Queen's Economics Department Working Paper No. 938

Choices, Beliefs, and Infectious Disease Dynamics

M. Christopher Auld

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-1996

Choices, Beliefs, and Infectious Disease Dynamics

M. Christopher Auld*

Department of Economics
Queen's University
Kingston, Ontario
Canada, K7L 3N6
auld@qed.econ.queensu.ca

First draft: February 1995
This version: October 1996

ABSTRACT

This paper develops a dynamic model of behavioural response to the risk of infectious disease. People respond to increased risk of infection by either making marginal adjustments in risky behaviour or by moving to a corner solution where perceived risk is zero. Individuals most prone to high-risk activity will tend to reduce activity less than low–risk people; very high risk people may exhibit "fatalism" and increase risky behaviour as the risk of becoming infected rises. Beliefs about the future course of the epidemic affect current behaviour even when utility is additively separable: pessimistic beliefs induce more risky behaviour. Simulations contrast the disease dynamics generated under these behaviours with those of standard epidemiological models and examine policy issues.

JEL Classification : I1, D8

Keywords : epidemiology, AIDS, uncertainty

---

* Thanks to Chris Ferrall for supervision and support. Dan Bernhardt, Allan Gregory, Debra Holt, Colin Telmer, and seminar participants at Queen's University, the Canadian Economics Association June 1995 meetings, the University of Victoria, and the University of Toronto made many helpful comments. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. Remaining errors are mine.

This study examines behavioural change to the risk of acquiring an infectious disease and how such responses affect the course of the epidemic. Although the results apply to other diseases, the specific disease in mind is the human immunodeficiency virus (HIV) — the causative agent of acquired immunodeficiency syndrome (AIDS). It is estimated that ten million people worldwide — one million of those in the United States — are infected with HIV (Tarantola *et al.* (1993)). Unlike many other diseases, people can choose their level of exposure to infection — as of 1992, over 90% of reported infections were due to either sexual contact or IV drug use (Centers for Disease Control (1993)). Understanding how behavioural responses to the risk of infection affect the way the epidemic moves through the population is therefore important for predicting the effectiveness of various policy interventions designed to limit the spread of the disease.[1]

This paper develops a dynamic model of behaviour during an epidemic. The model predicts that behaviour in the current period depends on the agent's beliefs about the probability a given contact is infected both in the current period and in future periods. Response to increases in the probability of meeting an infected partner varies across agent type. An increase in current risk per contact causes some agents to increase their desired number of contacts, others to reduce their risky contacts, and still others to cease having risky contacts altogether. Expectations about the epidemic's future affect current behaviour: more pessimistic expectations reduce the incentive to avoid risk in the present, which can spur the disease. These results imply that the higher–order moments as well as the mean of the distribution of rates of partner change will generally vary with disease prevalence, which is confirmed by empirical evidence.

---

[1] The epidemiological literature generally assumes away systematic behavioural response to the risk of infection. Sattenspiel (1990) notes, "There is too little attention paid to human cultural and behavioural factors in the formulation of models ... [Instead,] the majority of mathematical models are still being developed by ecologists, parasitologists, and mathematicians ... who tend to take human biology and behaviour as a constant."

The implications of these behavioural responses are explored via numerical simulation. Simulated agents are heterogeneous in an underlying taste parameter, which is calibrated to survey data from the San Francisco Men's Health Study (SFMHS). Three methods of expectations formation are examined: mypoic, in which behaviour is invariant to the risk of infection; adaptive, in which the agents' latest information on risk is assumed to prevail in the current and all future periods; and rational, in which agents correctly predict current and future risk. Overall, the simulations demonstrate two findings. First, disease dynamics generated by this model differ substantially from dynamics generated by the typical epidemiological model. Second, behavioural response and how expectations are formed can either spur or retard disease spread. The consequences of policy interventions are also simulated. Rapid public health efforts to transmit awareness of the disease and its modes of transmission are shown to reduce the total number of infections which occur. However, at the calibrated parameters, a partially effective vaccine is predicted to increase new infections due to offsetting behavioural changes. These results suggest that finding effective policies for dealing with the AIDS epidemic depends on both qualitative and quantitative knowledge of the interaction between disease dynamics and human behaviour.

Previous studies of models incorporating response to incentives to change behaviour in the presence of the risk of acquiring an infectious disease include Philipson and Posner (1993), Geoffard and Philipson (1995), and Kremer (1994b, 1996). This paper extends these models in six significant ways. First, agents solve an explicitly dynamic problem. The dynamic framework permits consistent modeling of behaviour when the risk of infection varies over time. Second, the analysis permits an explicit examination of how beliefs over future states of the world affect today's behaviour. I show that expectations that risk will be high in the future reduce incentives to reduce risky activity today, even when utility is additively separable across time. This effect alters both the initial spread of the disease — expectations of high risk can be

self-fulfilling — and has implications for certain public policy interventions, such as the announcement of partially effective vaccines. Third, decisions in this paper are over both whether to engage in any risky activity at all (as in Philipson and Geoffard (1996)) as well as how much risky activity to undertake, if any (as in Kremer (1996)). Fourth, I allow agents to periodically learn their infection status (a special case obtains when status is never revealed and the problem collapses to a static one). Thus, the model captures the feature that, in the developed world, testing for HIV infection is readily available and frequently done.[2] Fifth, I present simulations with a large number heterogeneous agents — with the distribution of the parameter of heterogeneity calibrated to SFMHS data at 100 points — which explore the impacts of various policy interventions and of several methods of forming beliefs. Finally, by modeling the problem in discrete rather than continuous time, the analysis lends itself naturally to empirical analysis of panel data on risky behaviour.

I. Background

There is considerable empirical evidence of behavioural responses to the risk of acquiring HIV. Most studies examing such response have focused on homosexuals, since in the developed world both the probability of meeting an infected partner and the probability of transmission are much higher for homosexuals than for heterosexuals (May, Anderson, and Blower (1990)). Philipson and Posner (1993) and Ahituv, Hotz, and Philipson (1993) both found that that condom use was responsive to the prevalence of AIDS. Becker and Joseph (1988) summarized empirical findings in cross–sectional and longitudinal studies of behaviour of high-risk individuals that documented substantial decreases in both number of partners and high-risk acts.

---

[2] Over 80% of SFMHS participants chose to receive test results by 1988 (personal communication, James Wiley, Survey Research Center, University of California at Berkeley.)

HIV prevalence in San Francisco reached between forty and sixty percent amongst the homosexual population in the eighties (Kellog *et al.* (1990)). Many studies have investigated behavioural change in this group (*e.g.*, Winkelstien *et al.* (1987a, 1987b) and McKusick *et al.* (1985a, 1985b)). McKusick *et al.* (1985a) report an increase in monogamy from 35% to 41% between November 1982 and November 1983, and McKusick *et al.* (1985b) report a decrease in mean number of partners per month from 6.7 to 3.2. These and other studies document a decrease in mean number of partners and an increase in the proportion of respondents reporting celibacy or a monogamous relationship. Few empirical investigations have examined changes in the higher–order moments of the distribution of rates of partner change, however.

Table 1 shows the mean, standard deviation, skewness and selected percentiles of the distribution of number of partners per six-month period in a sample of homosexual men in San Francisco prior to 1982, taken to be the pre-AIDS period, and at the end of 1983/beginning of 1984, when AIDS information was widely known and prevalence in the area was high.[3] The mean number of partners falls by roughly a third, consistent with the studies above, but the standard deviation actually increases substantially. Moreover, the skewness of the distribution increases, from 3.3 to 13.3.[4]

---

[3] These statistics were derived from the San Francisco Men's Health Study, a panel probability sample of men living in an area of San Francisco designed to study the history of AIDS, see Winkelstein *et al.* (1987a) for details of the sample design. The end of 1983/beginning of 1984 rates are taken from the first wave of the survey. The pre-1982 rates of partner change are averages constructed by taking the self-reported (in the first wave of the survey) lifetime number of partners in 1984, subtracting the number of partners reported between 1982 and 1984, and dividing by the number of years between the time the respondent reported beginning regular sexual encounters and 1982. The sample is restricted to homosexual or bisexual men reporting one or more partners per six-months in both periods.

[4] These figures are consistent with other studies. Schecter *et al.* (1988), for instance, reports annual number of partners in a sample of sexually active (more than eight partners at the first observation period) homosexual males in Vancouver falling from 32.8 to 20.3 between 1984 and 1986 while the standard deviation falls from 28.5 to only 24.2, the median decreases from 24.0 to 12.0, and the upper limit of the range increases from 200 to 208, indicating an increase in skewness (unreported). Amongst susceptibles, the mean falls from 33.5 to 22.0 while the standard deviation changes by

There are two key features of the data: (1) changes in number of partners are not proportional across people with different risk levels. The mean number of partners falls more than the standard deviation and skewness increases, suggestive of an increase, or much smaller decrease, in risky partners amongst some high-activity individuals. (2) Behavioural response occurs in two distinct ways: some people make incremental changes in number of risky partners whereas others completely abstain from risky sex, choosing celibacy, monogamy, or safe sex instead.

Consider how changes in the distribution of rates of partner change affects the basic epidemiological model, the susceptible – infected – recovered (SIR) model. The SIR model is a system of partial differential equations describing, in the canonical case, flows between the number of people susceptible to, infected by, and recovered from a disease (see Anderson and May (1991) for an extensive discussion). Many variants of this framework have been used to analyze the HIV epidemic (for example, Sattenspiel (1990), Castillo-Chavez *et al* (1991), Hyman and Stanley (1994), or Jacquez *et al.* (1994)). Behavioural change in this literature, when included in the analysis, is handled in an *ad hoc* manner. Rates of partner change are frequently assumed to change proportionally across people with different activity levels according to some exogenously specified function of calendar time (examples include Hethcote *et al.* (1991a and 1991b), Anderson *et al.* (1989), and Jacquez *et al.* (1994)). This approach is problematic for two reasons. First, it is incapable of predicting future behavioural changes to future changes in prevalence. Second, the assumption of proportional decreases in rates of partner change across activity levels is not innocuous; the shape of the distribution matters.

Analysis of the effects of behavioural change is complicated by the fact that the mean is not a sufficient statistic for the effect of distribution of rates of partner change on disease dynamics. In a neighbourhood around zero prevalence, it can be

---

the statistically insignificant amount 28.4 to 28.0.

shown (May and Anderson (1991)) that, under random matching, new infections are proportional to

$$c \equiv \mu + \sigma^2/\mu,$$

where μ and σ^2 are the mean and variance of the distribution of rates of partner change, respectively.[5] For illustration, suppose the mean and variance are functions of prevelance, P. Then a marginal change in P around $P = 0$ induces

$$\frac{dc}{dP} = \frac{d\mu}{dP}\left[1 - \frac{\sigma^2}{\mu^2}\right] + \frac{1}{\mu}\frac{d\sigma^2}{dP}.$$

If an increase in prevalence induces a mean-preserving spread in the distribution, then it spurs the disease. Further, even if the mean falls, if the variance falls by less than $(d\mu/dP)\mu[1 - \sigma^2/\mu^2]$, new infections still increase with prevalence. In contrast, proportional changes in rates of partner change would produce equiproportional changes in the mean and standard deviation, and hence new infections also decrease proportionately with these variables. Substituting the statistics in table 1 into the equation for c, the number of new infections before behavioural change was proportional to 82.23; after behavioural change, new infections are proportional to 138.01, suggesting that behavioural response may have spurred rather than retarded disease spread in the early phase of the epidemic.[6]

The intuition underlying these results is that changes in the distribution of rates of partner change affect the probability a given encounter is with a high risk person. Reductions in risky behaviour by those with relatively few partners, all else equal, actually spur the disease (Anderson *et al.* (1990), Whittaker and Rentin (1992), Kremer (1994a)). This is because a reduction in rate of partner change by a low risk individual increases the probability for everyone else of matching with a high

---

[5] This result obtains because as prevalence approaches zero, the probability of infection becomes proportional to number of partners.

[6] This conclusion is weak, however, since it ignores changes in risk per partner, specifically, increased condom use.

risk individual. Hence, behavioural change, if it occurs asymmetrically across risk groups, can have seemingly perverse effects: a reduction in mean rate of partner change can be consistent with more rapid spread of disease and higher steady–state prevalence (see Kremer (1994a) for an extended discussion). In general, analysis of behavioural change must consider changes in the distribution of activity, not just changes in that distribution's first (or even first and second) moments.

Studies that incorporate economic incentives into epidemic models include Philipson and Posner (1993), Geoffard and Philipson (1995), and Kremer (1994b, 1996). Geoffard and Philipson (1995) consider a population in which agents may take one of two actions: protect against transmission or do not protect. Agents determine whether to protect as the result of an optimal control problem. With heterogeneous agents, this model leads most notably to the prediction that the hazard rate from susceptibility to infection may be decreasing in prevalence, as opposed to the biological epidemiology result that more infecteds increase the risk of the remaining susceptibles.

Kremer (1994b, 1996) considers an SIR model in which the rate of partner change depends on prevalence according to an underlying utility maximization problem. Kremer assumes that agents never learn their infection status and analyses only steady–state prevalence, which allows him to collapse this inherently dynamic problem into a static one. Suppose people select a lifetime rate of partner change, s, to maximize $u(s) - p(s, \theta)$, where $u(\cdot)$ is the utility function and $p(\cdot, \cdot)$ is the lifetime probability of becoming infected, which depends on the selected rate of partner change and on the probability a given partner is infected, θ. Kremer seeks steady–state equilibria (θ cannot change over an agent's lifetime) in which the realized probabilities a given partner is infected induces agents to select the rate of partner change which supports those probabilities, for both homogeneous and two-group populations.

In this environment, Kremer shows that if activity is low enough that $p(\cdot,\cdot)$ can be approximated as linear in rate of partner change, then the desired rate of partner change is decreasing in the marginal probability of infection from another partner. If, instead, activity is high enough that $p(\cdot,\cdot)$ is concave in s, then an increase in the marginal probability of infection may induce "fatalism."[7]

Fatalism is defined as a situation in which an increase in the probability a given partner transmits the infection *increases* the number of risky partners selected. This counterintuitive effect occurs because an increase in the per–contact infection probability can decrease the marginal probability of infection for high–activity people. Consider someone who has ten partners per year at a per–contact risk of 10%. His probability of becoming infected by at least one partner is $1-(1-0.10)^{10}=0.65$. If he chooses an additional partner, that probability increases to 0.69 — the marginal probability of infection, evaluated at ten partners, is 0.04. If the per-contact risk of infection rises to 0.12, then his probability of becoming infected by one of ten partners rises to 0.72, but the marginal probability of infection from one more partner falls to 0.03: the person has less incentive to avoid further risk because he realizes he is more likely to already be infected by the first ten.

II. The Model

A. *Basic Setup*

Consider a person susceptible to an infectious disease who is in the market for non-steady, anonymous partners for each of a finite number T periods. The

[7] McCusick *et al.* (1985) argue that"It cannot be assumed that [people] will change their behaviour despite the fact that AIDS is lethal. Due to the apparently long period ... from exposure to diagnosis, many men are convinced that they have already been exposed to AIDS, [so] they may not be motivated to change their sexual behaviour." This is the earliest exposition of the intuition behind fatalism with which I am familiar.

person selects a number of sexual partners each period.[8] Assume that all information pertinent to assessing the probability that a partner is infected, particularly sexual history, is strictly private knowledge so that potential partners are indistinguishable. The number of risky contacts in the t^{th} period, denoted s_t, is a continuous, (for example, 1 partner per year is equivalent to 0.5 partners per six months) non-negative variable. There is a single infectious disease. At the end of each period, the agent is either infected by one or more of his period contacts or he remains susceptible. Uncertainty is resolved at the end of each period: whenever a decision is made, the agent knows whether or not he is infected.

In period t, the agent faces a probability of becoming infected by a given contact of θ_t. This probability is equal to the product of the probability a contact is with an infected person, denoted P_t, and ϕ, the probability that a contact with an infected transmits the disease: $\theta_t = \phi P_t$. When all agents choose contacts randomly from the pool (as in the simulations presented in section III) P_t, hereafter referred to as *effective prevalence*, is simply the proportion of total contacts offered that are from infecteds. The effective prevalence usually exceeds the proportion of infected agents, *prevalence*, since high-contact agents also have the greatest chances of being infected. The model is, however, applicable to other types of matching so long as the agent can change neither the probability of matching with an infected nor the probability of transmission.[9]

---

[8] I assume that transmission through over vectors, such as IV drug use, is similar enough to sexual transmission so as not to affect the model. As noted, the majority of reported HIV cases are from sexual contact.

[9] For instance, the extension of the simple proportionate matching assumption to heterosexual populations (see, for example, May and Anderson (1991)) would not change the optimization problem presented here. Other types of matching discussed in the epidemiological literature, such as when partially observable sexual histories lead to matching correlated correlated across activity levels (*preferred matching*) or matching by infection status (Philipson and Dow (1994)) can be incorporated under the heroic assumption that the agent can affect neither the probability a given partnership is with an infected nor the probability of transmission. In other words, the matching pattern, by activity level or infection status, must be specified exogenously.

Let h_t denote the agent's infection status, where $h_t = 1$ means the agent is infected and $h_t = 0$ denotes susceptibility. The period return is composed of infection-dependent benefits and costs, denoted $b(s_t, h_t)$ and $c(s_t, h_t)$, respectively. Define the net period return:

$$u(s_t, h_t) \equiv b(s_t, h_t) - c(s_t, h_t), \tag{1}$$

The period return is assumed to be twice continuously differentiable and strictly concave. It is further assumed to have a finite maximum, so that in the absence of an epidemic ($\theta_t = 0 \quad \forall t$) the agent simply chooses the number of partners each period which maximizes $u(s_t, 0)$. Assume that, for values of s_t where marginal utility is non–negative,

$$u(s_t, 0) \geq u(0, 0) = 0 > u(s_t, 1).$$

The agent weakly prefers having a positive number of partners over none – the utility of which is normalized to zero – and prefers having none to being infected. The period return is indexed by a parameter α such that a higher value of α implies higher marginal utility at any given s_t, although I will suppress the argument in order to simplify notation. A "high–risk" type refers to an agent with a relatively high value of α. The period return for an infected is assumed to have the same properties, but may have a different maximum than the return for susceptibles. In section III, the period return is given a functional form and the distribution of α is calibrated to survey data. Note that the number of partners selected in any period has no direct effect on any other period's return; the utility function is additively separable.

The probability of becoming infected by one or more contacts is given by $p(s_t, \theta_t) \in [0, 1]$ and the probability of remaining susceptible is $q(s_t, \theta_t) = 1 - p(s_t, \theta_t)$. Assume that the probability of becoming infected is given by the usual formula for at least one success in s_t trials with per–trial success rate θ_t:

$$p(s_t, \theta_t) = 1 - (1 - \theta_t)^{s_t}. \tag{2}$$

The agent is assumed to act as if he knows present and future values of the per–contact infection probability θ_t.

I assume people 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" for remaining susceptible at the end of the modeled time span of B units, receiving 0 otherwise.[10] Future returns are discounted at a constant rate β.

Infected agents simply choose the satiation number of partners per period:[11]

$$s_t|_{h_t=1} = \text{argmax}_{s_t} u(s_t, 1). \tag{3}$$

Evidence suggests that factors such as aversion to affecting others are important considerations in the choices infecteds make (Cates *et al.* (1988), Schecter *et al.* (1988), McCusker *et al.* (1988)), so I allow the number of contacts selected by infecteds to be lower than the number of partners selected by susceptibles in the absence of disease.

The susceptible agent's problem is to maximize expected utility at each period t[12]:

$$\max_{s_t} EU_t = \sum_{i=t}^{T} \left(p(s_i, \theta_i) \left[\prod_{j=t}^{i-1} q(s_j, \theta_j) \right] \left\{ \sum_{k=t}^{i} \beta^{k-t} u(s_k, 0) + \sum_{k=i+1}^{T} \beta^{k-t} u(s_k, 1) \right\} \right) + \beta^{T-t+1} \prod_{i=t}^{T} q(s_i, \theta_i) B. \tag{4}$$

---

[10] This assumption prevents agents approaching the last period from having no incentive to avoid becoming infected.

[11] However, there is evidence that repeated exposure may accelerate conversion to AIDS, decreasing life expectancy (Phair *et al.* (1992)). I abstract from this fact. Also, since the period return is infection–status dependent, this specification can incorporate altruistic behaviour by infecteds.

[12] Equation (4) uses the notational convention that $\prod_{i=x}^{y}(\cdot) = 1$ if $y < x$.

which, due to the separable utility function, can be written much more compactly in recursive form. Bellman's equation for a susceptible in period t is:

$$V(0,T-t) = \max_{s_t} \{u(s_t,0) + \beta[p(s_t,\theta_t)V(1,T-t-1) + q(s_t,\theta_t)V(0,T-t-1)]\}, \tag{5}$$

where $V(h_t,\tau)$ denotes the value function of an agent with health state h_t with τ periods remaining in the "market." By (3), an infected agent receives $u^I \equiv \max_{s_t} u(s_t,1)$ per period each period until the end of the modeled time period,[13] so

$$V(1,T-t) = \frac{1-\beta^{t+1}}{1-\beta}u^I, \tag{6}.$$

which does not depend on any infection probabilities. For notational convenience, define Δ_t as the loss associated with becoming infected in period t,

$$\Delta_t \equiv V(0,T-t-1) - V(1,T-t-1) \tag{7}$$

That is, the loss to becoming infected in period t is equal to the difference in the value functions associated with susceptibility and infection in the following period.

B. *Discussion.*

Using (2), (5), and (7), a susceptible agent's problem in period t can be written,

$$\max_{s_t} u(s_t,0) - \beta(1-(1-\theta_t)^{s_t}))\Delta_t. \tag{8}$$

Intuitively, the agent can increase today's utility only at the cost of increasing the probability that he will lose Δ_t. The first-order condition is:

$$\frac{\partial u(s_t,0)}{\partial s_t} = \beta(-ln(1-\theta_t)(1-\theta_t)^{s_t})\Delta_t. \tag{9}$$

---

[13] Note that there is no disease-induced mortality in this model; incorporating it is straightforward but does not qualitatively affect the agent's problem.

The right-hand side is the marginal decrease in expected lifetime utility from increasing s_t, that is, it is the marginal cost of risky behaviour. The left-hand side is simply the marginal utility of risky behaviour. The agent sets s_t to equalize these quantities at an internal solution (note that there is no analytical solution for s_t), as shown in figure 1 for two values of θ_t.

Marginal utility is linear and decreasing in s_t, with intercept $(-ln(1-\theta_t))\Delta_t$, whereas marginal cost is convex to the origin. The solution to (14) does not necessarily yield the optimal s_t given there may be a local minimum or the interior solution may yield lower expected utility than setting $s_t = 0$. The second-order condition for a maximum is:

$$\Theta \equiv \frac{\partial^2 u(s_t, 0)}{\partial s_t^2} - \beta\Delta_t[-(ln(1-\theta_t))^2(1-\theta_t)^{s_t}] \leq 0. \tag{10}$$

Intuitively, this inequality requires that the marginal cost curve must strike the marginal utility curve from below for a maximum to be reached. Denote the value of s_t which satisfies the first and second order conditions s_t^*.

The following propositions characterize the agent's responses to changes in the interesting variables, infection risk over time. The proofs make use of the following lemma:

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

PROOF:

Part 1 follows directly from equations (4) and (5); infection risk today does not affect the value of being infected or susceptible in the future. To prove part 2, consider expected utility at period t evaluated at an arbitrary sequence of partnership choices, $\{\tilde{s}_t\}$. An increase in θ_{t+i}, $\quad 0 < i \leq T-t$, does not change the probability of becoming

infected prior to $t+i$. If $\tilde{s}_{t+i} > 0$, the probability of becoming infected in period $t+i$ rises. Expected utility in period $t+i+1$ is independent of θ_{t+i}. Since the probability of infection in $t+i$ is weakly higher, expected utility evaluated at $\{\tilde{s}_t\}$ is decreasing in θ_{t+i}. But $\{\tilde{s}_t\}$ is an arbitrary sequence, so expected utility is also decreasing in θ_{t+i} for the optimal sequence. QED.

The first proposition shows that pessimistic beliefs about the future course of the epidemic tend to increase current risky behaviour.

PROPOSITION 1. *An increase in a future per-contact infection probability will increase the optimal number of partners chosen in the current period, that is,* $\partial s_t / \partial \theta_{t+i} > 0, \quad i \in \{1, T-t\}$.

PROOF:

Differentiate (14) with respect to θ_{t+k}, $k \geq 1$ to find:

$$\frac{\partial s_t^*}{\partial \theta_{t+k}} = \frac{\beta}{\Theta}(-ln(1-\theta_t)(1-\theta_t)^{s_t})\frac{\partial V(0,t-1)}{\partial \theta_{t+k}} \geq 0 \tag{11}$$

The inequality holds by the Lemma. QED.

To see the intuition, consider one extreme case in which the agent knows the disease is going to be cured next period, $\theta_i = 0, i = t+1 \ldots T$. Risky behaviour today is potentially very costly, as the returns to partnerships in the future are large. At the other extreme in which future infection probabilities are expected to become and remain very high, risky behaviour today is less costly since there is not as much to lose. Equivalently, an increase in a future infection probability increases the probability the individual places on becoming infected in the future (more accurately, decreases expected future welfare because of increased chances of becoming infected at

any given activity level), which decreases incentives to remain susceptible today.[14] A more pessimistic view of future conditions decreases the expected net cost of becoming infected, which in turn decreases incentives to reduce risk in the current period. One implication of this reasoning is that the announcement of the future release of a partially effective vaccine will reduce new infections immediately. This holds even if, when the vaccine is released, people respond to the reduced risk of transmission by increasing risky behaviour: if the product $\theta = \phi P$ is expected to fall, expected welfare increases and current risky behaviour decreases.

The next proposition shows that an increase in current risk will tend to increase risky behaviour for very high–risk agents while reducing such behaviour for others.

PROPOSITION 2. (Kremer 1994b) *An increase in date t's per-contact infection probability has an ambiguous effect on he number of partners chosen at date t.*

PROOF:

Totally differentiate (14) with respect to θ_t to find:

$$\frac{\partial s_t^*}{\partial \theta_t} = \frac{\beta}{\Theta}(1-\theta_t)^{s_t-1}[1+s_t ln(1-\theta_t)]\Delta_t, \tag{12}$$

which is negative iff

$$1 + s_t ln(1-\theta_t) > 0. \quad \text{QED.} \tag{13}$$

If the cross-derivative $\partial^2 p/\partial s_t \partial \theta_t$ is positive, so the marginal cost of another partner is increasing in θ_t, then an increase in θ_t causes the agent to decrease risky contacts. Note that an increase in θ_t causes a decrease in number of partners if

$$\theta_t < \hat{\theta}_t \equiv 1 - exp(-\frac{1}{s_t}), \tag{14}$$

---

[14] A recent article in the *New York Times Magazine* (Green 1996) argues that reductions in risky behaviour occurred "at least while it was still believed that the disease might disappear momentarily," but that risky behaviour increased when "it became clear that no cure was imminent." This observation be interpreted as a pessimistic revision of expectations inducing more risky behaviour. An alternate economic explanation for the "relapse" to unsafe behaviour is given by Philipson and Posner (1993).

where $\hat{\theta}_t$ denotes the critical value of θ_t at which $\partial s_t^*/\partial\theta_t$ becomes positive.

This possibly counterintuitive result implies that if this period's per-contact infection probability rises then some agents may choose higher levels of risky behaviour.[15] This effect occurs because the *marginal* probability of remaining uninfected, conditional on s_t, is decreasing in θ_t for $\theta_t > \hat{\theta}_t$. As illustrated in figure 1, increasing θ_t increases the intercept and slope of the marginal cost schedule, implying (as is clear from the observation $\lim_{s_t\to\infty} p(s_t,\theta_t) = 1 \quad \forall\theta_t > 0$) that it crosses somewhere, as illustrated in figure 1. If the crossing point occurs at a value of s_t lower than the optimal s_t, then the counterintuitive result holds. Conversely, if the crossing point occurs at a value of s_t higher than the optimal s_t, then an increase in per–contact risk of infection causes the agent to decrease the number of risky contacts selected.

COROLLARY. *Increases in the per–contact probability of infection may increase variance in the amount of risky behaviour.*

From (14), we see that agents with high values of s_t are more likely to increase s_t in response to an increase in θ_t. Therefore, there exists some threshold level of activity (specifically, $\hat{s}_t = -(ln(1-\theta_t)^{-1})$ such that agents below this level decrease activity and vice versa. This in turn results in a spread of the distribution of s_t if any agents had activity levels greater than $\hat{s}_t$. Statistical methods that only examine changes in the mean of s_t may then underestimate the true magnitude of behavioural response, since some agents are increasing and some decreasing their exposure levels. Note that no specific per-contact probability of infection is required to induce fatalism;

---

[15] This result is identical to Kremer's (1994b) result in the case where the number of periods is one. In the dynamic case, it holds within any period rather than within the agent's lifetime.

even very low effective prevalence can produce increases in partners in response to an increase in risk if activity levels are high enough.[16]

Proposition 2 implies that behavioural response induced by increases in θ_t may have perverse effects on the spread of disease. As discussed in section I, increases in the variance of activity increases the rate of change of prevalence, since the probability of matching with one of the highest risk agents increases. Here, the agents who increase their activity levels are the most active ones, who are also the ones most likely to be infected. Since these agents play a disproportionately large role in spreading the disease, any increase in their activity levels will accelerate the epidemic.

PROPOSITION 3. *Changes in a steady-state per-contact infection probability have an ambiguous effect on number of partners selected.*

PROOF:

Define θ as the steady-state infection probability, that is, $\theta_t = \theta \, \forall t$. Then

$$\frac{\partial s_t^*}{\partial \theta} = \left(\frac{\beta}{\Theta}\right)\Bigg[(1-\theta)^{s_t-1}[1+s_t ln(1-\theta)]\Delta_t - ln(1-\theta)(1-\theta)^{s_t}\frac{\partial V(0,t-1)}{\partial \theta}\Bigg], \tag{15}$$

which is negative iff:

$$(1+s_t ln(1-\theta))\Delta_t - ln(1-\theta)(1-\theta)\frac{\partial V(0,t-1)}{\partial \theta} > 0. \quad \text{QED.} \tag{16}$$

Note that

$$\frac{\partial s_t^*}{\partial \theta} = \frac{\partial s_t^*}{\partial \theta_t} + \frac{1}{\Theta}(-ln(1-\theta)(1-\theta)^{s_t})\frac{\partial V(0,t-1)}{\partial \theta} \tag{17}$$

Since the last term is positive, we have

$$\frac{\partial s_t^*}{\partial \theta} > \frac{\partial s_t^*}{\partial \theta_t} \tag{18}$$

---

[16] For instance, if a period is one year, a person with 250 partners per year will become fatalistic at a per-contact infection probability of only 0.004.

An increase in a future probability increases s_t unambiguously; an increase in the steady-state probability increases both current and future probabilities, so the steady-state change will be greater (less negative, possibly) than changing only today's rate. Agents not exhibiting fatalism reduce their number of partners less, and fatalistic agents increase their number of partners more, relative to the case where only the current probability changes.

It is plausible that changes in the perceived current risk of infection will also change expected future risk of infection. Kremer (1994b) notes that (13) implies that the probability of infection at which fatalism begins is $1 - 1/e \approx 0.63$.[17] This probability will be lower, however, if expected future infection risk rises with current risk: from (18), $\partial s_t/\partial\theta = 0$ implies $\partial s_t/\partial\theta_t < 0$, implying the critical value of s_t at which fatalism begins is lower when steady-state prevalence changes. Generally, if increases in beliefs about current risk are associated with increases in expected future risk, disease-minimizing behavioural response will be diminished: those reducing their activity will do so by a smaller amount, a greater number will increase their risky behaviour, and those already exhibiting fatalistic responses will further increase their risk levels.

An implication of increased risky behaviour in response to more pessimistic forecasts of the future path of the epidemic is that statistical methods that overestimate future cases contribute to disease spread. For instance, Philipson and Posner (1993) note that the Centers for Disease Control predicted 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." To the extent that individuals are aware of such projections and base their own predictions upon them, a direct implication of proposition 3 is that these overly pessimistic forecasts increased risky

[17] To see this, exponentiate both sides, add $P(s_t, \theta_t)$ to both sides and simplify. See also Kremer (1996) for a case in which fatalism sets in at a lifetime infection probability of one-half.

behavior, thereby spurring the epidemic. A related implication is that educational efforts designed to minimize risky behaviour should not emphasize dire predictions of the future state of the disease.

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

$$u(s_t^*,0)+\beta[p(s_t^*,\theta_t)V(1,T-t-1)+q(s_t^*,\theta_t)V(0,T-t-1)] > \beta V(0,T-t-1). \quad (19)$$

If this inequality does not hold, the agent chooses to have no risky contacts, since he can get $\beta V(0,t-1)$ with certainty by setting $s_t = 0$, whereas he receives expected utility equal to the left-hand of (19) side at the interior solution. Substituting and rearranging,

$$u(s_t^*,0) > \beta p(s_t^*)\Delta_t. \quad (20)$$

The left-hand side is the gain in this period's utility and the right-hand side is the expected loss from setting $s_t = s_t^*$ rather than zero.

PROPOSITION 4. *1) An increase in the current infection probability decreases the probability of participation, whereas 2) an increase in a future infection probability increases it, and 3) a higher steady-state infection probability has an ambiguous affect on participation.*

PROOF:

Differentiating (20) with respect to θ_t and invoking the envelope theorem, the left-hand side does not change and the right-hand side changes by $-\beta\Delta_t\partial p(s_t,\theta_t)/\partial\theta_t < 0$. This establishes the part 1) of the proposition. Part 2) follows by differentiating with respect to a future infection probability, the right hand-side changes by $-\beta p(s_t,\theta_t)\partial\Delta_t/\partial\theta_{t+i} > 0, \quad i \in \{1,T-t\}$, which establishes that an increase in a future expected probability increases participation. If the steady-state probability

changes, participation increases iff $\Delta_t(\partial p(s_t,\theta)/\partial\theta) + p(s_t,\theta)(\partial\Delta_t/\partial\theta) < 0$, proving part 3). QED.

If the current per-contact probability of infection increases *ceteris paribus*, the cost of participating in the market 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, leading to fewer exits. When both change due to a change in steady state, participation could either increase or decrease. Non-participation can be interpreted in either of two ways: agents become celibate or monogamous (under the assumption that subjective risk under monogamy is negligible), or that the activities chosen with multiple partners have a subjective risk of zero, for instance, strict condom use. Which interpretation is used is irrelevant to the theoretical model, but affects the dynamics of the disease — if exit is into safe-sex partnerships and those participating in the market match not only with each other but also with those not participating who insist on condoms, disease spread will be retarded since the risk of matching with a high–risk, transmissive partner is diminished (Kremer (1994a)). If, on the other hand, those insisting on safe sex match only with each other, their exit increases risk in the pool of those still participating, accelerating the epidemic.

In the usual case where the lowest-risk agents are the first to cease participating in response to increasing effective prevalence,[18] proposition 4 implies that variation along the extensive margin in response to an increase in today's risk will spur rather than retard the spread of the disease (with either of the interpretations of non-participation). The lowest-risk agents are also the least likely to be infected; their exit increases the proportion of active agents who are infected, which can dramatically increase effective prevalence among agents still participating. On the other

[18] This case is guaranteed only if the utility function is specified in such a way that $\partial\Delta/\partial\alpha < 0$, for instance, by a specification in which disease-free indirect utility is normalized to be constant across α.

hand, expectations that conditions will be worse in the future cause more low-risk agents to participate, which can reduce the speed at which the disease spreads. The participation decision does imply an upper bound on prevalence in the long run: for any given steady-state effective prevalence, a proportion of agents never participate in the market for risky sex at all, and hence cannot become infected.

Recall that the per-contact infection probability is the product of the effective prevalence and the biological probability of transmission. Propositions 2 through 4, since they are based on comparative dynamics with respect to this product, hold for either effective prevalence or the transmission coefficient. This implies that, under non-fatalistic behaviour, a reduction in the transmission coefficient, due to, for instance, substitution towards safer forms of sex, less sharing of needles, or a partially effective vaccine, will be at least partially offset by an increase in activity (similar arguments are given by Brauer, Castillo-Chavez, and Velasco-Hernandez (1994) and Kremer (1994b), and the general argument closely resembles the theory of risk homeostasis (for example, Peltzman (1975)). Under fatalistic behaviour, a reduction in the transmission coefficient will have reinforcing second-order effects. That is, if $\partial s_t^*/\partial \theta_t > 0$ then $\partial s_t^*/\partial \phi > 0$, so a reduction in ϕ will cause a reduction in s_t^* for high-activity individuals. This effect is reinforced by the change in expectations due to a change in ϕ: if ϕ decreases not just in the current period but permanently, expectations that the future situation has improved induces further decreases in risky activity, by propositions 1 and 4. Hence, substitution towards safer contacts may, in and of itself, lead to reduced rates of partner change amongst the highest-risk individuals.

It is worth emphasizing that the driving assumption in the model is the fact that agents do not know whether a previous *within-period* contact has infected them. If the agent's infection status were revealed after each contact, the decision process collapses to a comparison of the benefits and costs of one more partner, conditional

on being susceptible, as opposed to the benefits of one more partner, conditional on having had some number of partners in the past with uncertainty about transmission. The formulation here provides a tractable approximation to a continuous time model where uncertainty about past contacts is revealed with a lag. The key point is that the marginal probability of infection for one more partner is not constant when uncertainty about past contacts is unresolved.

III. Simulations

A. *Setup*

The complexity of the modeled behaviour precludes analytical solutions for the time-path of the epidemic. I employ numerical simulation to explore the implications of rational behaviour and contrast them to the the assumptions of biological epidemiology and explore some implications of rational behavioural response. A detailed explanation of the simulation environment can be found in the Appendix. Briefly, in order to focus on behavioural considerations, the epidemiological environment is highly simplified: there is no disease-induced mortality, matching is purely random, and infectivity does not depend on time elapsed since infection.[19] Time is discrete; agents remain in the market for 30 periods, where each period represents six months of calendar time. The period return function is assumed to be quadratic,

$$u(s_t, 0) = \alpha s_t - s_t^2, \qquad u(s_t, 1) = \gamma\alpha s_t - s_t^2, \quad \gamma \in (0, 1).$$

Per–contact utility, and hence number of partners chosen relative to the satiation number in the absence of disease, is scaled down by a factor γ for infecteds, which allows the number of partners chosen by infecteds to reflect aversion to infecting others. This parametrization implies that an agent's value of α is identified by observing

[19] It is well-known that infectivity actually varies greatly with time since infections, see Jacquez et al (1994).

his rate of partner change in the absence of disease, since $\text{argmax}_{s_t} u(s_t, 0) = \alpha/2$. Agents are heterogeneous in α. There are 100 types per generation. The n^{th} type corresponds to the n^{th} percentile of the distribution of rates of partner change in the absence of disease in the SFMHS. Other parameters are chosen to be consistent with other studies when known (see Appendix), in an *ad hoc* fashion otherwise.

B. *Belief Formation*

The formation of expectations are crucial to the modeled behaviour. I examine three different classes of expectations: myopic, adaptive, and perfect foresight or "rational." Myopic expectations, here used synonymously with a discount factor of 0, imply no behavioral changes occur (when β is zero, agents select the satiation number of partners each period), corresponding to the assumptions frequently made in the biological epidemiological literature. This case is considered as a baseline with which to compare the impact of rational response on disease dynamics. Agents with adaptive expectations observe the effective prevalence in the previous period and assume that same probability holds in the current period and all future periods. Perfect foresight expectations refers to the case in which expectations are fulfilled: agents know today's and all future effective prevalences. See the appendix for details of the computation of the perfect foresight solutions. Perfect foresight differs from adaptive in that agents with perfect foresight will correctly predict future changes in prevalence and change their current behaviour in response, although in a non-cyclic steady state the two modes of expectation formation will be identical.

Consider the effect of expectations on disease dynamics. Figure 2 shows incidence paths for agents with myopic, adaptive, and perfect foresight expectations with a transmission coefficient of 0.125 and Figure 3 with a transmission coefficient of 0.05. With the lower transmission coefficient, both adaptive and perfect foresight expectations involve *faster* disease spread than the myopic (no behaviour change) case for

the first several years of the epidemic. Perfect foresight expectations involves faster spread than adaptive,[20] since, as per Proposition 1, agents who believe that the epidemic is going to become worse do not reduce their risky behaviour as much as if they are not forward looking. Figure 2, however, shows that with a higher transmission coefficient, the disease spreads fastest in the myopic population. In either case, steady-state levels are significantly higher when no behavioural change occurs, largely because of the significant steady-state non-participation in the economic solutions (Figure 4). These simulations demonstrate that behavioural response can either spur or slow spread of disease, depending on the underlying epidemiological and behavioural parameters.

C. *Policy interventions: information and vaccines*

Geoffard and Philipson (1995) argue that public subsidies to change behaviour are crowded out by increases in prevalence (in their model, both subsidies and increases in prevalence increase incentives to cease risky behaviour) and emphasize the role of information in the spread of HIV. Figures 5 and 6 consider the effects of the most important form of information: the existence of the disease and how it is transmitted. HIV spread rapidly in the homosexual population in the U.S. between 1981 and 1984, a time when the disease itself and its transmission modes were first being identified. When people are unaware an illness exists, they simply select the satiation number of partners. When information that the disease exists becomes common knowledge, behavioural response sets in. Figure 5 shows prevalence paths for four epidemics: the myopic case, and three with perfect foresight expectations with varying delays from the beginning of the epidemic to the time when agents simultaneously

---

[20] The framework here understates the difference between expectations in the early stages of the epidemic, since with adaptive expectations agents believe that today's effective prevalence is the same as it was last period whereas perfect foresight agents know today's probability, so when effective prevalence is rising adaptive agents systematically underestimate today's risk and behave more riskily.

learn of its existence. Prevalence decreases monotonically to its steady–state level upon release of information the

disease exists *if* it had already exceeded the steady-state level. The later information is released the greater the total number of infections that take place, despite the equivalent steady–state levels. Figure 6 illustrates that, because of behavioural change, new infections fall upon release of information to the level that would have prevailed had the disease been discovered earlier. Despite the larger pool of infecteds, excess infections due to ignorance of the disease are immediately curtailed by behavioural change upon release of the information. These effects corroborate Geoffard and Philipson's argument: steady-state prevalence is invariant to when people become aware of the disease, but total infections are an increasing function of time to the release of information.

Consider the effects of the release of a partially effective vaccine, modeled as a decline in the transmission coefficient, ϕ. Castillo– Chavez and Hadeler (1994) and Kremer (1996) argue that the release of a vaccine could increase infections if offsetting behavioural change is strong enough. This effect is illustrated in figure 7, displaying prevalence for an epidemic in which a vaccine which reduces the transmission coefficient from 0.125 to 0.025 is introduced twelve years after the epidemic begins.[21] Even with this dramatic reduction in per-contact transmission rates, prevalence rises for both adaptive and myopic expectations, while falling in the myopic case. This effect occurs for two reasons: variation on the intensive margin increases rates of partner change for most agents, and also because far more agents participate after the vaccine is introduced (figure 8). Effective prevalence actually falls (undisplayed) despite increased average activity levels due to the increase in participation by low-risk agents,

[21] Prevalence before the vaccine release is lower for the perfect foresight solution than the adaptive solution because the former anticipate, from the beginning, the release of the vaccine at 12 years and hence engage in less risky behaviour each period prior to the vaccine, by Propositions 1 and 4.

yet many more people at risk cause infections to rise. This is again suggestive that the participation decision is more important to disease dynamics than incremental changes in number of partners.

In summary, the simulations show that behavioural response can dramatically affect disease dynamics. Adverse selection, in which low–risk agents exit the market first and proportionately reduce risky behaviour the most, can accelerate the spread of disease during its early stages. This effect is exacerbated by expectations the epidemic is to become worse in the future. A partially effective vaccine, even one which dramatically reduces per–contact infectivity, is predicted to increase the number of infections which take place. The date at which the population becomes aware of the existence of the disease can dramatically affect the total number of infections which occur, although steady-state prevalence will be unaffected. Predicting the future course of the AIDS epidemic and the consequences of policy interventions could be accomplished by formal estimation of the parameters of the underlying model with a more realistic epidemiological environment.

IV. Concluding remarks and directions for future research

This paper demonstrates that rational response to an infectious disease can lead to perverse and counterintuitive effects for both the individual and the population. An individual's response to increased risk of infection may be to undertake even more risk. The tendency for low–risk agents to cease risky activities before higher risk agents, coupled with lower reductions or even increases in activity for high risk agents, tends to accelerate the spread of the disease. I also show that expectations about the future course of the epidemic affect current behaviour; pessimistic expectations decrease incentives to avoid risk. Public health efforts designed to minimize disease spread should not emphasize rapid spread of the disease or make dire predictions about the future of the epidemic.

Simulation of the model calibrated to epidemiological and behavioural data reveals that disease dynamics are affected greatly by behavioural response. Both the extensive and intensive margins affect spread and steady-state prevalence. Policy simulations show that public efforts to transmit information about the existence of the disease and its modes of transmission can decrease the total number of infections. However, the release of a partially effective vaccine, at the calibrated parameter values, is predicted to increase prevalence, mostly through an induced increase in the number of agents who participate in the market for risky sex.

The results presented here suggest other implications for the economic analysis of epidemics that remain to be explored:

First, it is not productive to attempt to estimate the elasticity of demand for risky sex to prevalence. The demand for risky sex is not analogous to the demand functions for commodities ordinarily encountered in economics. We can sum the demand for apples across consumers, but we cannot similarly aggregate the demand for risky sex because the entire distribution of rates of partner change affects the dynamics of the epidemic. For example, a reduction in number of partners for the top decile of individuals does not have the same implications as the same reduction for the least active decile. Consequently, empirical investigations of the response to changes in risk should attempt to measure not a single elasticity but a response function across activity levels. Further, exit from the market for risky partners is at least as important as changes at the margin, so response at both the intensive and extensive margins must be estimated.

Second, this paper implicitly assumes that the search costs of acquiring new partners are not affected by changes in others' rates of partner change. If, in fact, it becomes more costly to find partners when fewer partnerships are being offered in total, then increasing one's rate of partner change generates externalities other than increasing risk of infection to others — the higher level of activity makes it

easier for others to trade, which could generate multiple equilibria. If this effect is quantitatively important, it complicates empirical analysis of behavioural response.

Third, this paper takes the amount of calendar time between updates of infection status and the probability a partner is infected as fixed, but presumably both of these variables are endogenous. A seemingly apparent implication of this model is that subsidized testing would slow disease spread as it decreases risky behaviour amongst those finding they remain susceptible, increasing the incentive to avoid risky activity. However, the endogenous nature of testing may belie this conclusion: if low-risk agents choose to test more frequently than their higher risk counterparts, then the result of subsidized testing will be relatively less risk reduction by high-risk agents, which will accelerate the spread of the disease. Philipson (1994) argues that neither low nor high risk individuals will choose to test, since these groups are most certain about their infection status and therefore have the least to gain from testing. Subsidized testing, then, may not minimize the spread of the disease, as those at highest risk will tend to be amongst the last to be induced to test, and reductions in risk by those in the middle may actually spur the disease. More research is needed on the interrelation between rate of partner change and testing.

Finally, this paper, following both the biological and economic epidemiological literatures, takes the probability of transmission from risky sex to be parametric. Empirical evidence (for example, Lawrence *et al.* (1989)) falsifies this assumption: not only do people respond to risk of infection by changing their number of risky partners and exit into monogamy, they also change the sorts and number of acts performed with each risky partner they do have, which effectively endogenizes the transmission parameter. As Kremer (1994b) argues in the case of condoms, the mixing pattern between individuals practicing various levels of sex can affect the course of the epidemic. However, the incentives individuals have when selecting the level of risk in a given partnership are not well understood.

APPENDIX

Details of Simulation Environment

Time is discrete. Each period, a group of N agents is born into the "market" for randomly-matched sexual partners. The period return is quadratic, as described in the text. Agents are identical up to the taste parameter, α, which effectively indexes disease-free rates of partner change (recall that $s = \alpha/2$ in the absence of disease). Agents remain in this market for T periods, regardless of whether or not they become infected. The period return for infecteds is adjusted so that all types receive the same return, D, when infected, to simplify calibration. This normalization is equivalent to subtracting $\gamma^2\alpha^2/4 - D$ from the period return function for infecteds.

Each period, agents select a number of partners, s, according to the optimization problem discussed in section II. Let S denote the set of susceptible agents and I the set of infecteds. Then the probability a given contact is with an infected is

$$P = \frac{\sum_{i \in I} s_i}{\sum_{i \in I} s_i + \sum_{i \in S} s_i}.$$

Let ϕ_t denote the possibly time-dependent probability that an infected trader infects a susceptible on a given trade. Then the per-contact probability of infection in period t is:

$$\theta_t = \phi_t P_t.$$

New infections each period were assigned according to the above two equations and equation (2). Since the number of agents is finite, the endogenous aggregate variables follow a stochastic process. If there are N identical agents who face a period infection risk of θ, then the number infected would follow a binomial distribution with parameters (N, θ). Although this does introduce unwanted uncertainty into the

simulation, with a large number of agents the discrepancy between the expected proportion infected and the realized values of this variable will be small, as the variance of the proportion of successes is inversely related to N.

The simulations are calibrated to roughly match the situation in San Francisco from the late seventies through the late eighties, in order to provide some concreteness without the expense of formal estimation of all the parameters. The distribution of α was chosen to match the distribution of rates of partner change prior to the epidemic, as in Table 1. Specifically, 100 types, corresponding to the percentiles of the distribution of partners prior to 1982, are assigned, with 15 agents per type per generation. Each period is assumed to be six months of calendar time and agents remain in the market for thirty periods, so that there are (100)(150)(30)=600,000 individual agents in the market at any given time, with (30)(100)=3,000 unique dynamic optimization problems to be solved each period. The transmission coefficient, ϕ, was chosen to be consistent with epidemiological evidence while abstracting from variable infectivity. Grant *et al.* (1987) finds 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 2, were chosen in an *ad hoc* manner to loosely 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 77, compiled using the XLF compiler, and executed on an an IBM RS/6000 model 43P running at 133 megahertz. Each

simulation, or iteration on the perfect foresight solution, took roughly five minutes of CPU time. The agents' dynamic programming problems were solved by backwards induction on the Bellman's equation.

Computation of perfect foresight equilibria

The perfect foresight solutions were computed by the following algorithm.

1. Posit an arbitrary sequence of expected prevalences, P_0^e.

2. Simulate the epidemic that results from expectations P_0^e and store the corresponding realized effective prevalences, P_0.

3. Update the expected effective prevalences by using a linear combination of the expected and observed prevalence paths, $P_1^e = (0.5)P_0^e + (0.5)P_0$.

4. Continue iterating on expected and observed prevalence paths until a convergence criterion is met.

The convergence criterion used in the simulations in the paper was that the average deviation between expected and observed outcomes must be less than 0.005 and no particular period can have an error greater than 0.01. Since there is some stochastic variation in outcomes, attempting to use a more stringent convergence criterion leads to excessive numbers of iterations. The limits here produce convergence in an average of approximately ten iterations. Note there is no guarantee that this procedure will converge, but hundreds of trial simulations produced only one failure.

References

Ahituv, A., V. Hotz, and T. Philipson (1993) "Will the AIDS Epidemic be Self-Limiting? Evidence on the Responsiveness of the Demand for Condoms to the Prevalence of AIDS" Population Research Center discussion paper 93-3.

Anderson, R., R. May, G. Medley and A. Johnson (1986) "A Preliminary Study of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV)" *IMA J. Math. Appl. Med. Biol.* **3** 229-63.

Anderson, R., S. Blythe, S. Gupta and E. Konings (1989) "The Transmission Dynamics of the Human Immunodeficiency Virus type I in the Male Homosexual Community in the United Kingdom: the Influence of Changes in Sexual Behavior" *Philos. Trans. R. Soc. Lond [Biol]* **325** 45-98.

Anderson, R. and R. May (1991) *Infectious Diseases of Humans: Dynamics and Control* Oxford: Oxford U.

Becker, M. and J. Joseph (1988) "AIDS and Behavioral Change to Reduce Risk: A Review" *American Journal of Public Health* **78**:394-410.

Blower, S. and C. Boe (1993) "Sex Acts, Sex Partners, and Sex Budgets: Implications for Risk Factor Analysis and Estimation of HIV Transmission Probabilities" *Journal of the Acquired Immune Deficiency Syndromes* **6**:1347-1352.

Castillo-Chavez, C., ed., (1989) *Mathematical and Statistical Approaches to AIDS Epidemiology* Berlin: Springer-Verlag.

Castillo-Chavez, C., and K. Hadeler (1994) "A Core Group Model for Disease Transmission" Working paper, Yale University.

Castillo-Chavez, C., K. Cooke, W. Huang, and S. Levin (1991) "On the Role of Long Incubation Periods in the Dynamics of Acquired Immunodeficiency Syndrome (AIDS). Part 2: Multiple Group Models." in Castillo-Chavez , ed, (1989).

Centers for Disease Control (1993) *HIV/AIDS Surveillance*, Feb. (Year- End Edition). Atlanta: Centers for Disease Control and Prevention.

Geoffard, P. and T. Philipson (forthcoming) "The Empirical Content of Structural Models of Infectious Diseases: the Proportional Hazard Specification" *Biometrika.*

Geoffard, P. and T. Philipson (1995) "Rational Epidemics and their Public Control" Mimeo.

Grant, R., J. Wiley, and W. Winkelstein (1987) "Infectivity of the Human Immunodeficiency Virus: Estimate from a Prospective Study of Homosexual Men" *Journal of Infectious Disease*' **156**:189-193.

Green, J. (1996) "Flirting with Suicide" *New York Times Magazine* September 15.

Hart, G. *et al.* (1992) " 'Relapse' to Unsafe Behaviour among Gay Men: A Critique of Recent Behavioural HIV/AIDS Research" *Sociology of Health and Illness* **14**:216-232.

Hethcote, H., J. Van Ark, and I. Longini, Jr. (1991a) "A Simulation Model of AIDS in San Franscisco: I. Model Formulation and Parameter Estimation" *Mathematical Biosciences* **106**:203-222.

Hethcote, H., J. Van Ark, and J. Karon, (1991b) "A Simulation Model of AIDS in San Francisco: II. Simulations, Therapy, and Sensitivity Analysis" *Mathematical Biosciences* **106**:223-247.

Hyman, J. and E. Stanley (1994), "A Risk-Based Heterosexual Model for the AIDS Epidemic with Biased Partner Selection" in Kaplan and Brandeau, eds, (1994).

Jacquez, J., J. Koopman, C. Simon, and I. Longini (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.

Jager, J. and E. Ruitenberg (1988) *Statistical Analysis and Mathematical Modelling of AIDS* Oxford: Oxford.

Kaplan, E. and M. Brandeau, eds, (1994) *Modeling the AIDS Epidemic: Planning, Policy, and Prediction* New York: Raven.

Kellogg, T., W. Marelich, M. Wilson *et al.* (1990) "HIV Prevalence Among Homosexual and Bisexual Men in the San Francisco Bay Area: Evidence of Infection Among Young Gay Men." Abstract FC548, 6th International Conference on AIDS, San Francisco, June 1990.

Kaslow, R. and D. Francis (1989) *The Epidemiology of AIDS* Oxford.

Kremer, M. (1994a) "Can Having Fewer Partners Increase the Prevalence of AIDS?" NBER 4942.

Kremer, M. (1994b) "Integrating Behavioral Choice into Epidemiological Models of the AIDS Epidemic" mimeo.

Kremer, M. (1996) "Integrating Behavioral Choice into Epidemiological Models of the AIDS Epidemic" *Quarterly Journal of Economics* **111**: 549-573.

May, R., R. Anderson, and S. Blower (1990) "The Epidemiology and Transmission Dynamics of HIV-AIDS" in Graubard (1990).

McKusick, L. *et al.* (1985a) "AIDS and Sexual Behaviour Reported by Gay Men in San Francisco" *American Journal of Public Health* **75** 493-96.

McKusick, L. *et al.* (1985b) "Reported Changes in the Sexual Behavior of Men at Risk for AIDS, San Francisco, 1982-84 – the AIDS Behavioral Research Project" *Public Health Reports* **100** 622-29.

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.

Ng, J. and J. Orav (1990) "A Generalized Chain Binomial Model with Application to HIV Infection" *Mathematical Biosciences* **101**: 99-119.

Peltzman, S. (1975) "The Effects of Automobile Safety Regulation" *Journal of Political Economy* **83**: 677-725.

Phair, J. et al (1992) "Acquired Immune Deficiency Syndrome Occuring within 5 Years of Infection with Human Immunodeficiency Virus Type-1: The Multicenter AIDS Cohort Study." *Journal of Acquired Immunodeficiency Syndromes* **5**: 490-496.

Philipson, T. (1994a) "Private Vaccination and Public Health: An Empirical Examination for US Measles" mimeo.

Philipson, T. (1994b) "Public Sector Screening for Infectious Disease" in Kaplan and Brandeau (1994).

Philipson, T. and W. Dow (1994) "Infectious Disease Transmission and Infection-Dependent Matching" mimeo.

Philipson, T. and R. Posner (1993) *Private Choices and Public Health: the AIDS Epidemic in an Economic Perspective* London: Harvard University Press.

Sattenspiel, L. (1990) "Modeling the Spread of Infectious Disease in Human Populations" *Yearbook of Physical Anthropology* **33**:245-276.

Sattenspiel, L. and C. Castillo-Chavez (1990) "Environmental Context, Social Interactions, and the Spread of HIV" *American Journal of Human Biology* **2**:397-417.

Sattenspiel, L., J. Koopman, C. Simon and J. Jacquez (1990) "The Effects of Population Structure on the Spread of HIV Infection" *American Journal of Physical Anthropology* **82**:421-429.

Tarantola, D., J. Mann, C. Mantel, and C. Cameron (1993) "Projecting the Course of the HIV/AIDS Pandemic and the Cost of Adult AIDS Care in the World" in Kaplan and Brandeau, eds, (1993).

van den Hoek, J., H. van Hasstrecht, and R. Coutinho (1990) "Increase in Unsafe Homosexual Behaviour" *Lancet* **336**:179-180.

Winkelstein, W., D. Lyman, N. Padian et al (1987a) "Sexual Practices and Risk of Infection by the AIDS Assiciated Retrovirus: the San Francisco Men's Health Study" *Journal of the American Medical Association* **257**:326-330.

Winkelstein, W., M. Samuel, N. Padian, J. Wiley, W. Lang, R. Anderson, and J. Levy (1987b) "The San Francisco Men's Health Study: III. Reduction in Human Immunodeficiency Virus Transmission among Homosexual/Bisexual Men, 1982-86" *American Journal of Public Health* **76**:685-689.

	Pre-1982	Late 1983/Early 1984
Mean	18.4	12.0
Std. Dev.	34.3	38.9
Skewness	4.6	15.2
Percentiles:		
5	0.4	1
25	1.99	2.0
50	6.4	4.0
75	19.7	10
95	79.17	50
99	178.3	120

Observations: 716
(*Source*: Computed using San Fransisco Men's Health Survey, Wave 1)

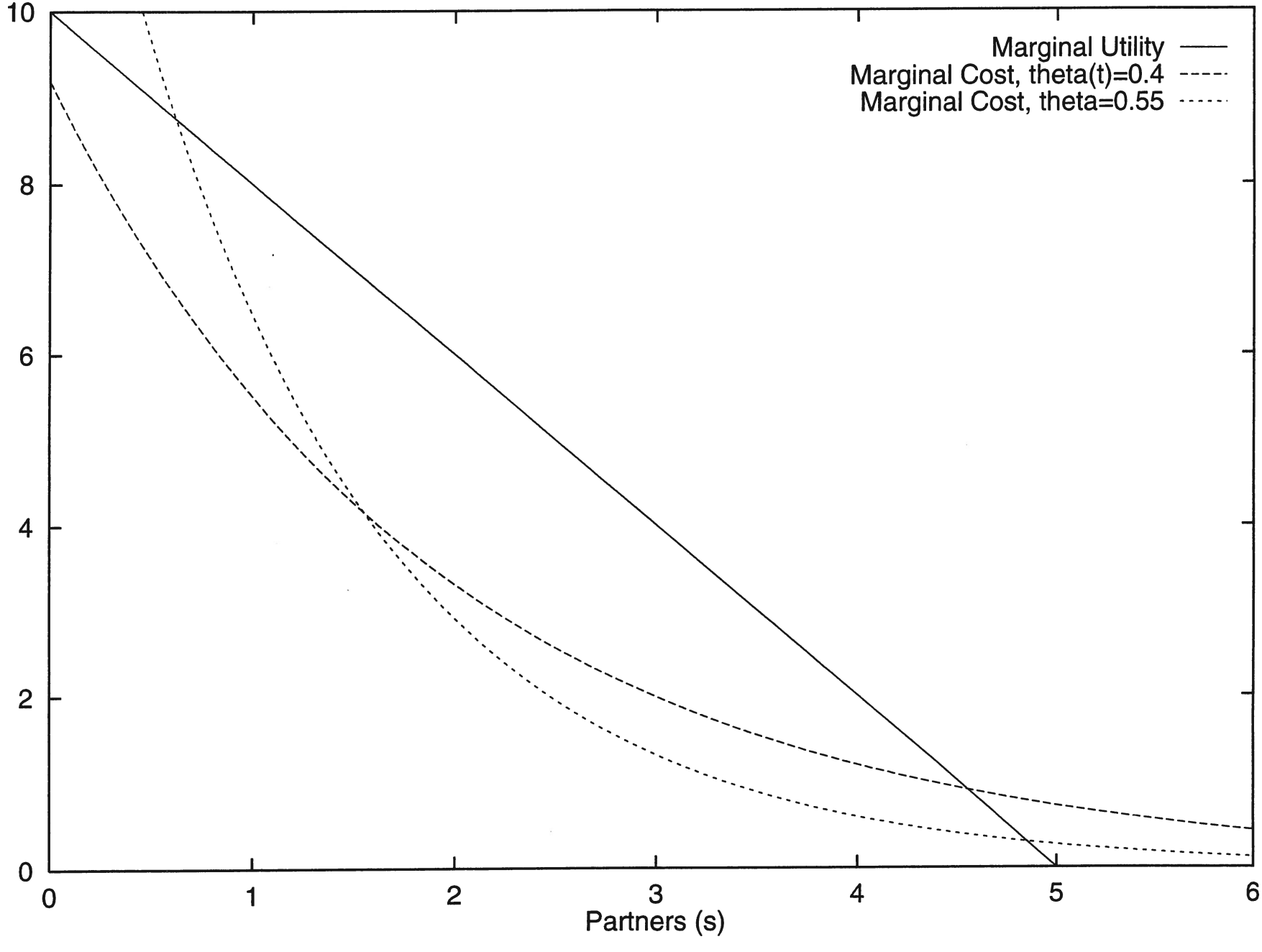
Table 1: Pre and Post AIDS Rates of Partner Change, SFMHS

Parameter	Description	Value
ϕ	transmission coefficient	0.125
γ	reduction in activity for infecteds	0.666
T	modeled lifespan, periods	30
N	number of agents at any time	600,000
B	utility bonus for exiting susceptible	250
D	per-period utility penalty for infection	-10
β	discount rate	0.975

Note: distribution of α calibrated using SFMHS

Table 2: Simulation Baseline Parameters

Figure 1: The Agent's Problem



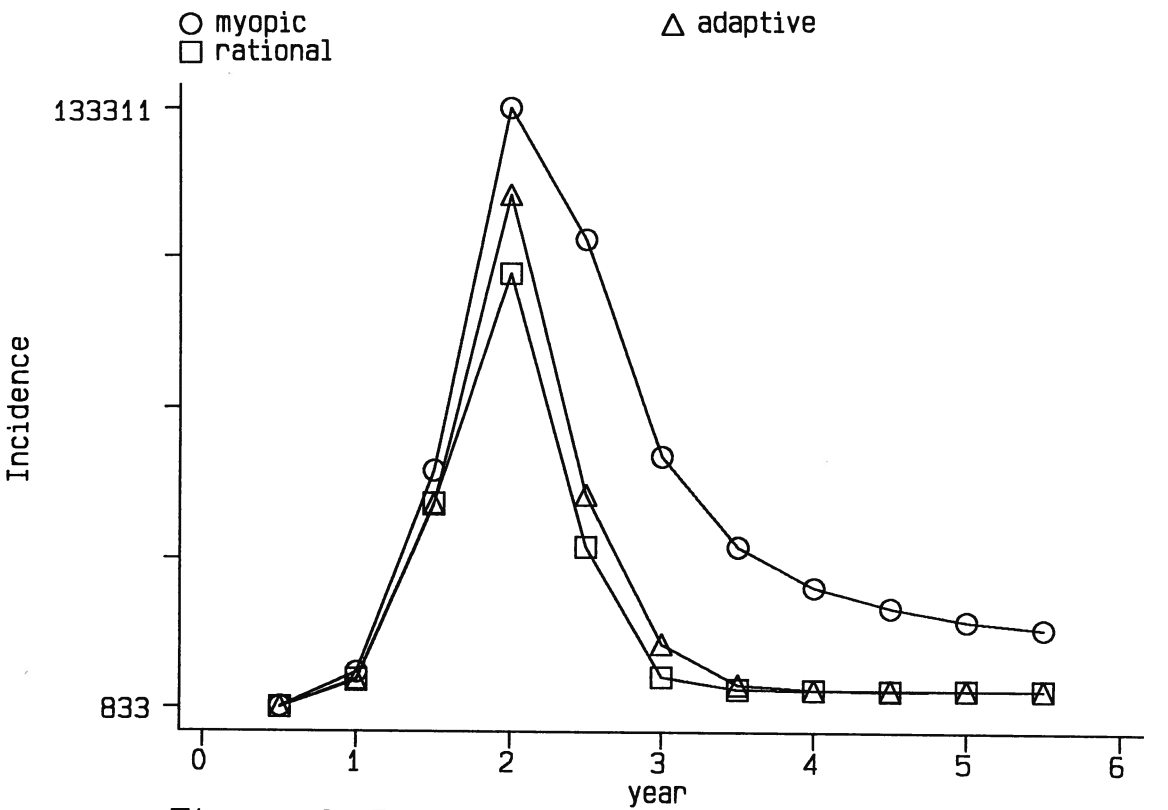


Figure 2: Incidence, Varying Expectations

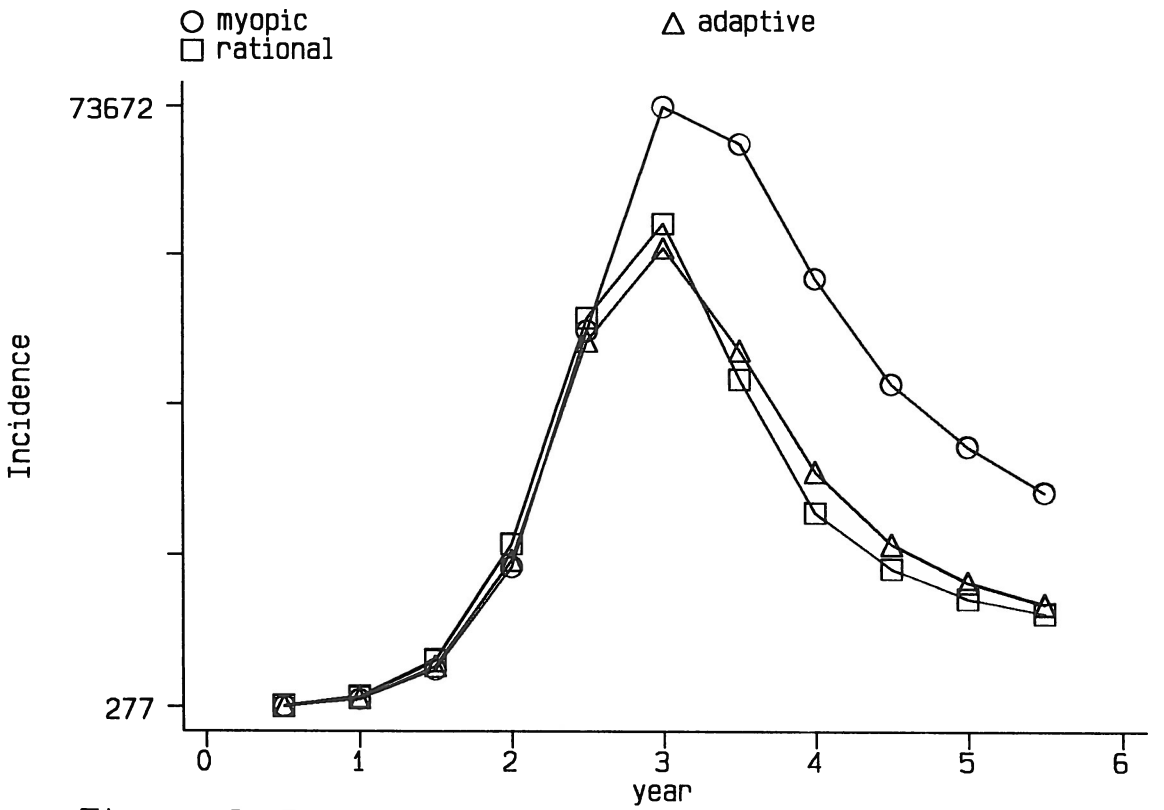


Figure 3: Incidence, Varying Expectations, phi=0.05

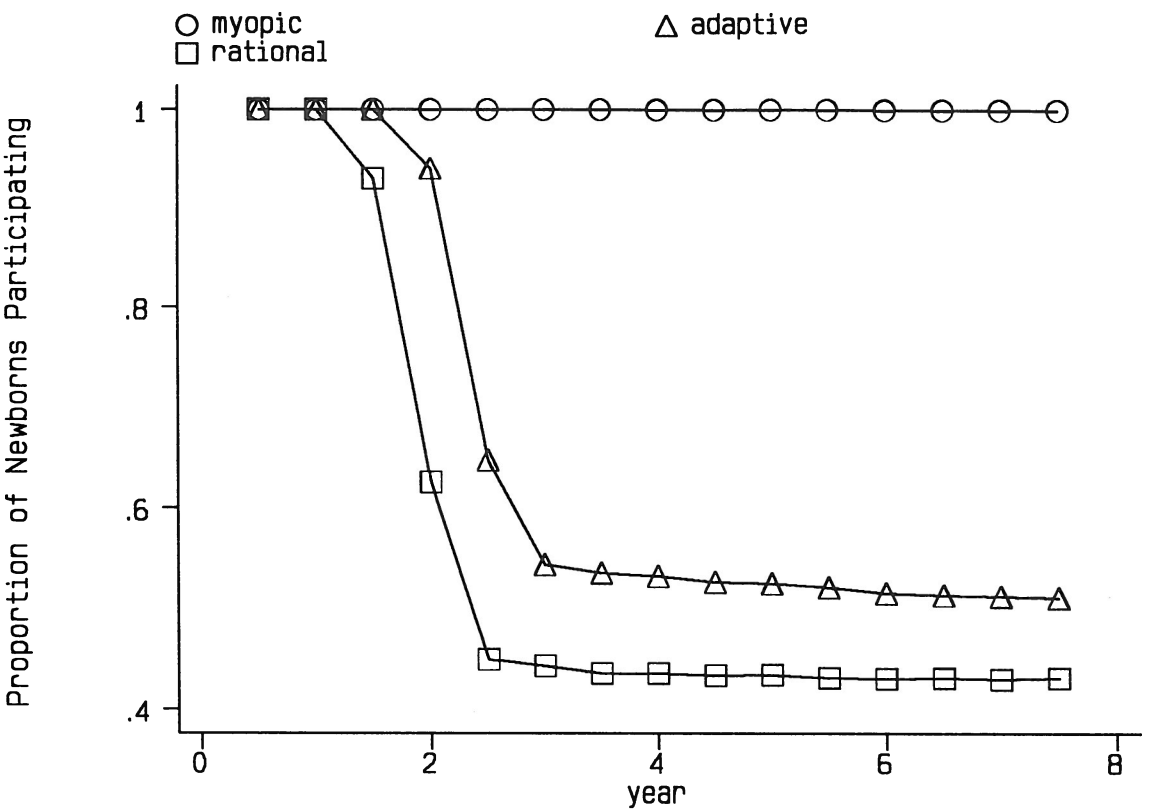


Figure 4: Participation

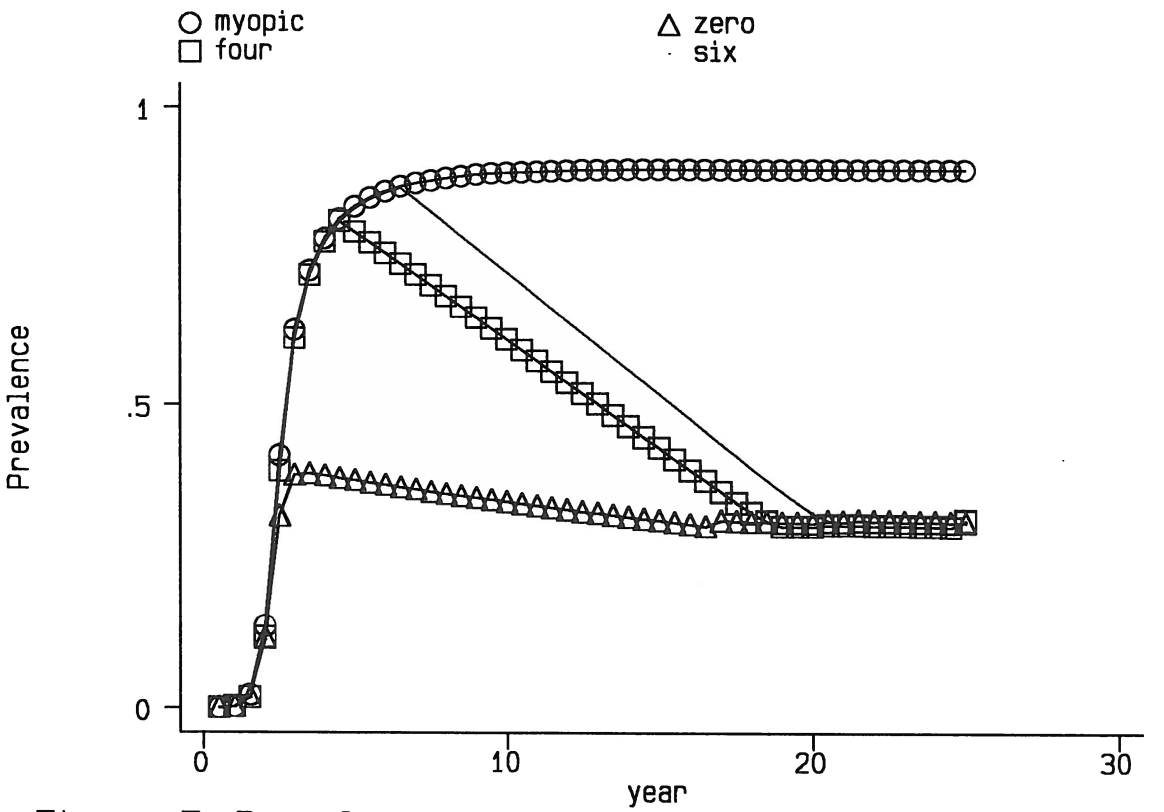


Figure 5: Prevalence vs Years to Awareness of Disease

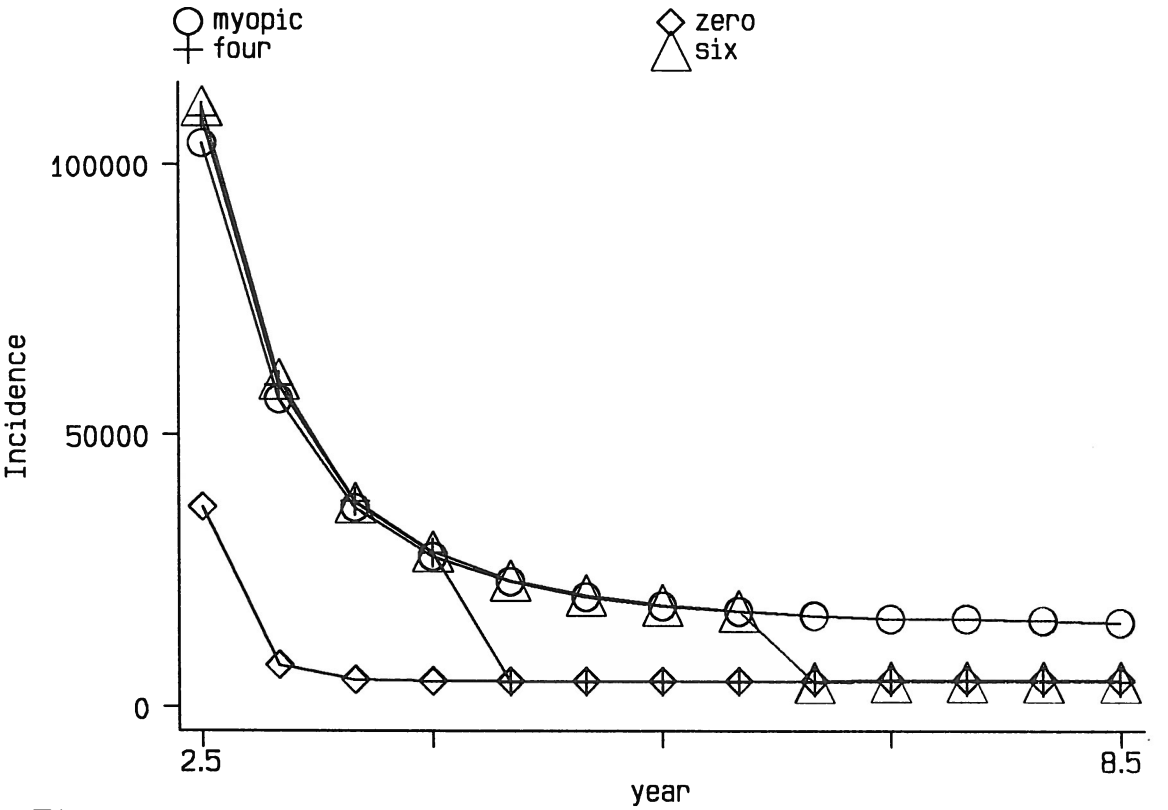


Figure 6: Incidence vs Years to Release of Information

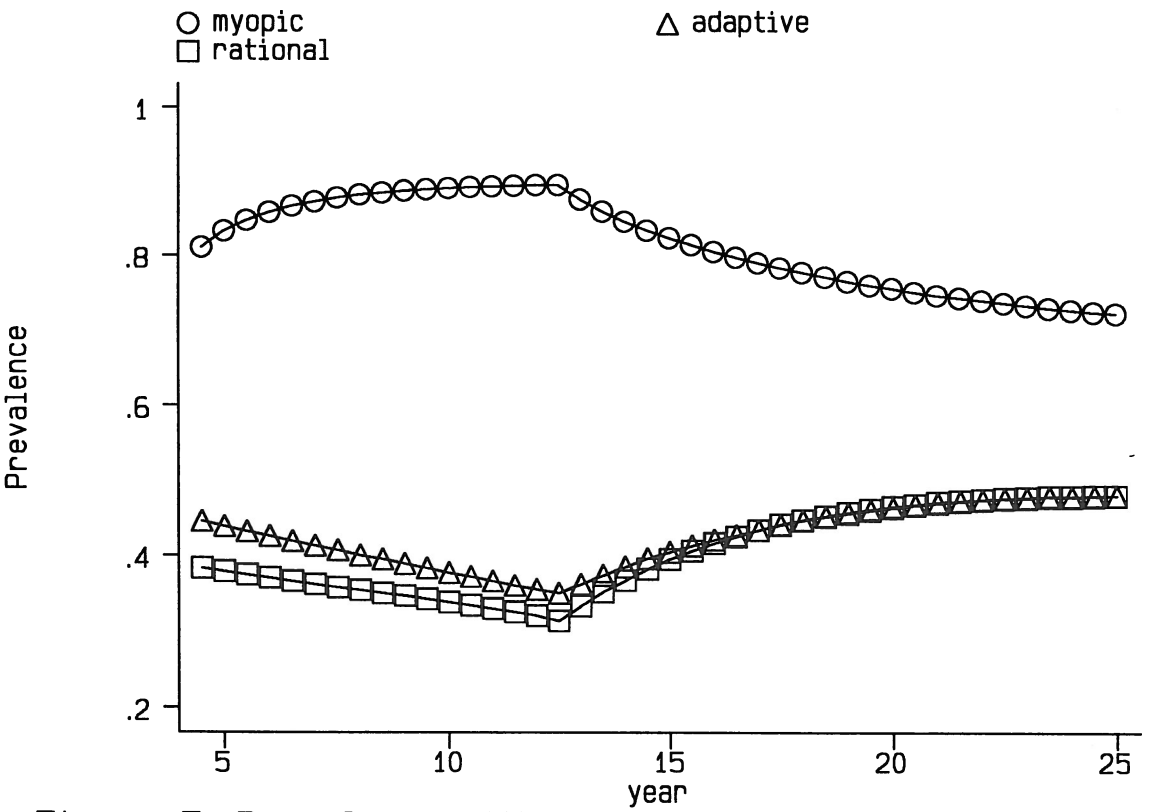


Figure 7: Prevalence with Vaccine Release at 12 Years

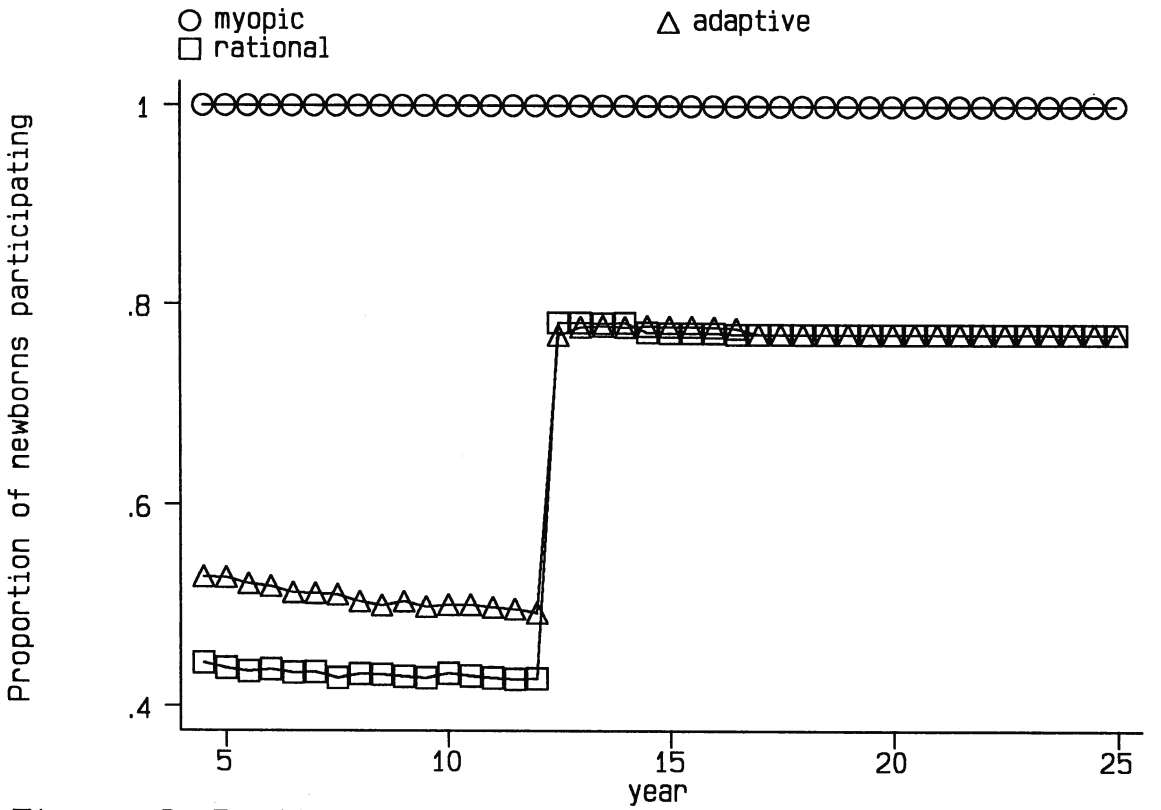


Figure 8: Participation with Vaccine Release at 12 Years