Mathematical Epidemiology and Welfare Economics

Mark Gersovitz

Abstract Economics provides theories of private behavior and government policy that can be integrated with mathematical epidemiology, as illustrated in a susceptible-infected-susceptible model of infection. Confronting infections, people decide on prevention and therapy with regard to consequences for themselves but not for others, the economic concept of an externality. Public policy can optimally offset the externality by subsidizing prevention and therapy at equal rates (or less practically, taxing infection). Absent such interventions, seemingly beneficial changes such as a decreased cost of infection can perversely lower welfare by worsening the externality, the economic concept of immiserization. Other issues discussed include uniqueness and stability of the optimal steady state and its response to parameter changes.

1 Introduction

This chapter discusses choices that affect people's health, choices about prevention and therapy made by individuals and governments in an environment of infectious diseases. Choice requires options. The range of options and their consequences provide the constraints on choice. Choice also presupposes objectives, the goals that individuals and governments pursue. This chapter discusses both constraints and objectives and how the decisions of individuals and governments to maximize their objectives subject to their constraints produce behavioral outcomes and the dynamics of infectious diseases. To fix ideas, this chapter illustrates these general principles with a model of transmission of a susceptible-infected-susceptible (SIS)

M. Gersovitz (✉)
Department of Economics, The Johns Hopkins University, 440 Mergenthaler Hall 3400 N. Charles St Baltimore, MD 21218, USA
e-mail: m.gersovitz@jhu.edu

P. Manfredi and A. d'Onofrio (eds.), *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*, DOI 10.1007/978-1-4614-5474-8_12,
© Springer Science+Business Media New York 2013

disease, drawing on [6].[1] Throughout, the discussion is abstract and is therefore an exercise in pure theory that is not yet ready to guide actual policy toward any particular disease.

Any approach to infectious diseases recognizes that a person who becomes infected and stays that way poses risks of infection to others. Presumably in many if not most such instances a person who poses such risks does not take into account in deciding about prevention and therapy the benefits to other people if that person is uninfected. Any such subsequent infections of other people, however, impose costs on them and their own attempts to avoid such infections also imposes costs on them. In modeling infectious diseases, one focus of economics is understanding the discrepancy between the costs and benefits as seen by self-interested individuals who make choices about prevention and therapy and the totality of these costs and benefits as seen by society as a whole. This discrepancy is termed an externality, for the costs and benefits that are external to the person making the decisions. Differences in the optimal choices as seen by individuals and by governments about prevention and therapy then provide the rationale for public policy to align private and social choices.

Certainly, epidemiologists are well aware that infectious diseases pose problems because the infection of an individual may have consequences for others. They use terms such as mass or community effects or herd immunity for phenomena that economists would term externalities. What primarily distinguishes these notions from the externality is that the economic concept of an externality is anchored in an explicit comparison between the incentives faced by rational individuals acting alone and by the policy maker acting for society as a whole. Furthermore, in acting for society as a whole, the policy maker's valuation of an individual's well-being does not differ from that individual's own valuation. Where the policy maker differs from individuals is in recognizing that individuals' actions taken together have consequences for people's well-being as a whole. This notion is made more precise in the succeeding sections in which the objective of the social planner and the representative private decision maker are congruent but their constraints differ.

Like any models, therefore, the one presented here has its assumptions, some of which may be unfamiliar, even uncongenial, to noneconomists. First of all, the decision makers both private and public are rational in that they maximize an objective subject to constraints. Second, because the problem is a dynamic one, individuals and policy makers have to form expectations about the future most especially about future values of the infection rate which determines the probability of infection. To avoid compounding the identification of any externalities with problems of myopia, decision makers are assumed to foresee perfectly the dynamics of the infection as summarized in the proportion of the population that is infected (a special case of rational expectations because the model is not inherently stochastic). Furthermore, to the extent that disease persists, it affects the well-being

[1]This chapter is not a survey of contributions by economists to the study of infectious diseases and does not undertake any literature review. Paper [3] provides such a survey and should be seen as a complement to this chapter.

of people into the indefinite future. Decisions taken today affect the future course of the disease and therefore affect the welfare of people in the future, including the welfare of people born after the decisions were made. The decision makers are therefore assumed to have objectives formulated in a way that provides a consistent accounting for present and future well-being so that decisions are made taking into account their future consequences. Finally, the model is simplified so that people do not care whether they have access to insurance against the costs of the disease or not, again to focus on externalities rather than problems in insurance markets. I sketch some alternative ways to formulate the model so that people do care about access to insurance, a topic for future research.

The next section lays out the accounting for people by disease status and the dynamics that move people from one status to another, the constraints on optimization. The following section introduces the objective of decision makers taking into account the costs of prevention, therapy, lost work time, pain, and suffering. As is conventional in modeling externalities, this chapter next discusses the problem of a hypothetical social planner who can directly control all preventive and therapeutic actions. Maximization of the objective function subject to the constraints provides the optimal solution to the social planner's problem and some of its properties. The next section looks at decisions by individuals, their deviations from the social planner's choices, and hence the existence of externalities and the role for public interventions to achieve the social planner's optimum.

2 The Dynamics of Infection and the Constraints on Choice

In the SIS model of infections transmitted from person to person, the total number of people (N) is the sum of the number who are susceptible (S) or infected and infectious (I):

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

The proportions of these groups in the population are denoted by s $(= S/N)$ and i $(= I/N)$ so that $s + i = 1$. The birth rate of the population is ε. For simplicity no deaths occur at all. An assumption that people die at the same rate regardless of whether they are infected or not would just introduce an extra parameter of no interest to the questions of this chapter. Specifying a death rate that depends on infection status is contrary to the assumption of an SIS infection; [6] discuss a model in which infection raises the probability of death. The net change in the population is therefore

$$\dot{N} = \varepsilon N. \tag{2}$$

The number of susceptibles changes over time according to

$$\dot{S} = \varepsilon N - \alpha S i + \beta I. \tag{3}$$

Given Eq. (2), the first term of the right-hand side embodies the assumption that all newborns are susceptible. The second term reduces the number of susceptibles by those people who become infected. Under the assumption of random (or homogeneous) mixing, the probability per contact of a susceptible person's meeting an infected (and infectious) person is the proportion of infected people in the population, $i = I/N$. The product, Si, is the number of susceptibles who do so. The factor α is a composite term incorporating both the rate of contact and the inherent infectiousness of an infected (or susceptibility of a susceptible). The third term, βI, is the addition to the susceptible pool resulting from the recovery of infecteds at rate β. Eqs. (1)–(3) can be solved for the change in the proportion of susceptibles:

$$\dot{s} = -\alpha s(1-s) + (\beta + \varepsilon)(1-s). \tag{4}$$

So long as the symbols α and β represent constant parameters, this equation is a classic model in mathematical epidemiology. In traditional epidemiological modeling, α and β are invariant (or exogenous) within any model. In particular, they do not vary with the prevalence the infection, i, or equivalently in an SIS model, the variable s. A simple representation of behavior in such a model is to assume that in some places or times, α and β take specific values but that in other places or times, they take different values. For instance, there may be immutable customs that differ among communities and affect the ease of becoming infected and of recovering with corresponding consequences for α and β. But an economist would not think of these differences as constituting a behavioral model of the transmission dynamics of an infectious disease. In such circumstances there would not be any choices that individuals could make and consequently none that could or should be affected by government policy.

Behavior enters the model if people make choices about preventive effort and therapeutic effort that affect α and β. Most importantly, these choices respond to the state of the infection (summarized by s in an SIS model), because the risk of infection is proportional to the infection prevalence, $i = 1 - s$, and this risk shifts the costs and benefits of prevention and therapy. A summary representation of these notions would be to make α and β functions of s, so that Eq. (4) would be modified to:

$$\dot{s} = -\alpha(s)s[1-s] + [\beta(s) + \varepsilon][1-s]. \tag{5}$$

Philipson [9] termed this type of dependence of α and β on s the prevalence elasticity of behavior, and he emphasized it as the touchstone of an economic approach to epidemiology.

Certainly, there are properties of Eq. (5) that are worth exploring. Such properties include whether the steady-state value of s is unique, whether it is stable, and the conditions under which these properties obtain. Some of these issues are discussed in [4]. But if Eq. (5) is the starting point for an investigation, little can be said about the functional forms of $\alpha(s)$ and $\beta(s)$, how they differ according to whether individuals are making all the decisions or governments are also intervening, and therefore the social desirability of the choices made by individuals and the role for public interventions.

Thus it is desirable to move back from Eq. (5) and rebuild it from the underlying components of the problem, the objectives, and constraints faced by maximizing decision makers and their consequent choices. The first step is to recognize that α and β do not depend directly on the prevalence of the infection, $i = 1 - s$, but rather on the inputs of preventive effort and therapeutic effort which are in turn choices.

Either input may be targeted in the sense that only a proportion of the population generates costs associated with the input. Let θ^j, $j = a, b$ be the proportions of the population that generate either preventive or therapeutic costs associated with an infectious disease. The θ^j are termed targeting functions; in general, they depend on s. 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 ($\theta^b = 1 - s$). 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. The type of targeting may be a choice variable, but in this chapter it is a technical given. For example, in the case of a respiratory infection such as a cold or influenza, θ^a could plausibly take values of 1, s, $1 - s$, and $s(1 - s)$. In the first case, everyone wears a mask, in the second only the uninfected do so, in the third only the infected do so (as in Japan), and in the fourth only in matchings involving an uninfected and an infected person do people wear masks. For therapeutic interventions, the simplest case is targeting exclusively at the infected so that $\theta^b = (1 - s)$. If it is difficult to diagnose the disease, cheap to treat, and treatment does not have important side effects, then mass treatment may be adopted with $\theta^b = 1$. Such targeting has been tried for sexually transmitted diseases.

In all these targeting schemes for prevention and for therapy, it is the level of these health inputs per targeted person that affect the parameters of the model directly, with prevention lowering α and therapy raising β. The number of units of preventive effort is denoted by a and the number of units of therapeutic effort is denoted by b so that $\alpha(a)$ and $\beta(b)$. Thus a and b determine, respectively, the rate of new infections and the rate of transition back to being susceptible. The preventive and therapeutic interventions exhibit positive but diminishing marginal products, i.e., $\alpha' < 0$, $\alpha'' > 0$, $\beta' > 0$, and $\beta'' < 0$. For many if not all diseases there is scope for undertaking additional preventive and therapeutic interventions although they are marginally less and less productive.

All types of targeting considered in this chapter will produce a dynamic equation of the form

$$\dot{s} = -\alpha(a)s[1 - s] + [\beta(b) + \varepsilon][1 - s]. \tag{6}$$

Other types of targeting would not. For instance, if only some of the susceptibles are targeted by prevention, then there would be two groups of susceptibles and their dynamics would have to be tracked separately and the model would no longer have only one state variable. But for the types of targeting considered in this chapter, Eq. (6) is operative and the problem is then to show how a and b depend on s.

3 The Well-Being of Individuals and the Consequences of Infectious Diseases

In evaluating situations, economics usually starts with the utility function, a relationship between consumption and well-being. It provides both a component of predictive theories based on the hypothesis that individuals maximize utility and a method of evaluating any situation from a social perspective based in individuals' own evaluation of that situation. When the situation is one involving health, a general formulation of the utility function would make it depend on health status as well as consumption. This approach differs from that of cost-effectiveness analysis which explicitly sidesteps the question of the utility of health; paper [5] discuss criticisms of cost-effectiveness.

A person who is susceptible (and therefore healthy) and who consumes commodities other than those involved in health of quantity c_s has utility function $\Psi^s(c_s)$, $(\Psi^s)' > 0$, and $(\Psi^s)'' < 0$, with these derivative conditions representing the assumption that marginal utility of consumption is positive but diminishing. A person who is infected and therefore ill consumes other commodities of quantity c_i and has utility $\Psi^i(c_i)$ with the same derivative properties. The costs of being infected, therefore, affect the utility of individuals in two ways. First, there are monetary costs. People may have less to spend on the consumption of other commodities if they become infected because they spend on therapy and because their ability to earn income is impaired. They may also be continuing to spend on prevention. At the same time, they may be receiving insurance payments and paying insurance premiums. If people are uninfected (and susceptible), they may be spending on prevention and paying insurance premiums. Second, people who are infected experience pain, suffering, and other physical impairments which affect their level of well-being. Formally, these latter consequences of illness are represented by the fact that the functions $\Psi^k(c_k)$ $k = i, s$ are superscripted and are therefore potentially completely different functions of the argument. About all that can be said is $\Psi^i(c) < \Psi^s(c)$, the utility of being infected is less than the utility of being susceptible if the amount of consumption is identical. Importantly, there is no presumption about the relative magnitudes of the marginal utilities of consumption, the $(\Psi^k)'$, if consumption is identical, i.e., $\left(\Psi^i(c)\right)'$ and $(\Psi^s(c))'$ bear no necessary relation to each other.

A specialization of these utility functions is: $\Psi^i(c_i) = \Psi(c_i) - h$ and $\Psi^s(c_s) = \Psi(c_s)$, $h > 0$ in which case the marginal utilities of consumption are the same if the value of consumption is the same and only the term h represents the utility loss from pain, suffering, and other physical impairments. Nonetheless, there are potentially important effects of the dependence of utility on health status even in this restricted specification.

A yet further simplification is to assume that all the effects of illness can be represented by money and can therefore be subtracted from available income (or have lowered income beforehand in the case of lost work opportunities). In this case, utility depends only on consumption and the utility function is the same regardless of disease status, $\Psi(c)$. There is potentially a long list of these monetary costs.

Suppose prevention costs p_a per unit and each targeted person receives a units. There are $\theta^a N$ such people so that the total cost of prevention is $p_a a \theta^a N$. Similarly, if therapy costs p_b per unit, the total cost of therapy is $p_b b \theta^b N$. These costs may be direct monetary costs, such as the cost of a drug, or they may also involve time in which case monetary estimates could be made and they could be included in the relevant p_j, $j = a, b$.

Some of the costs of actually being infected and ill are also monetary in a straightforward way. For instance, people may be unable to work and thereby lose income. It is more difficult to attach a monetary value to the costs of illness in terms of pain, suffering, and disability. The simplest cases to consider are ones in which it is possible to do so. If so, everyone who is ill experiences a monetary cost of p_I that includes the costs of missed work, pain, suffering, and disability and the total costs of being infected and ill are then $p_I(1-s)N$. Costs of therapy as discussed above are in addition.

If all these consequences of the existence of the disease can be measured in monetary terms then once one knows who pays, one can calculate the consequences for each individual's well-being. For instance, assume that everyone has income of V_0 available for consumption if there were no disease at all. Furthermore, assume that targeting is such that susceptibles undertake and pay for prevention ($\theta^a = s$) and the infecteds undertake and pay for therapy ($\theta^b = 1 - s$) as well as experience the costs of being ill valued at p_I. In particular, there is no health insurance or sharing of these burdens. In this case the consumption of commodities other than health of a susceptible person (c_s) is

$$c_s = V_0 - p_a a, \tag{7}$$

and the consumption of an infected person (c_i) is

$$c_i = V_0 - p_I - p_b b. \tag{8}$$

So long as the marginal utility of consumption is declining, there is a motivation for insurance. People would like to receive an assured level of consumption through insuring it rather than have a level of consumption that fluctuates with their health status. Insurance markets are, however, problematic and therefore complicate the analysis. The simplest assumption is therefore that the utility function, Ψ, is proportional to the level of consumption, c, so that motivations for insurance do not arise. In this case, utility maximization is identical to the minimization of the total costs of the infectious disease, the approach adopted in this chapter following [6] as well as other earlier analyses of the optimal control of infectious diseases. Part of the agenda for future research is to introduce more complicated versions of the utility functions into the analysis with the ultimate goal of providing an understanding of both externalities and insurance.

The next section shows how the social planner, an idealized decision maker for all people in the society, can use this notion of utility to specify objectives to maximize subject to the constraints of the infection process outlined in the preceding section. The important principle is that the social planner's valuation of any situation in terms

of the prevalence of infection and the availability of resources for consumption of commodities other than health is an aggregation of the valuations of the members of the society. The difference between the decisions of the idealized social planner and of individuals is not how the social planner values each person's well-being relative to how that person values his own well-being but rather in how the social planner accounts for externalities.

4 The Social Planner's Problem

The objective of government policy is to maximize people's well-being over the indefinite future, not just in the current period. Some method is therefore needed to aggregate people's well-being as represented by their utility in each and every period. In the case of cost minimization, the present discounted value of consumption is the obvious criterion because it takes account of the time value of money. Costs incurred at different times can be expressed in the common unit of present discounted value by using the interest rate. If the interest or discount rate is constant, the present discounted value of social welfare is

$$W = \int_0^\infty \{c_s S + c_i I\} e^{-rt} dt = \int_0^\infty \left\{V_0 N - [p_I iN + p_a a\theta^a N + p_b b\theta^b N]\right\} e^{-rt} dt \quad (9)$$

in which r is the discount rate.

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

$$H = N\left\{V_0 - \left[p_I(1-s) + p_a a\theta^a + p_b b\theta^b\right]\right\} + (\lambda_s N)[(1-s)(\varepsilon+\beta) - \alpha s(1-s)] + \lambda_N[N\varepsilon], \quad (10)$$

in which $(\lambda_s N)$ and λ_N are the current-value multipliers. Because $\lambda_s N$ is the multiplier on the change in s, it has the interpretation of the value of a unit increase in s on social welfare, W. The variable λ_s therefore has the interpretation of the value of an increase in s on the welfare of the average member of society, or alternatively as the value of an increase in one susceptible person on social welfare. Because W is measured in monetary units, e.g., dollars, λ_s is measured in monetary units and is like a price, but because it is not actually a price in a market, economists term it a shadow price. In Eq. (10) and what follows, the arguments a and b of α and β and their derivatives are suppressed for compactness when there should be no ambiguity.

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

$$p_a\theta^a = -\lambda_s \alpha' \, s(1-s), \quad (11)$$

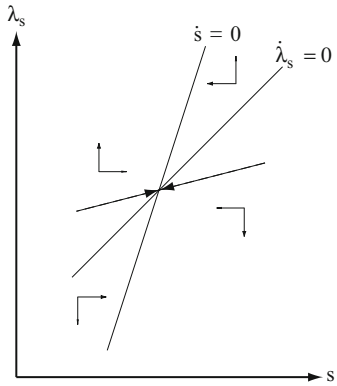

Fig. 1 The phase diagram of an SIS disease: a stable case

and

$$p_b\theta^b = \lambda_s\beta'(1-s), \tag{12}$$

Equation (11) 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 preventive effort as valued using the shadow price of s, λ_s. Equation (12) 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 monetary units.

Under the assumptions on the θ^j and on α' and β', the λ_s must be positive if Eqs. (11)–(12) are to hold. In addition, the dynamic equation for the multiplier implies:

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

Note that Eqs. (6) and (11)–(13) form a system of equations without the need for Eq. (2) or consideration of the variable N. Thus the steady state (states if there are more than one) of the system can be found by setting Eqs. (6) and (13) to zero and using Eqs. (11)–(12) to substitute (implicitly) for the variables a and b. Similarly, some qualitative properties of the dynamics of the system can be inferred from the phase diagram plotted in s-λ_s space, Fig. 1.

To simplify what follows, assume that $\theta^b = 1-s$, so that targeting of therapeutic interventions is restricted to the infected and therefore $\theta_s^b = -1$. Total differentiation of Eqs. (11)–(12) implies

$$a_\lambda \equiv \frac{\partial a}{\partial \lambda_s} = -\frac{\alpha'}{\lambda_s\alpha''} > 0, \tag{14}$$

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

$$b_\lambda \equiv \frac{\partial b}{\partial \lambda_s} = -\frac{\beta'}{\lambda_s \beta''} > 0, \tag{16}$$

and

$$b_s \equiv \frac{\partial b}{\partial s} = 0. \tag{17}$$

These expressions simplify the following discussion. They all have straightforward interpretations, except perhaps the indeterminate sign of the expression in Eq. (15) 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 infectious diseases. 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 the discussion has proceeded on the presumption that Eqs. (6) and (11)–(13) determine a path to a unique optimal steady state, but this need not be so. There may be bifurcations in the phase diagram in $s - \lambda_s$ space so that a Skiba point exists that divides the state space into different regions from which the system converges to different optimal steady states [8]. Paper [7] provides an analysis of this phenomenon in a model that is similar to the one in this chapter and paper [4] provides some further discussion of the model in this chapter. There do not seem to be general conditions that ensure uniqueness in terms of the underlying structure of the problem. In what follows, I assume that there is a unique path to a unique optimal steady state. By implicit differentiation, the slope of the locus in the s-λ_s plane that is obtained from setting Eq. (6) to zero is

$$\left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{s}=0} = \frac{\alpha + \alpha' s a_s}{\beta' b_\lambda - \alpha' s a_\lambda} = \frac{?}{+}, \tag{18}$$

and the slope of the locus from setting Eq. (13) to zero is:

$$\left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{\lambda}_s=0} = \frac{-(p_a a_s \theta^a_s + p_a a \theta^a_{ss} - 2\alpha\lambda_s + (1-2s)\alpha' a_s \lambda_s)}{[r - \alpha s + \beta + (1-s)(\alpha + \alpha' s a_s)]} = \frac{+}{?}. \tag{19}$$

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 the different cases of θ^a.

In two cases, $\theta^a = s$ and $\theta^a = s(1-s)$, 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 exceeds the population growth rate holds, $r > \varepsilon$, so that the society will not have infinite present discounted value and it makes sense to maximize the integral in Eq. (9). 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 because the characteristic equation of the linearized dynamic system has one positive and one negative real root (see Fig. 1). On this path, the variables s and λ_s move together toward the steady state, and b and β move with them so that therapeutic effort increases as the proportion of susceptibles is increasing as it approaches the equilibrium from below (or the reverse if from above). Preventive effort, a, and α move as determined by the relation between a and b as given by Eqs. (11)–(12), decreasing with the proportion of susceptibles if $\theta^a = s$ and varying with it if $\theta^a = s(1-s)$. 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. This chapter, however, does not discuss these divergent cases and is restricted to situations in which the structure of the model is such that policy takes the system to a unique optimal steady state in which the disease is endemic ($0 < s^* < 1$).

The two remaining cases, $\theta^a = 1$ and $\theta^a = 1 - s$, are slightly more complicated. When the model is linearized about the steady state both slopes may be positive, as in the preceding two cases and the foregoing analysis obtains. It may be that the slope of the $\dot{\lambda}_s = 0$ locus is positive and that of the $\dot{s} = 0$ locus is negative or both slopes may be negative, but the $\dot{\lambda}_s = 0$ locus is more negatively sloped. Both these cases are unstable and are not considered on the assumption that there is a unique stable steady state.[2]

The economic parameters of the model are the three prices, p_I, p_a, and p_b, and the interest rate, r. Changes in these parameters affect choices about prevention and therapy and the outcomes of welfare and the infection rate, i. Welfare rather than the infection rate is really what is important; it is, after all, what is being maximized.

For any change in the price parameters $x = p_I, p_a, p_b$, the effect on W is given by the dynamic envelope theorem (see [1], Chap. 9 and 14):

$$\frac{dW}{dx} = \int_0^\infty e^{-rt} \frac{\partial H}{\partial x} dt, \tag{20}$$

in which H is given by Eq. (10). By inspection of Eq. (10), it is immediately apparent that the effects of increases in all three prices is to lower welfare as would be expected.

Although secondary to the effects on welfare, the effects of the parameters on the steady-state infection rate, $i = 1 - s$, may also be of interest. The parameters p_I and r enter Eq. (13) for $\dot{\lambda}_s$ but not Eq. (6) for $\dot{s}$, nor do they enter Eqs. (11)–(12). 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 stable case of Fig. 1. The $\dot{\lambda}_s = 0$ locus shifts up and s^* rises;

[2]Paper [6] mistakenly stated (pp. 13–14 and Fig. 1c) that there could be a stable case with both isoclines negatively sloped and with the s-isocline more negatively sloped. This case would be stable were it possible, but it is impossible under the restrictions of the model as can be proved by careful collection of algebraic terms.

use of therapy increases but the change in prevention depends on targeting. The reverse change, a decrease in p_I, naturally produces the reverse change in s^* and the equilibrium infection rate, $i^* = 1 - s^*$, rises. The use of therapy declines, and in epidemiological terminology, there is disinhibition of therapy which may extend to prevention as well depending on targeting. On balance, disinhibition is so strong that the social planner adjusts the package of prevention and therapy in a way that leads to an increase in infection. But the well-being of members of society increases consequent on the decrease in p_I regardless as has already been established in the discussion of Eq. (20). The change in well-being and not that in the infection rate is what is ultimately important and avoiding disinhibition is not relevant to what the social planner should do. Indeed, if the social planner were arbitrarily ordered to keep the levels of a and b at the values that were optimal before the fall in p_I, welfare would not rise as much because it is suboptimal not to adjust a and b given the shift in costs relative to benefits.

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 therapeutic interventions is to decrease the amount used via Eqs. (11)–(12) and therefore either a and α or b and β are affected in both equations.

In the case of an increase in p_b, both the $\dot{s} = 0$ and the $\dot{\lambda}_s = 0$ loci shift up. 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 consequently 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.

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

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

rising with an increase in p_a if this expression is positive and falling if it is negative. The denominator is unambiguously positive regardless of θ^a so long as the steady state is stable. The sign of the numerator is ambiguous in all four cases; if $\theta^a = 1$ or $\theta^a = s(1-s)$, 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^*.

5 Decentralization

To this point the discussion has concerned 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 an externality.

In the model, 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 do not go through markets, like time and effort by the person at risk for infection or already infected. Some of these activities may even be entirely unobservable by third parties because they involve private and intimate behavior. Thus the government can subsidize condoms but not the act of safer sex. Some public health programs such as directly observed therapy short course (DOTS) in the case of tuberculosis and other diseases are attempts by the public health authorities to monitor and encourage patient compliance at least partially for its benefits external to the patient. The expenditure on such programs is, of course, a type of subsidy. In the case of any specific infectious disease and its control, these issues need detailed attention. In general, when a and b involve non-marketed and unobservable actions the subsidy/tax interventions may be infeasible or may have to be targeted only on the marketed components of preventive and therapeutic activities with limitations on their effectiveness.

The simplest way to illustrate the externality and its implications for policy is to assume that private decisions are made by a group of people termed a household, a construct that serves 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 the dynamics of the infection and therefore how the variable i will evolve and is fully forward-looking with regard to its future status as well as its current one. In its current decisions, the household takes account of the dynamics of the infection, 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 the rationale for government interventions does not

depend either on myopia or on a discrepancy between the social planner's and the representative agent's valuation of outcomes. Instead, the assumptions isolate the pure externality motivation for government intervention. To the extent that there are deviations from the preceding assumptions on the behavior of the household, there may be other important reasons for government interventions but they are not the subject of this chapter.

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 at any time, $i = 1 - s$, which equals the probability, π, that any random contact is with an infected person. The household neglects its effects on the aggregate infection rate because such an effect is too small for the household to take into the account as it affects the well-being of its own household members and because the selfish household does not care about its effects on all the other households even though these effects taken together are not negligible. 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 society from private decision makers each of whom is in one or another disease status at any one time and taking decisions about either prevention or therapy, with regard to their possible future status as well as their current one.

The dynamic equations of this version of the model are the same as for the social planner except that in Eq. (6) the term $\alpha s(1-s)$ 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.

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, 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, the household receives a lump-sum payment (possibly negative) per household member of T that it takes as exogenous to its own choices about prevention and therapy but that in fact equals $t_a p_a a^h \theta^a + t_b p_b b^h \theta^b$. A superscript 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:

$$\begin{aligned} H^h = N^h \left\{ V_0^h - [p_I i^h + q_a a^h \theta^a + q_b b^h \theta^b] + T \right\} \\ + (\lambda_s^h N^h)[(1-s^h)(\varepsilon+\beta) - \alpha s^h \pi] \\ + \lambda_N^h [\varepsilon N^h]. \end{aligned} \tag{21}$$

All functions of variables (θ, α, and β) 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.

Once again, assume that only the infected are targeted by therapies, so that $\theta^b = (1-s^h)$. Differentiation of Eq. (21) with respect to a and b implies:

$$q_a \theta^a = -\lambda_s^h \alpha' s^h (1-s) \tag{22}$$

and

$$q_b \theta^b = \lambda_s^h \beta' (1-s^h), \tag{23}$$

and the co-state equation is:

$$\dot{\lambda}_s^h = r\lambda_s^h - [p_I - q_a a^h \theta_s^a - q_b b^h \theta_s^b] + [\alpha(1-s)+\beta]\lambda_s^h. \tag{24}$$

Because the group is representative of society, s must equal s^h. Once this substitution is made, the only differences between Eqs. (11)–(13), the planner's problem, and Eqs. (22)–(24), the private problem, are the q_j and the $(1-s)$ term at the end of Eq. (24) rather than the $(1-2s)$ term at the end of Eq. (13). 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. Note that the s-isocline takes the same form for the problem of Eq. (21) as for the problem of Eq. (10) and therefore has the same position in Fig. 1. The λ-isocline has been altered from Eq. (13) to Eq. (24) and is lower in Fig. 1 than in the social planner's problem if there is no government intervention ($t_a = t_b = T = 0$). Consequently the steady state value of s is lower with households rather than the social planner making decisions, indicative of the diminished incentive to undertake prevention and therapy faced by households when acting without government interventions.

At this point, there are two natural goals to the investigation:

First is to establish what the government can do to induce private households to undertake choices that have been shown to be optimal for the social planner (in the preceding section). In particular, the goal is to find the taxes (or subsidies depending

on whether they are negative or positive) that induce this behavior, termed the decentralization problem. If the social planner's choices can be decentralized then the analysis of the effects of parameter changes on well-being and the infection rate is identical to that of the social planner and, in particular, disinhibition, if any, is a side effect of the optimal choices and is not relevant in guiding what should be done.

Second is to establish some of the consequences if the government does not intervene at all. For instance, there is the question of how welfare responds to changes in the parameters if the government does not intervene. In particular, can a change such as a decrease in the cost of infection, p_I, that improves welfare when the social planner is taking decisions instead lower welfare when decisions are made by households in the absence of government intervention to implement (decentralize) the social planner's solution? If it does so, it is because the direct effect of the parameter change (corresponding to what would happen in the social planner's problem) is overwhelmed by a worsening of the externality, the discrepancy between the social and private decisions. Immiserization is the economist's term for this type of outcome.

Turning to the first goal, the main result is that, in principle, 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 Eqs. (11)–(12) with Eqs. (22)–(23) shows that a property of successful decentralization of the social planner's problem is $t_a = t_b = t$. In other words, the government compensates for any differences between λ_s and λ_s^h in Eqs. (11)–(12) and (22)–(23). 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 Eqs. (22)–(23), a and b activities are affected to the same degree. At the steady state, the intervention is a subsidy (negative tax) at rate t^*:

$$t^* = -\frac{\lambda_s^* \alpha s^*}{p_I + \lambda_s^* \alpha s^*} < 0, \tag{25}$$

in which λ_s^* and s^* are the values from the planner's steady state [6]. Furthermore, for any non-steady-state s, the government must intervene with a subsidy [6]. 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. Note that this policy could also be implemented by taxing the condition of being infected (raising p_I) to raise λ_s^h to coincide with λ_s.

The second goal is to consider what happens if the government does not intervene to implement the social planner's solution, i.e., $t_a = t_b = T = 0$. Clearly, the first observation is that because the choices from the solution to the problem of Eq. (21) under these conditions are not the same as the solution from the problem of Eq. (10), the level of well-being under private decision making is lower than when the

social planner takes decisions. The social planner maximizes well-being; the private decision maker (household) ignores something that has to be taken into account to maximize well-being, namely that $\pi = (1 - s)$.

Furthermore, starting from a steady state, the effect of a change in a parameter, x, on welfare, W, is more complicated when the household takes decisions and the optimal subsidies have not been put in place. The equation determining the outcome is

$$\frac{dW}{dx} = \int_0^\infty e^{-rt} \left[\frac{\partial H}{\partial x} + \frac{\partial H}{\partial \pi} \frac{d\pi}{dx} \right] dt = \int_0^\infty e^{-rt} \left[\frac{\partial H}{\partial x} + \lambda_s N \alpha s \frac{ds}{dx} \right] dt. \tag{26}$$

This result is an extension of the dynamic envelope theorem to account for the fact that the household ignores its effect on s through its effect on π (the derivation is the same as in [1], Chaps. 9 and 14). The term $\partial H/\partial x$ has the same sign regardless of whether it derives from the social planner's or the household's problem, although the values at which it is evaluated differ between the two situations. Because the model has only one state variable, the state moves monotonically from one steady state to another so the sign of ds/dx is the same along the path as the sign of the difference between steady states.

As an illustration of what can happen if the social planner's solution is not implemented consider the effect of a decrease in p_I. This change increases H as it did in the social planner's problem. However, such a change also lowers s, through disinhibition as households slacken off on prevention and therapy taken together. Thus Eq. (26) raises the possibility that the net effect on welfare could be negative if disinhibition is extreme enough not just to lower s but to lower it so much that the first term of the right-hand side of Eq. (26) is dominated by the second term. At this point, I do not have an example that leads to immiserization based on a special case, but substitution for the terms in Eq. (26) makes it plausible that one exists. The reason immiserization may occur is that households make choices that are sub-optimal because they disregard their effect on others' welfare. A decrease in the cost of infection could worsen this discrepancy between the social planner's and the private solution and on balance lower welfare even though the direct effect of the decrease in the cost of infection is to increase welfare. The outcome is then immiserization, a perverse transformation of a seemingly beneficial change into an actual lowering of welfare. This outcome is, of course, only possible if the social planner's solution is not decentralized through taxes or subsidies. If the social planner's solution is implemented, a decrease in p_I does decrease s, but it cannot decrease welfare because the consequence for welfare is given by Eq. (20) for the social planner. One could say that the social planner is disinhibited or that households facing the optimal tax/subsidy package implementing the social planner's solution are optimally disinhibited, but they are not immiserized. This observation underlines the position that the goal of policy is the maximization of well-being inclusive of all costs and not the minimization of infection.

6 Conclusions

This chapter has tried to build a bridge between epidemiology and economics, as regards both terminology and substance. In this chapter, the epidemiological model is one in which infection spreads from person to person and people recover from the disease without becoming immune. Paper [6] presents results for models of infection from person to person that result in death or immunity as well as infections involving vectors. Gersovitz [2] presents a model of externalities and vaccinations that leads to explicit formulae for optimal subsidies in terms only of the underlying parameters of the models without need to calculate multipliers as in Eq. (25).

An important direction for further research in economic epidemiology is to make the models more realistic by capturing more aspects of infections and their transmission. A key part of such progress would seem to be a more explicit modeling of behavior by private decision makers and possibly the strategic interaction of people who know at least some of the people who are putting them at risk such as their sexual partners in the case of sexually transmitted infections or their family members in the case of tuberculosis. Another way to enrich the understanding from these models is to deal with motivations for public policy in addition to the externality. Questions of equity arise when people are in different situations especially because they have different incomes. If people are risk averse cost minimization is not an appropriate goal and it is necessary to substitute utility maximization with an explicit treatment of insurance markets and their attendant problems. The first step is to implement a version of utility maximization rather than cost minimization as sketched in Sect. 2.

References

1. Caputo, M.R.: Foundations of Dynamic Economic Analysis. Cambridge University Press, Cambridge (2005)
2. Gersovitz, M.: Births, recoveries, vaccinations and externalities. In: Arnott, R., Greenwald, B., Kanbur, R., Nalebuff, B. (eds.) Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz. MIT Press, Cambridge (2003)
3. Gersovitz, M.: Annu. Rev. Resour. Econ. **3**, 277–296 (2011)
4. Gersovitz, M.: Infection Control and Uniqueness in a Model of Susceptible-Infected-Susceptible (SIS) Diseases, mimeo, Johns Hopkins University(2012)
5. Gersovitz, M., Hammer, J.S.: The World Bank Res. Obser. **18**, 129–57 (2003)
6. Gersovitz, M., Hammer, J.S.: The Econ. J. **114**, 1–27 (2004)
7. Goldman, S. M., Lightwood, J.: Top. Econ. Anal. Policy **2**, 1–22 (2002)
8. Grass, D., Caulkins, J.P., Feichtinger, G., Tragler, G., Behrens, D.A.: Optimal Control of Nonlinear Processes: With Applications in Drugs, Corruption and Terror. Springer, Berlin (2008)
9. Philipson, T.: Economic epidemiology and infectious diseases. In: Cuyler, A.J., Newhouse, J.P. (eds.) Handbook of Health Economics. North Holland, Amsterdam (2000)