NBER WORKING PAPER SERIES

ECONOMIC EPIDEMIOLOGY AND INFECTIOUS DISEASES

Tomas Philipson

Working Paper 7037
http://www.nber.org/papers/w7037

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1999

I am thankful to participants at the Handbook Conference at the University of Chicago for comments, especially Charles Phelps and Peter Zweifel as well as seminar participants at Duke University and the University of North Carolina at Chapel Hill. Financial support is acknowledged from The National Science Foundation, The National Institutes of Health, and The Research Fellows Program of The Alfred P. Sloan Foundation. The views expressed in this paper are those of the authors and do not reflect those of the National Bureau of Economic Research.

Economic Epidemiology and Infectious Diseases
Tomas Philipson
NBER Working Paper No. 7037
March 1999
JEL No. I1

ABSTRACT

Infectious disease is currently the main cause of mortality in the world and has been even more important historically. This paper reviews recent research in economic epidemiology. Specifically, it discusses the occurrence of infectious diseases and the effects of public health interventions designed to control them. Several key points include: differences in the predictions regarding short- and long-run disease occurrence between rational and epidemiological epidemics, the nonstandard effects of interventions when epidemics are rational, the desirability and possibility of eradicating infectious diseases, as well as the components of the welfare loss induced by infectious diseases.

Tomas Philipson
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
t-philipson@uchicago.edu

1 Introduction

According to the 1997 WHO World Health Report, in 1996 there were about 52 million deaths world-wide. Infectious diseases caused about one-third of all deaths and represented the primary cause of mortality.[1] Historically, the share of world-wide mortality due to infectious diseases has been even greater, although data tends to be less reliable for earlier periods. Morbidity and mortality from infectious diseases such as tuberculosis, malaria, and acute respiratory infection have always been at the forefront of public policy in developing countries, where infectious diseases accounted for almost one half (45%) of mortality in 1996. Worldwide concern about infectious diseases has once again peaked with the onset of the most feared such disease of the twentieth century—the Human Immunodeficiency Virus (HIV) that causes Acquired Immune Deficiency Syndrome (AIDS).[2] Like most communicable diseases, especially those that are potentially fatal, HIV has incited an extensive governmental response, which has consisted of regulatory measures, subsidies for research, education, treatment, testing, and counseling. The broad scope of these and similar public interventions, and the private behavior they aim to change, makes the use of economic analysis in the study of their effects and desirability important.

Even though infectious diseases represent the primary cause of mortality worldwide, health economists are just beginning to understand their behavior and evaluate the many policies aimed at controlling their impacts. Indeed, currently, most research on the public control of infectious diseases is conducted outside economics in the field of epidemiology.[3] However, the evaluation of public health measures from an economic perspective is particularly important, since economic analysis separates the health effects of public policies from those of private decision making. In particular, it is often argued that the public, rather than private, control of infectious diseases is one of the main achievements of modern public health. Indeed, relying on standard arguments about the positive external effects of disease prevention, economists often echo such arguments for an active public role in the prevention of infectious diseases, such as AIDS.[4] On the other hand, economists have rarely attempted to explain patterns of disease occurrence or to evaluate public interventions in the context of a society with individuals who do the best they can given their constraints. Recent analysis in the rapidly growing field of economic epidemiology has cast considerably doubt

[1] The second leading cause at 28% was circulatory diseases, the third at 11% was cancers, and the fourth at 6% was respiratory diseases.

[2] The Global Programme on AIDS of WHO, as well as Mann et al (1992), summarizes the cumulative evidence on the prevalence of AIDS, and the mortality it has induced in the world. See also Bongaarts (1996). Bloom and Carliner (1988) discusses the financial impact of the epidemic in the US.

[3] An early economic treatment of public health issues may be found in Weisbrod (1961). The dominant form of epidemiological analysis is exemplified by the treatments and references contained in the works by Bailey (1975), Anderson and May (1991), Castillo- Chavez (1989), Brandeau and Kaplan (1993), and Geoffard and Philipson (1995).

[4] See for instance Stiglitz (1997, p. 15) who argues in a recent World Bank report that "an early and active government response encouraging safer behavior among those most likely to contract and spread the virus has the potential to avert suffering and save millions of lives."

on such textbook arguments by economists, on both theoretical and empirical grounds.

This chapter outlines the contributions made so far by economic epidemiology in explaining the occurrence of infectious diseases and helping to understand the effects of public health interventions. The discussion will focus on three general questions posed by this analysis:

1. How do economic and biological epidemiology differ in their predictions about the short- and long-run behavior of infectious diseases?

2. How do they differ in their predictions concerning the effects of public health interventions?

3. How do they differ in determining the welfare loss of a disease, and thus in the priorities for eradication and control which should be assigned to different diseases?

Section 2 begins by addressing the first question and considers the behavior of rational epidemics. It stresses the central interaction between the extent of disease, which is decreased by the demand for prevention, and the demand for prevention itself, which is increased by the extent of disease. At the heart of the analysis lies that an increase in the prevalence of infectious disease, *i.e.* the share of the population infected, induces growth in private prevention. Although epidemiological analysis surely discusses how various patterns of behavior affect disease occurrence, it does not analyze the implications of how behavior *change* in response to the new incentives created by the growth of a disease nor does it analyze the effects these changes have on the desirability of public health measures. Central to the study of rational epidemics is thus the *prevalence-elasticity* of private demand for prevention against disease. It represents the degree to which prevention rises in response to disease outbreak. The means by which preventive measures rise may differ across diseases. For example, the elasticity for vaccine-preventable diseases may represent the number of additional vaccinations induced by each new infection, while that for sexually transmitted diseases may represent the increases in the matching of sex partners who have the same infection status.

This type of prevalence elastic behavior has two major implications: first, growth of infectious disease is self-limiting because it induces preventive behavior; second, since the decline of a disease discourages prevention, initially successful public health efforts actually make it progressively harder to eradicate infectious diseases. We discuss a very general result concerning the inability of private markets to eradicate disease when demand is prevalence-elastic. This robust result does not depend on the market structure under which vaccines are produced or on how expectations are formed about future levels of prevalence. The result stems from the existence of barriers to disease eradication on both the demand side and the supply side. On the demand side, as the disease disappears, so too does the demand for vaccines; the subsequent decline in vaccinations allows the disease to return. On the supply side, a patent-protected

producer of vaccines has a special dynamic incentive to increase mark-ups: if the vaccine eradicates the disease, the demand for the monopolist's product is eradicated as well. Put simply, if there were fortunes to be made in disease eradication, we would have more of it.

By considering the private incentives for preventive behavior, the economic approach provides different predictions than that of epidemiology for the behavior of epidemics. In particular, the two approaches predict a different relationship between the hazard rate into infection and the prevalence of a disease. In epidemiological analysis, this hazard rate is an *increasing* function of prevalence. In other words, the larger the fraction of infected people in the population, the larger is the fraction of uninfected people who become infected in the next period. A higher prevalence increases the chance that a susceptible individual will meet an infected individual. Economic incentives imply that the hazard rate into infection may be a *decreasing* function of the prevalence of the disease, when private demand for prevention is prevalence elastic. As the stock of infected individuals grows, uninfected individuals face a larger risk of infection and hence raise their demand for prevention. The sharp reductions in infection hazards which often accompany outbreaks seem to provide empirical support for economic, rather than epidemiological, epidemics.

Section 3 addresses the second question by analyzing the impact of public health policies in general and their ability to eradicate infectious diseases in particular. From the perspective of the current population only, eradication is never Pareto optimal, because the benefits of lowering an almost extinct disease sooner or later fall below the cost of vaccinating more people. However, the missing market is dynamic: future generations cannot pay vaccine producers for the benefit they derive from the producers' product. Public interventions like price subsidies and mandatory vaccination may complete this missing market, but such measures have often failed to eradicate infectious diseases. Our analysis explains why eradication through such policies remains extremely difficult. Price subsidies alone will not bring about eradication for the same reasons that price reductions through increased competition will not. Both price subsidies and mandatory vaccination programs are limited in their ability to achieve eradication, because higher vaccination rates for individuals covered by any public program lower the incentive of those *outside* the program to become vaccinated. The prevalence-elasticity of demand lowers the *price*-elasticity of demand in the case of subsidies and thus lowers the total demand effect of increased mandatory coverage. As demand rises for those who are subsidized, demand falls among those who are not. In the extreme case total demand is inelastic to subsidies. This implies that classic economic justifications of Pigovian subsidies aimed at solving the private under-provision of vaccines may be highly ineffective due to these dynamic effects. In addition, the prevalence-elasticity of demand does not only affect the long-run results of such interventions but also their short-run results and thereby the Pareto optimal *timing* of public subsidy programs. Prevalence *competes* with public interventions in inducing protective activity, and this makes the timing of the public intervention a crucial factor in determining its economic efficiency. If the subsidy is not fast enough, the growth

in prevalence has already induced protection; the public sector thus ends up paying for behavior that would have been undertaken by the private sector of its own accord.

Section 4 discusses the third question concerning the welfare loss induced by a disease and the welfare effects of R&D to develop new methods of prevention or treatment. The implications for subsidization of R&D are quite different from those involved in subsidizing prevention as discussed above. In setting priorities among control efforts for many separate types of diseases, a major question facing public health authorities is the welfare loss inflicted upon a population by a given disease. The orthodox approach toward assessing disease burden has employed several cost-of-illness ("COI") measures, each of which is a product of prevalence and (possibly quality-adjusted) per-case severity of a disease. This approach has the seemingly self-evident implication that the more morbidity or mortality inflicted by a disease, the larger its welfare loss. In contrast, we argue that the more prevalence-elastic is the demand for prevention, the more this measure understates the total welfare loss. We interpret a disease as a random 'tax' on behavior which risks exposure, a tax which will distort individuals' consumption of risky behavior by inducing them to forego that otherwise valuable activity. Standard tax analysis argues that a tax imposes a burden *in excess* of the revenues collected by the public treasury if costly tax avoidance occurs. Similarly, if costly disease-avoidance occurs, a randomly collected disease tax on exposure imposes a burden beyond the case reports of disease incidence collected by the public health authority. However, cost-of-illness measures of the disease-induced loss, and indeed the measures used by public health authorities,such as The World health Organization (WHO) or The Center for Disease Control and Prevention (CDC), are implicitly "revenue-focused" in that they consider only the losses from morbidity and mortality and ignore the excess burden of disease prevention. The major point we make is that the standard cost-of-illness measures do not constitute a relatively large fraction of the total welfare loss when prevention is prevalence elastic. This is for the same reason that tax-revenue does not make up the major loss when tax-avoidance is elastic. For example, almost all loss inflicted by vaccine preventable diseases is from the excess burden. The case is similar for AIDS where the excess burden consists of the quantity of sexual consumption foregone from fear of infection. Many economists have argued that research expenditures on AIDS are excessive given its relatively small case load. However, few diseases have caused as much behavioral change as AIDS in terms of foregone sex which, if we were to believe biologists, is perhaps the most valued human activity. Large research expenditures to eliminate low-prevalence but behavioral diseases such as AIDS may be justified because their total welfare loss, the case-load revenue and the excess burden, is larger than for more common diseases. However, partial success in R&D may lead to cut-backs in prevention behavior that thereby offsets such medical advances.

Section 5 discusses existing empirical analysis of the type of prevalence-elastic behavior that underlies the theoretical analysis reviewed. We consider evidence for influenza, AIDS, and measles and attempt to estimate the prevalence-

elasticity of demand for prevention. We review the results of a particular US study that tracked the AIDS-preventive behavior of young individuals during the 1980s, which saw rapid growth in the disease, using panel data from the US National Longitudinal Survey of Youth (NLSY). The prevalence elastic behavior documented by this and other studies suggests the need to incorporate such responses into the formulation of public health policy aimed at limiting the occurrence of disease. Finally, section 6 concludes by outlining several directions for future research.

It should be noted that this chapter discusses a set of questions subjectively selected from the recent health economic research literature in this area. Naturally, a single review chapter cannot claim to cover everything that has been and is being done in a rapidly expanding area of research, and this chapter is no exception.[5] This is a deliberate choice, since we find it more useful to focus on a few ideas central to a research agenda, rather than provide a disjointed discussion of an exhaustive reference list.

2 Predictions about Disease Occurrence

This chapter first discusses the implications incentives have for the behavior of both short-run epidemics and the long-run occurrence of disease. Economic and biological epidemiology make different predictions about disease occurrence mainly due to their different predictions about the relationship between prevention and prevalence.

2.1 The Behavior of Rational Epidemics

Consider individuals who are classified into four health categories at a given time t: susceptible S_t, infected I_t, immune through recovery R_t, and outside the system. Normalizing the total population to unity, we refer to the fraction infected in the population, I_t, as the *prevalence* of the disease. A future path of prevalence is denoted $\underline{I}^t \equiv \{I_s;\ s \geq t\}$ and a future path of prices is denoted $\underline{p}^t \equiv \{p_s;\ s \geq t\}$, with the instantaneous demand for vaccines at time t for two such paths denoted $D(\underline{I}^t, \underline{p}^t)$. We denote by b and m the birth and mortality rates into and out of the system, respectively, by w the rate at which infected individuals are withdrawn naturally from infection into immunity, and by β the probability of transmission conditional on exposure to an infected person. The changes over time in the health of the population are determined by:

$$\left\{ \begin{array}{lcllcl} dS_t/dt & = & b[1 - D(\underline{I}^t, \underline{p}^t)] & - & \beta S_t I_t & - \quad mS_t \\ dI_t/dt & = & \beta S_t I_t & - & (w+m) I_t & \\ dR_t/dt & = & bD(\underline{I}^t, \underline{p}^t) & + & wI_t & - \quad mR_t. \end{array} \right\} \qquad (2\text{-}1)$$

[5] For example, we do not discuss the important literature on economic growth and AIDS: see for instance Cuddington (1993a-b) and Bloom (1997). Another area omitted is the statistical literature by economists on AIDS forecasting (see e.g. Hay and Wolak (1990,1994)). For alternative reviews on markets for vaccines, see Weisbrod and Huston (1987) and Pauly (1994).

The change in the fraction of susceptible individuals is due to the entry of newborn individuals who do not vaccinate. Exits are due to new infections and non-disease-related mortality. The change in the prevalence is due to the entry of new infections, while exits are due to immunity and infection-related mortality. New infections are caused by contact between susceptible and infected individuals under random matching, as in the term $\beta S_t I_t$. The change in the fraction of recovered immune individuals is due to the entry of newborn individuals who vaccinate, as well as those individuals recovering from infection, and the exits of agents through non-disease-related mortality.

It follows directly that the prevalence rises over time whenever

$$dI_t/dt \geq 0 \Leftrightarrow \frac{1}{w+m}\beta S_t \geq 1 \tag{2-2}$$

The factor βS_t is the rate at which infected individuals infect susceptible individuals with whom they come into contact, and the factor $1/(w+m)$ is the average time of infection. For the infected stock to grow, the average number of secondary infections by an infected individual must be above unity, so that an infected individual more than replaces himself by the time he exits the infected population. When there are only susceptible individuals, which is the relevant case when a disease is to be eradicated, the secondary infections generated by a single new infection is denoted $\rho \equiv \beta/(w+m)$, so that the disease can take off in a completely susceptible population only when this rate is above unity.

2.2 Implications for Private Disease Eradication

A major technology aimed at limiting such diseases has been vaccines. Although the introduction of a vaccine usually produces a sharp drop in the occurrence of a disease, the eradication of vaccine preventable diseases predicted by many at the time of these inventions has not been achieved except for smallpox.[6] Of the roughly forty vaccines on the market, only the smallpox vaccine has eradicated its target disease. Diseases such as measles, tuberculosis, and different types of influenza persist, despite explicit governmental efforts to eradicate them, and recent attempts to develop a vaccine against HIV or AIDS raise important questions about the causes behind these difficulties.

The prevalence-elasticity in private markets, coupled with rational demand for vaccines, represent powerful forces which make it difficult for private markets to achieve eradication. We call the demand for vaccines *prevalence dependent* if, when prices are positive in the future, demand vanishes for low enough prevalence. That is, for any strictly positive price path $\underline{p}^t$, there is a prevalence path $\underline{I}^t(\underline{p}^t)$ below which demand vanishes: $D(\underline{I}^t, \underline{p}^t) = 0$ for all $\underline{I}^t \leq \underline{I}^t(\underline{p}^t)$. It can be shown that prevalence dependent demand requires the simple condition that the benefits of vaccination not be large "enough" when prevalence levels are low enough. If demand is prevalence-dependent and if the prevalence goes to zero for

[6] See, for instance, Stephen Plotkin and Edward Mortimer (1988), and The World Bank (1993).

any future prices, there must be a time t_0 after which the prevalence is driven down to a level which generates small enough demand. As fewer individuals vaccinate after t_0, the population becomes increasingly susceptible. However, when an infection can regenerate itself in a susceptible population, which occurs when $\rho \geq 1$, this implies that the prevalence increases again, making eradication infeasible. In other words, the disease cannot be eradicated under positive prices when $\rho \geq 1$.

Since this argument holds for any prices, it implies that, regardless of the market structure in which vaccines are produced, the disease is not eradicated, since prices are presumably above costs in the long run for any feasible market structure. In particular, although a vaccine monopolist is faced with a problem similar to that of using an exhaustible resource, the resource (prevalence) will never be exhausted. Naturally, if competition drives prices down to minimum average costs of production, then eradication is not achieved under this market structure either. Interestingly, this argument is not only robust to the type of market structure, but also to many forms of expectations: it is true under myopic as well as rationally formed expectations. The general difficulty with eradication thus comes from the demand-side of the vaccine market, rather than from the supply side.[7]

2.3 Rational Disease Dynamics of Epidemics

So far, we have not specifically investigated the demand for prevention and have merely outlined the implications of various properties of this demand. Geoffard and Philipson (1996) discuss an environment in which the rational protection behavior of an individual in an epidemic can be traced out simply. A version of this model is also discussed in Auld (1997). Given a utility function $u(h, d)$ over a binary demand for protection ($d = 1$) and the state variable h representing the susceptible (s) or infected (i) health state. Proceeding heuristically to illustrate the main ideas, the value function evaluated in the susceptible state may be written as

$$V(s) = max\{u(s,1) + \alpha V(s), \quad u(s,0) + \alpha[\beta I_t V(i) + (1 - \beta I_t)V(s)]\} \quad (2\text{-}3)$$

where α is the discount rate. This says that continued protection today implies susceptibility tomorrow, but if the individual does not protect, he risks becoming infected with a probability which increases in prevalence. This directly implies that the individual remains exposed as long as the current benefit of exposed activity outweighs the expected loss in the future due to risk of infection:

$$d = 0 \Leftrightarrow u(s,0) - u(s,1) \geq \alpha\beta I_t[V(s) - V(i)] \quad (2\text{-}4)$$

Protection in an epidemic is then characterized by a simple rule under which the individual engages in protection only after a reservation prevalence, denoted

[7]For a contrary view see, for instance, General Accounting Office (1994) for a discussion of why monopoly production of vaccines makes eradication infeasible.

by K, has been reached. Solving for the value function, this reservation prevalence can be shown to satisfy

$$d = 0 \Leftrightarrow I_t \leq K \equiv \frac{\alpha}{\beta} \cdot \frac{[u(s,1) - u(s,0)]}{[u(s,1) - u(i,0)]}. \tag{2-5}$$

In other words, there is a threshold prevalence below which an agent engages in transmissive behavior and above which he engages in protection. Intuitively, this reservation prevalence rises with the instantaneous cost of protection and the discount rate, and falls with the cost of infection and the probability of transmission conditional on exposure.

The dynamics for a closed population, which no agents enter or exit, can then be traced out by assuming that the reservation prevalence levels are distributed according to the cumulative distribution function $F(K)$. The epidemic starts with a prevalence level denoted $I(K,0)$ in the group with reservation level K and then takes off by infecting in a given period those who have not started to protect yet and have not been infected in the past

$$\begin{cases} I_0 & \equiv & \int I(K,0)dF \\ \dot{I}_t & = & \beta I_t Q_t G(I_t) \end{cases} \tag{2-6}$$

where $G(I_t) \equiv \int_{I_t \leq K}(1 - I(K,0))dF$ is the susceptible population choosing to engage in transmissive activity at prevalence I_t and $Q_t \equiv e^{-\int_0^t \beta I_s ds}$ represents the share not infected in the past even though they were exposed. This function G summarizes the behavioral response of the population to the growth in prevalence. For an epidemiological model in which behavior is exogenous or prevalence inelastic, the function would be constant, that is, an increase in prevalence would not cause any change in protective behavior. The degree to which the population's protective behavior responds to an increasing prevalence of a disease determines the epidemic behavior of the disease. More precisely, Geoffard and Philipson (1996) outline the conditions under which the growth of the epidemic reveals the preferences of the population, in the sense that the risk attitudes of the population implicit in F can be identified through the time path $\{I_t; t \geq 0\}$ of prevalence.

This characterization of behavioral response demonstrates how economic models generate implications observably different from those generated by epidemiological models. The key implication concerns the hazard rate into infection from susceptibility and its relationship to the prevalence of a disease. The hazard rate measures the propensity to be infected conditional on not being infected yet. With the inelastic behavior assumed by epidemiological analysis, this hazard rate is an *increasing* function of prevalence. In other words, the larger the fraction of infected people in the population, the larger is the fraction of uninfected people who become infected in the next period. As prevalence rises, so does the chance that a susceptible individual meets an infected individual. This is true across a wide variety of epidemiological models, since they all share the feature that the demand for exposure does not respond to prevalence.

More precisely, if the demand for exposure is prevalence inelastic ($G = 1$), then the hazard function $h(I_t)$ satisfies:

$$h(I_t) = \frac{-\dot{I}_t}{1 - I_t} = \frac{\beta I_t Q_t}{Q_t} = \beta I_t. \tag{2-7}$$

Therefore, the prevalence has a positive effect on the hazard rate of infection $\partial h/\partial I \geq 0$.

In contrast, the incentives of epidemics imply that the hazard rate into infection may be a *decreasing* function of the prevalence of the disease, because an increase in risk provides a larger incentive for susceptible individuals to increase protective behavior. In other words, although more and more people may be infected as the epidemic grows, the share of still uninfected that become infected falls. The hazard as a function of prevalence is now:

$$h(I_t) = \frac{-\dot{I}_t}{1 - I_t} = \frac{\beta I_t Q_t G(I_t)}{1 - I_t} \tag{2-8}$$

It follows that the relationship of prevalence on the hazard rate is now given by

$$\frac{I}{h}\frac{\partial h}{\partial I} = \frac{1}{1 - I} + \frac{G'}{G} \cdot I = \frac{1}{1 - I} - \eta, \tag{2-9}$$

where η is the absolute value of the elasticity of G. It follows that

$$\frac{\partial h}{\partial I} \leq 0 \Leftrightarrow \eta \geq \frac{1}{1 - I}. \tag{2-10}$$

If the elasticity of the prevalence response is large enough, the hazard decreases as a function of the prevalence. In particular, the elasticity needs to be larger the larger is the prevalence, since the counteracting positive effect is more pronounced. The more the behavior responds to the new risk, the lower is the hazard rate into infection among susceptible people who have not already been infected.

Figure 1 illustrates empirically the way hazard rates often fall and eventually level off as a result of a rising prevalence (see Geoffard and Philipson (1996) and Auld (1997)). The curves are estimated using data from the San Francisco Men's Health Study (SFMHS).[8] The SFMHS, a longitudinal epidemiological survey study, consists of 15 half-year waves of data for individuals living in San Francisco during the period 1983-92, and is designed to yield data on the natural history of AIDS. The respondents were interviewed to obtain information about behaviors, attitudes, and beliefs relevant to AIDS. The sample population consisted of English-speaking, unmarried men aged 25-54. The survey sample was a stratified two-stage sample of all households within the designated census

[8]See Geoffard and Philipson (1995) for details on the estimation and "Sampling Methods and Wave 1 Field Results of the San Francisco Men's Health Study," Survey Research Center Technical Report, University of California, Berkeley for the details on the sample design.

tracts. All eligible persons in each selected housing unit were included in the sample.

In the SFMHS, the duration of susceptibility is defined here as the number of waves for which the individual remains HIV-negative, that is, the wave at which he becomes infected. Such durations were observed under both left and right censoring. In other words, HIV-positive individuals entered the first wave and HIV-negative individuals were present at the end of the last wave.

FIGURE 1 INSERTED HERE

Figure 1 displays the prevalence rate and the hazard rate as functions of the waves constructed from these data. The horizontal axis indicates the waves at which the rates were measured. The estimated prevalence rates, I_t for wave t, are plotted on the increasing (dotted) line and depict the percentage of individuals in the survey who were HIV-positive at each wave. These prevalence rates were estimated by computing, for each wave, the fraction of individuals present throughout sampling who were HIV-positive in the sample. The hazard rates are plotted on the generally decreasing (solid) line and depict, for each wave t, the fraction of individuals present throughout sampling who were HIV-negative in wave $t-1$, but turned HIV-positive during wave t.[9] Due to the attrition of subjects, the observed prevalence rates are not equivalent to the observed hazard rates. The figure displays a negative relationship between the estimated hazard rates and prevalence levels: the estimated hazard function is decreasing and the cohort prevalence is increasing. The displayed relationship thus provides suggestive evidence for prevalence-elastic rational epidemics.[10] Moreover, the eventual flattening of the hazard rate around the 10th wave of the survey corresponds to a complete reduction of new infections. This common pattern of a sharp reduction in hazard rates into infection as epidemics grow is suggestive of the type of prevalence elastic behavior discussed.

2.4 The Positive Effect of Prevalence on Assortative Matching

In the case of sexually transmitted diseases (STDs) such as AIDS, one form of protective behavior may be the choice of safer partners. In the extreme case of perfect "assortative matching" on infection status, *i.e.*, when all HIV-positives match with other HIV-positives and HIV-negatives with other HIV-negatives, the growth of the disease is zero. Therefore, the degree to which incentives lead to growth in infection-dependent matching of partners determines the growth of a sexually transmitted epidemic.

Consider an environment which contains different classes of individuals whose risk of HIV infection is known, and where risk is defined as the percentage of the

[9] By focusing on the propensity of HIV-negatives to become infected, the hazard rate is different from the so-called *incidence rate*, which is the flow of new cases divided by the *whole* population size.

[10] However, it is well known that such negatively sloped hazard functions may also be due to unobserved differences among individuals (see, e.g., Heckman and Singer (1984)).

class infected by the AIDS virus. Such classes may be subpopulations stratified on gender, appearance, weight, race, or other observable demographic characteristics. Given this environment, determining who will engage in sexual activity with whom, and what type of activity they will engage in determines the growth of the disease. Specifically, it determines the demand for unprotected sex by pairs of individuals, each of whom have a different infection status. The matching of infected individuals with uninfected individuals is necessary to generate new cases of the disease. The general theory of matching markets predicts that among traders of different quality levels, high-quality traders will match among themselves and low-quality traders will match among themselves (Becker (1991)). In other words, low-risk individuals match up with other low-risk individuals, and high-risk individuals with other high-risk individuals. The complementarity in health status that generates this assortative matching stems from the fact that low risk individuals have more to gain by the choice of low-risk partners than do high risk individuals. This has the important implication that disease growth is slower than in the random matching case considered by epidemiological analysis. Since the disease grows due to sexual partnerships between negative and positive individuals, the economic matching incentives slow disease growth. Such matching may be interpreted to change the matching patterns over the disease, inducing more dependence across the statuses of partners as the disease grows.

For the case of HIV, Dow and Philipson (1996) estimated the extent both of such assortative matching, as well as the extent to which such matching reduces HIV incidence relative to the random matching assumed by epidemiological models. They estimate that on average HIV-positive individuals are more than twice as likely to have HIV-positive partners, and that this assortative matching reduced HIV incidence by about one-third. They use the San Francisco Home Health Study (SFHHS),[11] which is detailed enough to allow estimation of the joint infection status of a pair of partners. The fraction of matches with a given infection status at a point in time may be represented by a $2x2$ table T of the form:

$$T \equiv \begin{array}{|c|c|} \hline p_{00} & p_{01} \\ \hline p_{10} & p_{11} \\ \hline \end{array} \tag{2-11}$$

p_{ij} is the fraction of couples of infection status (i, j), where the respondent's infection status is given by i, while the respondent's partner has status j. If i or j is 0, the individual is uninfected. The *prevalence* of the disease among respondents, p, or partners, q, is then defined as the marginal probabilities of the table T: $p \equiv p_{10}+p_{11}$ and $q \equiv p_{01}+p_{11}$. The degree to which there is assortative matching on infection status is measured by the degree of positive dependence in

[11] The SFHHS is an epidemiological study designed to yield data on the prevalence of HIV and related risk factors in multicultural neighborhoods, including information about behavior, attitudes, and beliefs relevant to HIV. The sampled population included persons currently unmarried, aged 20-44, and residing in San Francisco census tracts with substantial proportions of blacks and Hispanics.

the table. It is measured by the ratio δ of the conditional probabilities of being with a positive partner for positive respondents versus negative respondents,

$$\delta \equiv \frac{p_{11}}{p} / \frac{p_{01}}{(1-p)}. \tag{2-12}$$

In other words, the dependence is measured by the fraction of positive partners of positive respondents, relative to the fraction of positive partners of negative respondents. Such matching is important for infection incidence since new cases are generated only by contact between individuals of different infection statuses. More precisely, the incidence in the table T (denoted $\zeta(T)$) is given by

$$\zeta(T) \equiv \beta[p_{01} + p_{10}], \tag{2-13}$$

where β, as before, is the probability of transmission and represents the likelihood of an infection conditional on a match between an infected and a susceptible individual.

For example, when there is no assortative matching as in epidemiological analysis, δ=1. The incidence is then given by $\zeta(p, q, \delta = 1) = \beta[p(1-q) + q(1-p)]$. The matching independence assumption is inconsistent with the assortative matching generated by the incentive of a non-infected individual to avoid infection. These incentives imply that $\delta \geq 1$. As is well known, assortative matching is not an assumption; rather it is *implied* by the fact that, *ceteris paribus*, everyone prefers their partner to be negative rather than positive. Moreover, the greater the morbidity or mortality of a disease, the stronger the incentive of negatives to search for and match with other negatives.

Figure 2 illustrates the differences in the incidence for different levels of assortative matching δ consistent with a given prevalence level for the simplest possible case, in which the prevalence levels of the respondents and partners coincide: $p = q$.

FIGURE 2 INSERTED HERE

The top triangle depicts the disease incidence under the minimum assortative matching consistent with the prevalence, the lower line depicts the incidence under maximum assortative matching, and the oval shaped line depicts incidence under infection-independent matching. In other words, the top triangle represents the largest number of matches between individuals of different infection status consistent with a given prevalence level. The bottom x-axis represents the lowest number, zero, since all negatives of one group can match with negatives of the second group when the two groups have the same prevalence levels. In the intermediate case, matches are infection-independent. For the SFHHS of gays in San Francisco, the HIV-prevalence in the sample was roughly one half. Since this rate is high, the figure implies that incidence predictions will be very sensitive to different levels of assortative matching. Dow and Philipson (1996)

estimated that an HIV-positive individual is more than twice as likely as an HIV-negative individual to have a HIV-positive partner;$\hat{\delta} > 2$. The incidence reduction implied by such matching varies between 40 and 25 percent so that in the figure the estimated incidence line is about a third of the way from the independence line.[12]

2.5 The Effect of Immunity on the Prevalence of a Disease

With many infectious diseases, an individual can only be infected once and is thereafter immune to future infections. In the case of such an immunity-bearing infection, an infection may be a good which one is willing to pay for, rather than a "bad" which one has to be paid to accept. This occurs if the infection is more severe when caught in adulthood, so that a 'front-loaded' early infection may be optimal. Many childhood infections have this feature, and indeed the "get-it-over-with" attitude of parents reflects the value of early infection. Absent manufactured artificial vaccinations, contracting and surviving a disease represents the only means of "vaccinating" one's self, although at a higher cost. Clearly, if there is no immunity conferred, front-loaded infections fail to be optimal. Moreover, if the conditional risk of mortality remains constant over an individual's lifetime, there is no incentive to receive immunity through infection. However, consider such immunity-bearing infectious diseases as chicken pox, measles, rubella, and mumps–often called child diseases, because of the low average age of infection. Before vaccination was available, immunity represented the only source of protection. Cohort studies of immunity-bearing diseases conducted before vaccine development invariably showed that the fraction ever-infected or currently infected in a given age cohort rises so rapidly that, as the cohort reaches its late teenage years, around 95% of its members have been infected. Nearly all of the remaining 5% share of the cohort escapes infection for the remainder of their lives, so that the growth rate of cohort prevalence goes to zero after the teenage years. This pattern held true for all low-cost child diseases such as mumps, rubella, and chicken pox. For more severe diseases, however, such as polio and diphtheria, cohort prevalence never escalates to nearly universal seropositivity. Since infection confers immunity at the cost of experiencing the disease, it may be interpreted as a purchase of immunity similar to the purchase of medically manufactured vaccines. For low-risk childhood diseases, it appears that the demand for such immunity starts out as high for young children, but goes to zero after the late teenage years. This property is consistent with the following two factors: first, as expected future lifetime falls, the benefits of lifetime immunity fall as well, so benefits are highest for young children; second, the price of immunity (the morbidity cost of infection) increases in age. For example, the risk of dying from childhood

[12]Observed assortative matching may be infection-induced, in that infection may transform a pair of partners with initially different infection status into a pair of partners with the same status. The paper discusses how to isolate this infection-induced assortative matching from incentive-induced assortative matching. Such couples, however, represent a small share of the SFHHS.

diseases rises with age. Both factors make early infection preferable to later infection. Indeed, consumers may be willing to "pay" for their children to be infected when vaccines are unavailable, as evidenced by the common practice of having children sleep in the same bedroom with a sick sibling, or so-called "measles parties." The apparent desire of parents for early immunity, whether through infection or vaccination, reflects the fall of immunity's benefits and the rise of infection cost with age. This incentive explains both the high rate of infection for young people, and the extremely low rate of infection for individuals past the late teenage years, after which the cost of infection increases steeply. Epidemiologists explain such patterns in the age structure of disease through school mixing patterns. Since infected children are more likely to meet susceptible children when they are in school, the hazard rate into infection increases in early school age. This assumes that the likelihood of infection conditional on exposure remains constant across ages, but that exposure patterns differ across ages. However, mixing is a choice governed by its costs and benefits. Parents let their children go to school in the potential presence of a disease when the purchase of immunity through infection is on balance valuable. Parents react more negatively when the infection bears no benefit as for diseases such as AIDS; indeed, children with AIDS have faced pressure not to attend school.

3 Rational Epidemics & Public Health Interventions

The previous sections discussed some of the positive implications of prevalence-elastic demand for disease occurrence. We will now discuss and evaluate several types of public health interventions in light of such behavior of epidemics and long-run prevalence.[13]

3.1 Public Price Subsidies

Economists have long offered qualitative arguments concerning the positive external benefits of vaccination, effects which result in the private under-provision of vaccines. This has provided a motivation for Pigovian subsidies aimed at correcting the under-provision. However, we find that the steady state price elasticity of demand for vaccines is reduced under prevalence-elastic demand, so that such subsidies become ineffective. Therefore, the Pigovian subsidies traditionally seen as resolving the under-provision problem of vaccines can be short-run, or out of steady state, arguments.[14] Under a prevalence responsive demand, the relatively low price elasticity may limit the efficacy of Pigovian subsidies; in the extreme case, such subsidies may not raise total demand at all.

In the steady state, the fraction of individuals in each of the three health states remains constant over time at levels (S, I, R), now denoted without time indices. Denote by $D(I, p)$ the demand for vaccination under a constant future prevalence path at level I, where $D_p \leq 0$ and $D_I \geq 0$. The positive sign of D_I we refer to as the *prevalence response* of demand. The benefit of vaccination rises with prevalence. It can also be shown that for each stationary price, there is a unique steady state prevalence denoted $I(p)$, which is increasing in price and locally stable.[15] Therefore, unlike many other dynamic economic systems, our

[13] In predicting the effects of public health measures on disease occurrence, the economic approach differs from the epidemiological mainly in assuming that there is demand in the *private* sector for disease prevention, which may or may not be advanced by efforts undertaken in the *public* sector. The epidemiological approach fails to consider both the possibility of privately provided disease prevention and the possibility that public interventions may be rendered ineffective by private responses. Epidemiological models tend to credit reductions in disease prevalence entirely to public interventions. Philipson and Posner (1993) provide a basic qualitative discussion of the impact of private incentives on the AIDS epidemic and the reduced role of public intervention. For a critique of this view, see Kremer (1995,1996).

[14] Most standard treatments of public finance or health economics discuss these Pigovian subsidies for so called under-provided vaccines. See for instance Stiglitz (1988), Fuchs (1989), and Phelps (1992) and Pauly (1994). Brito *et al.* (1991) argue that a 100 percent vaccination rate is not Pareto optimal. However, this does not seem to be unique to the externality of vaccines–Pigovian subsidies may improve efficiency, but banning exposure-inducing behavior altogether may carry efforts too far. Recent analysis by Francis (1997) Xu(1998), and Hsu (1998) analyses the role of subsidies in a dynamic setting. Ainsworth and Over (1997) discuss a particularly interesting application of these ideas to the problem of AIDS in developing countries.

[15] See Geoffard and Philipson (1997). We ignore the steady state $(S, I, R) = (1, 0, 0)$ with zero prevalence.

model will not exhibit cycles, even with myopic demand: vaccination-induced cycles can only occur with a lagged prevalence response of demand.

Using the unique and positive relationship between prevalence and price, vaccine demand can be written as a function of price *alone*, as $\bar{D}(p) \equiv D(I(p), p)$. The total effect of a price increase on this demand then consists of not only the standard direct negative effect, but also the indirect and positive effect through increased prevalence:

$$\bar{D}_p = D_p + D_I \cdot I_p. \tag{3-1}$$

The indirect positive effect depends on the degree to which prevalence rises with price. This effect, in turn, falls with the prevalence response in demand D_I, as can be seen in the following steady state relationship:[16]

$$I_p = -\frac{1}{D_I + \frac{w+m}{b}} \cdot D_p. \tag{3-2}$$

As price increases, demand decreases, causing prevalence to increase. The rise in prevalence creates a counteracting feedback by causing demand to increase. This feedback limits the impact of price on prevalence. This counteracting effect becomes larger as demand becomes more prevalence responsive. Consequently, the total effect of price on demand falls with the prevalence response of demand D_I, as can be seen by substituting I_p into the expression for $\bar{D}_p$:

$$\bar{D}_p = \left(\frac{1}{1 + \frac{b}{w+m} D_I} \right) \cdot D_p. \tag{3-3}$$

The total price effect turns out to be the partial effect discounted by a factor which falls with the prevalence response of demand. In sum, the larger is the prevalence response of demand, the less it responds to price.

When public subsidies vary with prevalence, as denoted by $s(I)$, we call them counter-cyclical if $s'(I) < 0$, or pro-cyclical if $s'(I) > 0$. Without a doubt, the majority of public sector subsidy programs, whether international, national, or sub-national, are pro-cyclical. Once we consider prevalence dependent subsidies, the total steady-state demand function becomes

$$D^s(I, p) \equiv D(I, p - s(I)). \tag{3-4}$$

The total prevalence response of demand is then

$$\frac{dD^s}{dI} = D_I - s_I D_p. \tag{3-5}$$

An increase in prevalence affects subsidized demand directly, and indirectly by raising the level of subsidies under a pro-cyclical policy, or lowering the level under a counter-cyclical policy. An increase in subsidization induces a decrease

[16] See Geoffard and Philipson (1997).

in price and *vice-versa*. One can also show that the larger is the prevalence response of demand, the lower is the steady state prevalence I; formally, I is lower whenever D_I is uniformly larger.[17] This directly implies, therefore, that a pro-cyclical subsidy policy lowers steady state prevalence more than a counter-cyclical policy. Furthermore, since demand varies less with price when the prevalence response of demand is larger, pro-cyclical subsidies mitigate the effects on quantity of a less competitive market structure.

For a competitive market facing a stationary subsidy s, the market eradicates the disease only if the subsidy covers the minimum average cost of production. On the other hand, a monopoly producer of vaccines faces an unconventional incentive to keep the disease alive: if the disease is eradicated, so is the demand for the monopolist's product. Consider a monopolist who faces constant marginal costs c. Clearly, the monopoly price can never fall below the subsidy level since demand is completely price inelastic at such a price. If the monopoly price equals the subsidy, the vaccine is free to consumers who demand it universally, as in $\bar{D}(0) = 1$. Universal demand is assumed to lead to eradication after a length of time T. Given a discount rate α, the profits gained from eradication equal:

$$\Pi_E = \int_0^T (s - c)\bar{D}(0)e^{-\alpha t}dt \tag{3-6}$$

When the monopoly price p exceeds the subsidy s, we do not have universal demand. In this case, profits are given by:

$$\Pi_N = \int_0^\infty (p - c)\bar{D}(p - s)e^{-\alpha t}dt \tag{3-7}$$

If there exists a price p such that $\Pi_N > \Pi_E$, the monopolist will not eradicate the disease. This condition is equivalent to $(s-c)\bar{D}(0)(1-e^{-\alpha T}) > (p-c)\bar{D}(p-s)$. This condition demonstrates the crucial point that eradication is less likely to be profitable the more prevalence-elastic is demand. High prevalence elasticity implies that $\bar{D}_p$ will be low, and thus $\bar{D}(p-s)$ will not be much lower than $\bar{D}(0)$. A monopolist facing inelastic demand will never find it profitable to eradicate, because the short-term increase in quantity offered by eradication will not be large enough to compensate for the zero future profits offered by eradication.

The monopolist chooses eradication provided that the current profits from universal vaccination exceed the loss in future profits from the elimination of the disease. However, the important point here is that eradication is less likely to be profitable the more responsive is demand to prevalence. This is so because eradication is less profitable the less demand responds to price and, as discussed, $\bar{D}_p$ falls with D_I. When demand is price inelastic, it never pays to eradicate because the monopolist earns a loss both before and after the disease is eradicated: a loss after eradication (as discussed) because his product is valueless, and a loss before eradication because raising the price will increase current profits. In addition, if future profits are not heavily discounted, so that α is high, the cost of

[17] For a proof, see Geoffard and Philipson (1997).

eradication is higher. In sum, if demand is highly responsive to prevalence or if discounting is moderate, subsidized eradication is not profitable.[18]

3.2 Mandatory Vaccination

Virtually all observed mandatory vaccination programs are partial—they do not cover whole populations or even whole age groups. Therefore, private decisions to vaccinate outside of public programs remain an important component of the total demand for vaccination. The total demand when a public program covers a fraction f of the population is given by

$$D_T(I, p, f) = f + (1 - f)D(I, p). \tag{3-8}$$

The first term is the mandatory demand in the program, while the second term is the private demand outside the program. Partial mandatory programs crowd out the private demand for vaccination outside the program, in the sense that some individuals would vaccinate in the absence of the program, but will not vaccinate in its presence.[19] More precisely, the marginal effect on demand of increased public coverage is

$$\frac{dD_T}{df} = [1 - D] + (1 - f)D_I I_f. \tag{3-9}$$

The first term is the direct positive effect resulting from the increased public coverage of individuals who otherwise would not have vaccinated. The second term is the indirect negative effect on private demand by individuals not covered by the public program. It reflects the negative effect on demand exerted by a decrease in prevalence. This term increases in the prevalence response of demand. Therefore, the higher the prevalence elasticity, the less effective are mandatory vaccination programs at raising total demand.

3.3 The Pareto-Optimal Timing of Epidemic Interventions

Critics of the public response to the AIDS epidemic commonly charge that public health interventions occur "too late" into the epidemic. Table 1 below summarizes the speed at which state governments in the U.S. responded to the AIDS epidemic, as measured by expenditures on prevention programs. The table is computed using data on the time elapsed before states implemented programs

[18] Subsidized *suppliers*, rather than demanders, may not be prevalence elastic under so called supplier-induced demand. Supplier subsidization was undertaken in England in 1990 when general practitioners received bonuses if they achieved prespecified immunization targets for their patients. The fraction who achieved the targets increased from 55% at the start to 85% at the end of the program in 1992 (Principal Medical Officer, Department of Health, England, 1994). This policy raises the interesting question of whether health activities with positive external effects, such as those in many areas of public health, may be efficiently provided by supplier-induced demand.

[19] This crowding out effect is one possible interpretation of the relatively low *pre*-school vaccination rates in the U.S., given the mandatory vaccination required in public schools.

for education about AIDS prevention, the main type of public preventive expenditure program in the U.S. A row of the table corresponds to a given fiscal year. The first column (Total Left) refers to the number of state governments that had not implemented an education program at the start of the fiscal year; this represents the survival rate of non-implementation. The second column (Starters) indicates the number of governments that started an education program during that fiscal year; this represents the exit rate from non-implementation. The third column (Survival) indicates the fraction who survived, or the fraction of state governments which had not implemented an education program by the start of the next fiscal year. Lastly, the 95% confidence intervals are reported for this survival curve.

TABLE 1 INSERTED HERE

The table reveals the frequently discussed slowness of the public sector's response (see, *e.g.*, Shilts (1987)). This is illustrated by the fact that, even in 1988, well into the national epidemic, more than *half* of the states had not established prevention programs. The common focus on the calendar timing of the public response is misleading, however, since it is disease prevalence, not disease duration, that determines incentives for protection in the private sector. The rapidity of public-sector response should be judged in relation to prevalence levels, since the latter drive private incentives for protection.

Figure 3 below depicts the resulting relationship between state implementation of education programs and state prevalence levels. Specifically, the figure shows the share of US state governments as a function of disease prevalence in the state that had not yet adopted an AIDS education expenditure program. It is the survival curve of the duration until program adoption where the duration is measured in terms of the prevalence level.[20]

FIGURE 3 INSERTED HERE

The figure illustrates that disease prevalence increases adoption of public programs, and that the rate of adoption rises with prevalence, as evidenced by the convexity of the survival curve. While calendar response times do not affect private incentives, the slow calendar response and the rapid prevalence response of governments suggest that calendar response times may also be uncorrelated with the public sector incentives.[21]

What determines the Pareto timing of an intervention into an epidemic caused by a population that is heterogenous in the willingness to bear risk? To gain insight into the efficient timing of a subsidy program, consider first the effect of a complete subsidy to a *single* individual, which will not affect overall

[20] We measure AIDS prevalence as AIDS cases reported to the Centers for Disease Control, *per* 100,000 state residents.

[21] Indeed, the unconditional correlation between calendar duration and prevalence duration turns out to be close to zero. For example, the District of Columbia responded first in terms of calendar time, but last in terms of prevalence.

prevalence I_t. The impact of the subsidy on the behavior of this single individual depends on whether the prevalence has already reached that individual's threshold prevalence K. If the prevalence has reached this threshold prevalence, the subsidy has *no* effect on that individual. This implies that a subsidy to a heterogeneous population affects only those individuals with a larger threshold prevalence . More precisely, the prevalence and the public subsidy program are *competing* incentives influencing the individual demand for protection, which implies that subsidy programs may not be Pareto-optimal if they are undertaken when prevalence is too high. If they come in too late, the subsidies may be irrelevant for a large share of the population for whom prevalence itself has already induced protection. This may appear counter-intuitive, since it suggests that the larger the case load, the smaller is the role for government intervention through subsidies. However, as prevalence rises, the private incentive to undertake protective conduct also rises, thereby lowering the effect of government subsidies.

Consider the aggregate effect of a typical subsidy program which is financed by non-eligible individuals. Let $G_n(p, I)$ and $G_e(p, I)$ denote the share exposed in the dynamic model for non-eligible and eligible individuals of the program, respectively. For a program involving a percentage subsidy of s for protective activity, there is a potential for Pareto improvement only if non-eligible individuals are willing to pay eligible individuals to demand protection, and they are willing to pay an amount at least equal to the program cost. The following proposition shows that this can never be the case if the subsidy program starts "too late," in terms of disease prevalence. More precisely, it can be shown that *for every subsidy level s, there exists a limit prevalence level $I(s)$, such that the subsidy program is not Pareto-improving if it starts after prevalence has reached $I(s)$. Moreover, the limit prevalence level $I(s)$ decreases with s.* The higher the proposed subsidy, the larger must be the benefit of the program. Therefore, the program must be undertaken at lower prevalence levels in order to have greater effects. Interestingly, at the later stages of an epidemic, lower subsidization levels outperform higher levels.

The revenue for such a subsidy program would come from non-eligible individuals. Self-interested individuals who have already been induced by disease prevalence to engage in protective behavior, however, will pay *nothing* for other individuals to engage in protective behavior, since they have nothing to gain from others' protective activity.[22] If a public program is not fast enough, the growth in prevalence will thus have eliminated the willingness of some individuals to pay for others to engage in protective activity.

[22]This cap on the external benefit from protective behavior may be particularly relevant to explicitly choice-based diseases such as HIV, since many individuals may find protective behavior to be costless, or even preferable. For example, monogamous married couples may have to be *compensated* to engage in transmissive activity with high-risk groups.

3.4 The Dynamic Welfare Effects of Eradication

Any evaluation of a disease eradication policy will depend crucially on the way dynamic, rather than static, externalities are considered. This is so because a main benefit of an eradication program eliminates the costs of disease prevention in the future. For the current population, eradication is never optimal, because the marginal costs of vaccination eventually outweigh the benefits of further decreases in disease prevalence. However, there is a missing intergenerational market: future generations cannot pay vaccine producers for the benefit of eradication, although collectively they benefit most from an eradication program[23].

To consider the dynamic value of an eradication programs, let the lower bound $\underline{B}$ on the willingness to pay for eradication by future generations be given by the net-present value of these future prevention expenditures; using the demand for vaccination $\bar{D}$, we can write this as:

$$\underline{B} \equiv \int_{t=T}^{\infty} \bar{D}(p) p e^{-\alpha t} dt = \frac{p\bar{D}(p)e^{-\alpha t}}{\alpha}. \tag{3-10}$$

On the other hand, the cost of any efficient eradication scheme cannot exceed the upper bound $\bar{B}$, the cost of universal vaccination for T periods (where T is the length of time necessary to achieve eradication under universal vaccination). The quantity $\bar{B}$ is given by:

$$\bar{B} \equiv \int_{t=0}^{T} p e^{-\alpha t} dt = \frac{p(1 - e^{-\alpha T})}{\alpha}. \tag{3-11}$$

The upper bound must be the cost of subsidizing everyone, since vaccination demand goes to zero as prevalence goes to zero. The prevention expenditures avoided are thus larger than the required subsidies whenever

$$\underline{B} \geq \bar{B} \Leftrightarrow \bar{D}(p) \geq e^{\alpha T} - 1. \tag{3-12}$$

The discount factor α crucially determines the desirability of eradication. If the discount factor is zero then eradication is always desirable for all demand functions, since the value of the future prevention eliminated always exceeds the current cost of the eradication program. On the other hand, if there is full discounting, and α becomes very large, eradication is always dominated by the free-choice equilibrium. In addition, a longer eradication program naturally makes eradication relatively more costly, as does a low level of demand caused by high prices or other features of the disease.

It is useful to compare this dynamic social problem to the dynamic problem of a subsidized monopolist. In the social problem, the benefits of eradication accrue in the future, but in the monopolist's problem, eradication sends future profits to zero. The discount factor thus has different effects in the two problems: less discounting makes the monopolist less likely to eradicate, although he is

[23]There are other potential benefits of eradication not discussed here such as the elimination of a mutation or drug-resistent strain of a virus.

more likely to improve welfare. Thus, a deficit-financed eradication program, which spends beyond tax revenues during its operation but recoups the deficit in future generations, may improve welfare when discounting is at current market rates. Such a program would allow for the inter-generational transfers necessary to pay current generations to over-vaccinate for the benefit of future generations, who will not purchase vaccines from their manufacturers.

If eradication is to be achieved, there is still a choice between the regulatory approach of mandating demand and the fiscal approach of subsidizing it. Given the difficulty of subsidies relative to mandates the fact that larger distortionary taxes are needed for the latter only enforces the superiority of the regulatory approach. However, dependent on the cost of enforcing the regulations, and the distortionary taxes needed for that purpose as well, there may well be a trade-off between the two methods.

3.5 Public Intervention into Allocations of Information

Many public health interventions involve the dissemination or regulation of information. This is particularly true for sexually transmitted diseases such as HIV where screening, partner notification, education, confidentiality legislation, or surveillance reporting to the population are all part of the public arsenal. The effect of changes in the allocation of information, of "who knows what and when," cannot be studied using epidemiological analysis, because changes in the allocation of information do not change behavior in epidemiological analysis. This excludes explicit analysis of the effects and desirability of the bulk of public AIDS-prevention spending in the United States and abroad, on education, antibody testing, and counseling services for individuals at risk of acquiring or transmitting HIV. Virtually all fiscal institutions which have responded to the HIV-epidemic, whether international, national or sub-national, have included HIV education and testing in their prevention efforts.[24]

To evaluate public education programs, we must recognize that the information structure of a disease determines its growth and prevalence. We may regard the case of an "asymptomatic" disease such as HIV as a case of imperfect information–a person may harbor HIV even in the absence of any symptoms observable to his partner. Specifically, if we regard *symptomatic* disease as providing perfect information, asymptomatic disease corresponds to the case of asymmetric information. The classic paper by Akerlof (1971) first studied the effects of such information problems on the volume of trade in economic markets. If one interprets sex as an economic trade in the sense of being an activity which is *ex ante* beneficial to both parties involved, the problem of AIDS can be viewed as a problem of quality uncertainty among traders. In such a market, traders may learn their quality status through HIV testing. Incomplete

[24]Bloom and Glied (1991) first considered the incentives of private employment based HIV-screening, but were not concerned with the three main questions of this survey. They considered the lack of private demand for employer HIV-tests of employees for the purpose of reducing the costs of employment-based health insurance. They calculate that due to the low prevalence of the disease, the costs outweigh the benefits relatively more for smaller firms.

information will inefficiently restrict the volume of trade when traders forego mutually beneficial trades, which would take place under perfect information, for fear of trading with a bad partner. Similarly, HIV risk may cause pairs of susceptible individuals not to engage in mutually beneficial unprotected sex for fear of infection. In the extreme case of perfect information about infection status, this information theory predicts that infected individuals will engage in sexual activity only with other infected individuals and noninfected individuals will engage in it only with noninfected individuals–the degree of information determines the degree of assortative matching in the population. The privately determined disease growth rate would be much smaller under perfect information, zero under full assortative matching, indeed so small as to wipe out the disease.

Philipson and Posner (1995) and Boozer and Philipson (1997) analyze the private demand for information in this market and the effects of public subsidies on information acquisition. The theoretical analysis implies that mainly low-risk HIV-positive and high-risk HIV-negative individuals will alter their beliefs and behavior due to public information subsidies, and that the aggregate response to such a program may be small, because the effects across such risk groups are offsetting. This is analyzed empirically using a longitudinal survey that imitated a public HIV testing program by actually administering an HIV test as part of the survey–the San Francisco Home Health Study (SFHHS), collected by the AIDS Prevention Center at the University of California at San Francisco (UCSF) during 1988-89. The survey recorded the beliefs and sexual activity of individuals before and after the testing. Using this direct evidence on the effects of providing traders with private information, the analysis assessed the longitudinal impact of such a public information intervention. Consistent with the theoretical discussion, the study found that knowledge of one's HIV status *increased* the volume of sexual contact by 16 percent for high-risk HIV-negative respondents and had little effect on high-risk HIV-positives. Mechonlan (1998) provides an interesting analysis of the aggregate implications of these incentives for disease prevalence. He shows that not only does prevalence rise with subsidization of testing, but testing also reduces welfare.

The responsiveness of high-risk negative individuals and low-risk positive individuals calls into question the rationale for public education and testing programs. If such programs target high-risk groups, they do little to alter behavior favorably. Only those agents who actually learn something modify their behavior—the positive low-risk and negative high-risk traders. Indeed, the high-risk negatives subsequently engage in sexual intercourse with a greater number of partners, potentially placing them at greater risk of a new infection. The theoretical analysis and empirical results suggest that a public testing program can have unintended effects, particularly when focused on high-risk populations.

4 The Welfare Loss of Disease and Medical Technology

In setting priorities among control efforts across many separate types of diseases, a major question facing public health authorities is the welfare loss inflicted upon a population by a given disease. The orthodox approach has employed several cost-of-illness ("COI") measures, each of which is a product of prevalence and (possibly quality-adjusted) severity of a disease. This approach has the seemingly self-evident implication that the more morbidity or mortality inflicted by a disease, the larger its welfare loss. This section discusses the economic welfare loss incurred due to infectious diseases and then compares it to such common public health measures of this loss. The analysis is founded on the idea that standard welfare analysis of taxation may applied to the welfare effects of regulating diseases when interpreting exposure to disease as a valuable good, and the expected cost of incurring the disease, here becoming infected, as a "random tax" on the consumption of this good[25]. The principal result is that the cost-of-illness measures cannot comprise a relatively large fraction of the total welfare loss when private prevention is prevalence-elastic; the main welfare loss of a disease will then consist of the distorted behavior the disease induces.

If a disease is interpreted as a tax on the consumption of exposure to the disease, the disease 'distorts' consumption to involve preventive activity that would not have been undertaken in absence of the disease. This 'excess burden' of the disease tax represents the main difference between public health measures of disease cost, which consider only the direct cost imposed on those infected by a disease, and an economic measure of the disease, which includes the cost of the distorted preventive behavior. Since not all individuals who consume the good (expose themselves to the disease) contract the illness, the tax is paid randomly. Public health measures of the loss inflicted by a disease, such as *cost-of illness* measures, are invariably comprised of the product of the total number of cases and the cost of each case. This represents only the 'revenue' part of the disease tax, thereby ignoring the excess burden due to the distorted behavior. The case of polio illustrates this point–there are currently no infections in the United States but each child has to be vaccinated. According to a revenue-focused measure of welfare loss from disease, polio induces zero loss in welfare, although were it not for polio expenditures of time and money on its prevention could be used more productively.

To consider the welfare loss of a disease let $\theta \equiv u(s,0) - u(i,0)$ denote its per-period cost of infection as discussed in the individual decision problem in previous sections. Technological medical advances may reduce this cost of infection through new treatments. Let $I(\theta)$ be the prevalence under a given cost and denote by $D(\theta) \equiv D(I(\theta),\theta)$ the steady state demand for prevention. If L denotes the steady-state welfare loss it was shown to be given by:

$$L \equiv D(\theta)C + (1 - D(\theta))I(\theta)\theta \tag{4-1}$$

[25] See Philipson (1995).

where C is the cost of prevention. Here, the first term reflects the excess burden the disease imposes on individuals who do not consume exposure given the implicit disease-tax on such activity. The second term reflects the loss suffered by those who engage in exposure despite this tax, that is, individuals who will eventually make up the case-load. This is the revenue part of the total welfare loss of a disease. If the disease does not induce harm the welfare loss is zero; $\theta = 0$ implies $D(\theta) = 0$ and hence $L = 0$. Figure 4 shows the differences between the two types of welfare losses as a function of the cost of infection.

FIGURE 4 INSERTED HERE

The COI measures of the welfare loss of disease, usually a product of the prevalence and the quality-adjusted per-capita cost of a disease, focuses solely on the second term of this loss by ignoring the distortions induced by the disease. The COI measures of public health correspond to the revenues or 'Laffer'-curve L_{PH} of the disease. When demand is prevalence elastic, then the COI measure may take on the inverted U-shaped feature because a rise in the cost reduces prevalence so that their product may not be monotonic; the COI measure may claim that a more costly disease is welfare *improving.*

The effect of a reduced cost of infection through improved treatments, that is, reductions in the parameter θ, are determined by

$$-\frac{dL}{d\theta} = D_\theta[I\theta - C] - (1 - D)[I_\theta\theta + I] \tag{4-2}$$

The first term is reduction in the welfare loss due to the reduced distortions and the second term is the reduction due to the lowered average cost on those who choose to consume exposure. The special case of COI analysis applies when demand is completely inelastic. Then the reduction in the welfare loss would be the reduction in the COI measure by the corresponding reduction in the per-capita cost of infection

$$-\frac{dL}{d\theta} = -(1 - D)I \tag{4-3}$$

However, this highlights the difference in welfare effects of public efforts to change the cost of infection θ dependent on whether behavior is prevalence elastic. If new medical technology lowers the consequences of disease, more consumption of exposure will take place thereby partly offsetting the public health achievements For example, for AIDS the recent new treatments of protease inhibitors will reduce distorted behavior, here safe sex thereby having an offsetting effect on prevalence. AIDS causes an excess burden in terms of foregone sexual consumption[26]. The existence of this excess burden is indeed why the case-load is so small and the disease is self-limiting..Nonetheless, most public

[26]The asymptomatic nature of HIV may make this burden especially large, because, as discussed above, even two uninfected people may inefficiently choose to abstain from unprotected sex.

health authorities (*e.g.*, the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC)) remain focused entirely on the direct cost of the illness and ignore the excess burden. However, when behavior is prevalence elastic the excess burden will dominate the total welfare loss of disease for serious diseases.

5 The Evidence on Prevalence Elastic Behavior

Many of the discussed implications for the behavior of epidemics as well as for the impact of public health interventions stem from the positive relationship between preventive behavior and disease prevalence. This section discusses empirical estimates of this relationship.

Ahituv *et al.* (1996) investigates the degree to which the local prevalence of AIDS increases the demand for disease-preventing methods of contraception among young adults. Using data from the National Longitudinal Survey of Youth (NLSY), they find the use of condoms to be quite responsive to the prevalence of AIDS in one's state of residence, and they find this responsiveness to have grown over time. Using both cross-sectional and longitudinal evidence, they find that a one percent increase in the prevalence of AIDS in one's home state increases the propensity to use a condom by up to 50 percent. Small levels of growth in AIDS prevalence have thus resulted in enormous growth in the demand for condoms.

In the U.S., AIDS has been spread primarily through unprotected sexual intercourse, especially, although not exclusively, among homosexual males. Sexual activity accounted for transmission in over 70 percent of all U.S. AIDS cases diagnosed through June 1992; of these cases, 90.6 percent resulted from homosexual contact. Over the last several years, however, the number of new AIDS cases in which transmission occurred by a means other than sexual contact, such as intravenous drug use, has grown more rapidly than the number of new cases due to sexual transmission. In addition, transmission via heterosexual contact has increased over time. Ahituv *et al* examine the extent to which one form of protection from STDs, condom use, has responded to the increased prevalence of AIDS among young adults in the US. They focus on the demand for condoms particularly because of its suggested importance in stanching the spread of AIDS. Using data from several large cohorts of youth over the second-half of the 1980s, they present evidence on how local AIDS prevalence affects the demand for contraception or, more precisely, they estimate the elasticity of demand for condoms with respect to AIDS prevalence. While many other studies had examined the nature of sexual activity under STD risk, none had attempted to estimate prevalence elasticities. In addition, unlike those of prior studies, the data set used in this study spanned the entire 1980s, so that behavior before and after the epidemic could be compared. This is a major advantage of the NLSY over other sources of data on AIDS. The paper analyzes a subsample of the NLSY, namely individuals aged 25 through 27. For the sample years 1984, 1986, and 1990, 8924 respondents fall into this age category, and these

respondents generate a total of 11,351 person-year observations.

In sum, it is found that condom demand rose substantially among 25 through 27 year-old people over this period. In particular, while not finding differences in condom demand across regions in 1984, before AIDS was very prevalent, they find that condom demand became geographically heterogeneous as the AIDS epidemic spread, with a higher growth of condom use in states with higher AIDS prevalence rates. Moreover, the highest rates of increase occurred among sexually active, single men and single men living in urban areas, all of whom are generally expected to have higher exposure to the risk of HIV infection. Although other factors might account for this regional change, these patterns strongly suggest that young adults altered their level of protection against STDs in response to increased risks.

The use of personal and state-level characteristics as controls for regional demand does not alter, and sometimes strengthens, the conclusions of this analysis. Specifically, young adults exhibit a prevalence-elastic demand for condoms, and the elasticity rises among men, blacks, unmarried people, urban residents, and more sexually active individuals who face a greater risk of HIV exposure. Moreover, prevalence-elasticity has increased substantially over time: condom demand does not differ across states before the epidemic, but as the epidemic progresses, the interstate differences grow. Finally, the authors find that increases in the prevalence of AIDS in one's state of residence explain more than half of the dramatic rise in condom use which took place among young adults in the second half of the 1980s. The prior analysis treats the NLSY as a set of repeated cross-sections. Exploiting its longitudinal features, however, the rate of condom adoption is still found to be significantly higher for individuals living in areas with higher growth rates of AIDS prevalence.

Table 2 shows the *prima facie* evidence for the relation between differences in condom demand and differences in AIDS prevalence. It compares, within a given year, the demand for condoms broken down by quartiles of the AIDS prevalence rate in the respondent's current state of residence. These results are presented in Table 2, which displays the proportion of individuals using condoms by quartile as well as the p-value associated with the null hypothesis that condom use in the first quartile equals that in the fourth. Condom demand does not differ across quartiles of AIDS prevalence in 1984 for any of the demographic groupings. However, in subsequent years, the quartiles begin to diverge, with individuals in higher prevalence states using condoms more frequently. In both 1988 and 1990, condom use consistently increases across quartiles, and these increases are statistically significant.

TABLE 2 INSERTED HERE

Analysis of the longitudinal NLSY data reveals the same trend: the growth rate in condom use differs across quartiles of the growth rate of AIDS prevalence between 1986 and 1988, as well as between 1988 and 1990, but it does not differ across quartiles between 1984 and 1986. The higher the growth of risk, the more likely is growth of preventive behavior. In addition, condom use by

married individuals does not respond to movements in AIDS prevalence. This is consistent with the theory, since such individuals represent a control group which does not face increased risk of infection from increased rates of prevalence.

In order to investigate the robustness of these patterns, they presented results for four alternative empirical specifications of individual-level, reduced-form demand functions for condoms. In each, they control, in various ways, for personal and regional characteristics other than the state-level AIDS prevalence rates in order to better isolate the effects of regional and temporal changes in the full price of unprotected sex. They control for the pecuniary market price for condoms by using year and state dummies, so that any uncaptured differences in price must come from possible but unlikely within state and year variation in condom pricing. Table 3 shows the estimated prevalence elasticities broken down by year and by various demographic subgroups, for the most general empirical specification considered.

INSERT TABLE 3 HERE

The significantly negative prevalence elasticities in 1984 and 1986 appear to be consistent with the claim of public health officials that low condom use contributed to the spread of the epidemic. However, these significantly negative coefficients become transformed into significantly positive coefficients for almost all demographic groups except married men, as claimed.

Although this evidence relates to the demand for transitory protection against infectious disease, the same type of patterns characterize the demand for permanent protection through vaccination. Using data from the 1991 National Health Interview Survey, Mullahy (1997) provides evidence on the prevalence-elasticity of influenza vaccine demand. Each year in the US, anywhere from 10 to 50 million individuals contract influenza. In a typical year, about 20,000 of these individuals, 90% of whom are elderly, die from complications of the disease. During influenza season, the hospitalization of elderly patients reaches an annual high, while influenza accounts for more than 10% of visits to doctors' offices during peak flu season. Mullahy finds a high correlation between the demand for vaccines and mortality risk, calculated as the product of disease prevalence and the probability of mortality conditional on infection; high-risk groups such as health care workers, individuals in areas with more recent outbreaks, or the elderly demand vaccines at relatively larger rates. This relationship is illustrated for the elderly in Figure 5, which gives the age-profile of flu vaccine demand.

INSERT FIGURE 5 HERE

Less than 5 percent of people below age 50 demand the vaccine, but demand rises sharply for individuals older than that, until about half the population over 70 demands vaccination. The much higher risk of mortality faced by the elderly explains this dramatic change in demand; the risk of mortality conditional on infection rises very sharply in age and in fact resembles the age-profile of demand, as one would expect.

Also investigating the demand for vaccines, Philipson (1996) shows that the prevalence-elasticity of demand for the child combination vaccine against measles, mumps and rubella (MMR) parallels the prevalence-elasticity of HIV and influenza vaccination.[27] The paper used data from the National Health Interview Survey (NHIS) on US measles vaccination between 1984 and 1991, as well as a child health supplement from the 1991 NHIS; these data are combined with measles case loads in the respondents' states of residence, as reported by CDC's *Morbidity and Mortality Weekly.* Due to a measles outbreak between 1989 and 1991, the data display sufficient variation across states and time to permit the estimation of prevalence-elasticity for the MMR vaccine. The paper identifies the main dependent variable of interest as the duration represented by the age in months at which a child obtains her first MMR vaccination. The study finds that regional case load, a time varying covariate, has a large positive impact on the hazard rate into vaccination. Using a proportional hazards model, the effects on duration of a large set of control variables is investigated. These controls include several measures of public health policies aimed at stimulating demand in the respondent's state of residence. Prevalence exerts a large and highly significant positive effect on the hazard rate into vaccination. Indeed, disease prevalence affects the hazard rate more than any other determinants of vaccination completion, and this relationship persists across a variety of specifications. The results of this study have also been confirmed by Conroy and Fische (1996), who also find strong support for a prevalence-elastic demand using the same type of data from NHIS.

We have summarized a wide variety of evidence for the prevalence-elasticity of demand for prevention of infectious disease. A major drawback of the studies is that they do not identify why demand is prevalence elastic. In particular, empirically it is not known how information gets transmitted during the growth of a disease. Part of the transmission occurs in the private sector through mass media, but part of it also is attributable to public efforts at surveillance. Future empirical work should better attempt to assess the process by which information is disseminated. If such prevalence-elastic demand is identified in future studies, our arguments demonstrate that it has many important implications for both the behavior of infectious diseases as well as the effects of public health interventions aimed at controlling their spread.

6 Concluding Remarks

This paper has reviewed recent economic analysis of public health policies and infectious disease. The discussion focused on three general questions which reveal how the economic approach offers unique insights into the behavior of disease and optimal public health interventions, insights which epidemiology cannot provide. We first examined the difference between the disease behavior predicted by the economic approach and that predicted by epidemiology. We concluded that incentives for prevention make epidemics self-limiting, because

[27] See also Goldstein *et al* (1996).

the prevalence of a disease raises the incentives for preventive behavior. Second, we examined the different implications for public health policy offered by the economic approach and the epidemiological approach. The economic approach yields the insight that public intervention often provides less benefit than predicted by epidemiology, because private incentives counteract its effects. Lastly, we compared the different ways in which economics and epidemiology evaluate the cost of disease. Here we found that epidemiology focuses only on part of the cost, and misses the sizable costs of disease-avoidance. This focus causes epidemiology to stress the prevalence of a disease as the key determinant of its cost. We found this focus misplaced, because the total welfare loss imposed by a disease has more to do with the severity of the disease and the steps people take to avoid it.

The growth in economic analysis of infectious disease has been substantial but many important areas remain open for future work. For example, future work could examine how the endogenous matching of infected and uninfected individuals would change the predictions generated by dynamic models of disease growth. Second, researchers should more fully explore how asymmetric information about the disease status of sex partners reduces the volume of sexual trades and thereby limits the growth of STDs such as AIDS. Third, economic analysis should be brought to bear on the unique issues surrounding disease control in low-income countries.[28] Empirically, much more work is needed on the sources and determinants of the prevalence-elasticity of demand for prevention; such analysis could help reveal why certain subpopulations engage in less prevention than others as seems to be the case in many African countries. One major difference between rich and poor countries is that in poor countries, AIDS is spread though *markets*, namely prostitutes. Therefore, sexual behavior may be less prevalence-elastic if the market responds more in price than in quantity to the lowered quality of supply implied by more infected prostitutes.

More generally, the burgeoning field of economic epidemiology offers several promising directions for future research. Of course, epidemiology, despite its name,[29] comprises more than just the study of infectious diseases, and economic analysis of other areas within the field seem useful avenues for future research. One such area is the design and analysis of clinical trials.[30] Although there exists a substantial literature on evaluation of social programs through methods of random assignments of treatments in econometrics, the special features of clinical trials are plentiful and have not been addressed. A second area is the examination of the normative approach to disease control advocated by epidemiologists, for infectious and non-infectious diseases. This approach consists of discovering and estimating the most important covariates of disease[31]

[28] However, see the chapter by Paul Gertler in this *Handbook*.

[29] Epidemiology stems from two Greek words; the word for epidemics (epidemia) and the word for the study of (logos). I thank Charles Phelps for this extension of my Greek vocabulary.

[30] See, e.g., Philipson and DeSimone (1997) and Philipson and Hedges (1998).

[31] So called 'risk-factors' which consist of the significant regressors in equations with disease occurence as the dependent variable.

and then, in a great leap of faith, recommending private or public intervention aimed at controlling the most empirically important covariates. Our analysis suggests that there is little role for public prevention of non-communicable diseases, and that the crowding out of prevention limits the benefits for communicable diseases as well. It seems natural to inquire into the conditions under which this canonical approach makes sense from the standpoint of Pareto optimality. It seems clear that high-risk targeting often favored by such analysis may not be efficient since a high risk group, by definition, has high costs of prevention relative to benefits. A primary example of this is the continued advocacy of partner reduction programs for the African prostitutes who make up the so called 'core groups' of the HIV epidemic in those countries. Since earnings are proportional to partners for these workers, such programs are the equivalent of grant-reduction programs for academics, which presumably would have a similarly low take-up rate.

To end on an optimistic note, it seems that epidemiology is an area where economics as a social science may successfully compete with and outperform the natural sciences. Economics can provide theories which explain more phenomena and assess more precisely disease-control efforts more precisely. Despite its recent growth, the field remains in its infancy relative to its possibilities. The global importance of world-wide mortality caused by infectious disease ensures that such research will pay valuable dividends by improving the understanding of the way infectious diseases spread, and the ways individuals and institutions can control them.

References

[1] Ahituv, A., J. Hotz, and T. Philipson. 1996. "Is AIDS Self-Limiting? Evidence on The Prevalence Elasticity of The Demand for Condoms." *Journal of Human Resources* 31(4):869-898.

[2] Ainsworth, M., and A. M. Over. 1997. *Confronting AIDS: Public Priorities in a Global Epidemic.* World Bank,. Oxford University Press.

[3] Akerlof, G. 1971. "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 89: 488-500.

[4] Anderson, R., and R. May. 1991. *Infectious Diseases of Humans: Dynamics and Control.* London: Oxford University Press.

[5] Auld, M. C. 1996. "Choices, Beliefs, and Infectious Disease Dynamics." Working paper 938, Department of Economics, Queen's University.

[6] Auld, M. C. 1997. "Behavioral Response to the AIDS Epidemic: Structural Estimates from Panel Data." Working paper. Department of Economics, Queen's University.

[7] Bailey, N. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications.* 2nd ed. Hafner Press.

[8] Becker, G. 1991. *A Treatise on The Family.* Cambridge, MA: Harvard University Press.

[9] Bloom, D. 1997. "AIDS and Growth." Working Paper No ?, *National Bureau of Economic Research.*

[10] Bloom, D. E., and G. Carliner. 1988. "The Economic Impact of AIDS in the United States." *Science* 604: 239.

[11] Bloom, D. E., and S. Glied. 1991. "Benefits and Costs of HIV Testing." *Science* 252: 1798-1804.

[12] Bongaarts, J. 1996. "Global Trends in AIDS Mortality." *Population and Development Review* 22(1): 21-45.

[13] Boozer, M., and T., Philipson. 1997. "Public Intervention into Markets of Asymmetric Information: The Case of HIV", forthcoming, *Journal of Human Resources.*

[14] Brandeau, M., and E. Kaplan. 1994. *Modelling the AIDS Epidemic : Measurement, Planning, and Policy.* Raven Press.

[15] Brito, D., E. Sheshinski, and M. Intriligator. 1991. "Externalities and Compulsory Vaccinations." *Journal of Public Economics* 45: 69-90.

[16] Castillo-Chavez, C. 1989. *Mathematical and Statistical Approaches to Aids Epidemiology.* Springer-Verlag: Heidelberg.

[17] Conroy, P., and R. P. H. Fishe. 1996. "An Economic Analysis of Vaccination Rates for Preschool Children." Working Paper. Department of Economics, University of Miami.

[18] Cuddington, J. T. 1993a. "Further Results on the Macroeconomic Effects of AIDS: The Dualistic, Labor-Surplus Economy." *The World Bank Economic Review* 7(3): 403-417.

[19] Cuddington, J. T. 1993b. "Modeling the Macroeconomic Effects of AIDS, with an Application to Tanzania." *The World Bank Economic Review* 7(2): 187-189.

[20] Cuddington, J. T., and J. D. Hancock. 1994. "Assessing the Impact of AIDS on the Growht Path of the Malawian Economy." *Journal of Development Economics* 43: 363-368.

[21] Cuddington, J. T., J. D. Hancock, and C. A. Rogers. 1994. "A Dynamic Aggregative Model of the AIDS Epidemic With Possible Policy Interventions." *Journal of Policy Making* 16(5): 473-496.

[22] Dow, W., and T. Philipson. 1996. "The Implications of Assortative Matching for The Incidence of HIV." *Journal of Health Economics* 15(6): 735-752.

[23] Francis, P. J. 1997. "Dynamic Epidemiology and the Market for Vaccinations." *Journal of Public Economics* 63: 383-406.

[24] Fuchs, V. 1989. "Health Economics." In J. Eatwell, M. Milgate, and P. Newman, Eds. *The New Palgrave: A Dictionary of Economics.* New York: Norton.

[25] General Accounting Office. 1989. *GAO/PEMD 89-13 AIDS Forecasting: Undercounts Weaken Estimates.* Washington, DC: GAO.

[26] General Accounting Office. 1994. *Vaccines for Children: Critical Issues in Design and Implementation.* Washington, D.C.: US Government Printing Office.

[27] Geoffard, P., and T. Philipson. 1995. "The Empirical Content of Canonical Models of Infectious Diseases." *Biometrika* 82(1): 101-114.

[28] Geoffard, P., and T. Philipson. 1996. "Rational Epidemics and Their Public Control." *International Economic Review.* 37(3): 603-624

[29] Geoffard, P., and T. Philipson. 1997. "Disease Eradication: Public vs Private Vaccination." *American Economic Review* 87(1): 221-31.

[30] Goldstein, K. P., T. Philipson, H. Joo, and R. Daum. 1996. "The Effect of Epidemic Measles on Immunization Rates." *Journal of the American Medical Association* 276(1): 56-58.

[31] Hay, J. W., and Wolak, F. 1990. "Bootstrapping HIV/AIDS Projection Models: Back Calculation With Linear Inequality-Constrained Regression." *Stanford Hoover Institute Working Paper in Economics:* E-90-5.

[32] Hay, J. W., and Wolak, F. 1994, 'A Procedure for Estimating the Unconditional Cumulative Incidence Curve and Its Variability for the Human Immunodeficiency Virus, Appl. Statist., v 43, No 4, 599-624.

[33] Heckman, J., and B. Singer. 1984. "Econometric Duration Analysis." *Journal of Econometrics* 24: 63-132.

[34] Hsu, H.-P. 1998. "A Theory of Vaccination". Department of Economics. University of California, Berkeley.

[35] Kremer, M. 1995. "AIDS: The Economic Rationale for Public Intervention." World Bank, Washington, DC.

[36] Mann, J., D. Tarantola, and N. Netter. 1992. *AIDS in The World.* Cambridge: Harvard University Press.

[37] Mechonlan, S. 1998. "The Effect of Testing in AIDS Prevalence." Mimeo. Department of Economics, Northwestern University.

[38] Mullahy, J.. "It'll Only Hurt a Second? Microeconomic Determinants of the Demand for Flu Vaccines" forthcoming, *Health Economics.*

[39] New Zealand, Ministry of Transport. *Land Transport Pricing Study: Safety Externalities*, May 1996.

[40] Pauly, M. 1994. *A Study of The Economic Underpinnings of Vaccine Supply.* Department of Human Health and Services, Contract No: 282-92-0044.

[41] Phelps, C. 1992. *Health Economics.* New York: Harper-Collins.

[42] Philipson, T. 1995. "The Welfare Loss of Disease and The Theory of Taxation." *Journal of Health Economics* 14: 386-396.

[43] Philipson, T. 1996. "Private Vaccination and Public Health: An Empirical Examination for US Measles." *Journal of Human Resources* 31(3): 611-630.

[44] Philipson, T., and J. DeSimone. 1997. "Experiments and Subject Sampling." *Biometrika* 84.3: 618-632.

[45] Philipson, T., and L. Hedges. 1998. "Subject Evaluation in Social Experiments." *Econometrica* 66.2: 381-409.

[46] Philipson, T., and R. Posner. 1993. *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective.* Cambridge and London: Harvard University Press.

[47] Philipson, T., and R. Posner. 1995. "A Theoretical and Empirical Investigation of The Effects of Public Health Subsidies for STD Testing." *Quarterly Journal of Economics* 110: 445-74.

[48] Philipson, T., and R. Posner. 1996. "The Economic Epidemiology of Crime." *Journal of Law and Economics* 39(2): 405-436.

[49] Plotkin, S., and E. Mortimer. 1988. *Vaccines.* Philadelphia: Saunders

[50] Shilts, R. 1987. *And The Band Played On.* New York: Random House.

[51] Stiglitz, J. 1978. *The Economics of the Public Sector*, 2nd ed. New York: Norton.

[52] Stiglitz, J. 1997. Introduction in *Confronting AIDS: Public Priorities in a Global Epidemic.* World Bank Policy Research Report. Oxford University Press.

[53] Xu, X. 1998. 'Technological Improvements in Vaccine Efficacy, Individual Incentives to Vaccinate, and Economic Welfare' Working Paper, Department of Economics, University of California, Berkeley.

[54] Weisbrod, B. A. 1961. *Economics of Public Health: Measuring the Economic Impact of Disease.* Philadelphia, PA: University of Pennsylvania Press.

[55] Weisbrod, B., and J., Huston. 1987. 'Benefits and Costs of Human vaccine Programs: Assessing The Literature', Working Paper, Department of Economics, Northwestern University.

[56] World Bank. 1993. *Investing In Health.* World Bank Development Report. London: Oxford University Press.

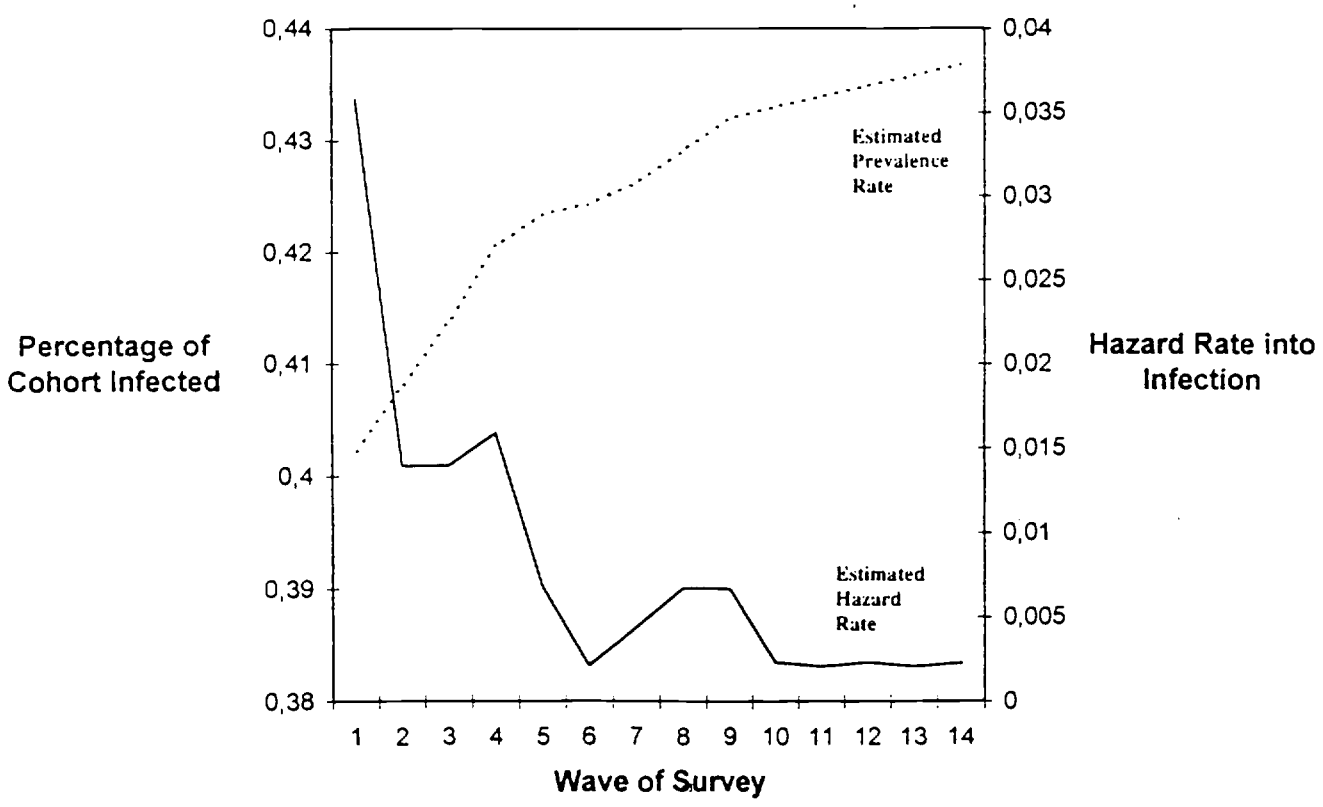

FIGURE 1 Estimated Prevalence and Hazard Rate

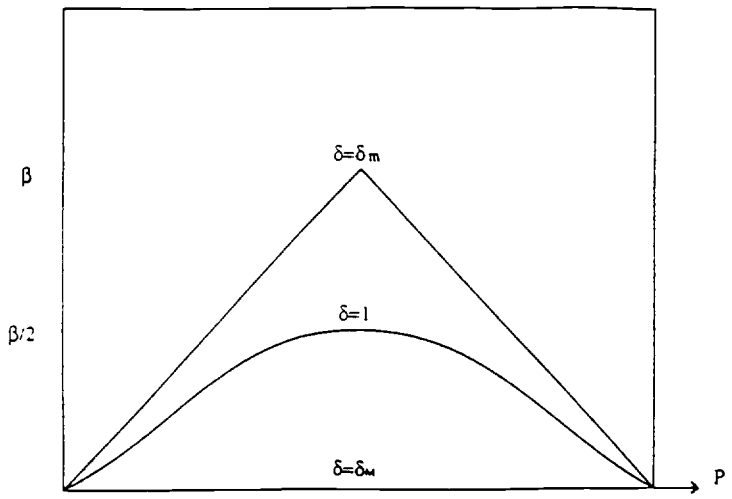

FIGURE 2: Assortive Matching and Incidence Under Same Prevalence

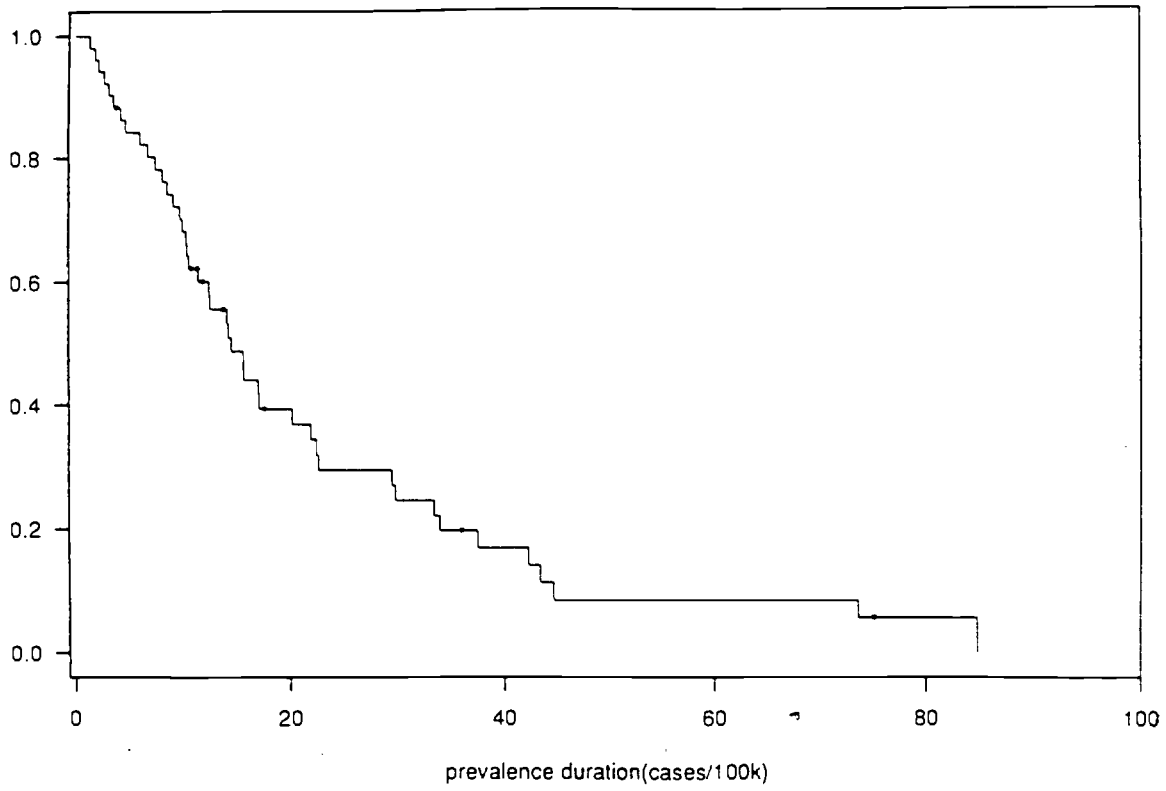

FIGURE 3: Survival in State Government Spending

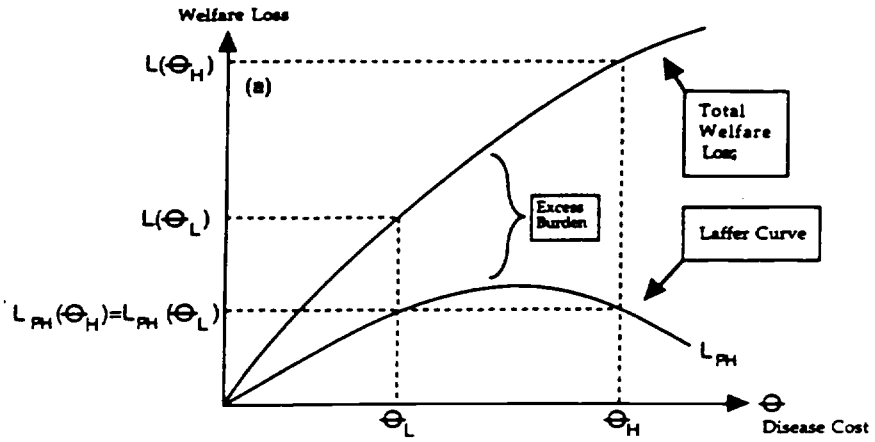

FIGURE 4: Disease Cost and Welfare Losses

FIGURE 5:

Flu Shot Propensity Age Profile
Jagged: Sample Mean by Age
Smooth: Predictions from Quartic Regression

Table 1:*Duration in years before start of main aids prevention program*

Fiscal Year	Total Left	Starters	Survival	[95% Conf.	Interval]
1986/87	51	12	0.7647	0.6229	0.8589
1887/88	39	7	0.6275	0.4801	0.7437
1888/89	32	5	0.5294	0.3847	0.6548
1889/90	27	18	0.1765	0.0869	0.2916
1990/91	9	1	0.1569	0.0734	0.2688
1992/93	8	0	0.1569	0.0734	0.2688

Source: Computed using data from the AIDS Policy Center, Intergovernmental Health Policy Project, George Washington University.

Table 2:
Condom Use (Proportion) by Quartiles of State-of-Residence Prevalence per 100,000 Populaton for 25-27 Year Olds [Data Source: NLSY, Selected Years]

Quartile/ p-values	Men and Women	Men Only	Single Men	Single Men in Urban Areas	Single, Sexually Active Men	Married Men	Married Men in Urban Areas
1984							
First	0.085	0.105	0.106	0.103	0.145	0.104	0.132
Second	0.082	0.108	0.074	0.048	0.085	0.146	0.150
Third	0.075	0.095	0.104	0.082	0.121	0.080	0.051
Fourth	0.098	0.118	0.092	0.106	0.122	0.166	0.165
p-values	0.481	0.676	0.714	0.934	0.667	0.232	0.586
	0.651	0.893	0.787	0.415	0.709	0.317	0.176
1986							
First	0.075	0.067	0.055	0.076	0.078	0.078	0.087
Second	0.080	0.099	0.088	0.103	0.117	0.112	0.095
Third	0.067	0.072	0.078	0.087	0.109	0.064	0.073
Fourth	0.072	0.105	0.085	0.092	0.101	0.136	0.155
p-values	0.823	0.100	0.296	0.632	0.577	0.123	0.126
	0.833	0.217	0.680	0.884	0.806	0.196	0.223
1988							
First	0.103	0.128	0.163	0.178	0.184	0.081	0.077
Second	0.112	0.146	0.158	0.118	0.161	0.127	0.151
Third	0.131	0.174	0.222	0.228	0.256	0.086	0.076
Fourth	0.158	0.191	0.225	0.248	0.228	0.118	0.113
p-values	0.001	0.013	0.067	0.064	0.329	0.310	0.431
	0.004	0.061	0.073	0.002	0.117	0.478	0.229
1990							
First	0.120	0.152	0.213	0.222	0.253	0.082	0.132
Second	0.159	0.214	0.244	0.245	0.268	0.174	0.177
Third	0.178	0.232	0.309	0.334	0.293	0.129	0.121
Fourth	0.194	0.266	0.305	0.325	0.341	0.208	0.220
p-values	0.000	0.000	0.024	0.025	0.103	0.001	0.074
	0.001	0.001	0.058	0.038	0.367	0.009	0.143

Notes: The categories of states represent the quartile of the annual population-weighted dirtribution of the state prevelence-per-100,000-population. For each quartile, the entry in the proportion of observations who used condoms last month. The p-value entries take the form $p1$; $p2$, where $p1$ is the p-value associated with the hypothesis that the incidence of condom use in the first quartile is equal to that in the fourth, and $p2$ is for the test of the null hypothesis of no difference in condomn use across the four quartiles. All results are based on weighted statistics from the NLSY.

Table 3

Average Estimated Prevalence Elasticity for the Probability of Condom Use, Evaluated for Selected Subsamples at Mean Characteristics of 25–27 Year Olds[a]

Population Group	1984		1986		1988		1990	
	All States	High-Prevalence States	All States	High-Prevalence States	All States	High-Prevalence States	All States	High-Prevalence States
All men and women	−0.0306***	−0.1047**	−0.0362***	−0.0991***	0.0592*	0.1605	0.0848*	0.2170*
All men	−0.0401***	−0.1378**	−0.0487***	−0.1324***	0.0784*	0.2048*	0.1099*	0.2674*
White single men	−0.0456***	−0.1563**	−0.0483***	−0.1337***	0.0863*	0.2196*	0.1226**	0.2953*
Black single men	−0.0484**	−0.1384**	−0.0582**	−0.1423**	0.1194**	0.3238*	0.1766**	0.5045**
Hispanic single men	−0.0588**	−0.0908**	−0.0634**	−0.0941**	0.1073**	0.1723*	0.1531*	0.2279*
White single men in urban areas	−0.0518***	−0.1670**	−0.0529***	−0.1459***	0.0982*	0.2422*	0.1422**	0.3203**
Black single men in urban areas	−0.0529**	−0.1421**	−0.0630**	−0.1453**	0.1318**	0.3308*	0.2031**	0.5074**
Hispanic single men in urban areas	−0.0596**	−0.0911**	−0.0650**	−0.0969**	0.1118*	0.1793*	0.1617*	0.2414*
White married men in urban areas	−0.0351**	−0.1333**	−0.0510***	−0.1450***	0.0626	0.1632	0.0955	0.2091
Black married men in urban areas	−0.0333***	−0.1574**	−0.0522**	−0.1550***	0.0978*	0.2563	0.1325*	0.3022*
Hispanic married men in urban areas	−0.0418**	−0.0737**	−0.0576***	−0.0921***	0.0730	0.1080	0.0930	0.1347

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

a. All calculations were done using parameter estimates for Model 4 in Table 4. The formula used for the estimated effect of a 1 percent change in prevalence on the probability of using a condom for the jth population group is:

$$\frac{1}{N_j}\sum_{i \in \text{group}\,j} \frac{\partial P(Y_i = 1 | \hat{\alpha}' X_i)}{\partial \ln \text{Prev}PC_i}$$