Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Health Economics 22 (2003) 361–377
JOURNAL OF HEALTH ECONOMICS
www.elsevier.com/locate/econbase

Choices, beliefs, and infectious disease dynamics

M. Christopher Auld*

Department of Economics, Faculty of Social Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alta., Canada T2N 1N4

Received 26 July 2001; received in revised form 19 June 2002; accepted 8 November 2002

Abstract

A dynamic model of risky behavior in the midst of an epidemic is discussed. The key result is that pessimistic expectations over the future of the epidemic induce more current risky behavior. Numerical simulation of equilibrium epidemics shows that this effect can accelerate spread of the disease in an epidemic's early stages and that the effect of policy interventions, such as preventative vaccines, may depend on whether the intervention was anticipated.

JEL classification: I1; D8

Keywords: Epidemiology; AIDS; Dynamic programming

1. Introduction

Concern over infectious disease has recently increased because of worldwide morbidity and mortality attributable to acquired immunodeficiency syndrome (AIDS) and its causative agent human immunodeficiency virus (HIV). HIV, a virus transmitted via blood, semen, or cervical secretions, attacks a component of the immune system called T-lymphocytes. Morbidity is initially low, but eventually immunological abnormalities, repeated opportunistic infections, and death result (May and Anderson 1991). The World Health Organization estimates that there were 40 million people infected with HIV as of December 2001, and the illness has killed some 20 million people worldwide (UNAIDS, 2002). The majority of HIV infections occur as a result of sexual activity. In the Western world, men having sex with men is the most common mode of transmission, followed by other behaviors such as intravenous drug use and heterosexual contact (UNAIDS, 2002). Since transmission can be prevented by choosing not to engage in such risky behaviors, studying incentives and

* Tel.: +1-403-220-4098; fax: +1-403-282-5262.
E-mail address: auld@ucalgary.ca (M.C. Auld).

0167-6296/02/$ – see front matter
PII: S0167-6296(02)00103-0

choices during an epidemic is critical to understanding both disease dynamics and the likely efficacy of public health interventions aimed at limiting disease spread.

Economic epidemiology applies tools from microeconomics and econometrics to model choices during epidemics.[1] This study meshes forward-looking rational response to changes in risk with a simple model drawn from biological epidemiology governing disease spread conditional on behavior. Agents sequentially choose rates of risky sexual partnerships, and may choose to abstain and avoid risk altogether. A key result is that a belief that the epidemic is going to become worse tomorrow leads to more risky behavior today. The implications of the model are explored using numerical simulations of epidemics calibrated to survey data on a population of homosexual men in San Francisco in the early 1980s.

A partially effective prophylactic vaccine (a vaccine which reduces probability of infection) illustrates many of the main ideas presented in the analysis. Prophylactic vaccines are considered by many clinical researchers to be the most promising HIV medical intervention under development. Several potential vaccines are currently undergoing trials (Anderson et al., 1995; Frey, 1999; Esparza, 2001). However, whether such a vaccine is desirable depends not only on its biological effect, but also on how the altered incentives it induces change behavior. It is possible that releasing a prophylactic vaccine will cause more risky behavior and the net effect will be to spur rather than retard the epidemic.

The argument that a prophylactic vaccine may cause more risky behavior is similar to Peltzman's (1975) theory that making cars safer may lead to more accidents. A safer car leads to riskier driving; a vaccine which makes risky sex less likely to transmit a disease leads to more risky sex. This unintended consequence could reduce or even reverse decreases in prevalence stemming from the biological effects of the vaccine. Medical researchers recognize that such responses may occur and this concern has contributed to currently existing vaccines not being released (Blower and McLean, 1994; Castillo-Chavez and Hadeler, 1994; Bogard and Kuntz, 2002). Direct evidence from clinical trials on prophylactic vaccines also suggests that vaccination causes increases in risky behavior (Chesney et al., 1997; Johnson, 1999). The framework developed in this study allows modeling of epidemic dynamics following the release or announcement of such a vaccine. Since the model predicts that changes in expectations over the future of the epidemic will change current behavior, forward-looking agents who anticipate the release of a vaccine will change their behavior immediately. Static models of such interventions, even if they allow endogenous behavioral response, may then be misleading through a mechanism similar to that outlined in Lucas's famous (Lucas, 1976) critique of macroeconometric models: the relationship between policies and outcomes is not constant but rather depends, though dynamic optimizing behavior, on both current and future structural relations.

Simulated epidemics calibrated to behavioral and biological parameters reflecting the homosexual population in San Francisco in the 1980s allow experiments on the effects of policy interventions. In one scenario, a counterfactual 1993 prophylactic vaccine release is

[1] Previous analyses of models incorporating response to incentives in the presence of the risk of acquiring an infectious disease include Philipson and Posner (1993), Geoffard and Philipson (1996), Kremer (1996), Mechoulan (1999). Empirically, it is well-documented that epidemics induce behavioral responses, see for instance McKusick et al. (1985a,b), Becker and Joseph (1988), Ahituv et al. (1996), Philipson (1996), Philipson and Dow (1996, 1998), and Mullahy (2000). See Philipson (2000) for a survey of this literature.

unanticipated, and in another the 1993 release is announced to simulated agents in 1988. As illustrated in Fig. 3, in both cases, the parameters are such that incidence is lower after the vaccine is released than it would have been in the absence of a vaccine. However, whether incidence rises or falls between 1992 and 1993 depends on whether or not the release was anticipated. If the release is announced in 1988, agents adjust their expectations and immediately reduce their risky behavior, such that incidence is lower than it would otherwise have been between 1988 and 1992. In this case, when the vaccine is released in 1993 incidence rises over its level in 1992 due to changes in behavior caused by the change in transmission probability. Conversely, if the release is not anticipated, there is no change in behavior until 1993 at which time incidence falls. The sign of the contemporaneous response of incidence to the vaccine then depends on whether the release was anticipated. Empirical researchers might then conclude that the vaccine spurred the disease even when the opposite is true, because of the confounding effect of expectations.

The simulations also demonstrate that how expectations are formed can have significant effects on epidemic dynamics. Rational epidemics may transition more quickly in early stages but wind up at steady states with far lower prevalence than in the case where behavior does not change in response to the epidemic. Initial disease spread will be particularly rapid when individuals anticipate that the epidemic will become worse in the future, because such anticipation blunts risk-reducing behavioral change. Generally, since pessimistic expectations over the future of the epidemic increase current risky behavior, messages from public health officials that lead people to believe that the epidemic will become much worse in the future may spur the epidemic. Overly pessimistic forecasts of future infection rates have been common. For example, Philipson and Posner (1993) note that the Centers for Disease Control forecast in 1988 that, "365,000 AIDS cases would be diagnosed in the United States through 1992 ... in fact by the end of 1992 only 253,000 cases had been reported". Philipson and Posner and other authors have argued that such forecasts have turned out to be overly pessimistic partially because they failed to incorporate endogenous changes in behavior. The model investigated here adds that this failure may have had a dire consequence: the overly gloomy forecasts that resulted may have lead people to become more pessimistic regarding the future of the epidemic, blunting risk-reducing behavioral response and spurring the epidemic. An immediate public health implication is that officials should not focus on publicizing worst-case scenarios.

2. The model

The framework is a simple model drawn from mathematical epidemiology which describes outcomes given choices, and choices made in accord with dynamic optimizing behavior.[2] This section first describes the epidemiological environment taking choices as given, then defines the preference and belief structures leading to the equilibria discussed in Section 3.

[2] See Anderson and May (1991) for a detailed discussion of models of disease spread. The model here is, in epidemiological jargon, a discrete-time susceptible-infected (SI) model with type I mortality and homogeneous (random) mixing. The common assumption in such models that an individual's probability of infection is linear in rate of partner change is relaxed here since it has strong behavioral implications.

2.1. Disease spread conditional on behavior

Time is discrete, in order to capture the fact that agents do not know whether or not a given sexual contact has infected them, but rather learn through periodic testing. There is a continuum of heterogeneous agents. Each agent remains in a market for anonymously matched partners for T periods, regardless of infection status, and the horizon is denoted T^*. Individuals are heterogeneous in preferences, in cohort, and in infection status. Type, age, and infection status are assumed to be private information such that matching across agents is random.

Let N^t_{jka} denote the measure of individuals with infection status j ($j = 0$ denoting susceptibility, $j = 1$ denoting infection) with preference type $k \in \{1, \ldots, K\}$ who are of age $a \in \{0, \ldots, T-1\}$. The three-dimensional array N^t with representative element N^t_{jka} describes the demographic structure at the beginning of period t. All individuals are born susceptible; each period a measure N_k of individuals are born of each type k, that is $N^t_{0k0} = N_k$, $N^t_{1k0} = 0$. The laws of motion for the economy are defined by these birth and death processes and the transition probabilities between infection statuses.

For simplicity, infection statuses are resolved at the end of each period.[3] Since infection is an absorbing state, the probability an agent infected in period t becomes susceptible in period $t+1$ is zero. Let p^t_{ka} denote the probability a susceptible individual in period t of type k aged a transitions to infection at the end of period t. It follows that

$$N^t_{0ka+1} = N^{t-1}_{0ka}(1 - p^t_{ka}) \tag{1}$$

$$N^t_{1ka+1} = N^{t-1}_{0ka}\, p^t_{ka} + N^{t-1}_{1ka}. \tag{2}$$

The probability an agent becomes infected at the end of period t depends on how many partners the agent has in period t, the probability a given match is with an infected partner, and the biologically determined probability that the disease is transmitted conditional on a match between an infected agent and a susceptible agent, denoted ϕ_t. A susceptible agent faces a *per-contact probability of infection* in period t of

$$\pi_t = \phi_t P_t = \phi_t \left(\frac{\sum_{k=1}^{K} \sum_{a=0}^{T-1} N_{1ka} s_{1ka}}{\sum_{k=1}^{K} \sum_{a=0}^{T-1} (N_{1ka} s_{1ka} + N_{0ka} s_{0ka})} \right) \tag{3}$$

where s^t_{jka} denotes the number of partners an individual with infection status j of type k aged a has in period t.[4] That is, if a susceptible agent has exactly one partner in period t, the probability he becomes infected at the end of period t is the product of the transmission probability ϕ_t and the probability the partner was infected P_t. The probability a given partner is infected is not generally equal to the proportion of the population infected, but is rather

[3] It should be emphasized that assuming agents learn their infection status exogenously is unrealistic for diseases such as HIV for which infection does not necessarily lead to immediate morbidity and for which testing is voluntary. Further, behavioral response to risk of infection may include sorting over infection status (assortive matching) as prevalence rises, in contrast to the random matching assumed in this study. See Philipson and Posner (1995), Mechoulan (1999) for discussion of implications of endogenous testing for HIV infection, and Philipson and Dow (1996) for evidence on assortive matching.

[4] Number of partners per period is, for convenience, treated as a continuous non-negative variable.

the weighted average of number of partners and infection statuses over types given by P_t, implicitly defined above. The probability a susceptible individual who has s^t_{0ka} partners in period t becomes infected at the end of the period is given by the usual formula for at least one "success" in s^t_{0ka} trials with per-trial success probability π_t,

$$p^t_{ka} = 1 - (1 - \pi_t)^{s^t_{0ka}}. \tag{4}$$

Finally, the summary statistics of key interest can be written in terms of the notation introduced above. Prevalence of disease in period t is defined by

$$\text{prevalence}_t = \frac{\sum_{k=1}^{K}\sum_{a=1}^{T-1} N^t_{1ka}}{\sum_{k=1}^{K}\sum_{a=1}^{T-1}(N^t_{0ka} + N^t_{1ka})}. \tag{5}$$

Incidence of disease in period t is the number of infections that occur at the end of period t,

$$\text{incidence}_t = \sum_{k=1}^{K}\sum_{a=0}^{T-1} N^t_{0ka} p^t_{ka}. \tag{6}$$

2.2. Preferences

Preferences are defined over sexual partners and loss of health and other costs of infection. Utility is additively separable across periods, with the period return for a type k agent taking the form $u_k(s^t_{jka}, j)$. Future returns are discounted at a constant rate β. The period return is assumed to be twice continuously differentiable, strictly concave, and to have a finite maximum, so that in the absence of an epidemic the agent simply chooses the number of partners each period which maximizes $u_k(s^t_{0ka}, 0)$. The health and psychic costs of infection are modeled as a decrease in utility for any number of partners, $u_k(s^t, 0) > u_k(s^t, 1)$, and the utility of being susceptible with zero partners is normalized to zero, $u_k(0, 0) = 0$. Agents spend only a portion of their lives in the market for randomly-matched sexual partners. Since the model considers time spent in this market rather than the agent's entire lifespan, an agent receives a utility "bonus" of B units, representing the present discounted utility of future returns to the higher health of susceptibility than infection, for remaining susceptible at the end of the modeled time span, receiving 0 otherwise. This assumption implies agents approaching the last period will still have incentive to avoid becoming infected.

2.3. Beliefs

The trade-off between risky sex and loss of health from disease is determined by current and future values of the per-contact probability of infection faced by susceptibles. Agents' have beliefs over current and future per-contact infection probabilities which may or may not coincide with objective infection probabilities. Let the *subjective* per-contact probability of infection i periods in the future as of period t be θ^t_i, so that θ^t_0 is the subjective per-contact probability of infection in the current period. The objective probability of transitioning from susceptibility to infection at the end of period t is given by Eq. (4), whereas the subjective probability the agent bases decisions upon at the start of period t is given by

$$p(s^t_{0ka}, \theta^t_0) = 1 - (1 - \theta^t_0)^{s^t_{0ka}}. \tag{7}$$

The rule agents use to form beliefs is represented $\theta^t = \mathcal{F}(\pi)$, where θ^t is the T-vector with typical element θ_i^t and π may variously refer to past, current, and future objective infection probabilities. A *pessimistic* revision of beliefs refers to increases in one of the elements of θ^t, *optimistic* revisions the opposite.

3. Equilibrium

Given the economic environment presented in the previous section, an equilibrium constitutes a set of decisions for all agents in all time periods which are optimal given their beliefs and a sequence of per-contact realized infection probabilities that evolve according to the laws of motion described above given those choices, as formally defined in Section 3.2. This section first characterizes optimal behavior for each agent given their beliefs, then solves for equilibrium allocations by simulation.

3.1. Optimal behavior

An agent in this economy faces a tradeoff between the benefits of the marginal sexual partner and the costs associated with the marginal probability that partner infects him. This section characterizes optimal behavior in this environment. A key result is that susceptible agents' optimal decisions today depend on both current and anticipated future probabilities of matching with and being infected by an infected agent.

Suppressing indexing over infection status, types, and cohort such that s_t denotes a choice made by the agent in period t, an agent entering period t solves

$$\begin{aligned} \max_{s_t,\ldots,s_T} \quad & E_t \textstyle\sum_{\tau=t}^{T} \beta^{\tau-t} u(s_\tau, I_\tau) \\ \text{s.t.} \quad & \Pr(I_\tau = 1 | I_{\tau-1} = 1) = 1 \\ & \Pr(I_\tau = 1 | I_{\tau-1} = 0) = p(s_\tau, \theta_{\tau-1}^t) \\ & u_{T+1}|_{I_{T+1}=0} = B, \qquad u_{T+1}|_{I_{T+1}=1} = 0, \end{aligned} \tag{8}$$

where I_t indicates infection status. That is, expected discounted utility is maximized subject to the laws of motion for infection status and the terminal conditions. Solutions are characterized first for infected agents, then susceptible.

3.1.1. Infected agents

Consider first the problem faced by an agent who enters period t in the infected state. Since his choice of s_t affects only utility in period t, his problem collapses to a sequence of static optimization problems. Infected agents choose the satiation number of partners per period,

$$s_t^* = \arg\max_{s_t} u(s_t, 1). \tag{9}$$

Let $V(I_t, T-t)$ denote the value of entering period t with $(T-t)$ periods remaining and infection status I_t. Then

$$V(1, T-t) = \frac{1-\beta^{t+1}}{1-\beta} u^I, \tag{10}$$

where $u^I = \max_{s_t} u(s_t, 1)$.

3.1.2. Susceptible agents

Susceptible agents' behavior in period t governs the probability of entering period $t+1$ susceptible, so a susceptible agent's problem is inherently dynamic. Bellman's equation for a susceptible agent at the beginning of period t is

$$V(0, T-t) = \max_{s_t}\{u(s_t, 0) + \beta[p(s_t, \theta_0^t)V(1, T-t-1) + (1-p(s_t, \theta_0^t))V(0, T-t-1)]\}. \tag{11}$$

For convenience, define Δ_t as the net benefit of remaining susceptible in period t,

$$\Delta_t \equiv V(0, T-t-1) - V(1, T-t-1). \tag{12}$$

A susceptible agent's problem can then be written (dropping a term that does not depend on s_t)

$$\max_{s_t} u(s_t, 0) - \beta p(s_t, \theta_0^t)\Delta_t. \tag{13}$$

The fundamental tradeoff the agent faces is between current benefits from more sexual partners and a stream of future benefits obtained by remaining susceptible. Current benefits can be increased only at the cost of increasing the probability that Δ_t, the net value of remaining susceptible, will be lost. How large Δ_t is depends on expectations over the future of the epidemic, the benefits of another partner relative to the costs of infection (morbidity, mortality, and psychic), and the discount rate. The associated first-order condition is

$$\frac{\partial u(s_t^*, 0)}{\partial s_t} \equiv \beta \frac{\partial p(s_t^*, \theta_0^t)}{\partial s_t}\Delta_t, \tag{14}$$

where s_t^* denotes an interior solution for s_t. The right-hand side is the marginal decrease in expected lifetime utility from increasing s_t, that is, the marginal cost of risky behavior. The left-hand side is the marginal utility of risky behavior. It may, however, be optimal to choose the corner solution $s_t = 0$.

Changing only the current per-contact subjective infection probability has no effect on Δ_t. The problem then becomes analytically identical to the static framework considered by Kremer (1996). Following the results in that paper, it is then immediate that a change in only the current per-contact infection probability will have an ambiguous effect on current partners chosen if $p_t(\cdot)$ is strictly concave in number of partners. This result obtains because the marginal probability of infection is decreasing in θ_0^t beyond some critical value, so that further increases in θ_0^t beyond that value reduce the marginal probability of infection from one more partner, increasing optimal number of partners (a result referred to as "fatalism"). Since these results are known, I focus on dynamic aspects of the problem in the two propositions below. Response to changes in current or future per-contact infection probabilities can occur on either the intensive or extensive margins. Consider first how such changes affect the behavior of agents who choose interior solutions.

Proposition 1. *An increase in a future per-contact subjective infection probability increases the optimal number of partners chosen in the current period at an interior solution, that is,* $\partial s_t^*/\partial \theta_i^t > 0$, $0 < i < T-1$. *An increase in steady-state per-contact subjective infection*

probabilities has an ambiguous effect on optimal number of partners chosen in the current period.

To see the intuition, consider the extreme case in which the agent knows a fully effective vaccine, one removing the possibility of becoming infected, is to be made available next period ($\theta_i^t = 0, i = 1, \ldots, T - t$). Risky behavior today is very costly in expectation since the returns to being susceptible in the future are large. As future values of θ increase beyond this point, the value of remaining susceptible falls because the optimal probability of becoming infected rises and/or optimal future number of partners fall (loosely, a pessimistic revision induces the agent to believe he is more likely to become infected tomorrow, which reduces his incentive to take costly action to avoid infection today). Since the value of being infected is independent of the state of the epidemic but the value of being susceptible decreases as the disease spreads, the net value of remaining susceptible falls as expectations regarding the future of the epidemic become more pessimistic. Thus, holding the relationship between current risky behavior and risk of infection constant, an increase in anticipated future risk unambiguously induces *intertemporal* fatalism: Current risky behavior for all types of agents (weakly) increases in response to increases in *future* risk of infection per partner. However, a change from one steady state to another causes both an expectations effect and a change in the tradeoff between current partners and current risk of infection. Since the latter effect could increase or decrease number of partners selected, the overall effect of a change in steady state is also ambiguous.

Now consider the conditions under which an agent will participate in the market for randomly matched partners. The internal solution, s_t^*, will be selected if

$$u(s_t^*, 0) + \beta[p(s_t^*, \theta_0^t)V(1, T - t - 1) + (1 - p(s_t^*, \theta_0^t))V(0, T - t - 1)] > \beta V(0, T - t - 1). \tag{15}$$

If this inequality does not hold, the agent chooses to have no risky contacts, since he can receive $\beta V(0, T - t - 1)$ with certainty by setting $s_t = 0$, whereas he receives expected utility equal to the left-hand side of Eq. (15) at the interior solution. This condition is equivalent to

$$u(s_t^*, 0) - \beta p(s_t^*, \theta_0^t)\Delta_t > 0. \tag{16}$$

The first term is the gain in this period's utility and the second is the expected loss from choosing the interior solution rather than zero. The following proposition summarizes the effect of changes in current and future epidemic conditions on propensity to participate.

Proposition 2. *An increase in the current infection probability decreases the probability of participation, whereas an increase in a future infection probability increases it. A higher steady-state infection probability has an ambiguous affect on participation.*

Proof. see Appendix A. □

If the current per-contact probability of infection increases, the cost of participating rises while the benefits do not, inducing more people to exit. If a future probability increases, the costs of participating fall, since it is less beneficial to remain susceptible in the future,

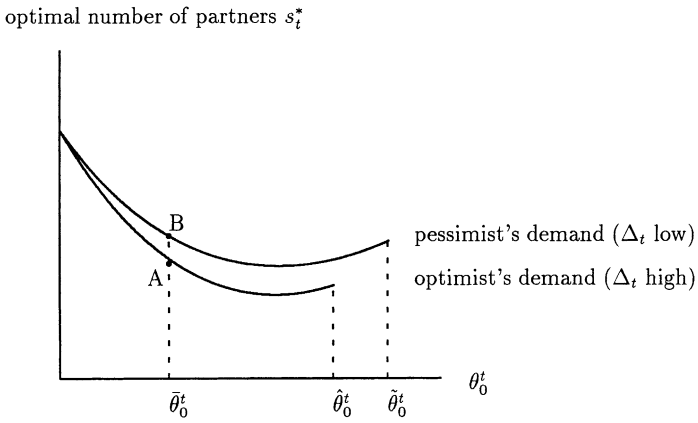

Fig. 1. Demand for partners as functions of current risk (θ_0^t) for agents with optimistic and pessimistic forecasts regarding the future course of the epidemic. Movement from point A to B shows change in current behavior, holding current risk per partner constant, caused by a pessimistic revision of expectations.

leading to fewer exits. When both probabilities change due to a change in steady state, participation could either increase or decrease.

Proposition 2 reinforces the implications of Proposition 1: a pessimistic revision of expectations reduces risk-averting behavior, on both the extensive and intensive margins. Such a revision induces some agents who would otherwise choose not to participate to enter the market for risky partners, and causes agents already participating to have more partners.

The optimal behavior of susceptible agents is summarized in Fig. 1, which displays "demand" for anonymously matched partners in period t as a function of beliefs about current per-contact subjective infection probabilities, θ_0^t. Holding θ_0^t constant at $\bar{\theta}_0^t$, a pessimistic revision of expectations shifts demand up, moving the agent's optimum from point A to point B. The threshold level of current risk at which the agent chooses to exit the market for anonymously matched partners increases with the pessimistic revision of expectations from $\hat{\theta}_0^t$ to $\tilde{\theta}_0^t$.

An implication of increased risky behavior in response to more pessimistic forecasts is that statistical methods that overestimate future cases could contribute to disease spread. To the extent that individuals are aware of such projections and base their own forecasts upon them, a direct implication of Propositions 1 and 2 is that these overly pessimistic forecasts increased risky behavior, thereby spurring the epidemic. It follows that public health efforts designed to minimize risky behavior should not emphasize dire predictions of the future state of the epidemic.[5]

[5] A referee has questioned the morality of public health officials deliberately distorting information in the manner of a "paternalistic and dishonest dictator". The positive implication of the model is that distorting information may, to the extent that the source retains credibility, be a policy tool to reduce risky behavior. But the normative implication that public health officials ought to engage in such manipulations does not necessarily follow. This issue is a special case of the broad question, 'Should policy makers deliberately mislead the public if the end result is beneficial?' which has been under debate since the dialogue regarding the "noble lie" in Plato's *Republic*. This paper does not suggest that policy makers *should* distort or deliberately withhold information, but it would seem warranted to suggest that emphasizing pessimistic forecasts may not be in the public's interest.

3.2. Simulation of equilibrium epidemics

Given a rule for expectations formation $\mathcal{F}$ and an initial demographic state N^0, an *equilibrium* of this economy is defined as:

1. A set of decisions s^t_{jka}, $t = 1, \ldots, T^*$, $j = 0, 1$, $k \in \{1, \ldots, K\}$, $a \in \{0, \ldots, T-1\}$ which are sequentially optimal (solve program (8)) given the sequence of beliefs $\theta^t = \mathcal{F}(\pi)$, $t = 1, \ldots, T^*$.
2. A vector of per-contact transmission probabilities π_t, $t = 1, \ldots, T^*$ which satisfy Eq. (3).

An equilibrium is defined as choices for all types of agents in each period which are optimal given their beliefs over per-contact infection probabilities (generated by $\mathcal{F}(\pi)$), and realized per-contact infection probabilities π_t which evolve consistently with the laws of motion defined in Section 2.1. Numerical simulation is employed to solve for equilibrium outcomes, as analytical solutions are not available in non-trivial cases. Details of the functional forms, numerical procedures used to solve for equilibria, and calibration to data are described in Appendix B.

3.3. Consequences of expectations on epidemics

Consider three different manners in which people could form beliefs (mappings $\mathcal{F}$) over the future of an epidemic. *Myopic* expectations refers to the case where the agent acts as if the probability of becoming infected is zero ($\theta^t_i = 0 \, \forall i, t$), which produces the same behavior as setting the discount rate to zero for any type of belief formation. These beliefs imply agents will not change their behavior in response to the epidemic, and is considered as a baseline with which to compare the impact of rational response on disease dynamics. Two types of rational response are considered, differing only in how expectations are formed. Agents with *adaptive* expectations update their beliefs each period according to the rule that current and all future per-contact infection probabilities are equal to the realized value of π in the previous period ($\theta^t_i = \pi_{t-1} \forall i, t$). These agents are then surprised by changes in π each period until a (non-cyclic) steady state is reached, so they are rational given beliefs, but form beliefs irrationally. Agents with *rational* expectations are rational given beliefs and form beliefs rationally. They act as if they know current and all future per-contact infection probabilities ($\theta^t_i = \pi_{t+i} \forall i, t$).

Fig. 2 illustrates the effect of different modes of expectation formation on epidemic dynamics. At the calibrated parameters, prevalence under adaptive or rational expectations converges to a much lower steady state than in the myopic case. This occurs largely because of response on the extensive margin summarized by Proposition 2: a large fraction of the population chooses to not enter the market for anonymously matched partners. The transition paths culminating in these steady states reflect the effects of expectations on behavior. Rational expectations epidemics transition faster initially because agents correctly forecast that the epidemic will get much worse, which attenuates risk-reducing behavior. Epidemics in which agents have adaptive expectations—responding to risk, but not correctly forecasting the future of the epidemic—spread somewhat slower than rational expectations epidemics. Finally, epidemics in which behavior is invariant to the spread of disease spread

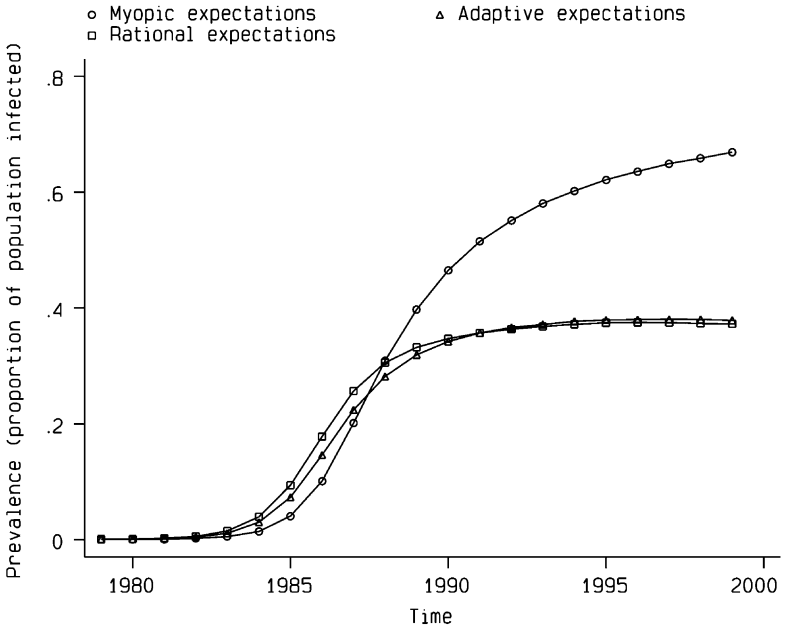

Fig. 2. Prevalence in simulated epidemics with myopic agents, agents with adaptive expectations, and agents with rational expectations.

slowest even though the eventual prevalence reached under these conditions is much higher than in either of the other two cases.

3.4. An imperfect vaccine and its announcement date

As discussed in the introduction, a prophylactic vaccine could either spur or retard an epidemic. Holding behavior constant, a vaccine must reduce incidence, but behavior will generally change not only in response to the release of such a vaccine, but also in response to the *anticipation* of the release of such a vaccine.

Suppose all simulated agents are vaccinated, modeled as a decrease in the biological transmission parameter ϕ, some time after the epidemic is under way. Announcing the *future* release of the vaccine affects expectations but not the current tradeoff between partnerships and probability of infection. Propositions 1 and 2 imply that agents will reduce their risky behavior, which is likely (depending on the distribution of response) to reduce new infections. *Releasing* the vaccine changes the current tradeoff between partnerships and infection probability; for most agents (those not exhibiting fatalism), the vaccination will cause more risky behavior. The effect of more optimistic expectations has already occurred prior to release when the vaccine is anticipated, but both the expectations effect (reducing risky behavior) and the effect of the vaccine itself (increasing risky behavior) occur together when a surprise vaccine is released. The magnitude and possibly even the sign of the contemporaneous effect of the release then depend on whether or not it was anticipated.

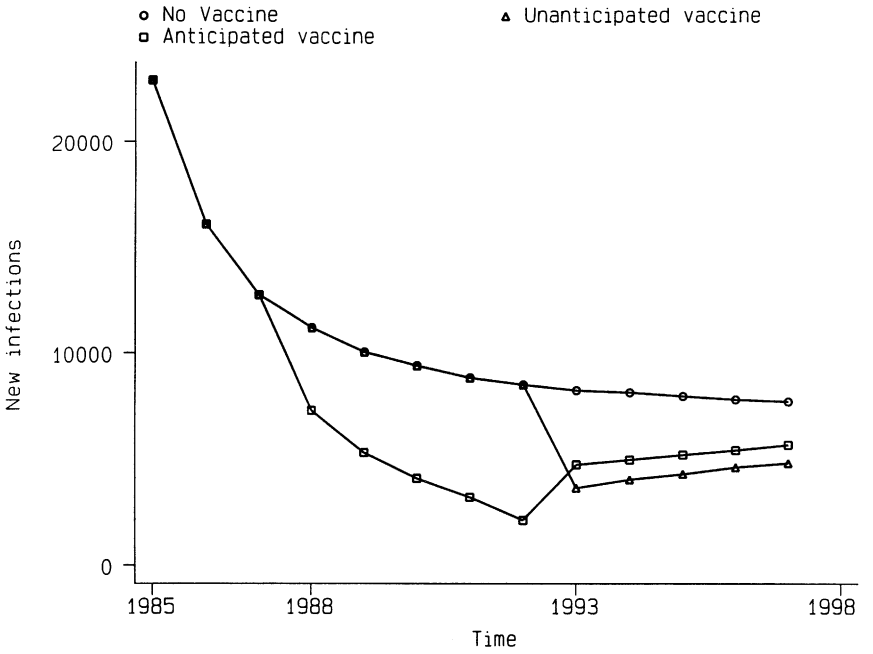

Fig. 3. Simulations of the release of a prophylactic vaccine. If the vaccine is announced at the start of 1988, incidence falls immediately, rising when the vaccine is released at the start of 1993. If the 1993 release is unannounced, incidence falls upon release. In both cases, incidence in 1993 and thereafter is lower than it would have been in the absence of a vaccine.

Fig. 3 shows incidence paths for epidemics in which a prophylactic vaccine is released. Three cases are considered: one in which the vaccine is never available as a baseline, one in which a vaccine is released in 1993 which surprises the simulated agents, and one in which the availability of the vaccine beginning in 1993 is announced at the beginning of 1988. Expectations are rational; after the vaccine is announced or released expectations are identical to realized outcomes.[6] The response of prevalence to the introduction of the vaccine depends on the distribution and magnitude of behavioral response and the effectiveness of the vaccine. In the example considered here, the vaccine is effective enough (reducing the biological transmission coefficient from 0.125 to 0.025) that the vaccine's release reduces prevalence.

In the case in which the release is announced 5 years prior to the event, risky behavior decreases or ceases for all types upon announcement, reducing incidence immediately. Incidence remains lower than it otherwise would have been for the interim period between announcement and release as a result of the optimistic revision of expectations. When the

[6] I have not included the myopic expectations case to simplify the figure. In that case, behavior does not change. Therefore, prevalence unambiguously decreases when the vaccine is introduced. At these parameter values, with myopic expectations steady-state prevalence decreases from about 72 to 41% after the vaccine is released, and from 44 to 40% under rational expectations.

vaccine is made available, however, expectations do not change but the contemporaneous relationship between number of partners and risk of infection does. Holding expectations constant, introducing the vaccine generates a large enough increase in risky behavior to increase prevalence. But if the vaccine surprises agents, both expectations and current risk per partner change, and the net effect is to reduce prevalence.

Hence, a vaccine effective enough to reduce prevalence despite risk-offsetting behavior could appear to have increased prevalence if the vaccine was anticipated. Incidence upon the vaccine's introduction in 1993 is lower with a vaccine than without regardless of whether the vaccine was anticipated, but whether incidence is higher or lower than it was in 1992depends on whether the vaccine was anticipated. The result generalizes to other interventions. For instance, new drugs or other treatments which effectively decrease the cost of becoming infected can be expected to increase risky behavior, but we should not expect to see large changes in behavior contemporaneously with the availability of the treatment if the treatment was anticipated.

4. Concluding remarks

This paper finds *homo economics* at risk of acquiring an infectious disease. He learns whether he is infected at intervals, and chooses how much risk to undertake during each interval. The key result is that pessimistic expectations over the future of the epidemic reduce incentive to avoid current risk. Several implications follow. If increases in contemporaneous risk are correlated with more pessimistic beliefs over the future of the epidemic, then reductions in risky behavior that would otherwise occur will be partially offset or reversed. People experiencing the initial stages of an epidemic may not engage in much protective behavior not only because current risk is minimal, but also because they correctly anticipate that the epidemic will get worse over time—pessimistic expectations can make an epidemic transition faster. We may also draw a public health conclusion: Concentrating attention on dire forecasts may have the unintended consequence of increasing risky behavior. A disquieting related observation is that widely publicized and, in the event, overly pessimistic forecasts in the first decade of the HIV/AIDS epidemic may have blunted risk-reducing behavioral change and increased the number of infections which occurred.

Would a partially effective prophylactic vaccine be effective in limiting HIV contagion? The answer depends on both the biological effectiveness of the vaccine on the magnitude and distribution of behavioral responses. Numerical simulation of equilibrium epidemics calibrated to survey data from San Francisco suggests that, for such a population, a vaccine reducing the probability of disease transmission by a factor of five unambiguously reduces disease spread. But the contemporaneous effect of the vaccine on incidence depends on whether or not its release was announced beforehand—infections rise in the period in which the vaccine was released if the release was anticipated, and fall otherwise. Expectations over future interventions and medical treatments can affect current behavior, complicating analysis of the effects of such changes in a manner not unlike stock prices reflecting expectations over future events complicates intervention analysis in that context.

Acknowledgements

I thank Dan Bernhardt, John Boyce, Curtis Eaton, Herb Emery, Chris Ferrall, Allan Gregory, Debra Holt, two anonymous referees and seminar participants at Queen's University, the Canadian Economics Association meetings, the University of Victoria, the University of Western Ontario and the University of Toronto for helpful comments. Support from the Social Sciences and Humanities Research Council of Canada and the Alberta Heritage Foundation for Medical Research is gratefully acknowledged. Remaining errors are mine.

Appendix A. Proofs

The proofs make use of the following lemma.

Lemma 1. *The net benefit of remaining susceptible,* Δ_t*, is*: (1) *independent of the current infection probability* ($\partial\Delta_t/\partial\theta_0^t = 0$)*, and* (2) *weakly decreasing in any future infection probability* ($\partial\Delta_t/\partial\theta_i^t \leq 0, 0 < i \leq T - t$).

Proof. Part (1) follows directly from Eqs. (11) and (12); infection risk today does not affect the value of being infected or susceptible in the future. Since by Eq. (10) the value of being infected is independent of all current and future values of θ_0^t, part (2) is true if the value of being susceptible is weakly decreasing in θ_i^t. $i > 0$. In any period $\tau \in \{t+1, \ldots, T\}$, the value of being susceptible at an interior solution is given by Eq. (11) evaluated at $t = \tau$, which can easily be shown to be weakly decreasing in θ_0^τ by total differentiation of the first-order condition. The result then follows from the observation that the value of entering τ susceptible is increasing in the value of entering $\tau + 1$ susceptible, which can be demonstrated by differentiation of Eq. (14). □

Proof of Proposition 1. Differentiate Eq. (14) with respect to θ_i^t, $0 < i < T - t$, to find:

$$\frac{\partial s_t^*}{\partial \theta_i^t} = \frac{\beta}{\Theta}\frac{\partial p(s_{it}, \theta_0^t)}{\partial s_i^t}\frac{\partial \Delta_{t-1}}{\partial \theta_i^t} \geq 0, \tag{A.1}$$

where

$$\Theta = \frac{\partial^2 u(s_t, 0)}{\partial s_t^2} - \beta\Delta_t\frac{\partial^2 p(s_t, \theta_0^t)}{\partial s_t^2} < 0$$

by the second-order condition. The inequality holds by the Lemma 1. The result with respect to changing steady states can be obtained as follows. In a steady state, $\theta_i^t = \theta\ \forall i$ for some constant θ. Differentiating with respect to θ yields

$$\frac{\mathrm{d}s_t^*}{\mathrm{d}\theta} = \frac{\partial s_t^*}{\partial \theta_0^t}\bigg|_{\theta_0^t=\theta} + \sum_{i=1}^{T-t}\frac{\partial s_t^*}{\partial \theta_i^t}\bigg|_{\theta_i^t=\theta} = \beta\left[\frac{\partial^2 p(s_t, \theta_0^t)}{\partial s_t \partial \theta_0^t}\Delta_t + \frac{\partial p(s_t, \theta_0^t)}{\partial s_t}\sum_{i=1}^{T-t}\frac{\partial \Delta_t}{\partial \theta_i^t}\bigg|_{\theta_i^t=\theta}\right] \tag{A.2}$$

The first term is positive unless $p(\cdot)$ is concave in s_t and s_t is large. The second term is negative by the Lemma 1. □

Proof of Proposition 2. Differentiating Eq. (16) with respect to θ_0^t and invoking the envelope theorem and the Lemma 1, the left-hand side changes by

$$-\frac{\beta\Delta_t \partial p(s_t, \theta_0^t)}{\partial \theta_0^t} < 0.$$

This result establishes the first part of the proposition. The result with respect to changes in expectations follows from differentiating with respect to a future infection probability, the left-hand side changes by $-\beta p(s_t^*, \theta_0^t)\partial\Delta_t/\partial\theta_i^t > 0,\ i \in \{1, T-t\}$, which establishes that an increase in a future expected probability increases participation. When a steady state θ changes, differentiation of (16) shows participation increases iff $-p(s_t, \theta)(\sum_{i=1}^{T-t} \partial s_t^*/\partial\theta_i^t) > \Delta_t \partial p(\cdot)/\partial\theta$. □

Appendix B. Simulation details

Functional forms and parameterizations are as follows. Agents remain in the market for 30 periods, where each period represents 6 months of calendar time. Preference types k are defined by values of α in the specification for the period return:

$$u_k(s_t, 0) = \alpha_k s_t - s_t^2, \qquad u_k(s_t, 1) = \gamma\alpha_k s_t - s_t^2 - D, \quad \gamma \in (0, 1) \qquad \text{(B.1)}$$

where D is a per-period utility loss representing disease-induced morbidity and γ scales marginal utility for infecteds, which allows the number of partners chosen by infecteds to reflect aversion to infecting others.[7] This form implies that an agent's value of α is identified by observing his rate of partner change in the absence of disease, since $\text{argmax}_{s_t} u(s_t, 0) = \alpha/2$. A continuum of agents is approximated by a large finite population. There are $K = 100$ types per generation. Each period denotes 6 months of calendar time and agents remain in the market for thirty periods, so that there are (100)(300)(30) = 900,000 individual agents in the market at any given time, with (30)(100) = 3000 unique dynamic optimization problems to be solved each period.

The simulations are calibrated the situation in San Francisco from the late seventies through the late 1980s in order to provide some concreteness without the expense of formal estimation of all the parameters. The distribution of α_k was chosen to match the distribution of rates of partner change prior to the epidemic observed in the San Francisco Men's Health Study (see Winkelstein et al., 1987 for details of this survey). Specifically, 100 types, corresponding to the percentiles of the distribution of partners prior to 1982 in the SFMHS, are assigned, with 15 agents per type per generation. Other parameters are chosen to be consistent with other studies when available, in an ad hoc fashion otherwise. The transmission coefficient ϕ, was chosen to be consistent with epidemiological evidence while abstracting from variable infectivity (Jacquez et al., 1994). Grant et al. (1987) report

[7] Empirical evidence suggests factors such as aversion to infecting others reduce partnership acquisition rates among infected individuals (McCusker et al., 1988).

Table 1
Calibration of simulation parameters

Parameter	Description	Value
γ	Activity reduction for infecteds	0.666
T	Modeled lifespan, in periods	30
K	Number of types (values of α)	100
N_k	Agents born type k each period	300
N	Agents alive each period	900,000
B	Utility of exiting susceptible	250
D	Per-period utility penalty for infection	10
ϕ	Biological transmission coefficient	0.125
ϕ_V	Transmission coefficient post-vaccine	0.025
β	Discount rate	0.950

Note: distribution of α calibrated to 1984 SFMHS data.

a per-partnership transmission rate of 0.102. The proportion of satiation partners chosen by infecteds, γ, was chosen to be close to the observed ratio of pre-AIDS partners to post-AIDS partners selected by men infected as of the first wave in the SFMHS. Amongst those testing positive in the first wave who had a positive number of partners in both periods, the mean number of partners in the post-AIDS period was 14.5, in the pre-AIDS period 25.7, for a ratio of 0.55. The discount factor β was chosen to be consistent with its usual range in econometric studies. The remaining parameters, displayed in Table 1, were chosen in an ad hoc manner to correspond to the early course of the epidemic in San Francisco. Epidemics were started by randomly infecting 0.1% of the population in the first period.

The simulations were written in Fortran 90, compiled using the XLF compiler, and executed on an IBM RS/6000 model 43P running at 200 megahertz. Each simulation, or iteration on the perfect foresight solution, took roughly two minutes of CPU time. The agents' dynamic programming problems were solved by backwards induction on Bellman's equation. The rational expectations equilibria were calculated by a fixed point algorithm. An arbitrary initial vector π_0^e was posited, which generates a vector of realized infection probabilities π_0 conforming to behavior given beliefs π_0^e. A convex combination of π_0^e and π_0 yields π_1^e, and so on, iterating until the convergence criterion that the mean deviation between expected and observed values of π_t must be less than 0.005 and no particular period can have an error greater than 0.01 was met. This criterion usually produced convergence in ten to fifteen iterations. Note there is no guarantee that this procedure will converge, but hundreds of trial simulations produced only one failure and no examples of equilibria exhibiting cycles. A variation of this procedure was used to solve epidemics in which a vaccine is introduced, with equilibria found in the manner above up to the vaccine's announcement date, then re-solved post-announcement taking the state of the world prior to announcement as the initial condition.

References

Ahituv, A., Hotz, J., Philipson, T., 1996. The responsiveness of the demand for condoms to the local prevalence of AIDS. Journal of Human Resources 31, 869–897.

Anderson, R., May, R., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.
Anderson, R., Swinton, J., Garnett, G., 1995. Potential impact of low efficacy HIV-1 vaccines in populations with high rates of infection. Proceedings of the Royal Society of London B: Biological Sciences 261, 147–151.
Becker, M., Joseph, J., 1988. AIDS and behavioral change to reduce risk: a review. American Journal of Public Health 78, 394–410.
Blower, S., McLean, A., 1994. Prophylactic vaccines, risk behavior change, and the probability of eradicating HIV in San Fransisco. Science 265, 1451–1454.
Bogard, E., Kuntz, K., 2002. The impact of a partially effective HIV vaccine on a population of intravenous drug users in Bangkok, Thailand: a dynamic model. Journal of Acquired Immune Deficiency Syndromes 29, 132–141.
Castillo-Chavez, C., Hadeler, K., 1994. A Core Group Model for Disease Transmission. Working paper, Yale University.
Chesney, M., Chambers, D., Kahn, J., 1997. Risk behavior for HIV infection in participants in preventive HIV vaccine trials: a cautionary note. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 16, 266–271.
Esparza, J., 2001. An HIV vaccine: how and when? Bulletin of the World Health Organization 79, 1133–1137.
Frey, S., 1999. HIV vaccines. Infectious Disease Clinics of North America 13, 95–109.
Geoffard, P., Philipson, T., 1996. Rational epidemics and their public control. International Economic Review 37 (3), 603–624.
Grant, R., Wiley, J., Winkelstein, W., 1987. Infectivity of the human immunodeficiency virus: estimate from a prospective study of homosexual men. Journal of Infectious Disease 156, 189–193.
Jacquez, J., Koopman, J., Simon, C., Longini, I., 1994. Role of the primary infection in the epidemics of HIV infection in gay cohorts. Journal of the Acquired Immune Deficiency Syndromes 8, 1169–1184.
Kremer, M., 1996. Integrating behavioral choice into epidemiological models of the AIDS epidemic. Quarterly Journal of Economics 111, 549–573.
Lucas, R., 1976. Econometric policy evaluations: a critique. In: Brunner, K., Metzer, A. (Eds.), The Phillips Curve and the Labor Market. North-Holland, Amsterdam.
Johnson, M., 1999. Belief of vaccine receipt in HIV vaccine trials: further cautions. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 21, 413.
McKusick, L., et al., 1985a. AIDS and sexual behavior reported by gay men in San Francisco. American Journal of Public Health 75, 493–496.
McKusick, L., et al., 1985b. Reported changes in the sexual behavior of men at risk for AIDS, San Francisco, 1982–1984—the AIDS behavioral research project. Public Health Reports 100, 622–629.
McCusker, J., et al., 1988. Effects of HIV antibody test knowledge on subsequent sexual behaviors in a cohort of homosexually active men. American Journal of Public Health 78, 462–467.
Mechoulan, S., 1999. HIV Testing: A Trojan Horse? Working paper, Northwestern.
Mullahy, J., 2000. It'll only hurt a second: microeconomic determinants of the demand for flu vaccines. Health Economics, in press.
Peltzman, S., 1975. The Effects of automobile safety regulation. Journal of Political Economy 83, 677–725.
Philipson, T., 1996. Private vaccination and public health: an empirical examination for US measles. Journal of Human Resources 31 (3), 611–630.
Philipson, T., 2000. Economic epidemiology and infectious diseases. In: Handbook of Health Economics.
Philipson, T., Dow, W., 1996. An empirical examination of the implications of assortive matching on the incidence of HIV. Journal of Health Economics 15, 735–749.
Philipson, T., Dow, W., 1998. Infectious disease transmission and infection-dependent matching. Mathematical Biosciences 148, 161–180.
Philipson, T., Posner, R., 1993. Private Choices and Public Health: The AIDS Epidemic in An Economic Perspective. Harvard University Press, London.
Philipson, T., Posner, R., 1995. A theoretical and empirical investigation of the effects of public health subsidies for STD testing. Quarterly Journal of Economics 110, 445–474.
UNAIDS, 2002. AIDS Epidemic Update December 2001. Joint United Nations Program on AIDS and World Health Organization.
Winkelstein, W., Lyman, D., Padian, N., et al., 1987. Sexual practices and risk of infection by the AIDS associated retrovirus: the San Francisco men's health study. Journal of the American Medical Association 257, 326–330.