DISCUSSION PAPER SERIES

No. 8111

INFECTION, ACQUIRED IMMUNITY AND EXTERNALITIES IN TREATMENT

Flavio Toxvaerd

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8111.asp**

ISSN 0265-8003

INFECTION, ACQUIRED IMMUNITY AND EXTERNALITIES IN TREATMENT

Flavio Toxvaerd, University of Cambridge and CEPR

Discussion Paper No. 8111
November 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8111

November 2010

ABSTRACT

Infection, Acquired Immunity and Externalities in Treatment*

This paper considers a model of infectious disease, such as swine flu, in which privately costly treatment confers immunity on recovered individuals. It is shown that under decentralized decision making, infected individuals ignore the externality that their treatment has on susceptible individuals and thus seek treatment only if it is privately optimal to do so. In contrast, a benevolent central planner who does take this externality into account in choosing the level of aggregate treatment, may choose to either eradicate the disease or to retard its eventual dissemination into the population even when individuals would not find it privately optimal to do so. The analysis shows that when immunity from future infection is obtained through recovery, treatment resembles vaccination in its effects on infection dynamics, but important differences remain. Vaccination is shown to more effectively curb infection than does treatment. Last, the inefficiency associated with decentralized decision making can be corrected through subsidized treatment offered on a first-come first-served basis.

JEL Classification: C73 and I18
Keywords: acquired immunity, economic epidemiology, externalities and treatment

Flavio Toxvaerd
Faculty of Economics
University of Cambridge
Austin Robinson Building
Sidgwick Avenue
Cambridge
CB3 9DD
Email: fmot2@cam.ac.uk

* I gratefully acknowledge very helpful feedback from Joshua Ross, Chryssi Giannitsarou and seminar participants at the University of Cambridge.

Submitted 02 November 2010

Infection, Acquired Immunity and Externalities in Treatment*

Flavio Toxvaerd†

October 2010.

Abstract. This paper considers a model of infectious disease, such as swine flu, in which privately costly treatment confers immunity on recovered individuals. It is shown that under decentralized decision making, infected individuals ignore the externality that their treatment has on susceptible individuals and thus seek treatment only if it is privately optimal to do so. In contrast, a benevolent central planner who does take this externality into account in choosing the level of aggregate treatment, may choose to either eradicate the disease or to retard its eventual dissemination into the population even when individuals would not find it privately optimal to do so. The analysis shows that when immunity from future infection is obtained through recovery, treatment resembles vaccination in its effects on infection dynamics, but important differences remain. Vaccination is shown to more effectively curb infection than does treatment. Last, the inefficiency associated with decentralized decision making can be corrected through subsidized treatment offered on a first-come first-served basis.

JEL Classification: C73, I18.

Keywords: Economic epidemiology, treatment, acquired immunity, externalities.

1. Introduction

In early July 2009, the UK Department of Health announced that in its battle against the swine flu pandemic, it had now entered a "treatment phase" under which treatment was to be the main policy instrument in controlling the outbreak of the disease. It stated that[1]

> "As swine flu spreads and more people start to catch it, it makes sense to move from intensive efforts to contain the virus to focusing efforts on treating the increasing number of people who have the disease."

*I gratefully acknowledge very helpful feedback from Joshua Ross, Chryssi Giannitsarou and seminar participants at the University of Cambridge.

†Faculty of Economics, University of Cambridge and CEPR. Address for correspondence: Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. Phone: +44 (0) 1223 335259; Fax: +44 (0) 1223 335475; Email: fmot2@cam.ac.uk; Web: http://people.pwf.cam.ac.uk/fmot2/.

[1]UK Department of Health (2009).

At the time of this announcement, a vaccine against infection was not yet widely available. Given that the number of new cases of infection was estimated to double every seven days[2], this raises the central question of what can be achieved through a policy that solely focuses on treatment of the disease, when recovery induces immunity against further infection. More generally, it raises the question of the exact nature of the external effects that individual treatment efforts have on the population at large. Surprisingly, this is an open question. This paper offers an answer based on a simple economic epidemiology model. Specifically, it considers the extent to which acquired immunity through treatment can be usefully employed as a policy tool to control infectious disease. It turns out that when treatment induces immunity after recovery, it has features in common with vaccination in that it increases the proportion of the population which is immune to infection. Importantly though, while vaccination works by shielding susceptible individuals from infection (and hence only indirectly influences disease prevalence), treatment directly reduces the proportion of infected individuals (and indirectly influences disease incidence).

It is shown that in a setup with decentralized decision making, each infected individual ignores the effects that treatment decisions have on susceptible individuals. This means that an individual decision maker's optimal policy is particularly simple, prescribing treatment if the cost is outweighed by the expected discounted net benefits of recovery. Importantly, this benefit is wholly independent of other individuals' decisions, so there is no strategic interaction.

With centralized decision making, a benevolent social planner directly chooses aggregate treatment levels with a view to maximize aggregate discounted expected welfare. In doing so, the planner explicitly accounts for the externalities that infected individuals' treatment decisions have on susceptible individuals. As a consequence, the planner values treatment more than do the individuals and hence will mandate treatment even when individuals would not choose any under decentralized decision making.

Under centralization, the optimal policy can be characterized by two distinct regimes (which depend on parameter values and initial conditions). In the first, the planner chooses to treat all and every infected individual and continues to do so in perpetuity. This means that the disease is eventually eradicated, leaving only recovered (and immune) individuals and susceptible individuals who are protected by herd immunity.

In the second regime, the planner initially treats all infected individuals and thereby slows down the spread of the disease for a period of time. Eventually, a point is reached

[2]UK Department of Health (2009).

after which the external benefits are too low to make costly treatment worthwhile and hence the planner ceases to treat infected individuals.

In the special case where recovery is only possible through treatment, the limiting distribution of health states across the population is such that whoever is not recovered is infected and remains so forever. If on the other hand there is a positive background rate of recovery (over and above that achieved through treatment), then in the limit the infection dies out and the population is composed of recovered and susceptible individuals only, as was the case in the regime with full eradication through treatment. An important difference is that in this scenario, fewer individuals remain susceptible than is the case in the eradication regime.[3]

It turns out that for sufficiently low treatment costs, decentralized decision making leads to the first-best optimal outcome, i.e. to the treatment decisions that a central planner would choose. This is despite the fact that individuals wholly disregard the positive externalities that their treatment has on the population as a whole. For higher levels of the treatment cost, private benefits from recovery do not by themselves justify treatment from a private perspective. But the social planner, who explicitly factors in the external benefits of treatment when deciding on aggregate treatment levels, may still mandate treatment even when the decentralized decision makers would opt for none. For such levels of treatment costs, private and social objectives do not coincide and thus decentralized equilibrium choices are inefficient.

There is a very simple policy that can remedy the inefficiency in private choices. A straightforward subsidy to treatment can fully align private and social objectives and thereby achieve the first-best outcome even under decentralized decision making.

Since the pioneering work of Bernoulli (1766), the formal modeling of infectious disease dynamics has formed an integral part of research on the evolution and control of epidemics and has contributed important insights that continue to inform public policy to the present day. The bulk of this research has focused on different protective measures such as vaccines, prophylaxis, isolation, quarantines and reductions in the rate of partner change. A smaller literature has considered the effects of treatment.

This paper sits in between the literatures on treatment and vaccination. The literature on treatment includes contributions by Sanders (1971), Sethi (1974), Sethi and Staats (1978) Goldman and Lightwood (1995, 2002), Rowthorn (2006) and Toxvaerd (2009). In all these analyses, treatment increases the rate of recovery, but individuals do not acquire immunity and thus make a transition back to susceptibility as in the classical susceptible-infected-susceptible (SIS) model. In these models, treatment therefore works by increasing the measure of susceptibles whereas in the present analy-

[3] That is, more individuals experience infection and eventual recovery and immunity.

sis, treatment works by increasing the measure of recovered individuals. Although this paper considers the effects of treatment, it has features in common with the literature on immunization through vaccines in economic frameworks, such as Barrett (2003), Barrett and Hoel (2007), Bauch (2005), Bauch and Earn (2004), Boulier et al. (2007) and Brito et al. (1991).[4] As emphasized above, there are important differences between the present analysis and those contained in the vaccination literature, as shall become clear in what follows. This is because vaccination draws individuals from susceptibility into immunity whereas treatment draws them from infectiousness into susceptibility, a difference that has interesting consequences.

The remainder of the paper is structured as follows: In Section 2, I set out the classical and economic versions of the susceptible-infected-recovered model. In Section 3, I solve the model under decentralized decision making while in Section 4, I characterize the optimal policy under centralized decision making. In Section 6, I compare the performance of treatment in reducing infection to that obtained through vaccination. Section 6 concludes.

2. The Model

The classical *susceptible-infected-recovered* (or SIR) model is simple to describe.[5] Time is continuous and runs indefinitely. A population $\mathcal{P} = [0, 1]$ consists of a continuum of infinitely lived individuals who can at each instant t each be in one of three states, namely *susceptible* or *infected* or *recovered*. The set of susceptible individuals is denoted by $\mathcal{S}(t)$ and has measure $S(t)$, the set of infected individuals is denoted by $\mathcal{I}(t)$ and has measure $I(t)$ and the set of recovered individuals is denoted by $\mathcal{R}(t)$ and has measure $R(t)$. Because the population size has been normalized to unity, these measures can be interpreted as fractions. Henceforth, $I(t)$ shall be referred to as *disease prevalence*.

At each instant, the population mixes homogeneously. This corresponds to pairwise random matching where each individual has an equal chance of meeting any other individual, irrespective of the health status of the two matched individuals. A match between an infected and a susceptible individual may infect the susceptible. The rate at which infection is transferred in such a match is denoted by $\beta > 0$. This parameter captures the infectivity of the disease. Recovered individuals are immune to further infection and also cannot carry the disease. Coupled with the assumption of homogeneous mixing, this means that the rate at which susceptible individuals become

[4] Early non-economic contributions include Anderson and May (1992) and Smith (1964).

[5] This model is also known as that of a general epidemic. See Kermack and McKendrick (1927) for the original treatment. For more recent expositions, see Anderson and May (1991), Daley and Gani (2001) and Keeling and Rohani (2008).

infected is given by the simple expression $\beta I(t)S(t)$.[6] This means that the rate of new infection, or *disease incidence*, is proportional to disease prevalence.[7]

Last, in the classical version of the model, individuals spontaneously recover at rate $\gamma \geq 0$. This means that on aggregate, the rate at which recovery occurs is $\gamma I(t)$.

It should be noted that even though each individual is subject to uncertainty (through the random matching that occurs at each instant and through the randomly evolving disease state of the individual), there is no aggregate uncertainty. That is, the population-wide distribution across health states and the evolution of this distribution is deterministic. This is true both in the centralized and in the decentralized versions of the model (in the latter case, for a given strategy profile of the decision makers).

For later use, I briefly analyze the classical SIR model. The dynamic system is described by the following equations:

$$\begin{aligned}
\dot{S}(t) &= -\beta I(t)S(t) && (1)\\
\dot{I}(t) &= I(t)\left[\beta S(t) - \gamma\right] && (2)\\
\dot{R}(t) &= \gamma I(t) && (3)\\
S(t) &= 1 - I(t) - R(t) && (4)\\
S(0) &= S_0 > \gamma/\beta, \quad I(0) = I_0, \quad S_0 + I_0 = 1 && (5)
\end{aligned}$$

The restriction that $S_0 > \gamma/\beta$ ensures that the epidemic can take hold in the population. With this assumption in place, the overall behavior of the system can be described as follows. The measure of susceptible individuals $S(t)$ decreases over time while the measure of recovered individuals increases over time. In contrast, the measure of infected individuals initially increases, peaks at $S(t) = \gamma/\beta$ and then tends to zero.

The SIR model cannot be fully characterized analytically. Nevertheless, the limiting distribution of health states can be characterized, which shall prove useful in the analysis of the economic model below.

Well-known steps lead to the central result that the final epidemic size is characterized by the equations

$$S(\infty) = 1 - R(\infty) = S(0)\exp\left(-R(\infty)\mathfrak{R}_0\right) \geq 0 \qquad (6)$$

[6] The term $\beta I(t)S(t)$ should be thought of as the rate at which susceptible individuals have contact with other individuals, multiplied by the probability of the contact being with an infectious individual, multiplied by the probability that the infection is transmitted in such a contact. See e.g. Keeling and Rohani (2008) for a detailed derivation.

[7] Note that while disease incidence is a flow value, disease prevalence is a stock value.

where $\mathfrak{R}_0 \equiv \beta/\gamma$ is the basic rate of reproduction.[8]

The basic rate of reproduction represents how many secondary infections are caused by the insertion of a single infected individual into a fully susceptible population. The second equation defines $R(\infty)$ implicitly and the first defines $S(\infty)$ as the residual, which is possible since $I(\infty) = 0$. The limiting proportions $S(\infty)$ and $R(\infty)$ are easily found for particular parameterization of the model.

There are two important insights that follow from this equation. First, in the limit the disease must die out in the sense that no infected individuals remain.[9] Second, and more importantly, when the disease dies out, there is generically a positive measure of susceptibles remaining in the population. This shows that what causes the disease to die out is not that there is eventually a lack of susceptibles that can be infected. Rather, it dies out because the measure of recovered individuals, which must grow over time, becomes so large that the contact between infected and susceptible individuals becomes too rare for the infection to be passed on. Infected individuals have increasingly long sequences of matches with recovered individuals and so, on expectation, will recover before having the opportunity to pass on the infection to a susceptible individual. The remaining susceptible individuals are said to be protected by *herd* (or *population*) immunity.

Note the central role played by the basic rate of reproduction. If $\mathfrak{R}_0 < 1$, then infection cannot take hold while if $\mathfrak{R}_0 > 1$, then infection first flares up and then tapers off. As will become clear in what follows, the optimal (centralized) control of the epidemic through treatment will work by modifying the magnitude of the basic rate $\mathfrak{R}_0$.

Having outlined the classical version of the SIR dynamic system, I now make a number of additions in order to turn it into an economically meaningful model. Since protective behavior is disregarded in the present analysis, the rate of infection cannot be directly influenced. Instead, I assume that the rate at which agents recover (and become immune to further infection) can be influenced through costly treatment. In particular, for some treatment intensity $\tau(t) \in [0, 1]$, the rate at which the individual transitions from $\mathcal{I}(t)$ to $\mathcal{R}(t)$ is given by $\tau(t)\alpha + \gamma$, where $\alpha > 0$ is interpreted as the efficiency of the treatment. This means that treatment increases the rate of recovery over and above the background rate γ. The treatment costs $c > 0$ per instant. Last, the individuals in the sets $\mathcal{S}(t)$, $\mathcal{I}(t)$ and $\mathcal{R}(t)$ earn flow payoffs $\pi_{\mathcal{S}}$, $\pi_{\mathcal{I}}$ and $\pi_{\mathcal{R}}$ respectively and discount the future at rate $\rho > 0$. It will be assumed that $\pi_{\mathcal{S}} \geq \pi_{\mathcal{I}}$ and $\pi_{\mathcal{R}} \geq \pi_{\mathcal{I}}$.[10]

[8] See Daley and Gani (2001) for details.

[9] Modulo atto fox.

[10] While it is natural to suppose that $\pi_{\mathcal{S}} \geq \pi_{\mathcal{R}}$, this assumption is not needed in the following analysis.

The economic version of the model inherits a number of simplifying assumptions from the classical model. First, there is only one disease and one level (or severity) of infection.[11] In particular, this rules out the possibility of superinfection by different strains of the disease. Second, the moment an individual is infected coincides with the onset of symptoms such as the welfare loss brought about by infection (i.e. the *incubation period* has zero length), so no infected individual acts under the mistaken belief that he or she is susceptible. Last, once an individual becomes infected, he or she immediately becomes infectious to other individuals (i.e. the *latency period* has zero length). Relaxing any of these assumptions constitute possible extensions of the present work.

3. DECENTRALIZED DECISION MAKING

Consider an individual's problem. For any fixed treating intensity $\tau(t)$, the health state of the individual follows a three-state continuous-time Markov process. Fortuitously, the actual problem to be solved by an individual can be considerably simplified by noting that in two of these states, the optimal choice is trivial. Since treatment is costly, it is trivially optimal for a susceptible or recovered individual to seek no treatment at all. The problem is therefore reduced to determining the optimal policy for an infected individual. Without loss of generality, consider an individual who is infected at $t = 0$. Since susceptibility is not feasible for this individual, all he is concerned with is the possible transition from the infected to the recovered state. Let $Q(t)$ denote the probability that the individual is *still* in the infected state at time $t \geq 0$. He then solves the following problem:

$$\max_{\tau_i(t)\in[0,1]} \int_0^\infty e^{-\rho t}[Q(t)\left(\pi_{\mathcal{I}} - \tau_i(t)c\right) + (1 - Q(t))\pi_{\mathcal{R}}]dt \tag{7}$$

$$s.t. \quad \dot{Q}(t) = -Q(t)\left[\alpha\tau_i(t) + \gamma\right], \quad Q(0) = 1 \tag{8}$$

This problem is equivalent to the following simplified problem, which differs only by the constant $\pi_{\mathcal{R}}$:

$$\max_{\tau_i(t)\in[0,1]} \int_0^\infty e^{-\rho t}Q(t)[\pi_{\mathcal{I}} - \pi_{\mathcal{R}} - \tau_i(t)c]dt \tag{9}$$

$$s.t. \quad \dot{Q}(t) = -Q(t)\left[\alpha\tau_i(t) + \gamma\right], \quad Q(0) = 1 \tag{10}$$

This objective is simply the expected, discounted utility for an individual pursuing treatment strategy $\tau_i(t)$. Note that in steady state, $\dot{Q}(t) = 0$ and so it must be that

[11] Thus the sets $\mathcal{S}(t)$ and $\mathcal{I}(t)$ are disjoint and exhaust $\mathcal{P}$.

$Q^* = 0$ eventually.

The associated current-value Hamiltonian for this problem is then[12]

$$H^D = Q(t)[\pi_{\mathcal{I}} - \pi_{\mathcal{R}} - \tau_i(t)c] - \vartheta(t)Q(t)\left[\alpha\tau_i(t) + \gamma\right] \tag{11}$$

where $\vartheta(t)$ is the multiplier. Differentiating the current value Hamiltonian with respect to the treatment rate $\tau_i(t)$ yields the following necessary condition for optimality (supposing that $Q(t) > 0$):

$$c + \vartheta(t)\alpha = 0 \tag{12}$$

The evolution of the multiplier is given by the following differential equation:

$$\dot{\vartheta}(t) = \vartheta(t)\left[\rho + \alpha\tau_i(t) + \gamma\right] + \left[\pi_{\mathcal{R}} - \pi_{\mathcal{I}} + \tau_i(t)c\right] \tag{13}$$

Setting $\dot{\vartheta}(t) = 0$ yields the steady state level of the multiplier:

$$\vartheta(t) = \frac{\pi_{\mathcal{I}} - \pi_{\mathcal{R}} - \tau_i(t)c}{\rho + \alpha\tau_i(t) + \gamma} \tag{14}$$

Substituting this in the optimality condition yields the following optimal bang-bang policy:

$$\tau_i(t) = 0 \quad for \quad c\,(\rho + \gamma) > \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) \tag{15}$$

$$\tau_i(t) \in [0,1] \quad for \quad c\,(\rho + \gamma) = \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) \tag{16}$$

$$\tau_i(t) = 1 \quad for \quad c\,(\rho + \gamma) < \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) \tag{17}$$

The optimal policy simply states that treatment is sought only if the expected discounted benefit (to the individual) is larger than the cost of treatment. If the benefit is large enough, then all infected individuals will always seek full treatment and the model reduces to the classical SIR model (but with an increased recovery rate). If the benefit is not large enough, then no infected individual will ever seek any treatment. The model then reduces to a classical SI model (for *susceptible-infected*, aka a *simple epidemic*) if the background rate of recovery $\gamma = 0$ and to the classical SIR model (aka a *general epidemic*) if $\gamma > 0$. In the SI case, all individuals eventually become infected and remain so indefinitely since they never seek treatment.[13] In the SIR cases, things become more complicated due to the possibility of herd immunity, as described in the model section.

[12] In the individual's problem, an admissible pair of functions $(Q(t), \tau_i(t))$ is such that for all $t \geq 0$, $Q(t)$ satisfies the differential equation for the state variable $Q(t)$ and $\tau_i(t) \in [0,1]$.

[13] See e.g. Daley and Gani (2001) for details of the SI model.

These finding are summarized as follows:

Theorem: Under decentralized decision making: (i) if $c\,(\rho+\gamma) > \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})$ then the equilibrium outcome coincides with that of the general epidemic if $\gamma > 0$ and with that of the simple epidemic if $\gamma = 0$; (ii) if $c\,(\rho+\gamma) < \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})$ then the equilibrium outcome coincides with that of the general epidemic with recovery rate $\alpha + \gamma$.

The comparative statics of the optimal decentralized policy are straightforward. The higher the discount rate ρ, the recovery rate γ or the treatment cost c, the less attractive does treatment become. Conversely, treatment becomes more attractive the higher the efficiency of the treatment α or the higher the health premium $(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})$.

It is interesting to note that the treatment decision is not strategic. This is because while infected individuals' treatment decisions do influence the prospects of the susceptibles, this influence is ignored since there is no feedback. Therefore infected individuals seek treatment if and only if doing so is privately worthwhile, a decision that is not influenced by other infected individuals' treatment decisions. Formally, the lack of strategic interaction follows from the absence of disease prevalence $I(t)$ in the individual's maximization problem.

In the case where no individual ever seeks treatment, the evolution of disease prevalence is given by the logistic growth equation

$$I(t) = \frac{\beta I_0}{e^{-t\beta}\beta + (1 - e^{-t\beta})\,\beta I_0} \tag{18}$$

In this setting, the equilibrium outcome has $\lim_{t\to\infty} I(t) = 1 - R_0$.[14]

It is notable that under decentralized decision making, each individual's problem is wholly independent of the aggregate evolution of the epidemic. Disease prevalence $I(t)$ only influences susceptible individuals and not infected or recovered individuals. But the only ones that can actually influence disease prevalence, through the evolution of disease incidence, are the infected individuals (collectively). But they have no direct incentive to do so. This observation is the key difference between the outcomes under centralized and decentralized decision making.

4. CENTRALIZED DECISION MAKING

The problem of the central planner is given as follows:

$$\max_{r(t)\in[0,1]} \int_0^\infty e^{-\rho t}[S(t)\pi_{\mathcal{S}} + I(t)\,(\pi_{\mathcal{I}} - r(t)c) + R(t)\pi_{\mathcal{R}}]dt \tag{19}$$

[14]Note that $R_0 = R(0)$ is the initial condition for the measure of recovered individuals and should not be confused with the basic rate of reproduction $\mathfrak{R}_0$.

The problem is solved subject to the following laws of motion for the measures of susceptible, infected and recovered individuals:

$$\begin{aligned}
\dot{S}(t) &= -\beta I(t)S(t) && (20)\\
\dot{I}(t) &= I(t)\left[\beta S(t) - \alpha r(t) - \gamma\right] && (21)\\
\dot{R}(t) &= I(t)\left[\alpha r(t) + \gamma\right] && (22)\\
S(t) &= 1 - I(t) - R(t) && (23)\\
S(0) &= S_0 > \gamma/\beta, \quad I(0) = I_0, \quad S_0 + I_0 = 1 && (24)
\end{aligned}$$

The problem solved by the central planner is similar to that of the decentralized individuals, but there are some notable differences. First, the planner aggregates the welfare of all individuals into its objective function. Second, the constraints take into account the fact that the planner directly controls the evolution of the aggregate variables through its choice of aggregate treatment. Therefore the fractions $S(t)$, $I(t)$ and $R(t)$ are endogenous for the planner whereas they are exogenous for any one individual.

In considering the overall effects of treatment, note the following useful analogy. Since recovery confers immunity on the (previously infected) individual, treatment may be interpreted as a kind of immunization at the aggregate level. Immunization transfers susceptible individuals directly into the recovered class; therefore it dilutes the effects of infection since the rate of contact between infected and susceptible individuals is reduced.[15] Treatment has a similar effect by transferring individuals from $\mathcal{I}(t)$ to $\mathcal{R}(t)$ rather than from $\mathcal{S}(t)$ to $\mathcal{R}(t)$ as is the case with immunization. Note however that from the perspective of the particular individual, treatment and immunization are quite different in that the former presupposes that the individual have a spell of infection while the latter does not.

Using the normalization to eliminate $S(t)$, the planner's current value Hamiltonian is given by

$$\begin{aligned}
H^C &= \left[1 - I(t) - R(t)\right]\pi_{\mathcal{S}} + I(t)\left(\pi_{\mathcal{I}} - r(t)c\right) + R(t)\pi_{\mathcal{R}} \\
&\quad + \lambda(t)I(t)\left[\beta\left(1 - I(t) - R(t)\right) - \alpha r(t) - \gamma\right] + \mu(t)I(t)\left[\alpha r(t) + \gamma\right] \qquad (25)
\end{aligned}$$

Note that $\lambda(t)$ and $\mu(t)$ are the costate variables associated with the laws of motion for infected and recovered individuals respectively.

Differentiating with respect to the treatment rate $r(t)$ yields the following necessary

[15] This is simply because a number of infected and/or susceptible individuals are matched with recovered individuals instead of each other.

condition for optimality (assuming that $I(t) > 0$):

$$c + \alpha\left(\lambda(t) - \mu(t)\right) = 0 \tag{26}$$

The evolution of the multipliers is governed by the following system of differential equations:

$$\begin{aligned}
\dot{\lambda}(t) &= \lambda(t)\left[\rho + \alpha r(t) + \gamma + \beta\left(2I(t) + R(t) - 1\right)\right] \\
&\quad -\mu(t)\left[\alpha r(t) + \gamma\right] - \left(\pi_{\mathcal{I}} - r(t)c\right) && (27)\\
\dot{\mu}(t) &= \rho\mu(t) + \lambda(t)\beta I(t) - \left(\pi_{\mathcal{R}} - \pi_{\mathcal{S}}\right) && (28)
\end{aligned}$$

Setting $\dot{\lambda}(t) = \dot{\mu}(t) = 0$, the system can be solved to yield the steady state pair $(\lambda(t), \mu(t))$.
The optimal policy is of the bang-bang type and given by

$$\begin{aligned}
r(t) &= 0 \quad for \quad c > \alpha(\mu(t) - \lambda(t)) && (29)\\
r(t) &\in [0,1] \quad for \quad c = \alpha(\mu(t) - \lambda(t)) && (30)\\
r(t) &= 1 \quad for \quad c < \alpha(\mu(t) - \lambda(t)) && (31)
\end{aligned}$$

This policy has a nice interpretation. Increasing the treatment rate has two effects, namely to increase the measure of recovered individuals and to reduce the measure of infected individuals. The marginal benefit of reducing the measure of infectives is $-\lambda(t) \geq 0$ while the marginal benefit of increasing the measure of recovered individuals is $\mu(t) \geq 0$. Thus the expression $\alpha(\mu(t) - \lambda(t))$ is simply the rate at which the total benefits of treatment accrue.

Observe that the source of the externality in this model is the effect that infected individuals have on the susceptible individuals. Specifically, an infected individual's failure to treat himself makes him a source of infection for the susceptible part of the population. To see this clearly, set $R(t) = 1 - I(t)$ and solve the system of steady state multipliers. Straightforward substitution then yields the following characterization of the optimal policy:

$$\begin{aligned}
r(t) &= 0 \quad for \quad c\left(\rho + \gamma\right) > \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) && (32)\\
r(t) &\in [0,1] \quad for \quad c\left(\rho + \gamma\right) = \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) && (33)\\
r(t) &= 1 \quad for \quad c\left(\rho + \gamma\right) < \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}}) && (34)
\end{aligned}$$

Thus when $S(t) = 0$, the central planner's problem coincides with that of the individ-

uals under decentralized decision making.[16]

The planner's objective can be decomposed into the terms $I(t)\left(\pi_{\mathcal{I}} - r(t)c\right) + R(t)\pi_{\mathcal{R}}$ and $S(t)\pi_{\mathcal{S}}$ respectively. The former term is the aggregate welfare of infected and recovered individuals and equals, on a per capita basis, the welfare that counts for individuals under decentralized decision-making. That is, in decision whether to treat themselves, infected individuals only consider these two possible welfare states. The latter term accounts for the welfare of the susceptible population and constitutes the source of externalities in this model. Since $\dot{S}(t) = -\beta I(t)S(t) < 0$, the term $S(t)\pi_{\mathcal{S}}$ can be usefully thought of as a decaying exhaustible resource. While the decay is unavoidable, the rate at which it occurs can be decreased through costly treatment.

Because of the positive externality that treatment has on susceptible individuals, the central planner always values treatment at least as much as individuals do in the decentralized setting. As a consequence, when treatment is privately optimal, i.e. when $c\left(\rho + \gamma\right) < \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})$, the decentralized equilibrium outcome coincides with that chosen by the central planner. The interesting setting is therefore the one in which treatment is socially but not privately optimal. For this reason, I impose the following conditions:

Assumption 1 $c > \alpha(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})/\left(\rho + \gamma\right)$.

Assumption 2 $c < \alpha(\mu(0) - \lambda(0))$ at $r(0) = 1$ and initial conditions $(S_0, I_0, 0)$.

Assumption 1 ensures that in the absence of externalities, i.e. at $S(t) = 0$, it is optimal not to seek any treatment. Assumption 2 ensures that treatment is optimal at the initial conditions $(S(0), I(0), R(0)) = (S_0, I_0, 0)$. If Assumption 1 is violated then full and perpetual treatment is trivially optimal whatever the state of the system, whereas if Assumption 2 is violated, then no treatment can ever be optimal.

To proceed with the characterization of the optimal policy, the following simple result shall prove useful:

Lemma: There is a unique critical measure of susceptibles

$$S^* \equiv \{S(t) \in [0,1] : \alpha(\mu(t) - \lambda(t)) = c \quad at \quad \{r(t) = 1, \quad \forall t \geq 0\}\} \tag{35}$$

at which the net benefit from treatment is zero.

Proof: Note that the net benefit from treatment

$$\alpha(\mu(t) - \lambda(t)) - c \tag{36}$$

[16] The function $(\mu(t) - \lambda(t))$ obtains its minimum $(\pi_{\mathcal{R}} - \pi_{\mathcal{I}})$ at $S(t) = 0$.

is increasing in the measure of susceptible individuals $S(t)$. Furthermore, $S(t)$ is weakly decreasing over time irrespective of the chosen (constant) policy $r(t)$. This follows from the fact that the evolution of the fraction of susceptible individuals can be expressed as

$$S(t) = S(0) \exp\left(\frac{-R(t)\beta}{r(t)\alpha + \gamma}\right) \tag{37}$$

Last, Assumptions 1 and 2 ensure that the net benefit from treatment is positive at $S(t) = 1$ and negative at $S(t) = 0$. The result then follows from continuity of $S(t)$ ■

Given this lemma, the following result follows immediately from the monotonicity of $S(t)$:

Theorem: The optimal policy under centralized decision making is given by

$$r(t) = 1 \quad if \quad S(t) > S^* \tag{38}$$

$$r(t) \in [0,1] \quad if \quad S(t) = S^* \tag{39}$$

$$r(t) = 0 \quad if \quad S(t) < S^* \tag{40}$$

This result succinctly characterizes the optimal policy in terms of the remaining measure of susceptible individuals in the population. As long as sufficiently many susceptible individuals remain, the optimal policy prescribes full treatment of all infected individuals. When the measure of susceptibles falls below a critical threshold, the optimal policy is to cease treatment entirely.

The next step is to determine if this critical threshold is reached or not. Specifically, the limit of the measure of susceptibles $S(t)$ must be found and then compared to the threshold S^*. To this end, recall that monotonicity and continuity of $S(t) \in [0,1]$ implies that the measure of susceptible individuals must converge to some limit $\lim_{t\to\infty} S(t)$ under any (constant) policy $r(t)$. From the analysis of the classical SIR model, the final measure of susceptible individuals can be described by the equations

$$\widehat{S} \equiv \lim_{t\to\infty} S(t) = 1 - R(\infty) = S(0) \exp\left(\frac{-R(\infty)\beta}{\gamma + \alpha}\right) \tag{41}$$

This limit is taken under the assumption that the policy of full treatment $r(t) = 1$ is pursued in perpetuity (i.e. under the policy $r(t) = 1, \quad \forall t \geq 0$). These equations are a straightforward modification of the corresponding classical equations.

Note that the limit $\widehat{S}$ is decreasing in the basic rate of reproduction $\beta/(\gamma + \alpha)$. This means that the fraction of the population that escapes infection decreases in the

infectiousness of the disease and increases in either the spontaneous rate of recovery or the efficiency of the treatment.

Using the modified final epidemic size equations, the different possible outcomes under centralized decision making can be classified as follows:

Full Eradication Regime. This regime corresponds to the case where

$$\widehat{S} > S^* \tag{42}$$

In this case, full treatment in perpetuity is optimal. Treatment is sufficiently effective in bringing down infection to make herd immunity take effect while it is still the case that significant positive external effects from treatment remain.

The limiting distribution has some individuals recovered (and immune) and the remaining individuals susceptible (and protected by herd immunity). Infection is thus fully eradicated.

Stemming the Tide Regime. This regime corresponds to the case where

$$\widehat{S} \leq S^* \tag{43}$$

In this case, the central planner starts by fully treating all infected individuals and continues to do so until the measure of susceptibles falls below the critical threshold. At this point, the planner ceases to treat infected individuals and lets the infection take its toll. Since $S(t)$ is monotone, there is a unique critical time t^* at which the optimal policy switches from full treatment to no treatment.

In this scenario, even though the planner initially fully treats all infected individuals, the external benefits from treatment erode so fast that herd immunity does not take effect before treatment becomes obsolete.

In this regime, the limiting distribution depends crucially on the background rate of recovery γ. If $\gamma = 0$ then in the limit some individuals have recovered and the remaining individuals are infected and remain so in perpetuity. There will thus be no susceptible individuals left, as is the case in a model of a simple epidemic. If $\gamma > 0$, then the limiting distribution is that of the standard SIR model with recovery rate γ, initialized at $(S(t^*), I(t^*), R(t^*))$. I.e. the disease is eradicated and some susceptibles may remain.

Figure 1 illustrates the optimal treatment policy and the different possible regimes. The curve connecting the points X and Z shows the marginal social benefit of treatment as a function of the fraction of susceptibles in the population $S(t)$. This function is upward-sloping in the measure of susceptibles because they are the ones who benefit

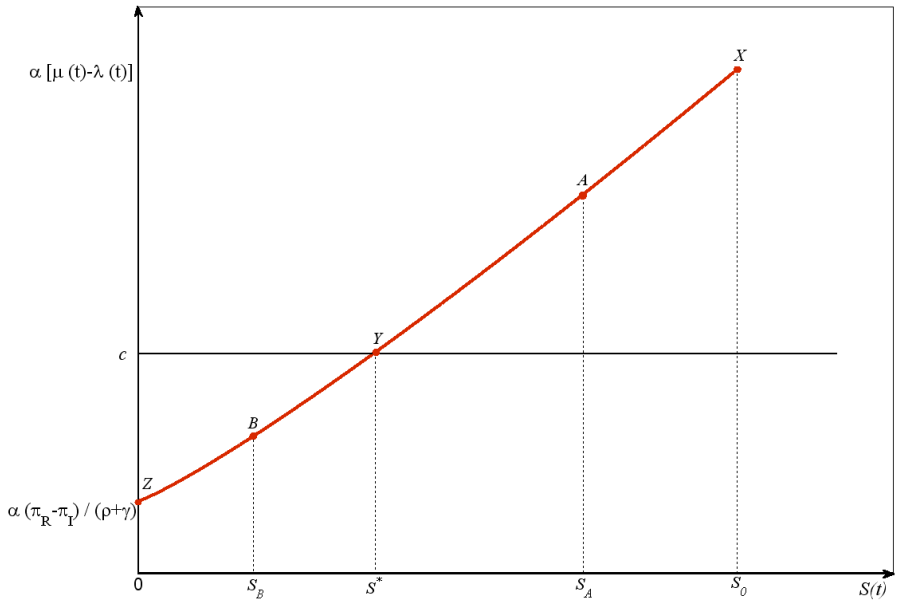

Figure 1: Treatment Regimes under Centralized Decision Making.

from the external effects of treatment. For initial value $S(0) = S_0$, corresponding to point X, the marginal benefit is larger than the marginal cost c (represented by the horizontal line). This is ensured by Assumption 2. Over time, the fraction of susceptibles must decrease (weakly) and so one moves leftward in the diagram. At point Z, which corresponds to the extreme case in which no susceptibles remain, there are no externalities from treatment and so private and social objectives coincide. At this point, the marginal cost of treatment outweighs the marginal benefit. This is ensured by Assumption 1. To summarize, on points between X and Y, treatment is socially optimal while on points between Y and Z it is socially optimal not to treat.

Turning to the two treatment regimes, consider the evolution of the measure of susceptibles under the policy of full and perpetual treatment of infected individuals (as long as any remain). Under this policy, the measure of susceptibles decreases over time and converges to some level $\widehat{S}$. The magnitude of the limit $\widehat{S}$ is determined by the epidemiological parameters α and β, the treatment efficiency γ and by initial conditions. The two regimes delineating the optimal policy simply corresponds to the location of the limit point $\widehat{S}$ relative to the critical threshold S^* (corresponding to point Y on the curve).

If $\widehat{S} = S_A$ and so movement along the curve ends at point A, then the policy of full treatment of all infected individuals succeeds in eradicating the disease early

enough for significant external effects to remain (and thus justify full treatment). If $\widehat{S} = S_B$ and so movement along the curve ends at point B, then even if full treatment is pursued, the measure of susceptibles eventually becomes so low that the external effects of treatment no longer justify the associated costs. In this regime, full treatment is therefore pursued initially, until point Y is reached. After this point, the planner ceases all treatment and there is a discontinuous decrease in the rate of recovery. Still, the optimal policy initially restricts disease incidence and thus prevalence.

Last, it should be noted that the comparative statics of the limit point $\widehat{S}$ are straightforward, as indicated above. Determining the comparative statics of the optimal policy is more delicate, as the critical threshold S^* depends on the economic parameters (c, $\pi_{\mathcal{S}}$, $\pi_{\mathcal{I}}$, $\pi_{\mathcal{R}}$ and ρ) and the epidemiological parameters (α, β and γ) in a complicated way through the multipliers $\mu(t)$ and $\lambda(t)$. In other words, when the parameters change the whole curve shifts around and hence so does the intersection point Y. This analysis cannot be done analytically and numerical analysis must be performed to further determine the effects of changes in these parameters.

5. Efficiency of Treatment as Immunization

As mentioned above, since recovery induces immunity, treatment may be viewed as a sort of vaccination. While it is true that the two interventions have similar effects, they differ in important respects. First, vaccination works by changing only disease incidence, and only indirectly by changing disease prevalence. In contrast, treatment directly influences disease prevalence and indirectly also disease incidence.

Second, vaccination is in some sense more efficient in reducing infection than recovery induced immunity. To see this, recall that if a proportion $p \in [0,1]$ of the population is successfully immunized, then the basic rate of reproduction is changed from $\mathfrak{R}_0 = \beta/\gamma$ to

$$\mathfrak{R}_0^V \equiv \frac{(1-p)\beta}{\gamma} \tag{44}$$

Similarly, treatment leading to immunity changes the basic rate of reproduction to

$$\mathfrak{R}_0^T \equiv \frac{\beta}{r(t)\alpha + \gamma} \tag{45}$$

With treatment, the rate of reproduction is minimized at $r(t) = 1$, at which level the disease may still spread. With immunization through vaccine, the rate of reproduction can be made arbitrarily small, entirely and instantly eradicating the disease at $p = 1$.

It is difficult to characterize the differences much further for a number of reasons. First, no mention of costs have been made. The way treatment has been modeled in the present analysis implies that it may have to be sustained for some time before recovery

is induced, while vaccination is modeled as a one-off measure. This difference may have implications for the cost-effectiveness of one or the other intervention. Specifically, it is difficult to compare the costs and benefits of the two interventions.[17] Second, even if vaccination were found to lead to a lower rate of reproduction, it should be recalled that vaccination does not directly alter disease prevalence whereas treatment does. In other words, treatment both reduces the measure of infective individuals and reduces the rate at which the disease propagates. In assessing the benefits of treatment, both these effects should be considered and quantified.

6. Conclusion

In this paper, I have considered an economic epidemiology version of the classical susceptible-infected-recovered model in which costly treatment may increase the rate of recover and confer immunity from future infection on the recovered individual. I found that in equilibrium, individuals adopt socially suboptimal treatment policies, leading to too little treatment and recovery. This is because decentralized and non-cooperative individuals disregard the socially beneficial external effects that treatment and recovery have on susceptible individuals (through their dampening effect on disease incidence). I show that depending on initial conditions and parameter values, the socially optimal policy may either involve full eradication of the disease through mass treatment of the infected part of the population, or treatment for a limited duration of time followed by complete cessation of treatment measures. In the latter case, if recovery can only be acquired through treatment, infection will be endemic.

In terms of public policy, there are two lessons to be drawn. First, treatment decisions should not necessarily be decentralized as individual agents may well find it privately optimal to avoid treatment even when society as a whole would benefit from increased treatment. Second, there is fortuitously a very simple way that the inefficiency in decentralized treatment decisions can be corrected. Namely, a simple subsidy to treatment will suffice to align private and public incentives to seek treatment. The only twist is that subsidies should be offered only until (and if) the critical level of susceptibles is reached, since after that point treatment is no longer socially desirable.

Another straightforward, if more controversial, policy conclusion is that in order to align private incentives with those of the general public, sanctions may be imposed on infected individuals. In incentive terms, this type of policy mirrors that of subsidizing treatment. In practice, sanctions seem only to be imposed on those individuals who refuse treatment.[18]

[17] One could compare the cost of a vaccine dose to the expected cost of treatment.

[18] See article *Some TB Patients Could be Forcibly Quarantined*, Associated Press, January 22, 2007.

REFERENCES

[1] ANDERSON, R. M. AND R. M. MAY (1991): Infectious Diseases of Humans: Dynamics and Control, *Oxford University Press.*

[2] BARRETT, S. (2003): Global Disease Eradication, *Journal of the European Economic Association*, 1(2-3), 591-600.

[3] BARRETT, S. AND M. HOEL (2007): Optimal Disease Eradication, *Environment and Development Economics*, 12(5), 627-652.

[4] BAUCH, C. T. (2005): Imitation Dynamics Predict Vaccinating Behaviour, *Proceedings of the Royal Society B*, 272, 1669-1675.

[5] BAUCH, C. T. AND D. J. D. EARN (2004): Vaccination and the Theory of Games, *Proceedings of the National Academy of Sciences*, 101(36), 13391-13394.

[6] BOULIER, B. L., T. S. DATTA AND R. S. GOLDFARB (2007): Vaccination Externalities, *B.E. Journal of Economic Analysis & Policy*, 7(1), (Contributions), Article 23.

[7] BRITO, D. L., E. SHESHINSKI AND M. D. INTRILIGATOR (1991): Externalities and Compulsory Vaccinations, *Journal of Public Economics*, 45(1), 69-90.

[8] DALEY, D. J. AND J. GANI (2001): Epidemic Modelling: An Introduction, *Cambridge Studies in Mathematical Biology.*

[9] DEPARTMENT OF HEALTH (2009): Swine Flu: From Containment to Treatment, leaflet published on July 2nd, 2009.

[10] GOLDMAN, S.M. AND J. LIGHTWOOD (1995): The SIS Model of Infectious Disease with Treatment, *mimeo.*

[11] GOLDMAN, S.M. AND J. LIGHTWOOD (2002): Cost Optimization in the SIS Model of Infectious Disease with Treatment, *Topics in Economic Analysis and Policy*, 2(1), 1-22.

[12] KEELING, M. J. AND P. ROHANI (2008): Modeling Infectious Diseases in Humans and Animals, *Princeton University Press.*

[13] KERMACK, W. O. AND A. G. MCKENDRICK (1927): A Contribution to the Mathematical Theory of Infections, *Proceedings of the Royal Society of London A*, 115(772), 700-721.

[14] ROWTHORN, R. (2006): The Optimal Treatment of Disease Under a Budget Constraint, in R. Halvorsen and D. Layton (eds), *Explorations in Environmental and Natural Resource Economics: Essays in Honor of Gardner M. Brown, Jr,* Edward Elgar.

[15] SANDERS, J. L. (1971): Quantitative Guidelines for Communicable Disease Control Programs, *Biometrics*, 27(4), 883-893.

[16] SETHI, S. P. (1974): Quantitative Guidelines for Communicable Disease Control Program: A Complete Synthesis, *Biometrics*, 30(4), 681-691.

[17] SETHI, S. P., P. W. STAATS (1978): Optimal Control of Some Simple Deterministic Epidemic Models, *Journal of the Operational Research Society*, 29(2), 129-136.

[18] SMITH, C. E. G. (1964): Factors in the Transmission of Virus Infections from Animals to Man, *The Scientific Basis of Medicine Annuals Reviews 1964*, 125-150.

[19] TOXVAERD, F. (2009): Recurrent Infection and Externalities in Treatment, *in preparation.*