

Market Structure and Communicable Diseases

Stéphane Mechoulan*

University of Toronto

August 2005

Abstract

Communicable diseases pose a formidable challenge for public policy. The current paper sets out a new conceptual framework for the analysis of treatments in the context of communicable diseases, under different market structures and epidemiological scenarios. Using numerical simulations, we present a number of results that yield fresh insights into the role of drug prices and regulations. First, we show when a monopoly's price and prevalence path converge to a nonzero steady-state. This expands on the intuition that a pharmaceutical company may have an incentive to keep a communicable disease alive to preserve its market. In contrast, a planner typically prices so as to quickly eradicate the disease, making welfare under mandatory universal treatment very close to the first-best outcome. If eradication is impossible, the planner still subsidizes treatments as long as the prevalence can be kept in check. The presence of drug resistance exacerbates the welfare difference between a monopoly and the social planner. Nevertheless, because the negative externalities from resistance compete with the positive externalities of treatment, we show that there are circumstances in which a mixed competition/monopoly regime does better than competition alone. Going beyond the case of communicable diseases, this has important implications for the design of many drug patents.

Keywords: communicable, disease, resistance

JEL Classification: I18, L12

*I thank Joe Altonji, Chris Auld, Gadi Barlevy, Don Dewees, Gérard Gaudet, Dino Gerardi, David Howard, Morton Kamien, Istvan Konya, Stuart Levy, Charles Mallory, Robert McMillan, John Panzar, Tomas Philipson, Paul Rubin, Nicolas Sahuguet, Aloysius Siow, Yossef Spiegel, Mark Stabile, Matt Turner, seminar participants at Guelph, Queen's, Health Canada, Waterloo, the third IHEA conference and the second CHESG workshop for helpful comments, suggestions and discussions. The comments of two anonymous referees improved an earlier draft considerably. Konstatin Rybakov provided exceptional research assistance. All remaining errors are mine. Affiliation: Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario M5S 3G7, Canada. E-mail: s.mechoulan@utoronto.ca

“Viruses, bacteria, fungi, and their human and non-human hosts are engaged, neck and neck, in a coevolutionary race, perpetually trying to devour, outcompete, or otherwise defeat each other.” (Daily and Ehrlich, p. 322)

1 Introduction

As the world becomes ever more interconnected, so the threat of communicable diseases has soared in recent years.¹ People travel more frequently, faster, and more widely than before, and they bring with them potentially hazardous micro-organisms. The recent outbreak of SARS has exemplified how a new virus jumping the species barrier in rural China could within weeks kill hundreds, sicken thousands, quarantine tens of thousands, and result in a multi-billion dollar crisis worldwide.

This paper examines the role of pricing and certain regulations in the context of communicable diseases. We propose a new conceptual framework for the analysis of treatments (as opposed to immunization²) under different market structures where (possibly competing) external effects are present. We thus provide fresh insights into a pharmaceutical firm’s profit maximizing strategy, and compare it with that of a social planner. Then, from a welfare comparison that also includes perfect competition and mandatory universal treatment outcomes we derive policy implications using various epidemiological scenarios.

What happens generally in terms of pricing in the presence of market power, externalities and heterogeneous agents in a dynamic environment is not well-understood. In particular, agent-based models of economic epidemiology are inescapably difficult to solve analytically beyond two periods or outside a steady-state. Hence the major

¹See www.who.int/infectious-disease-report/

²Immunization is a prophylactic step, therefore vaccines represent an investment in health and from an economic perspective fall in the specific category of durable goods. These entail a totally different approach and have different implications in terms of pricing and policy making that we shall not get into here. See Philipson (2000) and Gersovitz and Hammer (2003) for review articles on the economics of communicable diseases.

contribution of this paper is to formally solve the monopolist's and planner's dynamic problems using numerical simulations. As it turns out, this approach proves very useful, uncovering important regularities.³ We select the parameters (the discount factor, cost of the treatment and speed of disease transmission - and later on exogenous infection rate and drug resistance diffusion) within our model to meaningfully span the broadest range of qualitatively different cases.

Our key findings from this exercise are as follows. The monopolist's optimal price lies above the myopic price and the prevalence path typically converges to a non-zero steady-state. In contrast, the planner chooses to eradicate the disease if the cost of the treatment is low enough and the discount factor high enough. Consequently, welfare under mandatory universal treatment is generically very close to the first-best. Adding an exogenous infection rate to the model confirms the robustness of the planner's aggressive subsidization strategy: as long as the disease can be kept in check, prices remain below marginal cost even though eradication is then impossible.

Because a private firm would never extinguish an epidemic, drug resistance can exacerbate the welfare difference between the outcomes under a monopoly and a social planner. Yet, if the competing, negative externality of drug use outweighs its positive externality leading to overconsumption, the planner would need to price above marginal cost and the performance of the monopoly might dominate that under competition. We show this has important implications for the management of intellectual property rights here and also beyond the case of communicable diseases.

Economic theory has looked into communicable diseases in several ways. Brito *et al.* (1991) and Francis (1997) respectively analyze externalities associated with vaccines in a static environment with heterogeneous agents and under a dynamic

³“Undoubtedly, further progress will depend on more numerical work” (Gersovitz and Hammer, 2004, p. 25). See also Gersovitz and Hammer (2005) in the context of vector-borne infectious diseases (as opposed to those spread from person to person).

structure with homogeneous agents. Goldman and Lightwood (2002) apply cost minimization techniques in the canonical SIS epidemiology model. Barrett (2003) and Barrett and Hoel (2004) explore the issue of disease eradication through vaccines. Auld (2003) investigates the importance of rational expectations on the course of an epidemic. Gersovitz and Hammer (2004) provide a general framework for discussing the infection and prevention externalities that characterize communicable diseases and the role of government interventions. These studies have in common a market power-free environment.

Departing from that assumption, Geoffard and Philipson (1997) advance that a vaccine monopolist faces an unconventional incentive to keep a disease, and thus its market, alive. Their analysis shows that a steady-state of infection may be compatible with a constant price. However, the study left a number of issues unresolved, primarily the optimality of a constant price for the monopolist and of a constant subsidy for the planner. More recently, Kessing and Nuscheler (2005) reexamine the vaccine monopoly situation under income heterogeneity in a static framework and demonstrate how the monopolist may strategically leave the poor susceptible so as to increase the willingness to pay of the rich. Our approach builds on those works insofar as it considers treatment externalities under market power, and from an agent-based model characterizes the optimal prevalence and price paths in a dynamic framework and tells us more precisely what they depend on. It provides firm support for Geoffard and Philipson's (1997) main intuitions and for Kessing and Nuscheler's (2005) claim that their conclusions extend to a dynamic setting.

As for the growing literature on drug resistance (which may be thought of as a communicable disease within infectious micro-organisms),⁴ absence of market power is

⁴See Tisdell (1982), Doessel, (1988), Phelps (1989) and Fisher-Ellison and Hellerstein (1999), for the general approach and Laxminarayan and Brown (2001), Rowthorne and Brown (2002), and Laxminarayan and Weitzman (2002), for more in-depth studies of optimal allocation.

also usually assumed. Exceptions to this are Brown and Gruben (1997) who formally validate the generic intuition that a monopolist internalizes the externality when product effectiveness diminishes with use - as in the case of drug resistance, and Laxminarayan (2002) who calls for an extension of patent breadth for new antibiotics. We extend this literature by showing how a mixed competition/monopoly regime can perform better than competition or monopoly alone.

We defer to future work the analysis of a number of other interesting features, *e.g.*, we abstract from strategic buying arguments. Likewise, we do not consider quarantine-based policies or alternative allocation mechanisms such as drug cycling and treatment randomization. We also disregard the complex procedures by which a government agency idiosyncratically negotiates prices with different drug providers.⁵ Similarly we ignore the tension between R&D and markups to better focus on prices, profits, disease prevalence, and their welfare implications. Finally, while there may be no mechanism or set of incentives that would lead any agent (private or public) to achieve the first best outcome and while the pharmaceutical industry is usually oligopolistic, we believe that the social planning, monopolistic and competitive scenarios remain important benchmarks, which have been the focus of earlier work in economic epidemiology.

The rest of the paper is as follows. Section 2 lays out the analytical framework. Section 3 examines the dynamic problem of a treatment monopolist. Section 4 provides the counterpart analysis for a social planner. Section 5 discusses the welfare implications of the previous results. Section 6 revisits the model when there is an exogenous rate of infection or when the use of the drug spurs resistances. Section 7 concludes.

⁵See Horowitz and Moehring (2004).

2 Analytical Framework

2.1 Economic choice of the patient

In a population with size normalized to one, people have heterogenous preferences over being healthy. The value of being sick and untreated is normalized to zero. The taste parameter for being healthy θ is assumed to be distributed uniformly in the interval $[0, 1]$. The economic choice is between paying for a treatment or not.

Formally, letting u denote the utility function:

$$u = \theta \text{ (healthy) } \theta \in [0, 1]$$

$$u = 0 \text{ (sick untreated)}$$

$$u = \theta - p \text{ (treated)}$$

Note that the purchase of the treatment does not protect those who are healthy; therefore, only those who are sick might buy it. In other words, the individual gets treated if and only if she is sick at period t and the benefit exceeds the cost: $\theta > p_t$.

2.2 Biology

We now propose a simple transmission mechanism of the disease from one period to the next. We make a few key, simplifying assumptions. The disease strikes through person to person contacts upon random mixing. Sick people who did not purchase the treatment recover naturally within the period t .⁶ Reinfetion may occur immediately after so that neither treatment nor immediate natural recovery confers temporary immunity.⁷ Therefore contrary to vaccine models where demand may vanish altogether

⁶The model becomes drastically more complex if natural recovery from infection takes multiple periods or does not occur unless the patient is treated.

⁷Otherwise, the market value of being protected after being infected should enter people's willingness to pay for the drug. That value of being protected in the next period would depend on the probability of being infected in the future. Therefore, the *expected* prevalence would need to enter the formulation of the individual's current willingness to pay if we stick to perfectly rational and forward looking agents. I thank an anonymous referee for this remark. See Auld (2003) for a model where expectations over the course of an epidemic affect current behavior and therefore the spread of the disease itself. This is precisely the complication we avoid here.

under a given prevalence path (Geoffard and Philipson, 1997), there is always a positive demand for treatments for any positive prevalence. Yet most importantly, similar to a vaccine model, a patient who has bought the treatment at t cannot transmit the disease to others at $t + 1$, so it is in that sense that the economic choice influences its spread. In other words, anybody may become sick in any period, regardless of the past, but prevalence at $t + 1$ is still a function of the proportion of those sick who did not buy treatments in period t . This is the fundamental mechanism.

Our setup is clearly stylized and does not reproduce the exact transmission mode of any specific disease. Rather, we would like to capture important general features of communicable diseases diffusion under the constraint of analytical or computational feasibility. However, we can still interpret the framework concretely. Perhaps the simplest way to envision our problem is to think of a treatment for usually mild viral ailments like the common cold and influenza strains for which there are no vaccines, or bacterial illnesses such as strep throat, impetigo or the scarlet fever (group A streptococcal infections). The medication would alleviate the symptoms, accelerate recovery and prevent the transmission of the disease, but not a future relapse.⁸

Formally, let the proportion of sick be r and let $\gamma(\cdot)$ be the transmission function.⁹ Instead of considering a simple parameter γ , we assume γ to be endogenous, reflecting the idea that people take precautions to avoid a disease or that the disease affects certain categories of the population with more difficulty, perhaps because they are healthier. One simple way to do this is to assume $\gamma = \gamma(r_t)$. Intuitively, one would expect that if prevalence is close to zero, then $\gamma(0) = \gamma$ *i.e.*, the disease propagates naturally. On the other hand, the higher the prevalence the slower its spread, *i.e.*,

⁸Other diseases fit the model to various extents. Some kinds of metapneumovirus and rotavirus related illnesses share the main characteristics of person to person spread, natural recovery in most cases and possible reinfection shortly after.

⁹ γ may be interpreted as a policy instrument if we think of reducing the spread of a communicable disease through mandatory quarantine. However, we defer the analysis of quarantine policy to a further paper.

$\gamma' < 0$.¹⁰

2.3 Drug price-based disease transmsion

Define the indifferent patient by $\theta^* = p$. θ^* represents the proportion of the r sick who do not get treated. Therefore, $\gamma(r)rp$ is the size of the market in the next period and it depends on the key biological and economic parameters in a simple way. We therefore posit:

$$r_{t+1} = \min(\gamma(r_t)r_t p_t, 1) \tag{1}$$

This guarantees that prevalence never exceeds 100%.

Note that the above assumptions limit the smoothness of the prevalence path.¹¹ This is clearly because transition from sick state into healthy state is very fast. However, those assumptions eliminate extreme complications while keeping the necessary ingredients of a communicable disease transmission model, now combined with a typical industrial organization, agent-based structure.

3 Firm Problem

3.1 Two-period Case

This section is mainly illustrative and shows the mechanics of the model analytically. Let δ be the discount factor and let c be the cost of producing the treatment, with $0 < \delta < 1$, $0 < c < 1$. For simplicity in this two-period case, $\gamma(r) = \gamma$. Here and throughout the paper I assume no price discrimination. Let us assume that after the second period, we have competitive pricing so there are no profits to be made. The

¹⁰To keep things simple and to focus on the decision to buy treatments alone, conditional on being sick, we do not to model prophylactic behavior explicitly.

¹¹Prevalence swings unrealistically from one period to the next if the transmission function γ takes too high values. Attention is therefore limited to low values of γ .

firm solves:

$$\max_{p_1} \{r(p_1 - c)(1 - p_1) + \delta \max_{p_2} (p_2 - c)(1 - p_2)(\gamma r p_1)\} \quad (2)$$

Here, we assume for simplicity that the parameters are such that the condition $\gamma r p_1 < 1$ is satisfied. Solving backward:

$$p_2 = (1 + c)/2 > c \quad (3)$$

(myopic, or one-shot monopolistic pricing)

Therefore, the firm problem is

$$\max_{p_1} (p_1 - c)(1 - p_1) + \delta \gamma (1 - c)^2 p_1 / 4 \quad (4)$$

This leads to

$$p_1 = [1 + c + \delta \gamma (1 - c)^2 / 4] / 2 > p_2 \quad (5)$$

The firm prices above the one-shot monopoly price because of the negative externality of treatment - from the firm's perspective - on its future market.¹²

3.2 General Case

Solving for the monopolist optimal path in a three-period framework is relatively straightforward but shows the inherent intractability of the analytical approach beyond two periods. Yet, the multi-period case offers considerably more insights. Loosely speaking, the monopolist must find a path where current profit does not compromise future discounted profits too much. Here we are interested in solving the above problem for an arbitrary number of periods and in characterizing in which cases a firm would settle to a non-degenerate steady-state.

¹²Note that the result would not change if sick untreated people in period one were still sick in period two because on the average, people's willingness to pay would be lower, as in the standard durable good model; hence, the two effects would go in the same direction.

3.2.1 Framework

With a T period horizon the problem is written formally:

$$\text{Max}_{p_t} \sum_{t=1}^T \delta^t r_t (p_t - c)(1 - p_t) \text{ such that } r_{t+1} = \min(\gamma(r_t)r_t p_t, 1) \quad (6)$$

We solve the problem through the Bellman equation approach:

$$\begin{aligned} V_t(r_t) &= \max\{r_t(p_t - c)(1 - p_t) + \delta V(r_{t+1})\} \\ \text{such that } r_{t+1} &= \min(\gamma(r_t)r_t p_t, 1) \end{aligned} \quad (7)$$

given an initial r_0 and boundary condition $V_T(r_T) = [(1 - c)^2/4]r_T$.

The Bellman approach approximates the domain of the state variable with a grid and discretizes the objective function and law of motion. The procedure solves backwards a system $V_t(x_t) = \max[v_t(x_t, u_t) + \delta V_{t+1}(f(x_t, u_t))]$, where $v_t(x_t, u_t)$ is the objective function and $x_{t+1} = f(x_t, u_t)$ is the law of motion. This method is adapted to our problem because it can be applied to any function $v_t(x_t, u_t)$ and $f(x_t, u_t)$ and performs very well in dimension one or two.

The simulations results are presented for a fixed number of periods (50): T periods of optimization correspond to T iterations of the contraction mapping and therefore it is a good approximation of the infinite horizon model.¹³

3.2.2 Results

There are no analytic solutions for this problem, so therefore we carried enough simulations to encompass all relevant, qualitatively different cases. A natural candidate

¹³The convergence is exponential and the smaller the discount factor, the better the convergence. The problem exhibits the turnpike property (Intriligator, 1971, chapter 16) : the difference with the infinite horizon model occurs only as t approaches T so the conclusions of the finite horizon simulation extend to an infinite horizon.

for $\gamma(r)$ is $\gamma(r) = \gamma(1 - r)$ where γ is a parameter to be chosen.¹⁴ We have three free parameters (δ, c, γ) ¹⁵ and for each triplet selection, we show in the illustrative graphs the paths corresponding to three starting values of the prevalence r_0 . We restricted attention to high values of the discount rate (with an emphasis on the case $\delta = 0.95$ where a period can be interpreted as a year) since low discount rates can hardly be assigned any realistic economic meaning here. For brevity, we include in the appendix only representative pictures of qualitatively different cases. In each picture, the different paths drawn correspond to different starting points r_0 .

Several main features emerge from the simulations. The most important one is undoubtedly the fact that the monopoly price and prevalence paths typically converge to a unique steady state (figures 1a and 1b).¹⁶ It echoes Geoffard and Philipson's (1997) assertion that no cycle can occur in a vaccination model as long as demand responds to current - as opposed to lagged - prevalence. Another point is that the monopolist keeps prices above the myopic (or one-shot) price $(1 + c)/2$, just as in two-period case. The intuition for this result is the same as before. In Kessing and Nuscheler's (2005) model, the monopolist creates a high prevalence among the poor so that the rich be more likely to buy vaccines. Here, because we are dealing with treatments, this mechanism is even more direct: the monopolist prices high so that a lot of poor do not buy the treatments and increase the chance that some rich will be infected subsequently.

The starting prevalence is relevant : starting prices can be increasing or decreasing,

¹⁴Of course this implies $r_{t+1} \rightarrow 0$ as $r_t \rightarrow 1$ and bumpy paths for high values of γ ; this is why attention is limited to low values of γ . However, we have checked that smoother functions $\gamma(r)$ with $\gamma(0) = \gamma$ and $\gamma' < 0$ - such as $\gamma(r) = \gamma \exp(-r_t)$ - keep the key features of the solution. Further, it is worth noting that if we assume $\gamma(r) = \gamma$ for most values of γ the monopolist prices so as to make nearly everyone sick in every period.

¹⁵The different choices of γ are not derived from any known biological model and in that sense can be considered *ad hoc*.

¹⁶As the pictures show, given a certain triplet (δ, c, γ) , depending on the initial r_0 , the different paths converge more or less rapidly to the same steady-state.

as opposed, this time, to the two-period model. As γ increases, the convergence to the steady state is quicker and the prices get closer to the myopic monopoly price (figures 2a and 2b). The future market plays a less important role in the firm's strategy. High values of γ should be thought of as degenerate cases.

Increases in δ have little qualitative effects on the prevalence and price paths in general. As c increases, steady-state infection levels increase, prices increase, and obviously the difference between the optimal price and the myopic price decreases.

4 Public Intervention

4.1 Two-period Case

The rationale for intervention is the standard argument on the positive externality from protection, treatment, or vaccination. The study of the planner's behavior in a finite horizon case helps understand the issue at stake.

In this case, it is obvious that in the final period, price equals marginal cost because there is no positive externality involved. For simplicity in this two-period case, again we assume $\gamma(r) = \gamma$.

The government's problem is

$$\begin{aligned} \text{Max}_{p_1} r \int_{p_1}^1 \theta d\theta + (1-r)/2 + \delta[\gamma p_1 r \int_c^1 \theta d\theta + (1-\gamma p_1 r) \int_0^1 \theta d\theta] \\ -rc(1-p_1) - \delta\gamma p_1 r(1-c)c \end{aligned} \quad (8)$$

The solution is:

$$p_1 = \delta\gamma(c^2/2 - c) + c < c = p_2 \quad (9a)$$

Therefore, the social optimum price in period 1 is lower than the marginal cost. The intuition is the exact reverse of the argument given for the monopoly. Hence the distortion incurred by monopolistic pricing is decreasing over time in the two-period

case.¹⁷

4.2 General Case

We are mainly interested in characterizing those situations when a planner would choose to eradicate the disease.

4.2.1 Framework

With a T -period horizon the problem is written formally:

$$Max \sum_{t=1}^T \delta^t \{r_t [\int_{p_t}^1 (\theta - c) d\theta] + (1 - r_t)/2\} \text{ such that } r_{t+1} = \min(\gamma(r_t)r_t p_t, 1) \quad (10)$$

The same approach as above is used to solve this problem. Note that for the case $\gamma(r) = \gamma(1 - r)$ the disease is eventually eradicated if $c < 1/\gamma$ (recall $p_{planner} \leq c$) since it implies $r_{t+1} < r_t(1 - r_t)$ and for any ζ such that $0 < \zeta < 1$ the series $\{r_t\}_{t \geq 0}$ defined by $r_{t+1} = \zeta r_t(1 - r_t)$ converges to zero (for any $r_0 < 1$).

4.2.2 Results

The most striking aspect that emerges from the simulations is that for low enough values of the marginal cost c , the planner always achieves eradication within a few periods (figures 3a and 3b). For higher values of c , other considerations may take place. In the limit, if the future is highly discounted (low δ), then the planner would settle to a nonzero steady-state. In that case, the higher γ , the closer the price to marginal cost: we conclude that subsidization is fruitless when the disease is spreading so quickly that the planner has no control over prevalence. However subsidization does not only make sense if the planner can achieve eradication: we found that for low δ and low γ , the planner may settle for a non-zero steady-state with a price path

¹⁷Note that it has been implicitly assumed that taxation is neutral (*i.e.*, no welfare loss from raising revenue), since the positive difference $c - p_1$ must eventually be financed through taxes.

significantly below c . Although we could only produce this last result for low (and therefore, from a practical standpoint, implausible) values of δ , within this version of the model, we will see later it has valuable policy implications.

Finally, for high enough δ , we find that the disease is always eradicated: the present value of a disease-free world dominates all other factors. This aggressive subsidization strategy leading to eradication is on a par with Barrett and Hoel's (2004) intuition that a permanent high level of disease control cannot be optimal: the dividend of eradication for future generations is worth the short run cost of a big push. This outcome is also to be contrasted with Goldman and Lightwood's (2002) conclusion that in the SIS model, decreased levels of treatment are optimal for higher disease levels. As the authors put it, this appears as a consequence of the increased likelihood of reinfection in the presence of a larger infected proportion of the population. In this paper, we do have an extreme version of that proposition, but it only holds when the disease is out of control.¹⁸ Given the difficulty of eradicating a disease through vaccination with Pigouvian subsidies (Geoffard and Philipson, 1997) the analysis suggests that eradication with treatments may be easier to achieve than through vaccination alone if those treated do not spread the disease.

5 Welfare comparisons

5.1 Two period Case

It is clear that the planner's allocation strictly dominates the others (once again, if taxation is non distortionary and there is no excess cost of raising revenues), and that the competitive equilibrium dominates the monopoly's (because of the presence of the positive externality, combined with the usual welfare loss from market power).

¹⁸However, the optimal level of treatment is certainly sensitive to assumptions regarding length of immunity and policy makers should pay particular attention to that feature of the disease transmission mechanism.

However, it can be easily shown that the comparison between the mandatory universal treatment policy - being defined as one where all the sick are immediately cured (at marginal cost) - and the monopoly is ambiguous. Similarly, the ambiguity is also present when comparing perfect competition with the mandatory treatment case.

5.2 General case

We are interested in welfare comparisons between the monopolist, the planner, the competitive and the mandatory universal treatment cases. The results are presented for all starting possible values of infection levels r_0 . It is important to keep in mind that the scale chosen depends on the parameter δ .

In most cases, welfare under mandatory universal treatment and under social planning are very similar (figures 4, 5, 6). This should not be surprising since eradication is almost always optimal and achieved rapidly. Mandatory treatment is less likely to be close to optimal when δ is low and c is high, and the difference becomes more significant when the initial prevalence is high enough (figure 4). In that case, mandatory treatment is too expensive. It can indeed be less efficient than the competitive outcome and even the monopoly outcome (figures 4, 5). Surplus under mandatory treatment is always a linear function (contrary to surplus under different assumptions) since it has the effect of curing everybody immediately and therefore escapes the influence of the law of motion of the disease. The difference between surplus under monopoly and under perfect competition decreases with c (figure 6). Note that the U shape welfare curves arise only because a very high r_0 implies a very low r_1 .¹⁹

Finally, we checked for consistency purposes that welfare under perfect competition always lies between monopoly and social planner's levels. The lower the marginal

¹⁹I thank an anonymous referee for pointing it out. Other specifications of $\gamma(r)$ without the $r_{t+1} \rightarrow 0$ as $r_t \rightarrow 1$ feature show a strictly decreasing surplus.

cost c , the more welfare levels under competition get closer to the planner's, and obviously welfare under competition dominates monopoly outcomes in all situations.

6 Extensions

6.1 Exogenous Rate of Infection

For many diseases, eradication is simply impossible. This would happen if the pathogens responsible for them are airborne or so prevalent in the environment, for example through animal vectors, that it would require unrealistic sterilization measures.²⁰

6.1.1 Adding an exogenous component to the equations

Among the non-sick $(1 - r_t)$ at time t , we have a proportion $\gamma(r_t)r_t p_t$ that is going to be infected through the mechanism described in the previous section, and among those not infected through that mechanism, a proportion ϕ infected exogenously. It translates as:

$$r_{t+1} = \min(\gamma(r_t)r_t p_t + \phi(1 - \gamma(r_t))r_t p_t, 1) \tag{11}$$

$$\Leftrightarrow r_{t+1} = \min((\gamma_0 + \gamma_1(r_t))r_t p_t, 1) \text{ where } \gamma_0 = \phi \text{ and } \gamma_1(r) = (1 - \phi)\gamma(r)$$

6.1.2 Comparison of the results

It is most interesting to see if the planner still follows the same strategy, now that the final eradication outcome is unattainable by construction. Overall, the answer is that yes, the planner continues to heavily subsidize the treatment even if the steady-state of infection that follows is significantly different from zero (figures 7a,

²⁰This is true even in hospitals where hospital-acquired infections are responsible for thousands of deaths each year. See Gersovitz and Hammer (2005) for an analysis of pure vector-borne infections in developing countries. We do not consider an endogenous population of vectors here.

7b). Consequently, the planner's steady-state is close to γ_0 . Consistent with earlier results, the planner stops subsidizing and sets prices at marginal cost when it loses control of the infection. Only then is subsidization fruitless. This corresponds to cases when the discount factor δ is small and when either the exogenous parameter of infection γ is high or the cost c is high, or both. The intuition for the difference between our results and Goldman and Lightwood's (2002) remains the same.

The monopoly solution does not add additional qualitative insights. Now with an exogenous prevalence, the monopolist is able to maintain a steady-state of infection and follows the strategy described earlier. As for the welfare comparisons, evidently, the mandatory treatment policy (meaning in this case that everybody is treated each period) will be dramatically less efficient the higher the cost c and therefore more often dominated by the other outcomes (figure 8).

6.2 Resistances

We theoretically do not need communicable diseases to analyze drug resistance. However, because resistance can be conceptualized as the negative externality of current treatment on future treatment, it makes sense to study the phenomenon within the framework we have used so far.

Treatment effectiveness is a natural, economic resource. Similar to the notion of a stock that is being depleted, current treatment use lowers future effectiveness.²¹ Pathogens inevitably acquire resistances to survive treatments through natural selection. Resistant strains are usually not treatment-induced mutants. Rather, they often exist naturally but in insignificant proportions due to a so-called fitness deficit (or cost) compared to other germs of the same category.²² However, as treatments

²¹See for example Baumol (1996), Brown and Layton (1996), Cohen (1992), Levy (1997) etc.

²²See *e.g.*, Munroe (1997), for the same idea in the context of pest resistance.

destroy the susceptible strains, the proportion of resistant types increases.²³ The dramatic increase in the percentage of staphylococci strains that show resistance to *e.g.*, penicillin, tetracycline, and chloramphenicol²⁴ justify that firms must take drug resistance into consideration.²⁵

6.2.1 Modeling Resistance

We build on Fisher-Ellison and Hellerstein (1999) who use a transparent two-period model to allow for the externality of drug use. The following model assumes that in each period t , the use of the treatment destroys an endogenous proportion H_t of the non-resistant strains. The “hole” created in the flora is filled through the reproduction of the survivors, which now contain a higher proportion of resistant types. The equation governing the proportion of resistant strains Ω is thus given by:

$$\Omega_{t+1} = \min[\Omega_t[1 + (1 - \Omega_t)H_t], 1] \text{ with } \Omega_0 = \epsilon \quad (12)$$

Note that since H is a proportion we have $\Omega_{t+1} \geq \Omega_t$ and $\Omega_t \leq 1$.

The original steady-state proportion, absent treatment is ϵ . If H is constant, Ω is a sigmoid function, which is consistent with the biological literature on the topic (Austin *et al.*, 1999).²⁶

Here, we present results for the case where patients buy the treatment only if it is efficient, that is, if they do not carry the resistant type (which we assume can be found out subject to a costless test). The results are not qualitatively modified if

²³When treatments are suspended, the proportion of resistant types is theoretically supposed to move back to its initial, infinitesimal steady-state level. Yet, it is not known if such a reversal trend is a valid stylized fact at the aggregate level - see Levy (1997). For simplicity we ignore the modeling of fitness deficit.

²⁴Daily and Ehrlich (1996). See also Fisher (1994).

²⁵See Kline (1989).

²⁶In theoretical biology, modeling resistance through differential equations is the norm, but there is no commonly adopted law of motion. Ours is a very simplified and stylized approach.

we change the problem to allow patients who carry the resistant strain to buy the treatment. Note that the growth of resistance is not affected by either assumption.

It is not easy to find a compelling function H . An intuitive start, however, is to assume that the proportion of non-resistant strains destroyed is proportional to the number of treated individuals, *i.e.* those sick patients who do not have the resistant strain and are willing to pay for the treatment, so that :

$$H_t = \min[hr_t(1 - p_t)(1 - \Omega_t), 1] \text{ where } h \text{ is a constant.} \quad (13)$$

The maximization problems above are therefore modified by letting the population who buys the treatment at time t be: $r_t(1 - p_t)(1 - \Omega_t)$ and by adding the equation for resistance as an additional constraint of the monopolist's and planner's problems. Again, we choose different values for h to span as broad a range of qualitatively different cases as possible.

6.2.2 Results with resistance

It turns out that the main results do not differ qualitatively from when there is no resistance. 3D graphs, available upon request, can account for the different parametrizations of h in the former cases examined. The major discrepancy is that in the majority of cases where the planner achieves eradication, the difference in welfare is larger since eventually monopoly pricing leads to full resistance.

A key question is whether monopoly pricing ever yields higher welfare than competitive pricing. Intuitively, it is possible since we now have two external effects going in opposite directions while it is ambiguous which one dominates.²⁷ However, quite surprisingly, we find that the present value of welfare under monopoly is never higher

²⁷To assess this issue we only need to examine high values of the discount factor δ since otherwise the present value of welfare under competition would necessarily be higher than under monopoly, even if resistance is better controlled by a monopoly. Also, with a high δ the monopoly is more likely to internalize the resistance path in its pricing.

than under competition, except when h is very high, in which case resistance almost immediately increases to 100% and the treatment becomes obsolete. This is of course a pathological case, which indicates that the positive externality of treatment generically dominates the negative externality of resistance in our model. It goes without saying that this result would not necessarily (and actually would less likely) hold if we analyzed the resistance phenomenon for non-communicable diseases. There, the tension would simply arise from classical, market power-induced deadweight loss vs. negative externality internalization.²⁸

Still, we focus on finding cases, within our framework, where fine tuning the introduction of monopoly pricing improves welfare. If we characterize such situations, the conclusions would be all the more relevant for other diseases where there are no positive treatment externalities present. It turns out that introducing monopoly power after a period of competitive pricing may outperform competition in all periods. The idea for correctly introducing monopoly power is to choose periods when the difference between optimal pricing (the planner's choice) and competitive pricing is greatest. In such cases, the monopolist's price is closer to the planner's than to the competitive price, and the negative externality of resistance is so much better handled by the monopolist that it more than compensates for the loss of positive externality effects under competitive pricing.²⁹

An algorithm to find when to best introduce monopoly would be to compare welfare outcomes between $\sum_{p=0}^T C_T^p$ combinations of monopoly pricing periods within T periods. By solving the optimization problem backward choosing at each stage between competitive price and the optimal monopoly price, we can find a competitive

²⁸See Brown and Gruben (1997): their natural conclusion is that a monopoly may do better than competition when the welfare cost of failing to protect product effectiveness is greater than the welfare cost of market power.

²⁹However, the mixed regime does not improve on competition as much if we allow all patients to buy the treatment because the monopoly does not have an incentive to control resistance then.

surplus that is less than the optimal mixed regime surplus. This finding is important because it opens the possibility of adding patent life years for the inventor of a new drug - the only incentive for private R&D - yet without harming welfare *ex post*.

Patent policy usually does not differentiate between different types of discoveries. Yet, within the pharmaceutical industry, we may think about implementing patent structures specific to antibiotics, an idea explored by Laxminarayan (2002) who analyzes patent breadth. Here, we propose a patent scheme where patent duration is based on resistance levels. We refine the proposal of the now defunct U.S. Office of Technology Assessment (1995) which advocated extending antibiotic patent length to let patentees internalize the externality costs of resistance. Such an extension needs not kick in when the patent expires only.

Realistically, if the planner regrants monopoly rights after patent expiration, it will presumably be on the same model as the initial patent, *i.e.*, a certain time interval of monopoly power. In other words, after its initial patent expiration, a monopolist would get an extension on an interval $[t_1, t_2]$. The planner would then choose t_1 and t_2 to maximize welfare:

c at the interval $[0, t_1]$
 $p^{mix} = \{ p^{mon}$ at the interval $[t_1, t_2]$, where the price p^{mon} is the optimal monopoly price on the interval $[t_1, t_2]$
 c at the interval $[t_2, T]$

We find that welfare under pure competition is less than under the switching regime just described. Figure 10a provides a graphical illustration for the optimal choice of parameters (t_1, t_2) and figures 10b and 10c offer a comparison of the price and resistance paths under the competitive, monopoly and mixed regimes.

Obviously, the initial pricing phase should be monopolistic: an optimal patent scheme should maximize the *ex ante* welfare, that is taking into consideration the impact of expected discounted profits on R&D effort, which should therefore be rewarded early on. However, patent life may be covering not only the initial pricing

phase but also later phases, presumably when the sigmoid path of resistance is steepest. That is, when the externality of resistance is such that welfare under monopoly is closer to first best than to welfare under competition. Given that profits are discounted, if that period happens too far away in the future after regular patent expiration, it would not change existing patent structure much in terms of R&D incentives. Still, theoretically, we can envision a shorter initial patent to be reconducted at a later time: in other words, the initial patent phase would stop when the present value of profits under the (usual) remaining patent phase equals the present value of future profits under regranted patent, later on - when resistance is high enough. However, in practice, given that cutting on initial patent time is probably unrealistic, we suggest that patents for antibiotics follow a simple rule of thumb: keeping the initial patent length - as for any other drug - and regranted intellectual property rights to the innovator when the growth of resistance exceeds a certain threshold.

7 Concluding Observations

This work is a contribution to the economics of communicable diseases and drug resistance. We build a stylized model of disease transmission and conduct numerical simulations under different market and epidemiological environments. In particular, the paper characterizes a drug monopoly and a social planner's pricing strategies, their corresponding prevalence paths and welfare outcomes. The analysis expands with the introduction of exogenous infection rates and drug resistance. The main results are that the monopolist sets prices to reach a steady-state of infection, while the planner generally eradicates the disease, or subsidizes treatments when eradication is impossible or too costly, as long as it retains some control over prevalence. In the presence of drug resistance, regranted intellectual property rights for some time after initial patent expiration and a phase of competitive pricing can prove desirable when

the growth of resistance exceeds a certain threshold, thus providing both higher *ex post* welfare and greater incentives for private R&D.

The policy implications of this work are multiple. Since first best pricing generally leads to rapid eradication, should it be cost effective, public health decision makers should not hesitate to implement mandatory treatment programs for communicable diseases.³⁰ When a drug is under patent, knowing which steady-state price the firm is aiming at should help social security officials in bargaining over the determination of the drug price and of its coverage over time. Regarding the design of patents, this work shows that implementing original, resistance level-based patent schemes may be desirable. Independent monitoring of the evolution of drug resistance and cooperation of generics producers are therefore critical to the success of this idea.

Many extensions of this work are conceivable, for example by considering that pharmaceutical development is characterized by large fixed costs which subject price paths to a break-even constraint. In particular, an overlooked aspect of R&D on communicable diseases and drug resistance is that private innovation is only stimulated when there is a large enough demand so that a firm may earn positive profits.³¹ This may result in diverting R&D from slow growth diseases, in the planned obsolescence of previous drugs or in a firm sleeping on its patent if a communicable disease or resistance spreads fast enough.³²

³⁰The 2003 SARS epidemic exemplified the necessity of sanctions for people with symptoms but unsure of their condition or afraid to go to the hospital - the most dangerous places to be at the time.

³¹It can even be proved that if a new drug is brought to the market only once a certain resistance level for the old drug is reached, and if resistance grows asymptotically to that level, a policy discretely *increasing* resistance would be desirable at some point.

³²Implausible as it may sound, this seems to have happened in a related context when a major anti virus software company apparently withheld information about at least one big cyberthreat for hours after spotting it, possibly harming millions of Internet users.

References

- Auld, C. 2003. Choices, beliefs, and infectious disease dynamics. *Journal of Health Economics* 22: 361-377.
- Austin, D. J., K. G. Kristinson, and R. M. Anderson, 1999. The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance, *PNAS* 96: 1152-1156.
- Baumol, W. J., 1996. Antibiotics. Overuse and Other Threats. *Environment and Development Economics* 1: 346-49.
- Barrett, S. 2003. Global Disease Eradication. *Journal of the European Economic Association* 1: 591-600.
- Barrett, S. and M. Hoel. 2004. Optimal Disease Eradication. FEEM Working Paper No. 50.04.
- Brito, D., E. Sheshinski, and M. D. Intriligator, 1991. Externalities and compulsory vaccinations. *Journal of Public Economics* 45: 69-90.
- Brown G. and D. F. Layton, 1996. Resistance economics: social cost and the evolution of antibiotic resistance. *Environment and Development Economics* 1: 349-355.
- Brown, S. P. A. and W. C. Gruben, 1997. Intellectual property rights and products effectiveness. *Federal Bank of Dallas Economic Review* (4): 15-20.
- Cohen, M.L.1992. Epidemiology of drug resistance: implications for a post-antimicrobial era. *Science* 257: 1050-1055.
- Daily, G. C. and P. R. Ehrlich, 1996. Impacts of development and global change on the epidemiological environment, *Environment and Development Economics* 1: 311-346.
- Doessel, D. P.1998. The “ Sleeper ” issue in medicine, *International Journal of Social Economics* 25: 956-67.

Fisher, J. A. 1994, *The Plague Makers, How we are creating catastrophic new epidemics and what we must do to avert them* Simon & Schuster.

Fisher Ellison, S., and J. K. Hellerstein, 1999. The Economics of antibiotics: An Exploratory Study. in *Measuring the Prices of Medical Treatments*, ed. J. Triplett, The Brookings Institution.

Francis, P.J. 1997. Dynamic epidemiology and the market for vaccination. *Journal of Public Economics* 63: 383-406.

Geoffard, P-Y. and T. Philipson, 1997. Disease Eradication: Private versus Public Vaccination. *American Economic Review* 87: 222-30.

Gersovitz, M. and J. S. Hammer., 2003. Infectious Diseases, Public Policy, and the Marriage of Economics and Epidemiology. *The World Bank Research Observer* 18: 129-157.

Gersovitz, M. and J. S. Hammer. 2004. The Economical Control of Infectious diseases. *Economic Journal*, 114: 1-27.

Gersovitz, M. and J. S. Hammer. 2005 Tax/subsidy policies toward vector-borne infectious diseases', *Journal of Public Economics* 89: 647-674.

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): Article 4.

Horowitz, J. and B. Moehring. 2004. How Property Rights and Patents Affect Antibiotic Resistance, *Health Economics* 13: 575-583.

Intriligator, M. 1971. *Mathematical Optimization and Economic Theory*. Prentice Hall.

Kessing, S. and R. Nuscheler, 2005. Monopoly Pricing with Negative Network Effects: the Case of Vaccines, forthcoming *European Economic Review*.

Kline, J. D. 1989. Research and Development in the Pharmaceutical Industry:

the impact of Drug Resistance, PhD thesis, University of Wisconsin Madison.

Laxminarayan, R., and G. M. Brown, 2001. Economics of antibiotics resistance: A theory of optimal use. *Journal of Environmental Economics and Management* 42: 183-206

Laxminarayan, R. and M. L. Weitzman, 2002. On the implications of endogenous resistance to medications *Journal of Health Economics* 21: 709-718

Laxminarayan, R., 2002. How broad should the scope of antibiotics patents be? *American Journal of Agricultural Economics* 84: 1287-92.

Levy, S. B. 1992. *The Antibiotic paradox: How Miracle Drugs are Destroying the Miracle*. New York: Plenum Press.

Levy, S. B. 1997. *Antibiotic resistance : origins, evolution, selection, and spread*. Chichester. New York: J.Wiley.

Munro, A., 1997. Economics and Biological Evolution, *Environment and Resource Economics* 9: 429-449.

Office of Technology Assessment, 1995. Impact of Antibiotic-resistant Bacteria: A Report to the U.S. Congress. Washington, DC: Government Printing Office.

Phelps, C.E. 1989. Bug/drug resistance: Sometimes less is more. *Medical Care* 27: 194-203.

Philipson, T. 2000. Economic Epidemiology and Infectious Diseases, in *Handbook of Health Economics*, chap 33, Elsevier Science.

Rowthorne, R and G. Brown. 2002. Using Antibiotics when Resistance is renewable, in *Battling Resistance to Antibiotics and Pesticides: An Economic Approach*, edited by R. Laxminarayan, Resources for the Future Press.

Tisdell, C., 1982. Exploitation of techniques that decline in effectiveness with use, *Public Finance/Finance Publiques*, vol XXXVII: 428-37.

Appendix

Monopolist case
($c=0.3$; $\delta=0.95$; $\gamma=2$)

Monopolist case
($c=0.3$; $\delta=0.95$; $\gamma=4$)

Social planner case
($c=0.3$; $\delta=0.95$; $\gamma=4$)

Prevalence paths

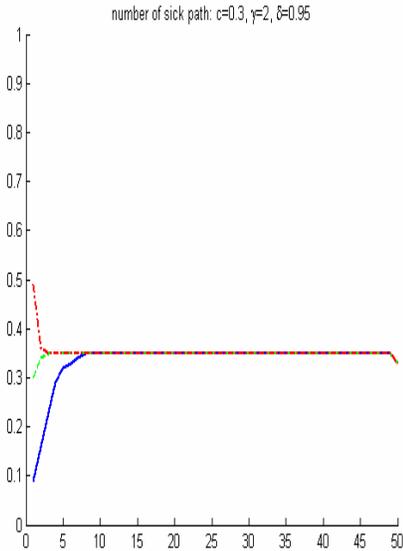


Figure 1a

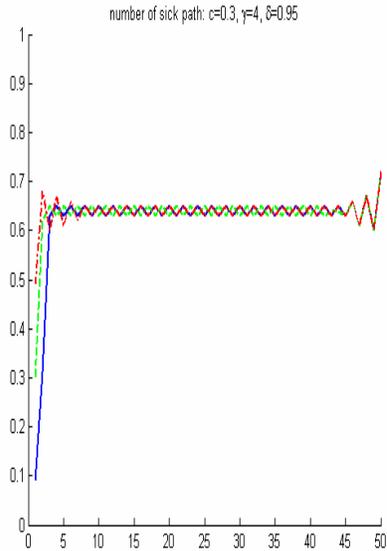


Figure 2a

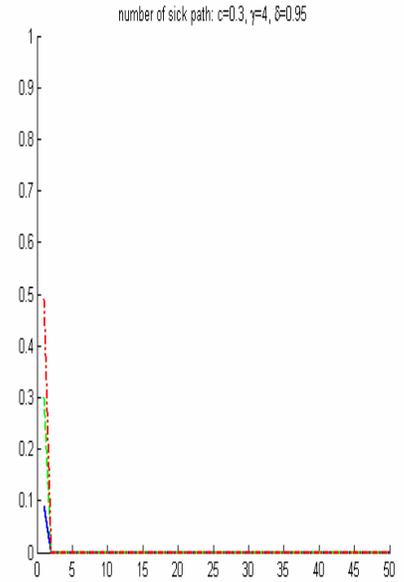


Figure 3a

Price paths

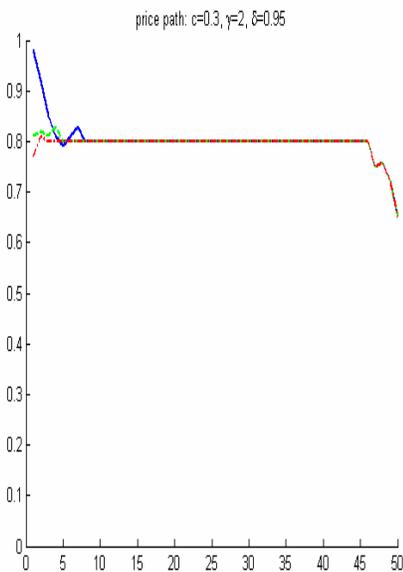


Figure 1b

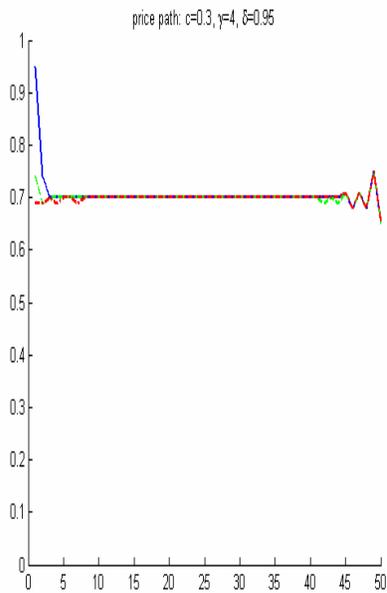


Figure 2b

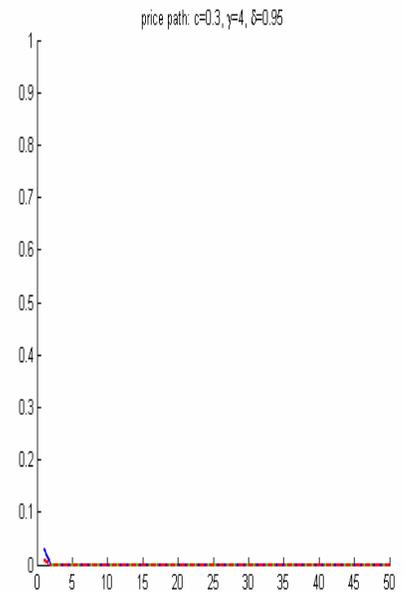


Figure 3b

Welfare Comparisons

dash-dot line: social surplus under mandatory treatment

dashed line: social surplus in monopoly case

dotted line: social surplus under competitive case

solid line: social surplus under social planning

$(\delta=0.1; c=0.3, \gamma=2)$

$(\delta=0.95; c=0.3, \gamma=2)$

$(\delta=0.95; c=0.6, \gamma=2)$

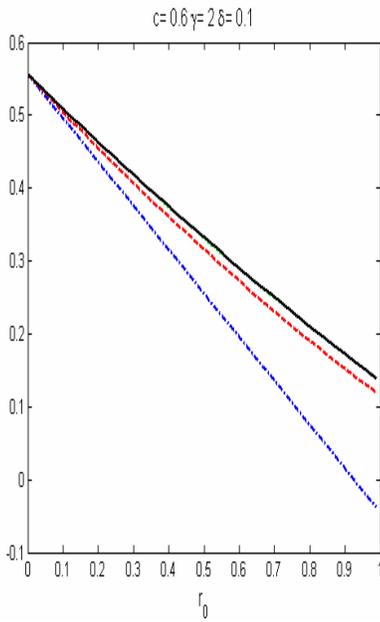


Figure 4

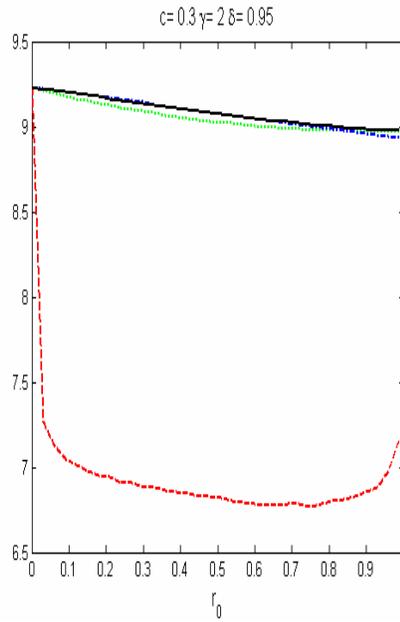


Figure 5

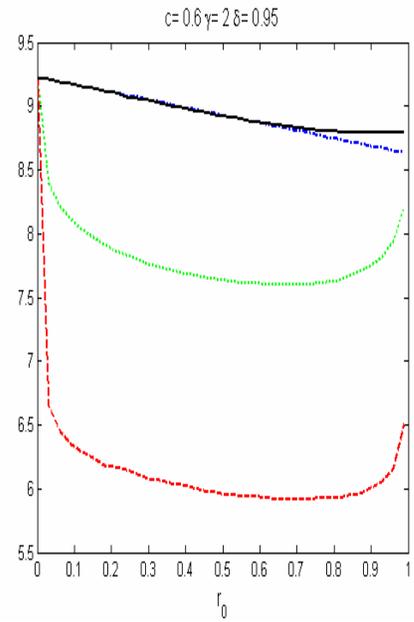


Figure 6

Adding an exogenous parameter of infection

Planner's basic choices

$(\delta=0.95; c=0.3, \gamma_0=0.2; \gamma_1=2)$

Prevalence paths

Price path

Welfare Comparisons
 $(\delta=0.95; c=0.5, \gamma_0=0.5; \gamma_1=1.5)$

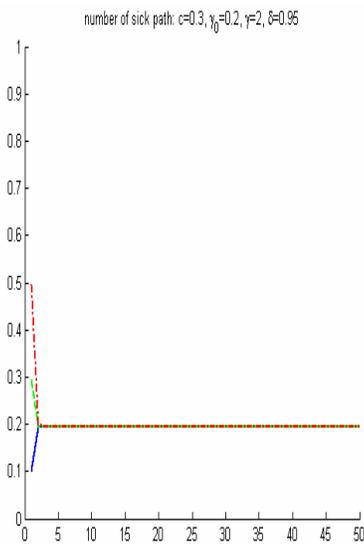


Figure 7a

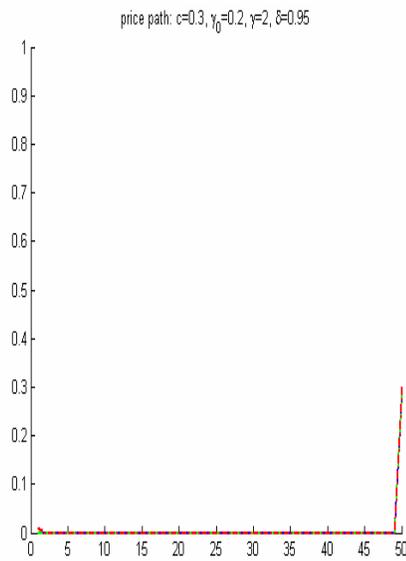


Figure 7b

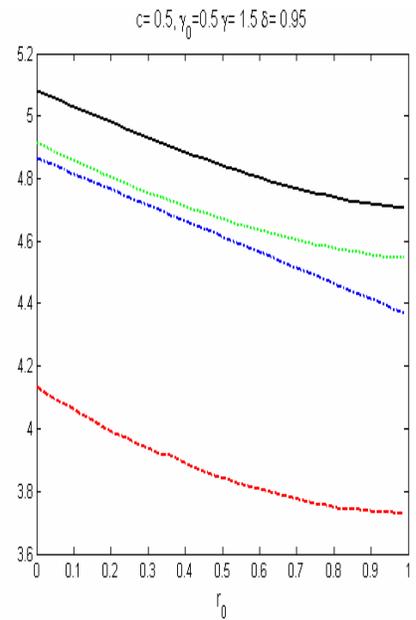


Figure 8

A case where a mixed monopoly/competition regime outperforms competition alone:

finding $[t_1^*, t_2^*]$ such that welfare is optimal with c on the interval $[0, t_1^*]$, p^{monopoly} on the interval $[t_1^*, t_2^*]$ and c on the interval $[t_2^*, T]$

Welfare under mixed regime

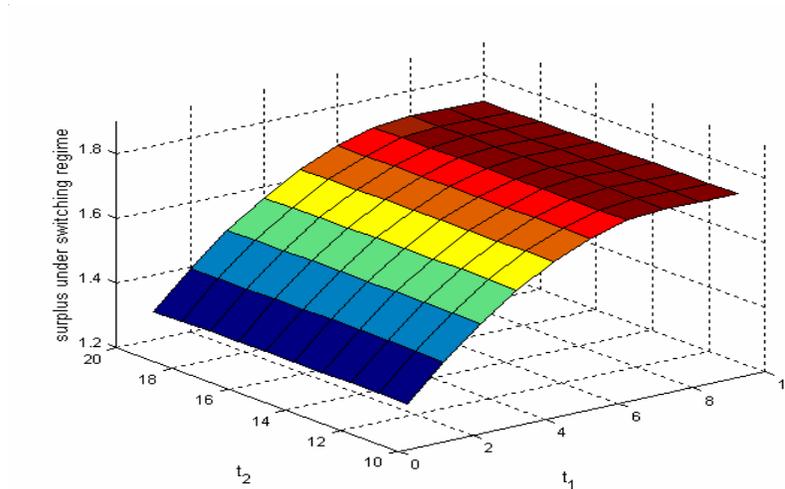


Figure 10 a: The optimal choice of parameters: $t_1=8, t_2=12$

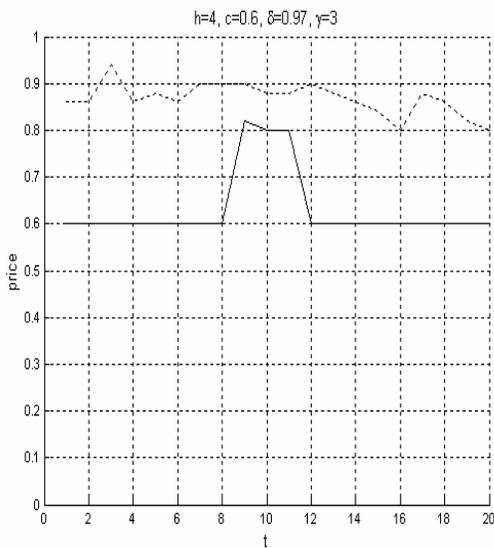


Figure 10 b: price path

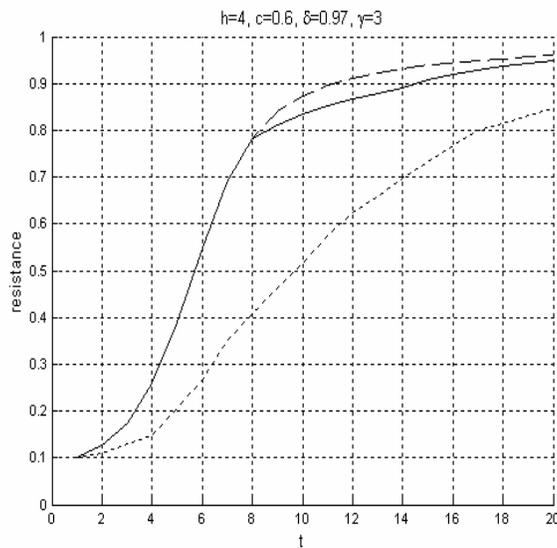


Figure 10c: resistance path under monopoly (dotted line), mixed regime (solid line) and competition (dashed line)