Optimal Control of Some Simple Deterministic Epidemic Models,” Journal of the Operational Research Society, vol. 29, no. 2 (February) pp. 129-136.

Wickwire, K. (1977). "Mathematical Models for the Control of Pests and Infectious Diseases: A Survey," Theoretical Population Biology, vol. 11, pp. 182-238.

Wiemer, C. (1987). "Optimal Disease Control Through the Combined Use of Preventive and Curative Measures," Journal of Development Economics, vol. 25, no. 2 (April), pp. 301-319.

Zeidan, V. (1984). “First and Second Order Sufficient Conditions for Optimal Control and the Calculus of Variations,” Applied Mathematics and Optimization, vol. 11, no. 2 (June), pp. 209-226.

Table 1
Targeting and the Relative Reliance on Preventive and Therapeutic Policies
[Values of Φ of Equation (14)]

Values of θ[a]		Values of θ[b] I. 1	II. (1-s)
A.	1	s	s(1-s)
B.	s	1	(1-s)
C.	(1-s)	s/(1-s)	s
D.	s(1-s)	1/(1-s)	1

Table A.1
The Policy Rules and Value Function

s	α	β	W
0.100	0.1846	0.0821	4566
0.200	0.1853	0.0823	5170
0.300	0.1860	0.0825	5833
0.400	0.1868	0.0827	6568
0.443	0.1872	0.0829	6911
0.500	0.1877	0.0830	7394
0.600	0.1888	0.0835	8338
0.700	0.1900	0.0841	9445
0.800	0.1917	0.0852	10801
0.900	0.1941	0.0874	12605

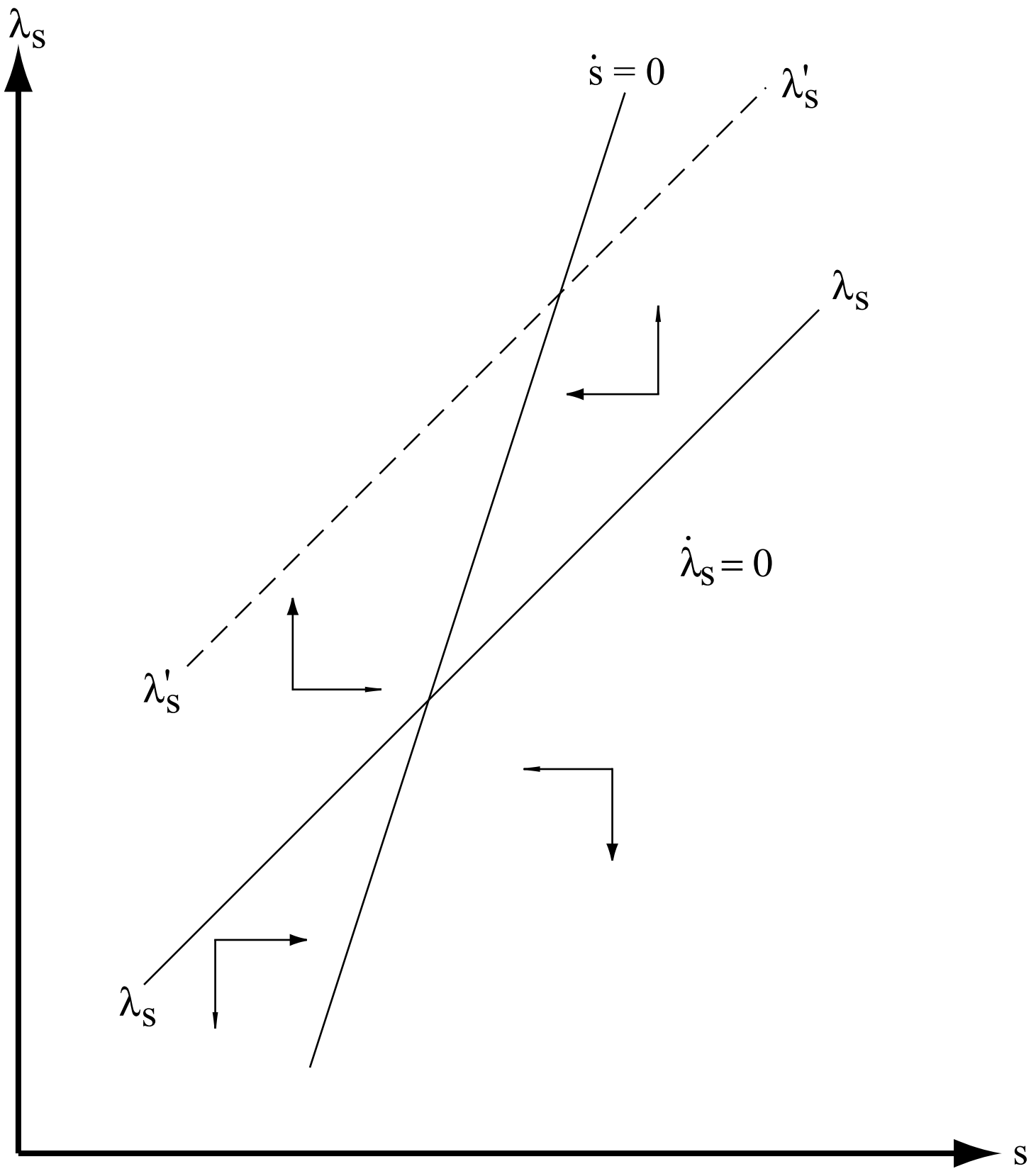

Fig. 1a
The Phase Diagram for an SIS Disease with Case IIA or D Targeting: A Stable Case

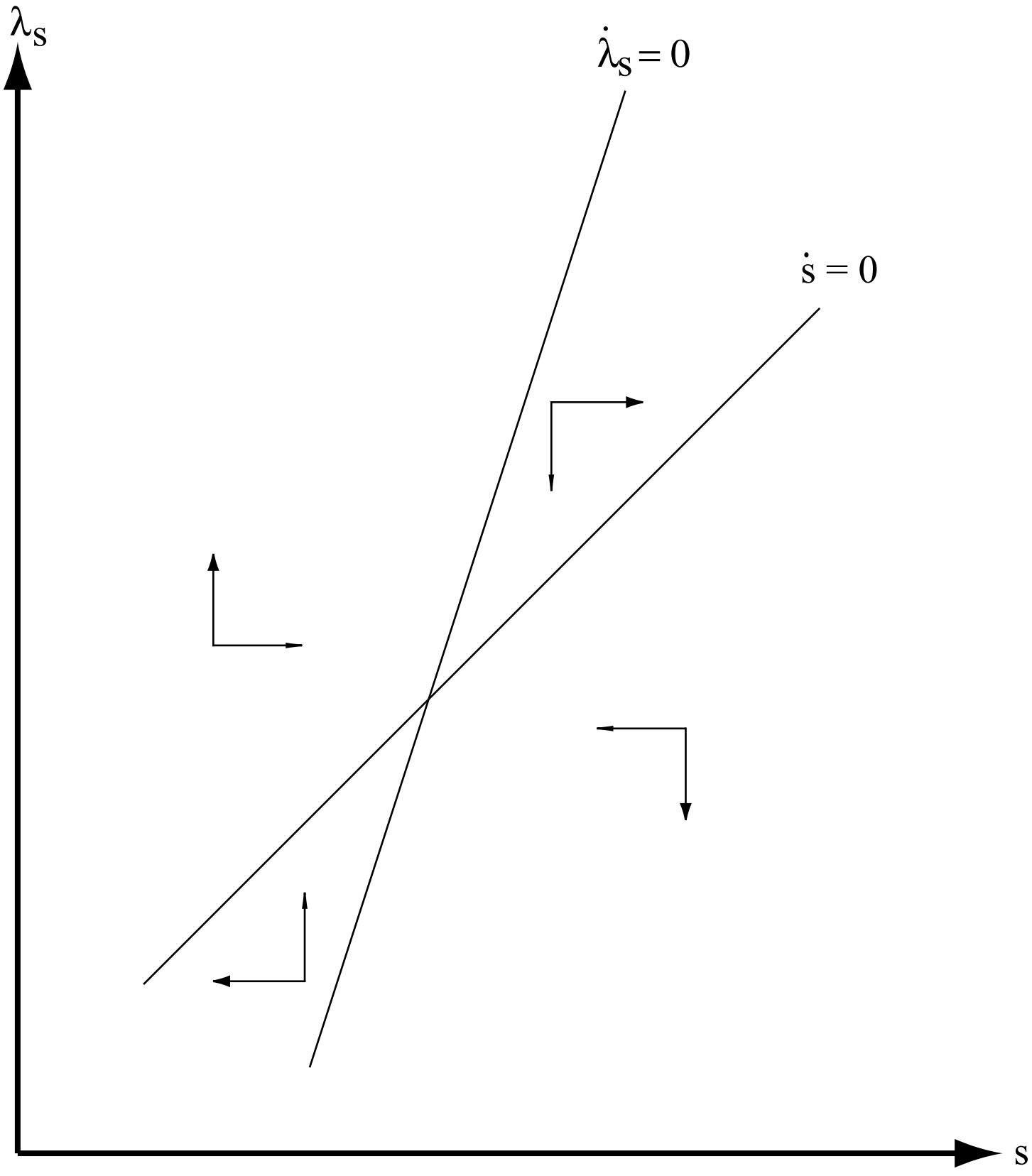

Fig. 1b
The Phase Diagram for an SIS Disease with Case IIA or D Targeting: An Unstable Case

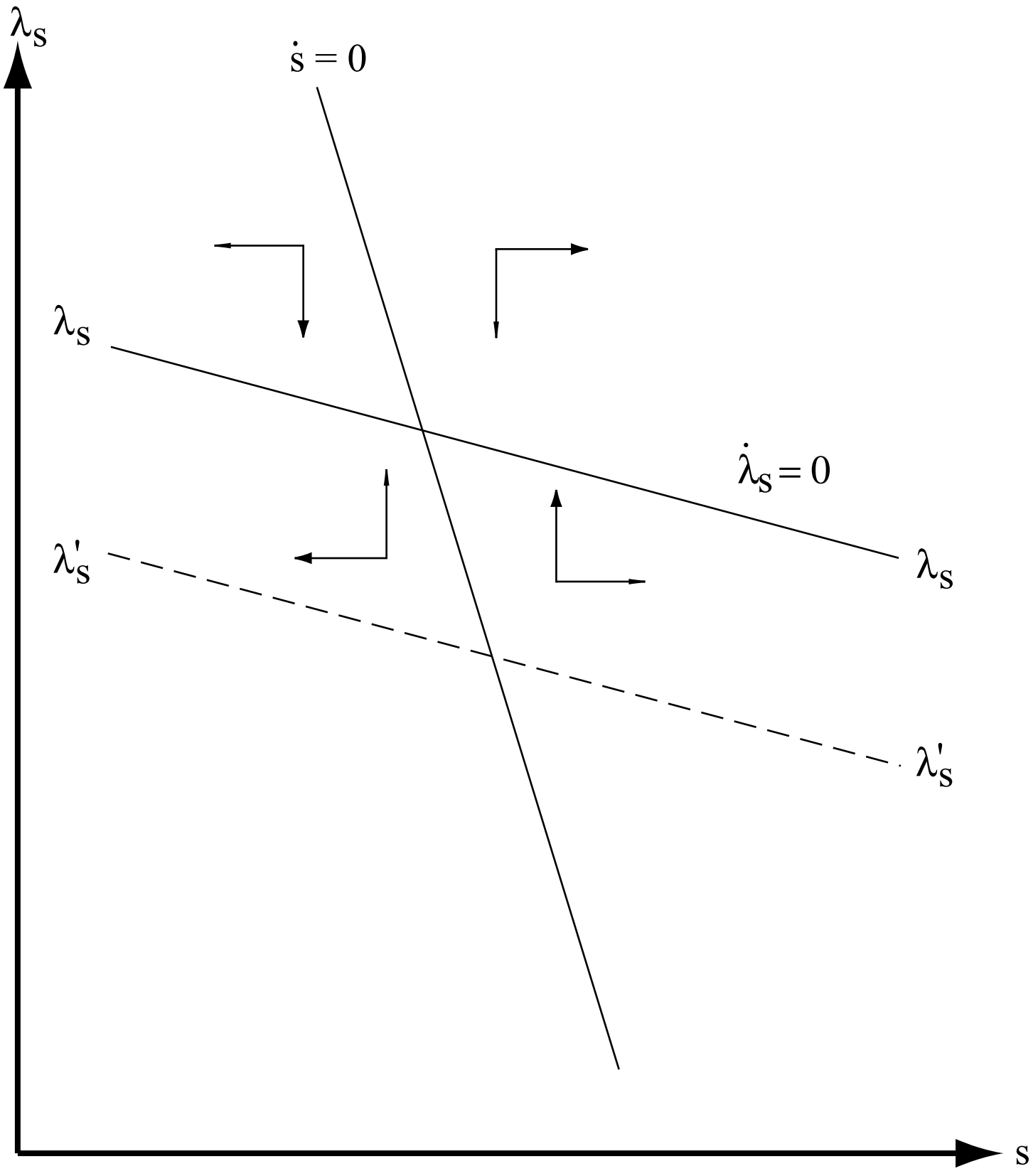

Fig. 1c
The Phase Diagram for an SIS Disease with Case IIA or D Targeting: A Stable Case