Abstract. This paper considers a susceptible-infected-recovered type model of an infectious disease, such as swine flu, in which costly treatment or vaccination confers immunity on recovered individuals. Once recovered, individuals become immune and thus indirectly protect the remaining susceptibles, who benefit from a measure of herd immunity. It is shown that under decentralized decision making, infected individuals ignore the externalities that their decisions have on susceptible individuals and thus seek treatment or vaccination 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 or vaccination, may choose to either eradicate the disease or to retard its eventual dissemination into the population. Optimal treatment is shown to involve intervention at early stages of the epidemic. In contrast, optimal vaccination defers intervention to later stages of the epidemic. Thus, while treatment and vaccination have superficial similarities, their effects and desirability at different stages of the epidemic are radically different.

JEL Classification: C73, I18.

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

1. Introduction

The Spanish Flu Epidemic, caused by the first known H1N1 influenza virus, wreaked havoc across the globe in 1918-1920, leaving death and devastation in its wake. An estimated 500 million people were infected by the virus and an estimated 50-100 million of those cases resulted in death (Johnson and Mueller, 2002 and Tauenberger and Morens, 2006). The dark memories of the 1918 epidemic resurfaced in March 2009, when the first cases of novel influenza A (H1N1), also known as swine flu, were diagnosed in Mexico and the USA. A new strain of the H1N1 virus, swine flu rapidly spread across the globe, causing an estimated cumulative incidence of 24% of the world’s population (Kerkhove et al., 2013). The disease was highly infectious, with an estimated basic rate of reproduction in the range 1.4-1.6 (Fraser et al., 2009). Fortuitously, the case fatality ratio was lower than initially feared, at 0.4% (Fraser et al., 2009). Swine flu can be treated with antiviral drugs, which reduce symptoms and speed up recovery.

I gratefully acknowledge very helpful feedback from Joshua Ross, Chryssi Giannitsarou, Stephen Kissler and seminar participants at the University of Cambridge, Faculty of Economics and Department of Plant Sciences.

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. Email: fmot2@cam.ac.uk.

1 For any rate above one, the epidemic will grow in the population if left unchecked.
recovery if taken early. Additionally, a vaccine against infection from swine flu has been available since November 2009. The epidemic was officially declared over by the WHO in August 2010.

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

“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.”

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, 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 and vaccination efforts have on the population at large. Surprisingly, in the context of treatment, 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). Since incidence and prevalence are functions of each other, the distinction between containment and treatment evidences in the health authority’s thinking is in fact a false dichotomy. This is because the latter is one of the possible instruments that can be used to contain the disease.

A central tool for public health policy in the face of epidemic outbreaks is so-called herd (or population) immunity. According to Fine (1993), herd immunity refers to “the indirect protection afforded to nonimmune individuals by the presence and proximity of others who are immune”. When a sufficiently high proportion of agents are individually immune, the population at large is said to be immune since the epidemic will eventually die out even without intervention. Agent-specific immunity can come about through different channels, namely via natural (or innate) immunity or via acquired immunity through spontaneous recovery, through vaccination or through treatment.

---

2 According to the CDC (2010), patients who are infected by swine flu show symptoms 1-4 days after becoming infected (with an average of two days); they then remain infectious for 5-7 days without treatment, or as long as symptoms persist. Once recovered, individuals become immune to further infection from that particular strain of the virus. Antiviral drugs such as oseltamivir (Tamiflu) or zanamivir (Relenza) reduce the time to recovery by 1.5-2.5 days. The vaccine against swine flu has an estimated efficiency of 79% (see Fielding et al., 2011).

3 UK Department of Health (2009).

4 UK Department of Health (2009).
induced recovery. Of these channels, the last two can be directly controlled and thus chosen with a view to maximize overall welfare.

It is difficult to underestimate the importance and attention given to the notion of herd immunity in public health policy and research. It has taken centre stage in the formulation of global infection control policy for more than half a century and continues to be the bread and butter of policy thinking (see Fine, 1993 for a historical overview of theory and practice). Rather than seeing to control herd immunity directly, as is customary in the epidemiology and public health literature, I will instead characterize the optimal use of the two instruments that can influence the build-up of herd immunity, namely treatment and vaccination. From this analysis, I will then back out how the optimal policies rely on herd immunity. The analysis shows that whether herd immunity is socially useful depends on the costs and benefits of inducing such immunity and furthermore, that this tradeoff may change across the different stages of the epidemic.

To better understand how herd immunity interacts with the formulation of optimal policy and decentralized equilibrium behavior, I consider the economic control of a susceptible-infected-recovered (SIR) model, a framework which is the basis of nearly all formal and empirical analyses of disease outbreaks such as that of N1H1 (see e.g. Coburn et al., 2009). I consider and compare two policies, namely costly treatment that increases the rate of recovery (and immunity) and costly vaccination. In contrast to other preventive measures such as condoms, reductions in promiscuity, quarantines and mosquito nets, treatment and vaccination are permanent and irreversible. More importantly, they act at the population level by leveraging the targeted individuals’ immune responses to create herd immunity and hasten the end of the epidemic outbreak. Thus acquired immunity (which can be achieved through spontaneous recovery, treatment or vaccination) can be thought of as a tool to control infection at the population level because of the existence of herd immunity.

I show 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. In fact, individuals’ privately optimal decisions turn out to be entirely independent of the state of the epidemic and thus be qualitatively different from the policy that maximizes aggregate social welfare.

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 infected individuals and continues to do so in perpetuity. This means that the disease is asymptotically 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 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 asymptotic eradication through treatment. An important difference is that in this scenario, fewer individuals remain susceptible than is the case in the eradication regime.\(^5\)

I then discuss the effects and desirability of vaccination in the context of the same model, with special attention paid to how this alternative policy makes use of herd immunity. I argue that despite superficial similarities, both equilibrium behavior and the socially optimal policy differ radically from that under treatment. In particular, while optimal treatment involves treating infected individuals at early stages of the epidemic, optimal vaccination calls for intervention either at late stages (in case there is no spontaneous recovery) or at the peak of the epidemic (in case there is spontaneous recovery). In the former case, the optimal policy has no reliance on herd immunity, while in the latter case, optimal vaccination policy is dictated by its evolution across the epidemic.

The literature on infection control, economic or otherwise, is large and varied. The bulk of this research has focused on different protective measures such as vaccines, prophylaxis, isolation, quarantines, reductions in the rate of partner change or the control of vectors with equipment such as mosquito nets. A smaller literature has considered the effects of treatment. Other than the literature on vaccination, the existing literature has not focused explicitly on the creation of herd immunity and its role in optimally managing the health of the population at the different stages of the epidemic.

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 analysis, 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. There is a very large literature on different aspects of the economic control of infectious diseases through vaccination. See Chen and Toxvaerd (2014) for a detailed review and synthesis of this literature. Of direct relevance to the work here is Francis (1997, 2007), who considers fully discounted economic models of vaccination and characterizes optimal and

\(^5\)That is, more individuals experience infection and eventual recovery and immunity.

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 outcomes under treatment with those under vaccination. Section 6 concludes.

2. The Model
To make the exposition self-contained, I will first outline the basics of two classical epidemiological models, namely the general epidemic and the simple epidemic models. Those readers familiar with these models can skip this subsection and go directly to the next subsection in which I introduce the economic parameters.

2.1. The SIR and SI Models. The classical susceptible-infected-recovered (or SIR) model is simple to describe. Time is continuous and runs indefinitely. A popu-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{sir_diagram.png}
\caption{SIR Model with Treatment and Vaccination.}
\end{figure}
Toxvaerd

The 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 $S(t)$ and has measure $S(t)$, the set of infected individuals is denoted by $I(t)$ and has measure $I(t)$ and the set of recovered individuals is denoted by $R(t)$ 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 infected is given by the simple expression $\beta I(t) S(t)$. This means that the rate of new infection, or disease incidence, is proportional to disease prevalence.

Last, in the classical version of the model, individuals spontaneously recover at rate $\gamma > 0$. 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 differential equations and initial conditions:

\begin{align}
\dot{S}(t) &= -\beta I(t) S(t) \\
\dot{I}(t) &= I(t) [\beta S(t) - \gamma] \\
\dot{R}(t) &= \gamma I(t) \\
S(t) &= 1 - I(t) - R(t) \\
S(0) &= S_0 > \gamma/\beta, \quad I(0) = I_0, \quad S_0 + I_0 = 1
\end{align}

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

\cite{2001} and Keeling and Rohani (2008).

\footnote{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.}

\footnote{Note that while disease incidence is a flow value, disease prevalence is a stock value.}
zero. At the peak of the epidemic, disease prevalence takes the value
\[ I \equiv I_0 + S_0 \frac{\gamma}{\beta} \left[ \log S_0 - \log \left( \frac{\gamma}{\beta} \right) + 1 \right] \]

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

\[ S(\infty) = 1 - R(\infty) = S_0 \exp (-R(\infty)R_0) \geq 0 \]  

where \( R_0 \equiv \beta/\gamma \) is the basic rate of reproduction.\(^{12}\) Since \( R(0) = 0 \), the cumulative incidence, i.e. the total case count across the epidemic, is given by \( R(\infty) = 1 - S(\infty) \). Intuitively, cumulative incidence is an increasing function of the infectivity parameter \( \beta \) and a decreasing function of the rate of spontaneous recovery \( \gamma \) (see Keeling and Rohani, 2008).

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 in terms of parameters and initial conditions 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.

In what follows, the notion of herd immunity will feature prominently. A measure of herd immunity, although imperfect, is the mass of susceptible individuals remaining at the end of the epidemic, \( S(\infty) \), since these individuals are not vaccinated yet at no risk of infection.\(^{13}\)

Throughout, I will maintain the following assumption:

**Assumption 0:** \( \beta > \gamma \geq 0 \).

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. 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 (i.e.

---

\(^{11}\)See Brauer and Castillo-Chavez (2012).

\(^{12}\)See Daley and Gani (2001) for details.

\(^{13}\)The reason that it is only a rough measure of herd immunity, is that it ignores the time profile of infection. In particular, this measure ignores early protection enjoyed by an individual who eventually becomes infected and recovers.
the chains of infection are broken). 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 $R_0 < 1$, then infection cannot take hold while if $R_0 > 1$, 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 $R_0$.

The classical susceptible-infected (or SI) model is simply the special case of the SIR model with $R(0) = \gamma = 0$. In this model, the measure of infected individuals grows according to a logistic growth equation, approaching one. As the measure of susceptible individuals is the remainder of the population, it tends to zero. In the SI model, the evolution of disease prevalence is given by the logistic growth equation

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

In this setting, the asymptotic distribution is characterized by $\lim_{t \to \infty} I(t) = 1 - R_0$ in the case where $R_0 > 0$, or with everyone infected otherwise. Typical dynamics for the SIR and SI models are illustrated in Figure 2.

2.2. The Economic Model. 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. I will consider the economic control of the SIR model via two instruments, namely treatment and vaccination. Treatment, which is costly, increases the rate at which agents recover (and become immune to further infection). In particular, for some treatment intensity $\tau(t) \in [0, 1]$, the rate at which the individual transitions from $I(t)$ to $R(t)$ 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 $\gamma$. The treatment costs $c_T > 0$ per instant per individual.

Turning to vaccination, denote by $v(t) \in [0, 1]$ the rate at which individuals are vaccinated. The vaccine offers imperfect protection, being subject to a failure rate $\phi \in [0, 1]$. Thus vaccination at rate $v(t)$ induces transition from $S(t)$ to $R(t)$ at rate $v(t)(1 - \phi)$, entirely bypassing the class of infected individuals $I(t)$. Vaccination costs $c_V > 0$ per individual.

Having described the available tools for managing the epidemic, I will describe the payoffs of the decision makers. It will be assumed that the individuals in the sets $S(t)$, $I(t)$ and $R(t)$ earn flow payoffs $\pi_S$, $\pi_I$ and $\pi_R$ respectively and discount the future at rate $\rho > 0$. It will be assumed that $\pi_S \geq \pi_I$ and $\pi_R \geq \pi_I$.

For simplicity, treatment and vaccination have been modeled as continuous variables, but they can alternatively be interpreted as randomizations over discrete choices.
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. 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.

In what follows, control variables with a subscript $i$ refer to those of individuals under decentralized decision-making while the control variables without such a subscript refer to the those of the planner under centralized decision making.

Using the notation introduced above, the relevant parameters for swine flu are of the order $\gamma = 1/6$, $\alpha = 1/12$ (CDC, 2010) and $\phi = 0.21$ (Fielding et al., 2011). Fraser et al. (2009) found a basic rate of reproduction in the range 1.4-1.6, which is somewhat lower than the estimated rate of 2-3 for the 1918 epidemic (Mills et al., 2004).
3. Disease Dynamics under Treatment

In this section, I first characterize equilibrium treatment behavior under decentralized decision-making and then characterize the optimal policy under centralized decision-making. As will become clear, the outcomes under centralized decision making will not only differ from the optimal outcome quantitatively, but also qualitatively. This difference will be shown to be intimately related to how the planner and the individuals view the benefits of herd immunity.

3.1. Treatment under 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 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 or she 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 \( t \geq 0 \). The individual then solves the following problem:

\[
\max_{\tau_i(t) \in [0,1]} \int_0^\infty e^{-\rho t} [q(t)(\pi_I - \tau_i(t)c_T) + (1 - q(t))\pi_R] \, dt \\
\text{s.t.} \quad \dot{q}(t) = -q(t)(\alpha \tau_i(t) + \gamma), \quad q(0) = 1
\]

The integrand is simply the expected flow payoff of an individual, while the differential equation governs the evolution of the transition rate between infection and recovery.

The individual’s problem is equivalent to the following simplified problem, which differs only by the constant \( \pi_R \):

\[
\max_{\tau_i(t) \in [0,1]} \int_0^\infty e^{-\rho t} q(t)(\pi_I - \pi_R - \tau_i(t)c_T) \, dt \\
\text{s.t.} \quad \dot{q}(t) = -q(t)(\alpha \tau_i(t) + \gamma), \quad q(0) = 1
\]

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 \( q^* = 0 \) eventually if \( \gamma > 0 \).

The associated current-value Hamiltonian for this problem is given by

\[
H_T^D \equiv q(t)(\pi_I - \pi_R - \tau_i(t)c_T) - \eta_T(t)q(t)(\alpha \tau_i(t) + \gamma)
\]

where \( \eta_T(t) \) is the costate variable. Since the state variable in the individual’s problem is the transition probability between infection and recovery, the costate variable can be interpreted as the shadow value of recovery.

\textsuperscript{18}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 \( \tau_i(t) \in [0,1] \).
Differentiating the current value Hamiltonian with respect to the treatment rate \( \tau_i(t) \) yields the following necessary condition for optimality (supposing that \( q(t) > 0 \)):

\[
c_T + \eta_T(t) \alpha = 0
\]  

(13)

This equation simply states that the marginal cost of treatment equals the marginal benefit of treatment. To see this, recall that \( \alpha \) is the rate at which treatment induces recovery and that the costate variable \( \eta_T(t) \) is the net benefit of recovery.

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

\[
\dot{\eta}_T(t) = \eta_T(t) [\rho + \alpha \tau_i(t) + \gamma] + [\pi_R - \pi_I + \tau_i(t)c_T]
\]  

(14)

It’s immediately clear that the net benefit of recovery is independent of the aggregate state of the system and of time. That is, conditional on being infected, the problem is stationary. But then it must be that \( \dot{\eta}_T(t) = 0 \), which implies that

\[
\eta_T(t) = \frac{\pi_I - \pi_R - \tau_i(t)c_T}{\rho + \alpha \tau_i(t) + \gamma}
\]  

(15)

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

\[
\tau_i(t) = \begin{cases} 
0 & \text{for } c_T(\rho + \gamma) > \alpha(\pi_R - \pi_I) \\
[0,1] & \text{for } c_T(\rho + \gamma) = \alpha(\pi_R - \pi_I) \\
1 & \text{for } c_T(\rho + \gamma) < \alpha(\pi_R - \pi_I)
\end{cases}
\]  

(16)\hspace{1cm} (17)\hspace{1cm} (18)

The optimal policy simply states that treatment is sought if and 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 \( \alpha + \gamma \)). 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.\(^{19}\) In the SIR cases, things become more complicated due to the possibility of herd immunity, as described in the model section.

These findings are summarized as follows:

**Theorem 1:** Under decentralized decision making: (i) if \( c_T(\rho + \gamma) > \alpha(\pi_R - \pi_I) \) 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 \( c_T(\rho + \gamma) < \alpha(\pi_R - \pi_I) \) 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 \( \rho \), the recovery rate \( \gamma \) or the treatment cost \( c_T \), the less

\(^{19}\)See e.g. Daley and Gani (2001) for details of the SI model.
attractive does treatment become. Conversely, treatment becomes more attractive the higher the efficiency of the treatment $\alpha$ or the higher the health premium $(\pi_R - \pi_T)$.

It is interesting to note that the treatment decision is not strategic in the sense that an individual’s privately optimal action depends on those of other individuals, as would be the case with treatment in a susceptible-infected-susceptible model (see e.g. Toxvaerd, 2010). This is because while infected individuals’ treatment decisions do influence the prospects of the susceptibles, this influence is ignored since there is no feedback from these decisions to the individual’s future welfare. 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.

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.

It is worth emphasizing that herd immunity is a good enjoyed exclusively by susceptible individuals, but provided by recovered individuals. An added twist is that under treatment, it is determined by the infected individuals how much of this good is provided. In other words, the benefits flow from the decisions of individuals in one class to individuals in another class to which the former can never return (and hence from which they will themselves never benefit).

To appreciate how inefficient equilibrium treatment can be, suppose that treatment is too costly to be privately optimal but sufficiently inexpensive for the social planner to want to treat. Furthermore, suppose $S(0) = 1 - \varepsilon$ and $I(0) = \varepsilon$, with $\varepsilon > 0$ arbitrarily small. It is clear that a desirable outcome would involve immediate treatment of the small group of infected individuals and eventual eradication. Yet however small $\varepsilon$ is, this would not happen in equilibrium.

### 3.2. Treatment under Centralized Decision Making.

The problem of the central planner is given as follows:

$$\max_{\tau(t) \in [0,1]} \int_0^\infty e^{-\tau t} [S(t)\pi_S + I(t) (\pi_T - \tau(t)c_T) + R(t)\pi_R]dt$$

(19)

In the planner’s problem, the policy $\tau(t) \in [0,1]$ can interchangeably be thought of as the proportion of infected individuals who undergo treatment or as the intensity of treatment that all individuals are subjected to.

The problem is solved subject to the following laws of motion for the measures of susceptible, infected and recovered individuals:
\[\begin{align*}
\dot{S}(t) &= -\beta I(t)S(t) & \text{(20)} \\
\dot{I}(t) &= I(t) [\beta S(t) - \alpha r(t) - \gamma] & \text{(21)} \\
\dot{R}(t) &= I(t) [\alpha r(t) + \gamma] & \text{(22)} \\
S(t) &= 1 - I(t) - R(t) & \text{(23)} \\
S(0) &= S_0 > \gamma/\beta, \quad I(0) = I_0, \quad S_0 + I_0 = 1 & \text{(24)}
\end{align*}\]

Note that in this formulation of the planner’s objective, no distinction is made between individuals who reach the recovered class \(R(t)\) through different channels (i.e. through spontaneous recovery, treatment-induced recovery or immunization).\textsuperscript{20}

The problem solved by the central planner is similar to that of the problem solved by individuals under decentralized decision making, 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, it is useful to make 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.\textsuperscript{21} Treatment has a similar effect by transferring individuals from \(I(t)\) to \(R(t)\), rather than from \(S(t)\) to \(R(t)\) 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 has a spell of infection while the latter does not. The differences make optimal treatment and immunization policies qualitatively different, as will be discussed further below.

Returning to the characterization of the optimal policy, using the normalization to eliminate \(S(t)\), the planner’s current-value Hamiltonian is given by

\[H^C_T \equiv \left[1 - I(t) - R(t)\right] \pi_S + I(t) \left(\pi_I - \pi_T - \tau(t)c_T\right) + R(t) \pi_R + \lambda(t)I(t) [\beta \left(1 - I(t) - R(t)\right) - \alpha r(t) - \gamma] + \mu(t)I(t) [\alpha r(t) + \gamma] \quad (25)\]

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

\[c_T + \alpha \left(\lambda(t) - \mu(t)\right) = 0 \quad (26)\]

\textsuperscript{20}In the literature, it is common to set \(\pi_S = \pi_R\). This simplification reduces the system to one with two state variables.

\textsuperscript{21}This is simply because a number of infected and/or susceptible individuals are matched with recovered individuals instead of with each other.
The evolution of the multipliers is governed by the following system of differential equations:

\[
\begin{align*}
\dot{\lambda}(t) &= \lambda(t) \left[ \rho + \alpha \tau(t) + \gamma + \beta (2I(t) + R(t) - 1) \right] \\
&\quad - \mu(t) \left[ \alpha \tau(t) + \gamma \right] - (\pi_T - \tau(t) c_T - \pi_S) \tag{27} \\
\dot{\mu}(t) &= \mu(t) \left[ \alpha \tau(t) + \gamma \right] - (\pi_R - \pi_S) \tag{28}
\end{align*}
\]

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

\[
\begin{align*}
\tau(t) &= 0 \quad \text{for} \quad c_T > \alpha(\mu(t) - \lambda(t)) \tag{29} \\
\tau(t) &\in [0, 1] \quad \text{for} \quad c_T = \alpha(\mu(t) - \lambda(t)) \tag{30} \\
\tau(t) &= 1 \quad \text{for} \quad c_T < \alpha(\mu(t) - \lambda(t)) \tag{31}
\end{align*}
\]

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. The optimal policy is thus dictated by a comparison of the marginal cost and the marginal benefit of treatment, the latter of which is a function of the state of the system and thus not constant.

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 itself, makes it a source of infection for the susceptible part of the population. To see this clearly, set \(R(t) = 1 - I(t)\) so that there are no susceptibles and thus no external effects. In this case, the problem of the planner is stationary and hence the multipliers must be constant. Straightforward substitution of the steady state multipliers into the planner’s Hamiltonian conditions then yields the following characterization of the optimal policy:

\[
\begin{align*}
\tau(t) &= 0 \quad \text{for} \quad c_T (\rho + \gamma) > \alpha(\pi_R - \pi_I) \tag{32} \\
\tau(t) &\in [0, 1] \quad \text{for} \quad c_T (\rho + \gamma) = \alpha(\pi_R - \pi_I) \tag{33} \\
\tau(t) &= 1 \quad \text{for} \quad c_T (\rho + \gamma) < \alpha(\pi_R - \pi_I) \tag{34}
\end{align*}
\]

Thus when \(S(t) = 0\), the solution of the central planner’s problem coincides with that of the individuals under decentralized decision making.\(^{22}\)

The planner’s objective can be decomposed into the terms \(I(t) (\pi_I - \tau(t) c_T) + R(t) \pi_R\) and \(S(t) \pi_S\) 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 deciding 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 welfare.

\(^{22}\)The function \(\alpha(\mu(t) - \lambda(t))\) obtains its minimum \((\pi_R - \pi_I)\) at \(S(t) = 0\).
source of externalities in this model. Since this externality is not taken into account by individuals, it follows that the planner always assigns at least as much value to treatment as does the individuals. For that reason, if individuals find it privately optimal to treat themselves, then so would the planner. As the optimal choices are always of the bang-bang variety, when the individuals choose to treat they in fact treat the same amount. But this means that in this case, the equilibrium outcome is socially optimal. This case obtains when treatment costs are low, when treatment if very efficient, when the health premium from recovery is large or when the future is heavily discounted.

Since \( \dot{S}(t) = -\beta I(t)S(t) < 0 \), the term \( S(t)\pi_S \) 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 \( c_T (\rho + \gamma) < \alpha (\pi_R - \pi_T) \), the decentralized equilibrium outcome coincides with that chosen by the central planner. The interesting case is therefore the one in which treatment is socially, but not privately, optimal. For this reason, I impose the following restrictions on the problem:

**Assumption 1:** \( c_T > \frac{\alpha (\pi_R - \pi_T)}{(\rho + \gamma)} \).

**Assumption 2:** \( c_T < \alpha (\mu(0) - \lambda(0)) \) at \( r(t) = 1 \ \forall t \geq 0 \) 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

\[
S^* \equiv \{ S(t) \in [0, 1] : \alpha [\mu(t) - \lambda(t)] = c_T \ \text{at} \ \{\tau(t) = 1 \ \forall t \geq 0\}\} \quad (35)
\]

at which the net benefit from treatment is zero.

To understand the definition in the lemma, recall that the benefit of treatment \( \alpha [\mu(t) - \lambda(t)] \) depends on the actual treatment choices over time (as can be seen from the laws of motion for the multipliers). The definition simply evaluates this benefit for the policy of perpetual treatment.

**Proof:** Note that the net benefit from treatment

\[
\alpha [\mu(t) - \lambda(t)] - c_T \quad (36)
\]
is increasing in the measure of susceptible individuals $S(t)$. Furthermore, $S(t)$ is weakly decreasing over time irrespective of the chosen (constant) policy $\tau(t)$. This follows from the fact that the fraction of susceptible individuals can be expressed as

$$S(t) = S_0 \exp \left( \frac{-R(t)\beta}{\tau(t)\alpha + \gamma} \right) \quad (37)$$

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 2:** The optimal policy under centralized decision making is given by

$$\tau(t) = \begin{cases} 
1 & \text{if } S(t) > S^* \\
[0, 1] & \text{if } S(t) = S^* \\
0 & \text{if } S(t) < S^* 
\end{cases} \quad (38)$$

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.

In other words, it is socially optimal to treat when external effects are sufficiently important. since these decrease monotonically over time, like an exhaustible resource, it is optimal to cease treatment when the external effects become sufficiently weak.

The next step is to determine whether 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 $\tau(t)$. From the analysis of the classical SIR model, the final measure of susceptible individuals can be described by the equations

$$\hat{S} \equiv \lim_{t \to \infty} S(t) = 1 - R(\infty) = S_0 \exp \left( \frac{-R(\infty)\beta}{\gamma + \alpha} \right) \quad (41)$$

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

Note that the limit $\hat{S}$ 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.

It should be noted that the threshold $\hat{S}$ is a function of the biomedical parameters only, while the threshold $S^*$ is a function of both the biomedical parameters and the economic (i.e. cost and preference) parameters.
Managing Population Immunity

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

**Asymptotic Eradication Regime.** This regime corresponds to the case where

\[ \hat{S} > S^* \]  

(42)

In this case, perpetual full treatment is optimal. This is because 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.

In the limiting population distribution, some individuals have recovered (and are immune) and the remaining individuals are susceptible (and protected by herd immunity). Infection is thus fully eradicated asymptotically. The measure of protected individuals is given implicitly by

\[ S(\infty) = 1 - R(\infty) = S_0 \exp \left( \frac{-R(\infty)\beta}{\alpha + \gamma} \right) \geq 0 \]  

(43)

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

\[ \hat{S} \leq S^* \]  

(44)

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 run its course. Since \( S(t) \) is monotone, there is a unique critical time \( t^* \) at which the optimal policy switches from full treatment to no treatment.

The reason is that although 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. The measure of protected individuals is given implicitly by

\[ S(\infty) = 1 - R(\infty) = S(t^*) \exp \left( \frac{-R(\infty)\beta}{\gamma} \right) \geq 0 \]  

(45)

In this regime, the limiting distribution depends crucially on the background rate of recovery \( \gamma \). If \( \gamma = 0 \) then in the limit some individuals have recovered and the remaining individuals are infected and remain so in perpetuity since treatment ceases beyond time \( t^* \). 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 \( \gamma \), initialized at \((S(t^*), I(t^*), R(t^*))\). I.e. the disease is eradicated and some susceptibles remain.

Figure 3 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 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 \( c_T \) (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 downward the marginal benefit curve. 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 $\tilde{S}$. The magnitude of the limit $\tilde{S}$ is determined by the epidemiological parameters $\alpha$ and $\beta$, the treatment efficiency $\gamma$ and by initial conditions. The two regimes delineating the optimal policy simply corresponds to the location of the limit point $\tilde{S}$ relative to the critical threshold $S^*$ (corresponding to point $Y$ on the curve).

If $\tilde{S} = S_A$ 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 $\tilde{S} = S_B$ 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.

The comparative statics of the limit point \( \hat{S} \) follow from those of the cumulative incidence \( R(\infty) \) and the fact that \( \hat{S} = 1 - R(\infty) \), with the straightforward modification that recovery occurs at rate \((\gamma + \alpha)\). In particular, it can be verified that \( \hat{S} \) is decreasing in the effective rate of reproduction under full treatment, to wit decreasing in \( \beta \) and increasing in \( \gamma \) and \( \alpha \). Determining the comparative statics of the optimal policy is more delicate, as the critical threshold \( S^* \) depends on the economic cost and preference parameters \((c_T, \pi_S, \pi_I, \pi_R, \rho)\) and the biomedical parameters \((\alpha, \beta, \gamma)\) in a complicated way, through the costate variables \( \mu(t) \) 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 on the optimal policy.

To align the individuals’ incentives with those of the planner, a simple subsidy to treatment can be introduced. Suppose that the planner would prefer treatment but that \( c_T(\rho + \gamma) > \alpha(\pi_R - \pi_I) \), so individuals do not voluntarily demand any. The planner can implement the first best outcome simply by offering a subsidy \( s_T \) such that \((c_T - s_T)(\rho + \gamma) < \alpha(\pi_R - \pi_I)\) for as long as the first-best prescribes treatment.

4. Disease Dynamics under Vaccination

In this section, I turn my attention to optimal and equilibrium outcomes under vaccination. This will enable a clear comparison with the outcomes under treatment. As emphasized in the introduction, both treatment and vaccination work by boosting the mass of recovered individuals, thereby inducing herd immunity to protect the remaining susceptible individuals. But the similarities turn out to be superficial only. In this section, I show that seemingly subtle differences between treatment and vaccination can have radically different equilibrium outcomes and that even the optimal policies may differ qualitatively.

4.1. Vaccination under Decentralized Decision Making. First, consider the individual’s vaccination problem, which is given by

\[
\max_{v_i(t) \in [0,1]} \int_0^\infty e^{-\rho t} \left[ (1 - p(t))(\pi_S - v_i(t)c_V) + p(t) \left( \frac{\rho \pi_I + \gamma \pi_R}{\rho(\gamma + \rho)} \right) \right] dt \tag{46}
\]

The relevant constraint is given by

\[
\dot{p}(t) = (1 - p(t))(1 - v_i(t)(1 - \phi))\beta I(t) \tag{47}
\]

In this formulation of the individual’s problem, \( p(t) \) denotes the probability of transition from susceptible to infected, rather than the probability of being in the infected state.\textsuperscript{23,24} To understand the maximization problem, note that the first term in the integrand is simply the flow payoff of a susceptible individual who vaccinates at

\textsuperscript{23}This is because when \( \gamma > 0 \), any infected individual eventually recovers. The probability of this happening is captured in the expected net present value obtained upon infection.

\textsuperscript{24}Note that conditional on becoming infected, the shadow price of infection for the individual is in fact independent of the aggregate state of the system and of time. Thus the costate variable is a constant.
rate \( v_i(t) \). The second term corresponds to the expected flow payoff of an individual who has just become infected. This term is slightly more involved and needs further explanation. Recall that once infected, an individual has some exogenous probability of recovering and thus transitioning from \( I(t) \) to \( R(t) \). This transition occurs at rate \( \gamma \). Until the transition occurs, say at some time \( T \), the individual earns flow payoff \( I(t) \), whereas after \( T \), the flow payoffs increases to \( \pi_R \). The second term in the integrand is thus simply the expected flow payoff of an individual who has just become infected and who will recover spontaneously at some random time \( T \) (which is governed by an exponential distribution with rate \( \gamma \)).

The current-value Hamiltonian for the individual’s problem is given by

\[
H_V^D \equiv (1 - p(t))(\pi_S - v_i(t)c_V) + p(t)\pi_F - \eta_V(t)(1 - p(t))(1 - v_i(t)(1 - \phi))\beta I(t)
\]  

where

\[
\pi_F \equiv \left( \frac{\beta \pi_I + \gamma \pi_R}{p(\gamma + \rho)} \right)
\]

Differentiating with respect to the vaccination rate \( v_i(t) \) yields the following necessary condition for optimality (assuming that \( p(t) < 1 \)):

\[
c_V = -\eta_V(t)(1 - \phi))\beta I(t)
\]  

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

\[
\dot{\eta}_V(t) = \eta_V(t)[\rho(1 - v_i(t)(1 - \phi))\beta I(t)] + (\pi_S - v_i(t)c_V - \pi_F)
\]  

Note that for the individual, conditional on becoming infected, the shadow cost of infection \( \eta_V(t) \) is a constant (i.e. it is time invariant). This means that \( \dot{\eta}_V(t) = 0 \). Together with the Hamiltonian condition, this means that the best response of the individual is of the bang-bang variety and characterized by the policy

\[
v_i(t) = \begin{cases} 
0 & \text{if } I(t) < I_D^* \\
[0,1] & \text{if } I(t) = I_D^* \\
1 & \text{if } I(t) > I_D^*
\end{cases}
\]

where the critical thresholds is given by

\[
I_D^* \equiv -\frac{c_V \rho}{\beta (1 - \phi)(\gamma + \rho)[\pi_S - \pi_F] - \beta c_V}
\]

Note for later reference that with decentralized treatment decisions, the cases \( \gamma = 0 \) and \( \gamma > 0 \) are qualitatively similar, with the critical threshold increasing in the recovery rate \( \gamma \). The reason for this comparative statics result is simply that as the rate of spontaneous recovery increases, the expected disease burden decreases.\(^{25}\)

To make the problem interesting, I impose the following restriction:

**Assumption 3:** \( I_D^* < \bar{I} \).

\(^{25}\)The case where \( \pi_S = \pi_R \) and \( \gamma = \phi = 0 \) corresponds to the model of Francis (2007).
If Assumption 3 were violated, then the epidemic peaks before the individuals (or the planner) would start vaccinating and thus the option to vaccinate would not be socially useful. Note that while $\overline{T}$ depends only on biological parameters, $I_D^*\mid_{\gamma=\phi=0}$ depends on both biomedical parameters and economic (cost and preference) parameters.

Under decentralized decision-making (and under the maintained assumptions), herd immunity never takes effect if the population is homogeneous and the vaccine is perfect. The reasons for this result are two-fold, namely (i) ex ante homogeneity and (ii) lack of strategic interaction in vaccination decisions. As shown in Chen and Toxvaerd (2014), these two features imply that in equilibrium, all individuals wait until the critical threshold is reached and then all simultaneously vaccinate themselves. But this in turn implies that when any given individual chooses to vaccinate, there are no non-vaccinated susceptibles left in the population who can benefit from the individual’s vaccination (through herd immunity). Having said that, it is clear that even minimal deviations from the homogeneous population, perfect vaccine setup, will reintroduce the scope for herd immunity. With heterogeneous individuals, some will decide to vaccinate earlier than others (i.e. have lower critical thresholds of disease prevalence), with the latter benefiting from the formers’ vaccination.\(^{26}\) Similarly, with an imperfect vaccine, even vaccinated individuals will benefit from the vaccination of others through herd immunity.

### 4.2. Vaccination under Centralized Decision Making.

Turning to centralized decision-making, the planner’s problem can be written as follows:

$$\max_{v(t)\in[0,1]} \int_0^\infty e^{-\rho t}[S(t)[\pi_S - v(t)c_V] + I(t)\pi_I + R(t)\pi_R]dt$$

(55)

The relevant constraints are given by

\begin{align}
\dot{S}(t) &= -S(t)[\beta I(t) + (1-\phi)v(t)] \\
\dot{I}(t) &= I(t)[\beta S(t) - \gamma] \\
\dot{R}(t) &= \gamma I(t) + S(t)(1-\phi)v(t) \\
S(t) &= 1 - I(t) - R(t) \\
S(0) &= S_0 > \gamma/\beta, \quad I(0) = I_0, \quad S_0 + I_0 = 1
\end{align}

The optimal policy for the planner turns out to be more complicated than that of the individual. In the special case $\gamma = \phi = 0$ and $\pi_S = \pi_R$, it is known that the planner’s optimal policy is of the bang-bang variety, with critical threshold exactly like in the individual’s problem (i.e. with threshold $I_C^* = I_D^*\mid_{\gamma=\phi=0}$). Equilibrium is thus socially optimal and there are neither external effects in equilibrium nor any herd immunity (see Francis, 1997 and Chen and Toxvaerd, 2014 for a detailed exposition of this property). In this case, $I(\infty) + R(\infty) = 1$, so no-one benefits from herd immunity.

Note that when $\phi > 0$, then there are external effects from vaccination even with homogeneity and without spontaneous recovery (i.e. with $\gamma = 0$).

In the case with $\gamma > 0$ and $\phi = 0$, the optimal policy has only been partially characterized (see Sethi and Staats, 1978 and Francis, 2007). The optimal policy is still

\(^{26}\)This point is also made in Fine et al. (2011).
of the bang-bang variety and is never singular (i.e. interior) for any positive interval of time.\textsuperscript{27} Furthermore, it is known that there can be at most two switches in policy and that the last is always from full vaccination to no vaccination. So the planner either vaccinates at the outset and then ceases doing so, or vaccination is initially delayed, then implemented at full force and then eventually ceased. Despite the difficulty in formally characterizing the optimal policy in this case, the main features seem intuitively clear. At early stages of the epidemic, the hazard is moderate and thus it is optimal not to vaccinate, as is the case when $\gamma = 0$. At some point, the optimal policy switches and full vaccination is implemented. This is done in part with a view to benefit from herd immunity (i.e. the planner values vaccination higher than do the individuals, who only maximize their own welfare and disregard the social benefits flowing from their vaccination). As vaccination is pursued, in conjunction with natural recovery, the mass of susceptibles becomes sufficiently large that the remaining non-vaccinated susceptibles are effectively protected by herd immunity. At this point, the optimal policy switches back and no further vaccination is pursued. The fully discounted problem has received surprisingly little attention. Exceptions include Sethi and Staats (1978), Francis (1997) and Francis (2007). Of these, none consider imperfect vaccines and only the latter contribution features spontaneous recovery. In short, while $I_C^* \leq I_D^*$, typically $I_C^* < I_D^*$. Let the time of vaccination be denoted by $t_j^* = \{\min t \geq 0 : I(t) = I_j^* \}$, $j = D, C$. The first-best outcome under vaccination can be achieved by offering individuals a vaccination subsidy $s_V$ at time $t_C^*$ so that they vaccinate at that time rather than at time $t_D^* > t_C^*$.

5. Comparisons to Non-Economic Control

To put this analysis into perspective, it is worthwhile considering a typical non-economic approach to infection control in this type of setting and to compare it to the optimal policies characterized in the previous sections. As noted in the introduction, the mathematical epidemiology literature has seen a multitude of analyses focused almost entirely around the basic rate of reproduction in different settings and on how the public health authority may bring about eradication by influencing this rate. For a discussion of this literature and the central role accorded to herd immunity and the basic rate of reproduction, see e.g. Fine (1993) and Fine et al. (2011).

First, recall the special role played by the basic rate of reproduction $R_0 = \beta / \gamma$. For $R_0 < 1$, the disease dies out over time while for $R_0 \geq 1$, the epidemic takes hold in the population. Perhaps not surprisingly, a large part of the traditional analysis of disease control takes $R_0$ as the implicit objective function. For example, the textbook analysis of imperfect (leaky) vaccination argues that if a proportion $v \in [0, 1]$ of the population is immunized but that the vaccine has a failure rate of $\phi \in [0, 1]$, then the basic rate of reproduction changes to

$$R_0^v \equiv \frac{[1 - v(1 - \phi)]\beta}{\gamma} = [1 - v(1 - \phi)]R_0$$

\textsuperscript{27}Sethi and Staats (1978) consider a finite time horizon and focus mostly on the undiscounted case.
In order to eradicate the disease, one must vaccinate at least a fraction

\[ v > \frac{R_0 - 1}{R_0(1 - \phi)} \]  

(62)

of the population. Indeed, this is the typical policy prescription in most epidemiology textbooks (see e.g. Keeling and Rohani, 2008).

Whether this is feasible, depends on parameter values. One can paraphrase the non-economic eradication policy with vaccination as follows:

- If \( \beta \in [0, \gamma) \), eradication is achieved without vaccination.
- If \( \beta \in [\gamma, \gamma/\phi) \), eradication is achieved with vaccination.
- If \( \beta \in [\gamma/\phi, \infty] \), eradication is not feasible with vaccination.

To see what this type of thinking would imply for a policy on treatment, suppose that treatment \( \tau \in [0, 1] \) increases the recovery rate to \((\gamma + \alpha \tau)\), where \( \alpha > 0 \) is the (finite) efficiency of treatment. In other words, treatment boosts the patient’s own immune defence and speeds up recovery. The basic rate of reproduction now changes to

\[ R_{0T} \equiv \frac{\beta}{\tau \alpha + \gamma} \]  

(63)

Again, in order to eradicate the disease, the policy must ensure that \( R_{0T} < 1 \), which is achieved by permanently treating infected individuals at some level

\[ \tau > \frac{\beta - \gamma}{\alpha} = \frac{R_0 - 1}{(\alpha/\gamma)} \]  

(64)

Since treatment is not infinitely efficient in inducing recovery, i.e. recovery is not instant, such eradication may not be feasible as treatment may fail force \( R_{0T} \) below one. One can paraphrase the non-economic eradication policy with treatment as follows:

- If \( \beta \in [0, \gamma) \), eradication is achieved without treatment.
- If \( \beta \in [\gamma, \gamma + \alpha) \), eradication is achieved with treatment.
- If \( \beta \in [\gamma + \alpha, \infty) \), eradication is not feasible with treatment.

The startling result is that under certain circumstances, it is not worthwhile to vaccinate or to treat at all, simply because full asymptotic eradication is not feasible. The obvious weakness of this analysis, is that it completely disregards the desirability of using each instrument at different stages of the epidemic (for the individual or for a planner) and therefore ignores the important tradeoffs that determine privately and socially optimal decisions. To make meaningful policy recommendations and sensible comparisons between instruments, one must explicitly consider the problems faced by decision makers under centralized and decentralized decision making. Furthermore, one must fully take into account that any decision, whether on treatment or vaccination, has both costs and benefits and that these must be appropriately balanced. The purely mechanistic approach inherent in the eradication-focused analysis above, may lead to socially undesirable outcomes and simplistic policy recommendations.
Figure 4: Basic Rates of Reproduction under Treatment and Vaccination.

The basic problem from a social perspective, is that individuals do not value interventions such as vaccination or treatment enough. This means that there are instances in which individuals may decide against intervention but where there would be social value from these interventions. By the nature of self-interested individuals, they will typically not take population immunity into account when making decisions. But it is certainly true that whenever individuals want to intervene, so would a benevolent social planner, because the latter always values the intervention at least as much as the former, whether herd immunity is created or not. In other words, interaction in the market by a central planner, should not be predicated on whether that intervention can achieve herd immunity. Rather, it should depend on whether that intervention can increase social welfare. Seen from this perspective, the achievement of complete herd immunity is a secondary issue.

Figure 4 shows the basic rates of reproduction under optimal treatment and vaccination, respectively. The left-hand side panel illustrates treatment under the stemming the tide regime, while the right-hand side panel illustrates vaccination when there is no spontaneous recovery (i.e. when $\gamma = 0$). For comparison, the textbook suggestion for a vaccination policy involves early vaccination to reduce the mass of susceptible individuals so that infection starts decreasing. After that, once herd immunity takes effect, vaccination should be stopped.

There is a stark contrast between the economic and non-economic control of the disease, best exemplified by the case of an imperfect vaccine. Under the non-economic approach, a higher failure rate $\phi$ would call for a higher critical vaccination threshold to be achieved (to ensure eradication). In contrast, the economic analysis shows that a higher failure rate can be understood as an effective increase in the cost of vaccination (or as a decrease in the benefits to vaccination), thereby reducing the desirability of vaccination. Thus, ceteris paribus, it is optimal to vaccinate at a lower rate. Intuitively, similar points would apply for a decrease in treatment efficiency $\alpha$. In a sense, the non-economic approach confounds feasibility with desirability.

The treatment of vaccination in Keeling and Rohani (2008) is a good example of the ambiguity on policy issues in the existing literature. While noting that it may be infeasible to eradicate the disease through vaccination, they note that some vaccination may still be desirable. Yet, they stop short of letting the desirability of vaccination be the guiding principle in formulation vaccine policy.
6. Conclusion

In this paper, I have considered the economic control of the classical susceptible-infected-recovered model of infectious disease. Two costly measures were considered, namely treatment and vaccination. Treatment may increase the rate of recovery and confer immunity from future infection on the recovered individual. I found that in equilibrium, if the treatment cost is sufficiently large, 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 the effects of herd immunity). I show that depending on initial conditions and parameter values, the socially optimal policy may either involve asymptotic eradication of the disease through mass treatment of infected individuals, or treatment for a limited duration of time, followed by a complete cessation of treatment measures. In the latter case, if recovery can only be acquired through treatment, infection will be endemic. Vaccination decisions turn out to be different to those of treatment, whether made by individuals or a social planner. Individuals will delay vaccination until the hazard of infection becomes sufficiently high. While this is also true of the socially optimal vaccination policy, the vaccination threshold of the planner is typically lower than the equilibrium threshold. Furthermore, depending on parameter assumptions, the socially optimal vaccination policy may involve more than once switch between no vaccination and full vaccination.

I have argued in this paper that while treatment and vaccination may have superficial similarities when treatment induces acquired immunity, outcomes are radically different across instruments under both types of decision-making. In the case of decentralized decision-making on treatment, equilibrium is entirely independent of the state of infection, while equilibrium vaccination decisions are state dependent. Under centralized decision-making, the optimal treatment policy has full treatment at the early stages of the epidemic and a possible switch to no treatment at later stages. In contrast, the optimal vaccination policy involves no vaccination at early stages of the epidemic and then a switch to full vaccination. In the case of spontaneous recovery, there may even be an addition switch back to no vaccination.

Looking at the comparisons from a different angle, with vaccination there are cases (namely when no spontaneous recovery is possible) in which both the planner and individuals have qualitatively similar behavior, but where different valuations lead to quantitative differences in outcomes. In contrast, with treatment, equilibrium behavior differs qualitatively from the socially optimal policy. The individual's decision problem is a simple and state independent comparison of costs and benefits, while that of the planner is a delicate function of the measures of susceptible, infected and recovered individuals.

Under decentralized decision making no self-interested individual would consciously contribute towards herd immunity. Having said that, the presence of herd immunity may well influence individual decision making, as is the case with vaccination when the population is heterogeneous.

A topic of both theoretical and practical importance, but which this paper has not treated, is the equilibrium and the socially optimal combination of treatment and vaccination. It’s immediately clear that treatment and vaccination at different stages of
the epidemic changes the value of such efforts. The simplest case to consider is that of decentralized decision making. For an individual, treatment is an imperfect substitute for vaccination, while vaccination is no substitute for treatment (once infected, it is simply too late to vaccinate). Thus for an infected individual, the presence or otherwise of vaccination is immaterial and cannot influence the vaccination decision. Next, consider the effects that the presence of treatment has on the privately optimal use of vaccination. Whether treatment is available or not, the individual will choose to vaccinate once the cost of doing so is justified by the expected future cost of infection. Now, the option to treat infection in the future must decrease the expected future cost of becoming infected - at least weakly so since the option to treat can always be rejected. This means that if the individual would choose not to vaccinate without the option of treatment, then it would continue not to do so with that option present. On the other hand, an individual who would choose to vaccinate without the possibility of treatment, may or may not choose to do so once treatment becomes an option. It depends on how much the treatment option lowers the infection cost, relative to the cost of vaccination. This completes the characterization of the hybrid model under decentralized decision making.

Next, turn to centralized decision making. Perhaps not surprisingly, this case poses a significant challenge that will not be taken on here. When studied in isolation, optimal policy suggests using treatment at early stages, while delaying vaccination to later stages. Once treatment and vaccination can be combined, it is by no means clear that their timing would reflect that pattern. Depending on costs, parameters and initial conditions, it is conceivable that an optimal policy would involve so much initial treatment that no subsequent vaccination would be desirable. On the other hand, it is also possible that an optimal policy would start out by vaccinating such a large fraction of the population, that subsequent treatment would eradicate the disease asymptotically. That is, an initial concentrated vaccination campaign may tip the scales from a stemming the tide regime to an asymptotic eradication regime. To make progress on this front, one must resort to simulations. This would also make it possible to enrich the model in different directions, such as accounting for incubation and latency periods and adding demographics (i.e. births and deaths) to the model.

In this connection it is worth noting that, while an individual is always restricted to vaccination first and then treatment later (if infected), the central planner would have no such restrictions at the population level. Of course, the same restriction would apply in vaccinating or treating any given individual, but at the aggregate level, the planner can make any number of switches between combinations of vaccination and treatment across the stages of the epidemic.

There are some overall lessons to be learned from the present analysis. First, rather than focusing on the mechanics of herd immunity and on bringing down the basic rate of reproduction, optimal treatment and vaccination policy should be formulated with a view to optimally trade off costs and benefits of each intervention. Once this has been achieved, it emerges endogenously whether and the extent to which herd immunity plays a role in the optimal policy. Second, it is not always optimal to suppress infection by lowering the basic rate of reproduction. This follows from the simple observation that under certain circumstances, the cost of suppressing infection (by lowering the basic rate of reproduction) outweigh the social benefits of doing so. Third, the intensity
of optimal interventions may well vary across different stages of the epidemic in non-obvious ways. In particular, depending on the available tools, optimal intervention may be either front-loaded or back-loaded. This finding contrasts with the common assertion in the public health literature that interventions such as vaccination should be made earlier rather than later in order to maximize impact.

It’s worth emphasizing the importance of properly formulated policy in the management of epidemics, in order to prepare for the next outbreak. This is relevant not only for H1N1 outbreaks but also for other diseases that fit the present framework, such as seasonal influenza, hepatitis B, norovirus, pertussis (whooping cough) and a host of other viral and bacterial infections. The tools available for a health authority depends on the specific infectious disease at hand. Unfortunately, there are many vaccine-preventable diseases for which there are no effective treatments available. Similarly, there are many treatable diseases for which there are no viable vaccines. This paper has taken a first step towards characterizing equilibrium behavior and optimal policy to control population immunity when either or both instruments are available.

\footnote{While for some of these diseases acquired immunity is not permanent, the model is still appropriate for the study of a single outbreak.}
REFERENCES


