

Journal of Public Economics 63 (1997) 383-406



Dynamic epidemiology and the market for vaccinations

Peter J. Francis

Room 10202, Executive Office of the President, Office of Management and Budget, New Executive Office Building, Washington, DC 20503, USA

Received September 1994; final version received September 1995

Abstract

The intent of this paper is to raise the question of how and when externalities can arise in the market for vaccinations. Static and dynamic models of such markets are developed and compared. A special case in a continuous-time dynamic framework is examined and is found to be efficient, i.e. there is no externality. This is in contrast to the results from the static formulation, raising doubts about the suitability of a static approach. These results indicate that the question of whether there are externalities in individual immunization decisions is more involved than is generally supposed.

Keywords: Externalities; Public health; Vaccination

JEL classification: H23; I11; I18

1. Introduction

It has been customary, in normative discussions concerning policy toward public health questions, to suppose that there is an externality associated with individual decisions to get vaccinations for protection from communicable diseases. The usual argument for the presence of an externality goes something like this: individuals will make vaccination decisions solely on the basis of considerations of their own welfare. A decision to get immunized, however, has the potential to enhance the welfare of others. This is because, by immunizing yourself, you also eliminate the possibility that someone else

will catch the illness from you; therefore, you are improving the chances that they will not get sick. Since individuals are presumed to make their choices on the basis of the effects of those choices on their own welfare only, vaccination decisions are therefore being made without the full social impact of the decisions being taken into account by the individual. Thus the potential for an externality and a sub-optimal outcome.^{1,2}

That reasoning has been used to justify various forms of government intervention in the market for vaccinations. The most common forms that such policies take are subsidies and the imposition of mandatory requirements for immunizations. One example of the former would be the recently floated proposal in the United States to make childhood vaccinations universally available at no cost; another example would be the various immunization programs of the World Health Organization.³ Examples of the mandatory-rule approach include requirements placed by some countries on international travellers, and the requirement for measles vaccinations that is imposed by many school districts in the United States.

To be sure, the externality argument is not the only justification that can be offered for government action on immunizations. World Bank (1993) offers several other rationales for government intervention in this area, including poverty alleviation and information problems.⁴ However, the most commonly used argument for government action on vaccinations does appear to center around the externality.

That argument does have an easy intuitive appeal. However, until relatively recently, there has been little formal analysis done to back it up. The purpose of this paper is to add to our understanding of the nature of the externality associated with individual immunization decisions. Broadly speaking, the presentation will consist of two components. The first will be an examination of a static model that is adapted from that given in Brito, Sheshinski and Intriligator (1991) [BSI (1991) hereafter]. The exposition will center on a diagrammatic presentation of the model that will make use of the sort of 'left-right' diagrams used extensively in Schelling (1978).

¹ See Stiglitz (1988, p. 120) for a discussion along those lines; see also Ojo (1991). Weisbrod (1961) contains a general discussion of externalities in health-related individual decisions that captures much of the spirit of this line of reasoning.

² Although it is usually clear what is meant by an externality, there are occasions when the distinction between potential and actual outcomes results in some uncertainty as to what the term implies. To avoid any confusion on this, I will explain how it is used in this paper. Externality, as used herein, refers to a case in which (i) there are possible spillover effects among agents that are unpriced, and (ii) the actual market outcome is suboptimal because of those spillovers. Therefore, by this definition, a situation in which it is possible for external effects to occur, but where the actual outcome is optimal, is not an externality.

³ World Bank (1993) contains a discussion of these.

⁴ See, particularly, pp. 55-57.

Conclusions will be drawn from this model for purposes of comparison with results to be derived in the other part of the paper.

The other component of the paper consists of an analysis in which individual choice is built into an explicitly dynamic epidemic model. A continuous-time framework is used, and individuals can choose to get immunized at any time. A very simple case is examined in which individuals are perfectly homogeneous, infinitely lived, and do not recover when ill. They are rational expected-utility maximizers who are fully informed as to the overall prevalence of the disease in the population. The principal conclusions that come out of the analysis are as follows.

- (1) All individuals will eschew vaccination up to a certain disease transmission threshold, whereupon all remaining uninfected individuals will choose to get vaccinated. (In subsequent discussion this will be referred to as the 'knife-edge' characteristic.) Of particular interest is the fact that the threshold condition can be expressed as an equation in which the proportion of the population that is infected must be at least as large as a function of parameters related to individual and disease characteristics. Thus we get from this case a strong form of what Philipson (1993) calls 'prevalence elasticity', meaning that demand for vaccinations is sensitive to the prevalence of the disease.
- (2) To assess the normative implications of a laissez-faire policy, the unregulated market outcome described above is compared with a 'first-best' outcome derived from solving a planner's problem in which the sum of utilities of the (homogeneous) individuals is the objective function. The time path of vaccination that comes out of this exercise is readily seen to be identical to that of the free-market outcome. Thus, in this particular case there is no externality. Also of note is the fact that this result is not consistent with the conclusions drawn from the static model presented in the previous part: that model finds a divergence between the market outcome and the welfare-maximizing vaccination level, even given homogeneous individuals. (Later in the paper I show how to reconcile this discrepancy.)

The significance of the second result is this: the fact that a mechanism exists for transmitting a disease among individuals does not necessarily mean that there is an externality associated with an individual decision to acquire protection from the disease. This paper is meant to be provocative rather than definitive: the second result should not be interpreted as meaning that there can never be an externality associated with immunization decisions. The example presented here is a far too special case to serve as a basis for any such generalization; moreover, there are strong indications from research in progress that introducing additional complexity into the epidemic model can cause an externality to arise. But these results naturally suggest a direction of inquiry: Under what conditions would there be an external effect associated with vaccinations?

The remainder of the paper is organized along the lines described above:

Section 2 contains the analysis of the static model; Section 3 contains the dynamic analysis; Section 4 shows how to reconcile the differences between the two approaches; and Section 5 offers concluding remarks and highlights some possibilities for further investigation. There are many.

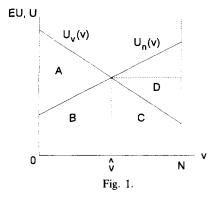
2. Static analysis

What follows is a presentation of a simple static formulation of the market for vaccinations. The proximate objective in presenting this is to provide a simple graphical exposition of analytics that have been, for the most part, devised by others. More specifically, this presentation has been inspired by the analysis contained in BSI (1991) and can be thought of as a simplified version of that. The diagrammatic form that is used here is based on one used extensively in Schelling (1978); again, there is little new.

The larger aim of this part of the paper is to show graphically some of the results that are derived analytically by BSI (1991) or are implied by their approach. These will serve as benchmarks for comparison with outcomes derived in the dynamic analysis which is presented in Section 3. It will then be apparent that omitting intertemporal considerations results in conclusions that, at least in some special cases, do not hold up when explicit dynamics are introduced.

We consider a static population of N individuals who are identical in all respects save one: they have different costs of vaccination, where these costs are measured in utility terms. (The special case where they are identical in that latter respect will be discussed later in the paper.) We assume, furthermore, that the more of them that choose to get vaccinated, the higher the ex ante expected utility of those who choose to do without an immunization. The rationale for this assumption is that the probability of falling ill is presumably smaller if more people are vaccinated. The homogeneity assumption is presumed to extend to susceptibility in the sense that any unvaccinated person has the same probability of falling ill—and therefore the same expected utility—as any other. Lastly, we assume that individuals will get vaccinated in an order corresponding to their vaccination costs, with a lower cost person always getting vaccinated before a higher cost one.⁵

⁵ This assumption is adopted in order to facilitate the diagrammatic identification of an equilibrium outcome. Without such a supposition it is not clear how one would find a schedule relating the number of people vaccinated to the net utility of the last person vaccinated. Also, it would not be obvious how such a market would find its way to a particular equilibrium if it was initially at some other state. The irreversibility of vaccinations means that it is possible to envisage scenarios in which individuals who would not choose to vaccinate if the system was already in equilibrium might wind up being vaccinated if the system was not initially at that state.



The situation is illustrated graphically in Fig. 1. The number of people vaccinated (v) is measured along the horizontal axis, while the vertical axis is in units of utility, or expected utility in the case of unvaccinated individuals. $U_{\rm n}(v)$ is a locus characterizing the expected utility of an unvaccinated individual given that there are v vaccinated individuals. It is upward sloping, in keeping with the assumption stated in the previous paragraph. Note particularly that it is an average schedule in the sense that it gives the expected utility of each unvaccinated person. $U_{\rm v}(v)$ is a schedule giving the actual utility of the vth individual vaccinated inclusive of vaccination cost. Its downward slope comes from the assumption of vaccinations occurring according to cost: a low-cost individual will have a higher net utility than a high-cost one, and since low-cost people are being vaccinated first, the last person vaccinated will always have a lower utility than those who were vaccinated before him.

Equilibrium occurs where the two schedules meet. This is so because individuals will choose to get vaccinated until the actual utility associated with getting vaccinated equals the expected utility of doing without a shot. Beyond that point, they will prefer to take their chances. The equilibrium number of vaccinations is denoted \hat{v} in Fig. 1.

Note that there is an implied assumption here that neither schedule lies wholly above the other. Should that occur, the market outcome would be at one of the boundaries. Either nobody would want to get vaccinated or everyone would, depending on which curve lay above the other. Presumably, such a state of affairs might arise if the utility cost of a vaccination was either very small relative to the expected utility of going unprotected, or very large. For instance, if the cost of vaccination was large relative to the expected utility of doing without, then it is the $U_{\nu}(\nu)$ schedule that might lie entirely below the $U_{\nu}(\nu)$ curve. A concrete example of such a situation would be the fact that yellow fever shots are not commonly bought by

Americans unless they expect to travel to tropical destinations.⁶

It is now possible to begin talking about social welfare. We take aggregate social welfare to be the unweighted sum of individual utilities, or expected utilities in the case of unvaccinated individuals (i.e. a utilitarian social welfare function). Then the welfare function $W(\nu)$ can be written as

$$W(v) = \int_{0}^{v} U_{v}(\tilde{\omega})d\tilde{\omega} + U_{n}(v)(N-v).$$
 (1)

If $v = \hat{v}$, then the two terms in Eq. (1) each correspond to areas in the graph depicted in Fig. 1. The integral term, which represents the sum of actual utilities of vaccinated individuals, is given by areas A and B. The $U_n(v)(N-v)$ term, which captures the expected utility of unvaccinated individuals, corresponds to areas C and D.

It is now possible to illustrate one of the results from BSI (1991). They found that forcing everyone to get immunized is inferior to the market equilibrium. This is easily seen in Fig. 1. From the previous paragraph, social welfare in the case of the market outcome is the sum of areas A, B, C and D. In the case of a mandatory vaccination scenario, v = N, and the areas that depict social welfare in Fig. 1 are just A, B and C. (Since there are no unvaccinated individuals in such a case, total welfare is just the area below the utility-if-vaccinated curve.) Area D therefore represents the amount of aggregate welfare by which the 100% vaccination scenario is inferior to the market outcome. Note, though, that if $U_v(v)$ is constant over the range $\hat{v} \le v \le N$, then area D vanishes and the market outcome has the same overall welfare as the mandatory-rule case. Thus, in the case where individuals are entirely homogeneous, the mandatory approach results in social welfare being the same as it would be if the market were left to itself (though still no better).

Neither the market outcome nor the mandatory outcome is socially optimal. The social optimum is found in the normal way: differentiate W(v) in Eq. (1) with respect to v and set the result equal to zero. This results in the following first order condition:

$$\frac{dW(v)}{dv} = 0 = U_{v}(v) + \frac{dU_{n}(v)}{dv}(N - v) - U_{n}(v).$$
 (2)

⁶ The graphical approach under consideration here does not directly cater for the possibility that different individuals have different likelihoods of falling ill. We are assuming explicitly that all unvaccinated persons have the same expected utility; thus, a situation where there are 'high-risk' and 'low-risk' groups—as with HIV, for example—is not readily accommodated within this framework. Some work has been done with such situations; see Lloyd (1991) and Bethwaite and Bethwaite (1991). This would be one of many possible extensions to both the static framework described in this section and the dynamic analysis in the following section.

This can be written:

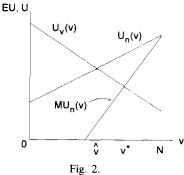
$$U_{\nu}(\nu) = U_{n}(\nu) - \frac{dU_{n}(\nu)}{d\nu}(N - \nu).$$
 (3)

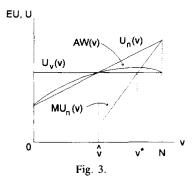
The right-hand side of Eq. (3) can be thought of as a marginal curve corresponding to $U_n(v)$, which is an average one. It is denoted $MU_n(v)$ in Fig. 2. Note that since $dU_n(v)/dv$ is positive by assumption, and since it must be the case that $v \le N$, $MU_n(v)$ lies everywhere below $U_n(v)$ except at v = N, where they are equal. The social optimum, denoted v in Fig. 2, occurs at the intersection of $MU_n(v)$ and $U_v(v)$ and lies to the right of \hat{v} . Thus, the equilibrium level of vaccination is less than the socially optimal one. This outcome is the result of the externality: individuals are not taking account of the impact of their vaccination decisions on the expected utility of other unvaccinated people. Note too that this conclusion—the suboptimality of the market outcome-will hold true even if individuals are homogeneous in all respects, including vaccination costs. This is illustrated in Fig. 3. Also shown in Fig. 3 is the curve denoting average social welfare, AW(v) = W(v)/N.

To recap, the conclusions that arise from this formulation of the vaccination problem are that:

- (1) Given a utilitarian social welfare function, the market outcome is superior to an across-the-board vaccination requirement except in a special case involving homogeneous individuals; in that case the two outcomes are equivalent from a social welfare perspective.
- (2) The socially optimal level of vaccinations is greater than the equilibrium level, even if individuals are homogeneous. That is to say, the market outcome is not optimal even in the homogeneous case.

We close this section with two observations. The first is that this kind of analysis is applicable to, and has been applied to, many types of problem. Schelling (1978) contains a chapter in which numerous cases of externalities





arising out of binary choices are examined, including a traffic congestion example that is formally quite similar to the foregoing.

The second point that needs stating—again—is that this model does not allow for intertemporal changes. In reality, infectious disease transmission is a dynamic process, and the incidence of a disease can generally be expected to vary over time. Since such diseases are transmitted by already-infected individuals, and since the number of individuals who are infected can vary over time, the risk faced by an unprotected person will in general vary over time too. In terms of the above presentation, this is tantamount to saying that the $U_n(v)$ curve can shift as time goes by. The purpose of the next section is to introduce such a consideration into the analysis.

3. Dynamic analysis

This section is devoted to a simple model of the market for vaccinations which incorporates explicit epidemic dynamics. Disease transmission is modeled as a deterministic phenomenon at the aggregate level; however, the individual's problem is still stochastic in the sense that, although the overall rate of increase in infection is assumed to obey a deterministic law of motion, each healthy unvaccinated individual is assumed to have an equal chance of falling ill at any given time. A continuous-time framework is used, and individuals can choose to purchase an immunization at any time.

The organization of this part of the paper is as follows. First, the individual's optimal-time-of-vaccination problem is analyzed. Next, the market equilibrium time path is derived; the epidemic laws of motion are discussed at this stage. After that, the socially optimal time path of immunization is found and compared with the unregulated market path. Finally, a comparison is drawn between the results of this section and those of Section 2.

3.1. The individual's problem

There are N individuals who are identical, have infinite lifetimes, and do not recover if they become ill. (The homogeneity of individuals means that this model corresponds to Fig. 3 in the previous section, except for the intertemporal character that is introduced here.) They experience utility at a rate \overline{u} when healthy and \underline{u} when sick $(\overline{u} > \underline{u})$. Vaccinations are only given to healthy unvaccinated individuals and confer permanent immunity. Vaccinated individuals receive the same utility as healthy unvaccinated individuals. Each individual has a discount rate δ and a cost of vaccination θ , the latter being measured in utility terms. (Both θ and δ are positive. Also, note that θ is assumed to be constant regardless of how many people are getting vaccinated at the time.) The present value of an individual's utility stream, W_i , assuming she falls sick at exactly time t, is

$$W_{i} = \int_{0}^{t} \overline{u} e^{-\delta \tau} d\tau + \int_{t}^{\infty} \underline{u} e^{-\delta \tau} d\tau$$
$$= \frac{1}{\delta} \left[\overline{u} - (\overline{u} - \underline{u}) e^{-\delta t} \right]. \tag{4}$$

We now assume that the individual believes he has a differentiable cumulative probability distribution of falling ill at or before time t; we denote this as P(t) and its time derivative as $\dot{P}(t)$. We suppose, furthermore, that the individual plans to get vaccinated at time T. Then the expected present value of that person's utility (before costs) is given by

$$E[W_i|T] = \int_0^T \dot{P}(t) \frac{1}{\delta} \left[\overline{u} - (\overline{u} - \underline{u}) e^{-\delta t} \right] dt + [1 - P(T)] \left(\frac{\overline{u}}{\delta} \right).$$
 (5)

The integral term in Eq. (5) deals with the outcome in which the individual gets sick before getting vaccinated. This happens with probability P(T). The second term represents the more favorable outcome in which the individual stays healthy until T; this happens with probability 1 - P(T).

The foregoing does not include vaccination costs. If the person remains healthy until T, she will incur a cost the present value of which is $e^{-\delta T}\theta$. Note, though, that this only occurs with probability 1 - P(T). Therefore, the individual's discounted expected utility inclusive of cost W_i^{θ} is

$$E[W_i^{\theta}|T] = \int_0^T \dot{P}(t) \frac{1}{\delta} [\overline{u} - (\overline{u} - \underline{u})e^{-\delta t}] dt + [1 - P(T)] \left(\frac{\overline{u}}{\delta} - e^{-\delta T}\theta\right).$$
 (6)

The appropriate problem for the individual is therefore

$$\max_{T} E[W_{i}^{\theta}|T] = \int_{0}^{T} \dot{P}(t) \frac{1}{\delta} \left[\overline{u} - (\overline{u} - \underline{u}) e^{-\delta t} \right] dt + \left[1 - P(T) \right] \left(\frac{\overline{u}}{\delta} - e^{-\delta T} \theta \right).$$
 (7)

The appropriate first-order condition is

$$0 = \dot{P}(T) \frac{1}{\delta} \left[\overline{u} - (\overline{u} - \underline{u}) e^{-\delta T} \right]$$

+
$$\frac{1}{\delta} \left[\delta^2 e^{-\delta T} \theta - \dot{P}(T) \overline{u} - \delta^2 e^{-\delta T} \theta P(T) \right] + e^{-\delta T} \theta \dot{P}(T),$$

which reduces to

$$\dot{P}(T)[(\overline{u}-\underline{u})-\theta\delta]=\theta[1-P(T)]\delta^{2}. \tag{8}$$

Eq. (8) is an expression involving T and known constants and functions, and so implies the optimal time of vaccination for an individual. It is conditional on knowledge of the probability distribution (over time) of falling ill. Of particular interest is the following rearrangement of Eq. (8), which gives an individual's threshold condition for getting vaccinated:

$$\frac{\dot{P}(T)}{1 - P(T)} \ge \frac{\theta \delta^2}{(\bar{u} - \underline{u}) - \theta \delta} \,. \tag{9}$$

Eq. (9) is critical to the analysis and bears some discussion.

The left-hand side of Eq. (9) is the hazard function associated with the individual's beliefs concerning P(T). Its interpretation is that it is the probability of a transition at exactly time t conditional on not having undergone the transition until that time.⁷ The transition in this case is, of course, falling ill, so the implication of Eq. (9) is that there is a threshold hazard level past which the individual prefers to get immunized.

The weak inequality is used in order to accommodate the possibility that the hazard may be above the individual's threshold right at the start. If Eq. (9) holds true with equality, then the individual will be indifferent between getting vaccinated and not doing so; if the left-hand side of Eq. (9)—the hazard—is strictly larger than the right-hand side, then the individual will purchase a vaccination. If the hazard is initially below the threshold and rises up to it, then the individual will get vaccinated when the threshold is reached. In this paper I will mainly restrict attention to the latter case, which is implied by condition (ii) below.

P(t) is increasing in t, so the denominator portion of the hazard expression will certainly be decreasing with time. $\dot{P}(t)$ can in principle either be rising

⁷ See Kiefer (1988) for a discussion of this.

or falling with t, and the same is in general true of a hazard function. However, for present purposes it will be convenient to assume that the hazard will be non-decreasing. This assumption is consistent with the epidemic dynamics which are developed below, and it will guarantee that the threshold rule will be obeyed: when the critical hazard is reached, agents will vaccinate. Delay beyond that point is obviously not a desirable choice since the situation can only get worse. However, it should be mentioned that this optimal-time-of-vaccination rule may not be valid in cases where the hazard rises and then falls. Such a case will not be considered in this paper.

Eq. (9) can be thought of as a *myopic* condition in the sense that it makes no direct reference to the subsequent evolution of the hazard. Its interpretation is that it represents a comparison of two alternatives: get vaccinated this instant, or wait one instant more before doing so. When it no longer makes sense to wait, the individual acts.

3.2. Equilibrium system

The equilibrium system is based on the individual's optimal time of vaccination rule (Eq. (9) above). What are also required are initial conditions and epidemic laws of motion. Taking the latter first, these are:

$$\dot{I} = \beta \frac{S(t)I(t)}{N} \tag{10}$$

and

$$\dot{S} = -\beta \frac{S(t)I(t)}{N} - \dot{V} , \qquad (11)$$

where S(t), I(t), and V(t) represent the number of susceptible (unvaccinated and uninfected), infected, and vaccinated individuals respectively. (An overdot denotes time derivative.) The explanation for these forms is reasonably straightforward. β , the infectiousness parameter, can be interpreted as a rate of contact among individuals ($\beta > 0$). If I(t)/N of the population is infected at time t, then the chance of an encounter with an infective during a contact at that time is just that fraction. Multiplying the contact rate, β , by the likelihood of encountering an infective, I(t)/N, and multiplying this in turn by the total number of people who could fall ill if they encounter an infective (i.e. S(t)) gives the total rate at which new infections are occurring. This is the basis for Eq. (10). The basis for Eq. (11) is simply that there are only two transitions that a susceptible can experience

in this formulation: either he falls ill or he gets vaccinated. The first term in Eq. (11) covers the former case, the second term the latter.⁸

Unfortunately, the model described above is not very realistic. For it to be a reasonable depiction of an actual situation, the disease in question would have to be one which (a) is incurable, and for which an infected person remains infectious always; (b) is not fatal; (c) has a vaccine available; and (d) spreads sufficiently rapidly that births and deaths can be ignored. If a vaccine for herpes becomes available, then that disease might conceivably be a candidate (for some populations); otherwise, additional complexity will have to be added to the above epidemic model in order for it to be a useful approximation of an actual disease. Nonetheless, the analysis is being developed using this special case because of its tractability and because it can give useful insights into the nature of the interaction between individuals without becoming cluttered with other mechanisms.

Epidemic models of this kind have appeared before in the economics literature. For example, Stoneman (1983) used such a form to model the diffusion of technological innovation among firms, while Shiller (1984) used epidemic forms to model the spread of information through a population of asset-market traders. It is a curious fact, therefore, that dynamic epidemic models have, until quite recently, almost never been used by economists to model epidemics!⁹

Concerning initial conditions, what follows will be based on the following three assumptions:

- (i) I(0) > 0. Nothing interesting will happen if there are no infectives in the population.
- (ii) I(0) is small. It will become clear below that, if I(0) is sufficiently large, then the equilibrium time path may entail the instantaneous vaccination of all susceptibles at t=0. Consideration will usually be restricted to cases where that does not occur, since otherwise nothing interesting will happen after t=0.
- (iii) V(0) is small. If there is a sufficiently large vaccinated population at t = 0, then it may be the case that none of the remaining susceptibles will ever wish to get vaccinated.

Apart from these three assumptions, I will pay little attention to initial conditions. It will soon become apparent that an individual's rule for when to get vaccinated can be succinctly stated in a way that does not directly

⁸ See, for example, Bailey (1975) or Anderson and May (1991) for a more elaborate discussion of the rationale for these functional forms. Note that there are many possible enhancements to such a model, including recovery with immunity and population turnover through births and deaths.

⁹ Recent articles in the economics literature that use epidemic models to address disease transmission include Wiemer (1987), Lloyd (1991), Geoffard and Philipson (1993), and Philipson (1993).

require specification of the initial conditions; the same is true for the solution to the social planner's problem. Although they affect the time path of the state variables of the system, the initial conditions do not affect decision rules that are framed in terms of those states.

To tie individual decisions to market outcomes, it is necessary to relate Eq. (9) to Eqs. (10) and (11). Eq. (9) is based on individual beliefs. By imposing an assumption of perfect (and costless) information, we can connect the individual's beliefs to the actual state of the world. To do this, we first recall the interpretation of a hazard function. As stated earlier, it represents the probability of a transition occurring at t given that it has not occurred until then. The transition in this case is falling ill, and we can derive the actual hazard faced by an individual directly from Eq. (10). This is done by simply noting that at time t the number of individuals who will fall ill is dI. Since they are coming out of a pool of size S(t), and since all susceptibles are assumed to be equally at risk, the true hazard, $\lambda(t)$ is simply

$$\lambda(t) = \frac{\dot{I}}{S(t)} = \beta \frac{I(t)}{N} \,. \tag{12}$$

The individual believes that her hazard at time t is given by $\dot{P}(t)/(1-P(t))$. By equating the individual's beliefs to the true hazard (specifically, for time T),

$$\lambda(T) = \frac{\dot{P}(T)}{1 - P(T)},\tag{13}$$

we get

$$\beta \frac{I(T)}{N} \ge \frac{\theta \delta^2}{(\bar{u} - u) - \theta \delta} \,. \tag{14}$$

Thus an individual decision as to when to get a shot depends directly on the fraction of the population that is infected.¹⁰

Before discussing how the equilibrium system will behave, I briefly discuss some of the limiting possibilities that can be inferred from Eq. (14). First, note that if either the discount parameter δ or the vaccination cost parameter θ is zero, then an individual will prefer to get vaccinated provided there are infectives present in the population, regardless of how small a proportion of the total population they represent. These cases are readily explained. Concerning the discount parameter, $\delta = 0$ implies that future utility is given the same weight as present utility (i.e. no discounting). Since these are infinitely lived agents, this means that the loss associated with falling ill has a present value that is unbounded. Provided there is a

¹⁰ This is consistent with results presented in Geoffard and Philipson (1995).

non-zero probability of falling ill, then the expected loss associated with going unvaccinated is negative infinity. Provided the cost of vaccination is finite, then the present value of getting vaccinated will be larger than the present value of going unvaccinated. Thus, the individual prefers vaccination. Concerning the cost parameter, setting it equal to zero means that vaccination is costless, so that if there is a non-zero probability of incurring a non-zero loss, then it would be preferable to take a costless vaccination to prevent that.

Next, notice that the left-hand side of Eq. (14) can never exceed β . Thus, if the value of the right-hand side of Eq. (14) is larger than β , then the individual will never wish to get vaccinated, even if he is the only uninfected person in the population. Lastly, note that the numerator on the RHS of Eq. (14) is always non-negative but that the denominator would be less than zero if $\overline{u} - \underline{u} < \delta \theta$. If that situation obtained, then the threshold, I(t), implied by Eq. (14) would be negative, which would not make sense. However, there is a sensible interpretation of that situation: it means that an individual would never choose to purchase a vaccination regardless of the hazard. This stems from the fact that the potential loss from falling ill, measured as a flow, has to exceed the flow-equivalent utility cost of the vaccination in order for it to make any sense to get vaccinated.

We are now in a position to see that this system will have a 'knife-edge' character to it in the sense that, once the threshold level of infection in the population has been reached, all remaining susceptibles will immediately get vaccinated. This is apparent from an examination of Eqs. (10) and (14). From Eq. (10) it is clear that I(t) can never decrease: since everything on the right-hand side of that equation is non-negative, $l \ge 0$. Initially, let us assume that I(t) is below the threshold level dictated by Eq. (14). Therefore, no vaccinations occur and the second term on the right-hand side of Eq. (11) is zero. Once the critical prevalence has been reached, individuals will start to want to get vaccinated. However, their doing so will not cause the left-hand side of Eq. (14)— $\beta(I(t)/N)$ —to drop below its threshold level. It is therefore apparent that, once the threshold has been reached, the fact that some agents go and get their shots will not remove the incentive for all of the other remaining susceptibles to do the same; thus, they will all go at once. (Note: A more formal proof that the time path implied by Eq. (14) is a unique equilibrium path appears in the appendix.)

This behavior is not really all that surprising. The risk of infection rises as the proportion of the population that is infectious grows. The vaccination of susceptibles will slow down the rate at which prevalence increases, but the instantaneous risk faced by any particular susceptible will not decline unless the number of infectives declines. Given the structure of this model, which does not permit the number of infectives to diminish, it should not be surprising that the conditions under which an individual wants to get

vaccinated, once attained, will apply to everybody and will not be reversed by the subsequent actions of others.

The threshold rule for vaccination—Eq. (14)—contains no explicit mention of initial conditions. However, this does not mean that the initial conditions have no effect on the system. The greater is the initial number of infectives, all else being equal, the sooner will the threshold prevalence level be reached. Likewise, a higher initial level of vaccinated individuals, i.e. V(0), will, ceteris paribus, result in a slower rate of growth of infections, and therefore a longer time before the threshold is reached. (The latter point stems from the fact that, for a given I(0) and N, a greater V(0) implies a smaller S(0) and thus, from Eq. (10), a smaller rate of change in I(t).) The point here is that although varying initial conditions does not affect the individual's optimal time-of-vaccination rule when that rule is stated in terms of the state variables of the system, the time implied by that rule is, in general, affected by the initial conditions.

We have seen how the individual agents, when left to their own devices, will all get vaccinated at the same time, in accordance with the condition stated in Eq. (14). It is natural at this stage to wonder about the normative implications of this result: in other words, is it optimal? In the next subsection I develop the social planner's problem which corresponds to the competitive equilibrium analyzed above, and compare that first-best outcome with the outcome obtained in this subsection.

3.3. The social planner's problem¹¹

The population is assumed to be exactly as described above. The planner's problem is to maximize the present value of the unweighted sum of utilities of all people in the population, i.e. to maximize a utilitarian social welfare function. As above, per-person vaccination costs are assumed to be constant regardless of the rate (r) at which they are given. However, in the initial set-up here I will assume that there is a maximum possible vaccination rate, r_{max} . The purpose in doing so will be to show clearly the 'bang-bang' character of the necessary conditions; the assumption of a maximum possible rate will then be dropped. The problem can be stated as follows:

$$\max_{r} \int_{0}^{\infty} e^{-\delta t} [(N - I)\overline{u} + I\underline{u} - \theta r] dt$$
 (15)

subject to

¹¹ Problems of this sort have received some attention in the biometrics literature. See Wickwire (1977) for a survey.

$$\dot{V} = r \,, \tag{16a}$$

$$\dot{I} = \beta \frac{I(N - I - V)}{N} \,, \tag{16b}$$

$$N \ge I + V \tag{16c}$$

$$r_{\max} \ge r \ge 0 \,, \tag{16d}$$

$$I(0), V(0)$$
 given. (16e)

Note that, in contrast to the previous section, the substitution S = N - I - V has been made at the start. Also note that the '(t)' notation has been dropped except in cases where omitting it could cause confusion.

The current-value Hamiltonian for this problem is as follows:

$$\widetilde{H} = (N - I)\overline{u} + I\underline{u} - \theta r + \mu_1 r + \mu_2 \left(\beta \frac{(N - I - V)I}{N}\right) + \mu_3 (N - I - V) + \lambda_1 r + \lambda_2 (r_{\text{max}} - r).$$
(17)

By the maximum principle, the necessary conditions for an optimal path are

$$\frac{\partial \tilde{H}}{\partial r} = 0 = -\theta + \mu_1 + \lambda_1 - \lambda_2 , \qquad (18a)$$

$$-\frac{\partial \vec{H}}{\partial V} = \dot{\mu}_1 - \delta \mu_1 = -\left(-\mu_2 \beta \frac{I}{N} - \mu_3\right), \tag{18b}$$

$$-\frac{\partial \tilde{H}}{\partial I} = \dot{\mu}_2 - \delta \mu_2 = -\left(\underline{u} - \overline{u} + \mu_2 \beta \frac{N - 2I - V}{N} - \mu_3\right), \tag{18c}$$

$$N - I - V \ge 0; \quad \mu_3 \ge 0; \quad \mu_3(N - I - V) = 0,$$
 (18d)

$$r \ge 0;$$
 $\lambda_1 \ge 0;$ $\lambda_1 r = 0;$ $r_{\text{max}} - r \ge 0;$ $\lambda_2 \ge 0;$ $\lambda_2(r_{\text{max}} - r)$
= 0, (18e)

and Eqs. (16a) and (16b).

Eqs. (18a)-(18c) can be written as

$$\theta = \mu_1 + \lambda_1 - \lambda_2 \,, \tag{19a}$$

$$\dot{\mu}_1 = \delta \mu_1 + \mu_2 \beta \frac{I}{N} + \mu_3 , \qquad (19b)$$

$$\dot{\mu}\left\{2\right\} = \delta\mu_2 - \mu_2\beta\left(\frac{N-2I-V}{N}\right) + \mu_3 + (\overline{u} - \underline{u}). \tag{19c}$$

This system of equations is rather opaque at first glance, but some insight is possible. Let us assume that, initially, μ_1 is sufficiently small so that $\lambda_1 > 0$. Then no vaccinations occur and they will not commence until $\mu_1 = \theta$.

The key question is: Is it possible for the optimal path to contain an r that is not equal to either zero or the upper bound over a finite non-zero time interval? For that to occur, $\lambda_1 = \lambda_2 = 0$ and it is therefore the case that $\mu_1 = \theta$ over the whole interval. With μ_1 constant, $\dot{\mu}_1$ and its time derivative, $\ddot{\mu}_1$ must equal zero, as must μ_3 . ($\mu_3 = 0$ because, over such an interval, I + V < N.) It will be shown, by differentiating Eq. (19b) and substituting, that the only way this can be true is if I(t) is at a particular value; in fact, the same one implied by Eq. (14). Unless all remaining susceptibles are vaccinated at exactly that time, I(t) will continue to increase and the maximum r will be invoked.

From Eq. (19b), setting $\dot{\mu}_1 = 0$ with $\mu_1 = \theta$ implies that

$$\mu_2 = -\frac{\delta\theta N}{\beta I} \,. \tag{20}$$

Also, $\ddot{\mu}_1 = 0$ means that

$$\dot{\mu}_2 I = -\mu_2 \dot{I} \tag{21}$$

or

$$\dot{\mu}_2 = \frac{-\mu_2 \dot{I}}{I} \,. \tag{22}$$

Next, we use Eq. (22) to substitute for $\dot{\mu}_2$ in Eq. (19c), and then substitute Eqs. (20) and (16b) into (19c). After simplifying, we are left with

$$\beta \frac{I}{N} = \frac{\theta \delta^2}{(\bar{u} - \underline{u}) - \theta \delta} \,, \tag{23}$$

which is identical to Eq. (14).

We now remove the constraint that $r_{\rm max}-r=0$. This system then takes on the 'knife-edge' character of the competitive outcome, and Eq. (23) suggests that the threshold condition in this case will be the same as in the competitive case. As a check on this, the social planner's problem can be restated as an optimal time-of-vaccination problem analogous to the individual's problem above:

$$\max_{T} \int_{0}^{T} e^{-\delta T} [(N-I)\overline{u} + I\underline{u}] dt + I(T) \int_{T}^{\infty} \underline{u} e^{-\delta t} dt + (N-I(T))$$

$$\times \int_{T}^{\infty} \overline{u} e^{-\delta t} dt - (N-I(T) - V(0))\theta e^{-\delta T}.$$
(24)

The idea is that at T all remaining susceptibles are vaccinated at a cost of θ per person; the problem is then to find the T that maximizes social welfare.

The method of solution is the same as for the individual's problem, and it does indeed turn out that the necessary condition that comes out of this problem is identical to Eq. (14). Thus, in this case the competitive equilibrium time path is identical to the socially optimal path, and there is no externality.

3.4. Summary

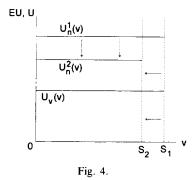
We have seen that, given homogeneous individuals, there is no divergence between the competitive equilibrium time path of vaccinations and the socially optimal one. This stands in marked contrast to the results derived in Section 2, where it was seen that there is a divergence between the market outcome and the first-best outcome, even in the case of homogeneous individuals. In the next section it will be argued that the reason that this discrepancy arises is that the static model is just that—i.e. static—and that it is necessary to take changes over time into account. Once that is done, it is possible to reconcile these approaches.

4. Reconciling the two approaches

The approach presented in Section 2 overlooks the fact that infectious disease is transmitted from person to person, and that this process occurs over time. Any one person's risk of infection is directly related to the number of infectious people in the population. That number will vary over time, so the expected utility of going unvaccinated will vary over time. At any given time, the instantaneous probability of falling ill does not depend on the number of people who are vaccinated; it depends on the number of infectives. The rate at which the overall number of infections increases is affected by the number of people vaccinated, so in that sense vaccinations do matter, but at any given time what matters to an individual is his own risk of infection. This is determined by the prevalence of the disease, not the prevalence of vaccination.

It is possible to bring these considerations to bear in the graphical analysis of Section 2. By doing that we can reconcile that approach with the results developed in Section 3. Fig. 4 shows the state of affairs at two times, t_1 and t_2 . The horizontal axis now gives the number of people who choose to get

¹² The dynamic model used in this paper is based on the assumption that disease is transmitted by direct contact between individuals. Epidemic models that deal with vector-borne diseases do exist, but are not considered here. (A vector-borne disease is one that is transmitted between human beings by another biological agent such as a mosquito.) Wiemer (1987) is one example. For a general discussion of these, see Bailey (1975).



vaccinated at a given time. The maximum number of people that can get vaccinated at any point in time is the number of susceptibles; this number, too, will vary over time. The expected utility of any given susceptible is the same as any other at any given time; thus, $U_n^t(v)$ is horizontal for each t. (The t superscript refers to the time.) This will be true regardless of how many persons get immunized at that instant. $U_v(v)$ is horizontal because of the assumed homogeneity of the population; it is also assumed to be invariant over time.

It can now be seen that this formulation of the graphical presentation will give the same qualitative results as in Section 3. As time progresses, $U_n^t(v)$ will shift downwards. This is because the number of infectives, and hence the risk of infection, is increasing. The number of susceptibles will be decreasing as well. When $U_n^t(v)$ hits the $U_v(v)$ schedule, all susceptibles will be exactly indifferent between getting vaccinated and not doing so; when the $U_n^t(v)$ line moves ε below the $U_v(v)$ schedule, all will get vaccinated. Thus, we have the 'knife-edge' behavior of Section 3. Furthermore, the optimal character of this time path can be seen by noting that, with $U_n^t(v)$ horizontal, the marginal schedule associated with it will coincide with it. Thus, the marginal social benefit of getting vaccinated equals with the marginal benefit of doing without a vaccination when $U_n^t(v)$ meets $U_v(v)$, and that is the point at which people get vaccinated.

5. Concluding remarks

The main conclusions of this paper are as follow.

- (1) There exists at least one special case in which there is no externality associated with individual vaccination decisions.
- (2) A simple static formulation of the market for vaccinations, such as that of Section 2, gives results that are different from those that arise when

explicit dynamics are considered. However, these differences can be reconciled by considering the static schedules of Section 2 as things that can, and do, shift over time.

In closing, I wish to emphasize that the analysis presented here represents the beginnings of what promises to be a rich line of inquiry. There are many directions in which there is a possibility of fruitful investigation. A nonexhaustive list follows:

- (1) Augmenting the epidemic model. There are numerous possible enhancements here. For example, we could introduce recovery with and without immunity, population turnover through births and deaths,¹³ and vector-borne diseases.
- (2) Variable vaccination costs. In this paper vaccination costs were always assumed to be invariant with respect to the rate at which they were being purchased. It would be quite reasonable to suppose that these costs might in fact vary directly with that rate.
- (3) Imperfect vaccines. In this paper vaccines were always assumed to be perfectly effective, instantly effective, and permanent. It might be reasonable to relax any or all of those assumptions under some circumstances. For instance, some vaccinations require periodic boosters, while others have been known to be less than 100% reliable.
- (4) Imperfect information. Although this is a difficult issue to address, it seems clear that an assumption of perfect information is rather strong. To see this, we need only reflect on the fact that there are many communicable-disease situations in which there is a latency period during which the infective himself is not aware that he has the disease. If individuals are not always aware that they themselves are infectious, then it must surely be debatable as to whether anyone can know exactly how many infectives there are in a population at any instant.
- (5) Non-competitive markets. This work has been based in part on implicit assumptions that the market price of a vaccination, and the market interest rate too, were set in competitive markets so that they reflected the true social costs associated with them. There are many situations where it might be reasonable to question those assumptions, such as situations where there are private or government monopolies at work, or market distortions of other kinds. This might be of particular relevance in developing countries.
- (6) Spatial heterogeneity. There might be various kinds of epidemic heterogeneity. Epidemic models that have a spatial character could be used to look at situations in which the contact rates between individuals are not uniform. (Bailey, 1975, and Anderson and May, 1991, are good references on these models.) This type of variability could arise for reasons of

¹³ Some work has been done on models involving recovery with immunity and births and deaths. See, for example, Francis (1995) and Philipson (1993).

geography, in which case the heterogeneity is literally spatial; or it could arise for social or economic reasons. Sexually transmitted diseases are often modeled as being centered in very active sub-populations. (See, for instance, Lloyd, 1991.) Also, a society that is segregated according to socioeconomic status or race might be treated as spatially heterogeneous.

- (7) Heterogeneity in susceptibility. Another form of epidemic heterogeneity is variable susceptibility. The idea here is that some individuals may be less resistant to a disease, and therefore more likely to become infected and spread it to others. Care should be taken to distinguish between this sort of heterogeneity and the 'economic' sort which, say, varies over the population.
- (8) Behavioral variability. Individuals might alter their behavior as a result of their immune status or the availability of insurance or subsidized health care.
- (9) Stochastic epidemic models. It is usually the case that a contact between a susceptible and an infected person would result in infection of the susceptible less than 100% of the time. This uncertainty means that the rate of increase of infections is more realistically modeled as a random variable. Epidemic models of this kind are examined in detail in Bailey (1975) and Anderson and May (1991); see also Wickwire (1977).

Acknowledgements

This paper was initially written at Cornell University, and is related to my dissertation (Francis, 1995). It does not necessarily reflect the views of the Office of Management and Budget. Many people have provided useful insights, including Kaushik Basu, Valerie Bencivenga, Larry Blume, Carlos Castillo-Chavez, Jon Conrad, Gary Fields, Dan Hosken, Nick Kiefer, Rob Masson, Tapan Mitra, Amanda Newton, Uri Possen, Bruce Smith, seminar participants at Cornell and Bates College, and two anonymous referees. Any remaining errors are my responsibility.

Appendix

What follows is a proof that the 'knife-edge' time path of vaccination found in Section 3 is in fact a unique equilibrium path in the sense used by Lucas, Romer, and others. In this context, this equilibrium concept can be characterized as follows: take a path of vaccinations, V(t), to be given. Given this path, we consider the problem each individual would solve if she expected that V(t) would follow that path. When the path of vaccinations

that arises from the individuals' (expected) utility-maximizing behavior coincides with the assumed V(t), then V(t) is an equilibrium path.

In this case, that general statement can be boiled down to something much simpler. Individuals have only one decision to make in this environment: when to get vaccinated. Since they are the same, they will all do so at the same time. (Excepting, of course, those that have already fallen ill, since a vaccination will not do them any good.) This means that the time path of vaccinations, V(t), can be succinctly characterized as follows:

$$V(t) = V(0), \quad \forall t < \hat{T} , \tag{A1a}$$

$$V(t) = N - I(t), \quad \forall t \ge \hat{T},$$
 (A1b)

$$= N - I(\hat{T}), \tag{A1c}$$

where V(0) is the number of vaccinated individuals at t = 0 and \hat{T} is the time at which all remaining susceptibles get vaccinated. N is, by assumption, constant, and N = S(t) + I(t) + V(t). Eq. (10) in the main body of the paper is reproduced here as Eq. (A2):

$$\dot{I} = \beta \frac{S(t)I(t)}{N} \,. \tag{A2}$$

(Note: Throughout this appendix assumptions (i)-(iii) of Section 3 will be maintained.)

Next, note the following implications of Eq. (A2):

$$\dot{I} \ge 0, \quad \forall t \ge 0,$$
 (A3)

$$\dot{I} = 0, \quad \forall t \geqslant \hat{T}$$
 (A4)

$$\dot{I} > 0$$
, when $S(t) > 0$. (A5)

Eq. (A3) follows from the fact that β and N are positive and S(t) and I(t) are non-negative. Eq. (A4) holds because S(t) = 0, $\forall t \ge \hat{T}$, which can be seen by substituting Eq. (A1b) into N = S(t) + I(t) + V(t). Eq. (A5) holds because β and N are positive and because, given that I(0) > 0, I(t) > 0, $\forall t \ge 0$.

Next, we consider the individual's threshold rule for getting vaccinated (Eq. (14), repeated here as eq. (A6)):

$$\beta \frac{I(T)}{N} \geqslant \frac{\theta \delta^2}{(\bar{u} - \underline{u}) - \theta \delta}$$
 (A6)

This rule is framed in terms of I(t); however, we know from Eq. (A5) that, provided S(t) > 0, I(t) is strictly increasing with time. By the inverse function theorem there is a one-to-one correspondence between I(t) and t provided S(t) > 0, and in that case an increase in t implies an increase in I(t), and vice

versa. This means that we can consider an increase in the threshold *time* of vaccination, \hat{T} , as being equivalent to an increase in the threshold I(t)/N.

Now, we denote the actual threshold I(t) of an individual, i.e. the one defined by Eq. (A6), as I^* . Corresponding to I^* , and implied by it, there is a threshold time of vaccination; we denote this as T^* . We will now consider in turn whether an assumed \hat{T} that is greater than T^* and one that is less than T can be equilibrium paths. First, we suppose that all of the remaining susceptibles are going to get vaccinated at a time $\hat{T} > T^*$. The question is: If this were so, would an individual want to get vaccinated at this \hat{T} ? The answer is clearly no, because $\hat{T} > T^*$ implies an $I(\hat{T}) > I(T^*) = I^*$. This violates the individual's threshold rule, Eq. (A6). We now consider the case when $\hat{T} < T^*$. In that case, all remaining susceptibles get vaccinated at \hat{T} , and the number of infecteds is $I(\hat{T}) < I(T^*) = I^*$. However, once that occurs, I = 0, $\forall t$ as per Eq. (A4), so I(t) will never reach I. Therefore, by Eq. (A6), the individual would never wish to get vaccinated at all, and $\hat{T} < T^*$ cannot be an equilibrium path. Having eliminated $\hat{T} > T^*$ and $\hat{T} < T^*$, the only possibility left is $\hat{T} = T^*$, and since it is obviously consistent with Eq. (A6), it is a unique equilibrium path. Q.E.D.

References

Anderson, R. and R. May, 1991, Infectious diseases of humans: Dynamics and control (Oxford University Press, Oxford).

Bailey, N.T.J., 1975, The mathematical theory of infectious diseases and its applications, 2nd edn. (Charles Griffin & Co., London).

Bethwaite, J.G. and P.B. Bethwaite, 1991, An economic evaluation of vaccination strategies against hepatitis B virus: Health care workers and homosexual men, New Zealand Economic Papers 25, 19–38.

Brito, D.L., E. Sheshinski and M.D. Intriligator, 1991, Externalities and compulsory vaccinations, Journal of Public Economics 45, 69–90.

Francis, P.J., 1995, Dynamic epidemiology and the market for vaccinations, unpublished Ph.D. dissertation, Cornell University.

Geoffard, P.-Y. and T. Philipson, 1995, The empirical content of canonical models of infectious diseases: The proportional hazard specification, Biometrika 82, no. 1, 101-111.

Geoffard, P.-Y. and T. Philipson, 1993, Market structure and disease eradication: Private vs. public vaccination, unpublished working paper.

Kiefer, N.M., 1988, Economic duration data and hazard functions, Journal of Economic Literature 26, no. 2, 646-679.

Lloyd, P.J., 1991, AIDS: An economic approach to the choice of policies, The Economic Record, June, 126–138.

Ojo, K., 1991, Economics of public policy in health care systems, Journal of Policy Modeling 13, no. 4, 595-598.

Philipson, T., 1993, The demand for immunity and the effects of public health, Paper presented at the Econometric Society North American Summer Meetings, 24–28 June.

Schelling, T.C., 1978, Micromotives and macrobehavior (W.W. Norton, New York).

Shiller, R.J., 1984, Stock prices and social dynamics, Brookings Papers on Economic Activity 2, 457-510.

Stiglitz, J.E., 1988, Economics of the public sector, 2nd edn. (W.W. Norton, New York).

Stoneman, P., 1983, The economic analysis of technological change (Oxford University Press, Oxford).

Weisbrod, B.A., 1961, Economics of public health (University of Pennsylvania Press, Pittsburgh, PA).

Wickwire, K., 1977, Mathematical models for the control of pests and infectious diseases: A survey, Theoretical Population Biology 11, 182-238.

Wiemer, C., 1987, Optimal disease control through combined use of preventive and curative measures, Journal of Development Economics 25, 301-319.

World Bank, 1993, World development report (Oxford University Press, Oxford).