Are Drugs More Profitable Than Vaccines?

Michael Kremer

Harvard University
The Brookings Institution
The Center for Global Development
National Bureau of Economic Research

Christopher M. Snyder

George Washington University

October 5, 2003

Abstract: In a simple representative consumer model, vaccines and drug treatments yield the same revenue for a pharmaceutical manufacturer, ceteris paribus. Revenue equivalence breaks down if consumers vary in probability of contracting the disease; since drug treatments are sold only to people who have contracted the disease, there is less asymmetric information to prevent the firm from extracting consumer surplus with drug treatments than with vaccines. For appropriate distributions of risk of infection, the ratio of drug treatment to vaccine revenue can be arbitrarily high; we calculate that the ratio is about two to one for empirical distributions of HIV risk. However, if consumers also vary in income, if price discrimination is impossible, and if income covaries negatively with risk of infection, vaccines may be more profitable than drugs. For example, if the ability to price discriminate internationally broke down, incentives to develop vaccines against HIV/AIDS and tuberculosis might exceed incentives to develop drug treatments. Revenue equivalence also breaks down if vaccines interfere more than drug treatments with the spread of the disease. An integrated economic epidemiological model implies that the ratio of steady-state revenue for drug treatments to that for vaccines is highest for rare diseases. The case for subsidizing research and development on vaccines relative to drug treatments is strongest when the distribution of risk among consumers is skewed and when the disease is rare.

JEL codes: O31, L11, I18, D42

Kremer: Department of Economics, Harvard University, Littauer Center 207, Cambridge MA 02138; email: mkremer@fas.harvard.edu. Snyder: Department of Economics, George Washington University, 2201 G Street N.W., Washington DC 20052; email: csnyder@gwu.edu.
The authors would like to thank Bryan Boulier, Glenn Ellison, Corinne Langinier, David Malueg, and seminar participants at the 2002 IAEN Symposium on the Economics of AIDS/HIV in Developing Countries, the 2002 International Industrial Organization Conference, and the 2003 NBER Summer Institute for helpful comments, and Heidi Williams for excellent research assistance.

To-do list

· Insert a proposition and Chris's proof on the convexity/concavity result. Possibly include a more extensive discussion on when the gap is large. Reference the new figure (3). One place for this might be right before Proposition 6 (directly before the paragraph, "An obvious corollary...").

· Competition issue. At a minimum, should explain when the results are likely to go through.

1 Introduction

It is conventionally believed that pharmaceutical manufacturers prefer to develop drug treatments rather than vaccines. Patricia Thomas, journalist and author of a widely publicized book on the search for an AIDS vaccine (Thomas 2001), notes,

> Private companies find vaccines less financially rewarding than drugs. In 2001, the global marketplace for therapeutic drugs exceeded $300 billion, whereas worldwide vaccine sales were only about $5 billion .... It is not hard to understand why major pharmaceutical companies, capable of developing drugs and preventive vaccines, generally invest in drugs that patients must take every day rather than shots given only occasionally. Drug company executives have investors to answer to, after all. (Thomas 2002)

The case of HIV is consistent with this conventional belief that pharmaceutical firms are more inclined to invest in drug treatments than in vaccines: although it may certainly be in part due to differing degrees of scientific difficulty, drug treatments for HIV/AIDS have been developed, but as yet there is no HIV vaccine.

Thomas' explanation of why firms prefer drug treatments to vaccines—that is, that drug treatments are administered more frequently, allowing firms more opportunities to extract revenue—appears to be widely held (for example, see also Rosenberg 1999). Yet from the perspective of neoclassical economics, this explanation seems odd because a rational consumer would pay the expected present value of the stream of benefits in an up front lump sum for the vaccine, and thus by this argument vaccines and treatments should yield equivalent revenues.

While behavioral economics may provide reasons why people are willing to pay more for a cure than for prevention, in this paper we examine two ways in which the view that firms are biased toward developing drug treatments can be reconciled with standard neoclassical economics. We also characterize the circumstances in which the gap between drug and vaccine revenue will be large.

In Section 2, we show that if one moves from a representative consumer model to a more realistic model with heterogeneous consumers, revenue equivalence between vaccines and treatments breaks down. It is indeed realistic to suppose that consumers are heterogeneous in their ex ante probabilities of contracting the disease. Take as an example the case of HIV: people engaging in unprotected sex with a high number of partners or sharing needles during intravenous drug use have a higher risk of contracting HIV than do those not engaging in similar behaviors. We show in this paper that treatments extract revenue from heterogeneous consumers more effectively than vaccines. Since vaccines are administered before consumers contract the disease, there is no basis on which the firm can discriminate among the consumers. If the firm attempts to charge a high price for the vaccine, only consumers at high risk of contracting the disease will buy it, but this segment is often only a small fraction of the population. On the other hand, at the point when treatments are administered, the firm has better information about consumers; in particular, the firm at least knows which consumers have the disease and which do not. The firm can use this information to charge high prices to all consumers who contract the disease, regardless of whether they come from the small segment of the population at high risk or the large segment of the population at low risk.

A simple example suffices to illustrate this point. Suppose there are 100 total consumers, ninety of whom have a ten percent chance of contracting the disease and ten of whom have a 100 percent chance. Suppose consumers are risk neutral and are willing to pay \$100,000 to be cured of the disease if they contract it. A monopolist selling a vaccine could either charge \$100,000 and sell to the ten high-risk consumers or charge \$10,000 and sell to all 100 of them. Either way, the monopolist's revenue is \$1,000,000. A monopolist selling a treatment would, in expectation, sell to the nineteen consumers contracting the disease (all ten of the high risk consumers as well as an average of nine consumers from the low-risk group) at a price of \$100,000 for a total revenue of \$1,900,000, almost twice the revenue from a vaccine.

In Section 2.2 we develop a formal model in which the probability of contracting a disease for consumer i, x_i, is a random variable with distribution function $F(x_i)$. We prove that for any distribution with a non-trivial amount of consumer heterogeneity, a treatment yields more revenue than a similarly effective vaccine. We prove that there exist distributions of consumer heterogeneity for which the ratio of treatment to vaccine revenue is arbitrarily high. While our results are proved in the simplest possible setting in which vaccines and drug treatments produce the exact same social benefits, given the substantial gap in revenue between the two, it is straightforward to argue by continuity that there will exist a broad range of cases in which the social benefit from a vaccine exceeds that from a drug treatment, yet the revenue advantage of the drug treatment will induce the firm to develop a drug treatment rather than a vaccine. Our results suggest that the gap in revenue realized from the development of a vaccine versus a drug treatment will be especially large in the case of skewed distributions of risk.

In Section 2.3 we show how our results from Section 2.2 can be applied to estimates of two actual distributions of expected risk in populations to bound the treatment/vaccine revenue gap. Such estimates can be used to calculate the subsidies needed to induce firms to develop vaccines rather than treatments where the relative social benefit of vaccines is large relative to treatments.

In Section 2.4 we extend our model from Section 2.2 to examine the case in which there are multiple sources of consumer heterogeneity. We use two examples to illustrate that if income covaries negatively with risk of infection vaccines may be relatively more profitable than drug treatments. This work implies that if the ability of firms to price discriminate internationally broke down, incentives to develop vaccines for diseases such as HIV/AIDS and tuberculosis might exceed incentives to develop drug treatments.

In Section 3 we consider a dynamic model and reveal additional disadvantages of vaccines relative to drug treatments in terms of rent extraction. Because vaccines cause greater

reductions in disease transmission than drug treatments, it is more difficult for developers to capture the full social benefit of their medicine. We examine the effects of disease transmission on pricing and research and development (R&D) decisions by embedding an economic model within a standard dynamic epidemiological model, which we use to solve for the optimal price and profits for both vaccines and drug treatments. We show that the steady-state flow of revenue for drug treatments is greater than for vaccines, and thus that R&D expenditures will be distorted towards drug treatments rather than vaccines. The fraction of social benefits that the private developer of a vaccine captures declines with disease prevalence, implying that the gap in revenue between a vaccine and a drug treatment will be especially large in the case of rare diseases.

Sections 2 and 3 focus on the case of private markets for pharmaceuticals; however, in practice, governments are often large purchasers. In Section 4, we argue that if the prices the government pays for vaccines and drug treatments are influenced by the threat point of profits the firm could realize on the private market if bargaining breaks down, then to the extent that vaccines are less profitable than drug treatments on the private market, they will also be less profitable when sold to the government.

A preference for developing drug treatments over vaccines may have detrimental consequences for developing countries since much more medical infrastructure is required to deliver drug treatments than to deliver vaccines. Vaccines are more suited for use in developing countries for several reasons: vaccines do not require prior diagnosis; do not need to be taken on a long term basis but instead require only a few doses; do not have side effects that need to be monitored; and can more easily be delivered by personnel with limited medical training. As an illustration, consider that while three-quarters of the world's children receive a standard package of cheap, off-patent vaccines through the World Health Organization's Expanded Program on Immunization (Kim-Farley et al. 1992), it is estimated that only 50,000 of the 30 million people with HIV in Africa have access to antiretroviral therapies

(Moeti, World Health Organization, 2003), at least in part due to difficulties with safe and effective delivery of the drugs (World Health Organization 2001).

To our knowledge, our comparison of drug treatments and vaccines in the static model with heterogenous consumers is new in the literature. Our analysis of the bounds on the profitability of drug treatments relative to vaccines in Propositions 5 and 6 also represents a contribution. One might look to the industrial organization literature for a related result since, as we argue, the relationship between drug treatments and vaccines in our model is analogous to the relationship between price discrimination and uniform pricing. However, the industrial organization literature provides bounds on the social welfare from price discrimination relative to uniform pricing (Malueg 1993) but not bounds on a monopolist's relative profits, to which our results apply.

Our dynamic extension is based on a standard epidemiological model. Geoffard and Philipson (1997) use a related model to show that if a vaccine is produced by firms with market power and sold on the private market, the disease will not be eradicated in the steady state. Our work differs because we explicitly derive the optimal monopoly price and profits for vaccines, as well as the fraction of vaccines' social value appropriated by developers. Furthermore, our analysis of drug treatments, our comparison of drug treatments to vaccines, and our result that R&D expenditures will be distorted towards drug treatments in a dynamic model are new. This allows us to show that R&D expenditures will be most distorted towards drug treatments when the disease is rare.

While we focus on the positive comparisons between incentives for R&D on vaccines and R&D on drugs, it is also worth briefly discussing the normative implications. In this model, incentives to invest in drugs are exactly optimal, so showing that incentives to invest in vaccines are less than incentives to invest in drugs implies incentives to invest in vaccines are too low. In more general models, for example with patent races, incentives to invest in R&D may be either too high or too low. However, the empirical literature suggests that social

returns to R&D exceed private returns, so a natural inference from the positive conclusion that incentives to invest in vaccines are less than incentives to invest in drugs is that welfare could potentially be improved through measures to increase R&D incentives for vaccines.

Although we have referred to vaccines and drugs, the key distinction in this paper is between medicines administered after a disease has been contracted ("drug treatments") and medicines which prevent healthy people from ever contracting the disease ("vaccines"). Although preventative vaccines are designed to protect individuals against initial infections, therapeutic vaccines are used with the objective of improving the immune system of individuals who are already infected. Likewise, some drug treatments, such as chemoprophylaxis against malaria, are used for prevention. We therefore refer to "preventatives" and "treatments."

2 Static Model

2.1 Homogeneous Consumers

A monopoly pharmaceutical manufacturer, called the firm, has the choice of developing alternative medicines for a disease affecting a population of consumers. The timing of the model is given in Figure 1. First, the firm chooses which of the alternative medicines to develop: a preventatitive or a treatment. To fix ideas, we will suppose the firm's choices are mutually exclusive: it will develop either a preventatitive or a treatment but not both.[1] Let $k_j \in [0, \infty)$ be the present discounted value of the fixed cost of developing medicine j, where $j = v$ for the preventatitive and $j = t$ for the treatment. Let $c_j \in [0, \infty)$ be the present discounted value of the cost of administering medicine j to an individual consumer. Note that the treatment may be administered later in a consumer's life than a preventatitive, and

[1] Given the normalizations adopted later in the paper, we will show (Propositions 4 and 9) the firm does not prefer to develop both a vaccine and a drug treatment.

so the nominal cost of the treatment may be discounted more heavily than the preventatitive, but such discounting is reflected in the terms c_v and c_t since they are expressed as present discounted values. Let $e_j \in [0,1]$ be the efficacy of medicine j, that is, the probability that medicine j prevents the consumer from experiencing any harm from the disease. Let $\sigma_j \in [0,1]$ be the probability that a consumer experiences side effects from medicine j and $s_j \in [0,\infty)$ the present discounted value of the harm from the side effects conditional on experiencing them. Let $P_j \in [0,\infty)$ be the present discounted value of the price the firm receives for medicine j.

Interpreting P_j as a net price the firm receives for medicine j allows for a consistent representation of the legal/liability costs associated with side effects. Assuming a caveat emptor regime in which the consumer bears the liability for harm, consumers' willingness to pay will be reduced by the harm they expect from side effects, and P_j will reflect a discount for this lower willingness to pay. Assuming a caveat venditor regime in which the firm bears liability for harm, P_j can be interpreted as the price the firm receives after subtracting off payments it makes to consumers for damages. Other exogenous legal/liability costs can be embodied in k_j if the costs are fixed or in c_j if the costs vary with the number of consumers who receive the medicine.

Before pursuing any medicine, consumer i learns the probability that he or she will contract the disease, $x_i \in [0,1]$. To capture the notion that consumers are homogeneous, we will assume that x_i takes on a single value, which is public information for consumers and the firm.[2] Whether or not consumer i contracts the disease is represented by Bernoulli random variable d_i, where $d_i = 1$ indicates i contracts the disease, an event which occurs with probability x_i, and $d_i = 0$ indicates i does not contract the disease, an event which occurs with probability $1 - x_i$. Without loss of generality, assume d_i is public information,

[2]The case in which consumers are homogeneous but in which the firm does not know x is formally identical to the case of heterogeneous consumers drawn from a distribution known to the firm. We will treat this case in Section 2.2.

observable not only to consumer i but also to the firm.[3]

As Figure 1 shows, the key difference between a preventatitive and a treatment hinges on when the medicine is administered relative to the realization of d_i. A preventatitive is administered before d_i is realized and a treatment is administered after.

Suppose consumers are risk neutral. If a consumer contracts a disease and has not had a preventatitive or does not receive a treatment, he or she experiences harm $h \in [0, \infty)$ in present discounted value terms. Normalize the mass of consumers to unity.

First, consider the firm's profit from a preventatitive. A consumer's expected net surplus from a preventatitive is $x_i h e_v - \sigma_v s_v - P_v$. That is, with probability e_v the preventatitive is effective and provides a benefit to the consumer in that expected harm $x_i h$ is avoided. From this benefit, the expected harm from side effects $\sigma_v s_v$ and the price P_v have to be subtracted to yield net consumer surplus. The profit maximizing price extracts all this surplus; hence $P_v^* = x_i h e_v - \sigma_v s_v$. Since consumers are of unit mass, the firm's maximum profit from the preventatitive is

$$P_v^* - c_v - k_v = x_i h e_v - \sigma_v s_v - c_v - k_v. \tag{1}$$

Next, consider the firm's profit from a treatment. The consumer will only purchase the treatment if he or she contracts the disease. Conditional on contracting the disease, the consumer's net consumer surplus from the treatment is $h e_t - \sigma_t s_t - P_t$. The profit maximizing price extracts all this surplus; hence $P_t^* = h e_t - \sigma_t s_t$. Since consumers are of unit mass, and a fraction x_i end up contracting the disease, the firm's maximum profit from the treatment is

$$x_i(P_t^* - c_t) - k_t = x_i(h e_t - \sigma_t s_t - c_t) - K_t. \tag{2}$$

[3]To see that this assumption can be made without loss of generality, consider two cases. First, if the firm has developed a vaccine rather than a drug treatment, the firm does not make any decisions conditional on d_i, so it is immaterial whether it can observe d_i. Second, if the firm has developed a drug treatment rather than a vaccine, the firm can indirectly observe who has contracted the disease by observing who demands the drug treatment.

Using expressions (1) and (2), we can characterize which medicine the firm chooses to develop.

Proposition 1. In the homogeneous consumer model, the firm strictly prefers to develop the preventatitive over the treatment if and only if (1) strictly exceeds (2), strictly prefers to develop a treatment over a preventatitive if and only if (2) strictly exceeds (1), and is indifferent if (1) equals (2).

In view of Proposition 1, it is straightfoward to perform comparative statics analyses on the various parameters. Ceteris paribus, the firm tends to prefer to develop a preventatitive over a treatment if it is cheaper to develop (i.e., k_v is low relative to k_t) or cheaper to produce (c_v is low relative to c_t). The firm tends to prefer a preventatitive if it involves less severe side effects (σ_v and s_v are low relative to σ_t and s_t, respectively). The firm tends to prefer a preventatitive if it is a more effective cure (e_v is high relative to e_t).

Obviously this model does not exhaust the list of factors that might lead the firm to prefer preventatitives over treatments or vice versa. However, it would be straightforward to extend the model to consider alternative factors, and we will briefly mention a few here. First, if consumers were assumed to be risk averse, preventatitives would become relatively more profitable, since they would provide insurance to consumers for which consumers would pay a premium. Second, the effect of assuming consumers face liquidity constraints is less clear, depending on the nature of the constraint assumed. If the liquidity constraint is a constraint on lifetime expenditures, say because the consumer has access to relatively efficient credit markets, then the liquidity constraint may bind less with preventatitives than with treatments. To see this, recall that we found the equilibrium price for the preventatitive to be $P_v^* = x_i h e_v - \sigma_v s_v$ and for treatment to be $P_t^* = h e_t - \sigma_t s_t$. Adopting the ceteris paribus assumptions that $e_v = e_t$, $\sigma_v = \sigma_t$, and $s_v = s_t$, it is evident that $P_v^* < P_t^*$ for all $x_i < 1$. Hence, conditional on contracting the disease, total payments are lower with preventatitives. This type of lifetime liquidity constraint would bias the firm in favor of preventatitives. If, on the other hand, the liquidity constraint were a per-period constraint, say because the

consumer does not have access to credit, then the liquidity constraint may bind less with treatments since the total payment with treatments may be spread out in installments (with a payment for each separate treatment) whereas the total payment for the preventatitive would need to be paid in a lump sum at the time the preventatitive is administered. This type of liquidity constraint would bias the firm in favor of treatments.

The conclusions drawn from Proposition 1—that is, that the firm prefers cheaper, more effective medicines associated with fewer side effects—are both intuitive and well-known. To focus on the more subtle issues that are the focus of this paper, we will normalize certain variables so that the firm is indifferent between developing preventatitives and treatments in the homogeneous consumer model. In particular, throughout the remainder of the paper, we will normalize $k_j = c_j = \sigma_j = 0$ and $e_j = 1$ for $j = v, t$. That is, we will assume that both medicines are costless to develop and produce, have no side effects, and are perfectly effective. The following revenue-equivalence result for the case of homogeneous consumers is an immediate corollary of Proposition 1.

Proposition 2. Assume $k_j = c_j = \sigma_j = 0$ and $e_j = 1$ for $j = v, t$. Then the firm is indifferent between developing the preventatitive and the treatment in the homogeneous consumer model.

We will show in the next subsection that treatments are more profitable than preventatitives if consumers vary in risk of infection.

2.2 Heterogeneity in Risk of Infection

In this subsection, we will adopt the preceding model with one modification. As before, consumer i learns the probability that he or she will contract the disease, $x_i \in [0, 1]$, before pursuing any medicine. Now, however, we assume x_i is a random variable distributed according to a nontrivial cumulative distribution function $F(x_i)$. Each consumer in the population has a type given by an independent draw from this distribution. Variable x_i is private infor-

mation for the consumer; the firm only knows the distribution from which x_i is drawn. We are attempting to capture the fact that the consumer's background and/or actions put him or her into a risk category that he or she can observe more accurately than can outsiders. For example, engaging in unprotected sex with multiple partners or in intravenous drug use would put a person at higher risk of contracting HIV, but such behaviors would be difficult for a firm to monitor accurately enough to be able to charge a discriminatory price. Likewise, frequenting mosquito-infested tropical regions increases the chance of contracting malaria, but again may be difficult to monitor accurately.

Normalize $k_j = c_j = \sigma_j = 0$ and $e_j = 1$ for $j = v, t$ as before. That is, both medicines are costless to develop and produce and both are perfectly effective. These normalizations allow us to concentrate on the revenue generated by each medicine in a heterogeneous consumer model.

Consider first the firm's profit maximization problem if it decides to develop a preventatitive. Given that consumers' types x_i are private information, the firm is forced to charge a uniform price. Since consumers are risk neutral, consumer i will buy the preventatitive if the price P_v is less than the expected harm from the disease, hx_i, which represents i's probability x_i of contracting the disease times the harm h from the disease conditional on contracting it. Thus there exists a cutoff type $\hat{x}_v = P_v/h$ such that consumer i weakly prefers to buy if and only if $x_i \geq \hat{x}_v$. The firm's expected revenue from the preventatitive, also equal to its profit given the assumption of zero costs, is $\int_{\hat{x}_v}^{1} P_v dF(x_i)$. Substituting $P_v = h\hat{x}_v$ and rearranging, the firm's profit from the preventatitive is $h\hat{x}_v[1 - F(\hat{x}_v)]$. The firm will choose $\hat{x}_v$, which is equivalent to choosing P_v, to maximize profit; thus, we can write the monopoly profit from preventatitive as

$$\Pi_v = \max_{\hat{x}_v \in [0,1]} \left\{ h\hat{x}[1 - F(\hat{x}_v)] \right\}. \tag{3}$$

Next consider the firm's profit maximization problem if it decides to develop a treatment.

Any consumer who has contracted the disease (i such that $d_i = 1$) would be willing to pay a price up to the avoided harm h. The firm's optimal price for the treatment fully extracts consumer surplus: $P_t^* = h$. Of course consumers will only pay P_t if they happen to contract the disease, which occurs for consumer i with probability x_i. The maximum revenue (and, equivalently, the maximum profit) from the treatment is therefore

$$\Pi_t = \int_0^1 h x_i dF(x_i) = hE(x_i), \tag{4}$$

where $E(\cdot)$ is the expectations operator.

Before formally examining the profits from the preventatitive, Π_v, and the treatment, Π_t, we can gain intuition by analyzing the graphical illustration in Figure 2. Note that the axes on the graph in Figure 2 have been formatted with 1 minus the cumulative distribution $F(x_i)$ on the vertical axis in order to allow for a more intuitive interpretation of the curve in the figure as the demand curve (we will adopt a similar format in later illustrations of the same concept). The preventatitive involves charging a uniform price to all consumers (graphically, this is the area of the twice-shaded rectangle in the figure). Of course the firm would choose the price optimally, so Π_v can be seen in the figure as the rectangle of greatest area that can be inscribed in the demand curve. On the other hand, Π_t is the entire once-shaded area under the demand curve. To see this, note each type x_i pays h for the treatment conditional on contracting the disease, which occurs with probability x_i, thus producing an expected revenue of hx_i for each consumer. Integrating over consumers with respect to their density gives revenues (and, equivalently, profit) Π_t. No matter how the twice-shaded rectangle is inscribed, $\Pi_t > \Pi_v$. Formally, we have the following proposition, proved in the Appendix.

Proposition 3. If the population of consumers with a positive probability of contracting the disease is nontrivially heterogeneous (that is, at least two distinct subintervals of $(0, 1]$ have positive measure), then $\Pi_t > \Pi_v$. Hence the firm's profit from developing a treatment is higher than from developing a preventatitive.

A few remarks about the proposition are in order. First, note that the proposition holds for general distributions, including discrete, continuous, and mixed. Second, note that a two point distribution in which one of the points is $x_i = 0$ is effectively homogeneous, because the relevant population for revenue considerations includes only those consumers with a positive probability of contracting the disease, and this relevant population would in this case then have a single-point distribution.

Further intuition for Proposition 3 can be obtained by reconsidering the problem of medicine choice in terms of price discrimination. A preventatitive constrains the firm to charge a uniform price both from an ex ante and an ex post perspective. A treatment also constrains the firm to charge a uniform price from an ex post perspective; that is, all consumers who contract the disease pay the same price. From an ex ante perspective, however, consumers' expected payments for a treatment are not uniform. High risk consumers will pay for the treatment with high probability, thus leading to a high expected payment from an ex ante perspective; the opposite is true for low risk consumers. A treatment tailors the ex ante expected price to the value consumers place on avoiding the disease. From an ex ante perspective, treatments effectively allow the firm to engage in third degree price discrimination, whereas preventatitives result in a uniform pricing situation. It is a general result in the industrial organization literature that monopolists are able to extract more rent from consumers using third degree price discrimination than using uniform prices (see, e.g., Varian 1989), just as illustrated by the firm considered here.

We have implicitly assumed that the firm develops one of the two medicines but not both. Given the normalization $k_v = k_t = 0$, so that the medicines are costless to develop, it might be thought the firm could do better by developing both and using them in a complicated mixed-bundling scheme. In fact, as the next proposition shows, the firm does prefer to develop both, justifying our focus on exclusive development. The proof of Proposition 4, provided in the Appendix, relies on the fact that the firm extracts 100 percent of social

welfare with a treatment, so a preventatitive would provide no additional benefit.

Proposition 4. The firm does not strictly prefer developing both a treatment and a preventatitive to developing a treatment alone.

We have shown that the firm earns more revenue from treatments than from preventatitives, raising the question of how much more revenue treatments can extract. We will answer this question in a series of propositions, starting with the case in which x_i is a discrete random variable of arbitrary form, and building from there.

Suppose that consumers fall into R risk classes indexed by $r = 1, \ldots, R$. Within each risk class r, consumers have the same probability x_r of contracting the disease. Consumers observe their risk class, but the firm cannot. We will arrange the risk classes without loss of generality such that $0 \leq x_1 \leq \cdots \leq x_R \leq 1$. Let $m_r \in (0, 1)$ be the mass of consumers in risk class r and normalize the mass of the total population such that $\sum_{r=1}^{R} m_r$ is equal to one. Note that this setup captures the case in which an individual i's probability of contracting the disease x_i is a discrete random variable of arbitrary form. The next proposition shows that the number of risk classes determines a tight upper bound on the amount the profit from a treatment exceeds that from a preventatitive, and this proposition will serve as a useful building block for subsequent results.

Proposition 5. For any $\epsilon > 0$, there exist distributions of consumers in R risk classes such that $\Pi_t/\Pi_v > R - \epsilon$. That is, we can find distributions of consumers in R risk classes such that the profit from a treatment can be made arbitrarily close to R times the profit from a preventatitive. Moreover, R is an upper bound on Π_t/Π_v.

In the proof of Proposition 5, contained in the Appendix, we construct a distribution of consumers in which the masses of the R risk classes $\{m_r\}_{r=1}^{R}$ decline geometrically. Further, we specify probabilities $\{x_r\}_{r=1}^{R}$ such that the firm earns the same profit whether it sells to all consumers at a low price hx_1, to all consumers but the lowest risk class at a higher price hx_2, etc., on up to selling to the highest risk class alone at price hx_R.

We note that Proposition 5 has a straightforward corollary in the simplest possible case of consumer heterogeneity, that is, the two type case with a low risk class and a high risk class. The example from the Introduction (with 100 consumers, 90 of whom have a ten percent chance of contracting the disease and ten of whom have a 100 percent chance) is such a case. As noted in the Introduction, the treatment produces higher profit than the preventatitive by a factor of 1.9. Proposition 5 implies that a treatment can be as much as twice as profitable as a preventatitive in the two type case, but no more. The example given in the Introduction approaches our bound of two, and we can come closer to the bound with examples in which the size of the high risk pool as well as the probability of contracting the disease in the low risk pool are reduced. For example, consider a population of 100 consumers, 99 of whom have a one percent chance of contracting the disease, and one of whom has a 100 percent chance. Then it can be shown, given the assumption from the Introduction that the harm from the disease is 100,000, that a preventatitive produces a profit of 100,000 while treatment produces a profit of 199,000, very nearly twice as much profit.

The two type case provides important insights into the settings in which firms will strongly prefer treatments to preventatitives. First, our results suggest that the gap in revenue between developing a preventatitive and developing a treatment will be especially large in the case of skewed distributions of consumer risk: distributions in which there exist a large segment of the population with a very small probability of contracting the disease and a small segment of the population with a large probability of contracting the disease will create the largest relative incentives for the firm to develop treatments.

An obvious corollary of Proposition 5 is that there exist distributions of consumer types such that treatments are arbitrarily more profitable than preventatitives. This can be seen by taking the limit as R approaches infinity in the proposition. Stated formally, we have the following proposition.

Proposition 6. For any finite bound $M \in (0, \infty)$, there exist distributions of consumers such that $\Pi_t / \Pi_v > M$.

By themselves, Propositions 3 and 6 do not raise public policy concerns. The propositions were proved under maintained assumptions which guarantee that the social benefit from preventatitives and treatments are equal, so no problems arise if the firm is biased toward developing treatments because of better rent extraction properties. Given the substantial profit advantage that treatments potentially have over preventatitives, it is easy to see by continuity that there will exist a broad range of cases in which preventatitives are socially more beneficial than treatments and yet the firm is still biased toward developing treatments. Because a preventatitive is administered at an early stage, it may be more effective in preventing the disease's spread, may reduce the harm the disease causes an individual, and indeed may increase the probability of curing the disease as compared to a treatment. Yet, if the revenue extraction advantage of a treatment is great enough, the firm will still have an incentive to develop a treatment rather than a preventatitive.

2.3 Applications to Empirical Distributions

In this section, we apply our theoretical results to estimates of actual distributions of HIV risk in certain populations. We consider two examples, one using a nationally representative sample but containing only crude risk information, and another using a non-representative sample but containing more detailed information on risk. The first example uses nationally-representative data from the United States Center for Disease Control National Health and Examination Survey (NHANES) in 1999-2000 to divide the population into risk classes based on the number of sexual partners in a twelve month period.[4] The second example involves a smaller population for which we have more detailed data, data from a study of projected HIV

[4] We do not allow for separate risk factors based on sex of the respondent or of his or her partners and, due to data limitations, we assume that all men who reported having male partners had no female partners (and analogously for women with female partners).

risk which surveyed a random sample of individuals living in poor neighborhoods with high drug use in Houston, Texas (Bell and Trevino 1999). The result from the previous subsection that there are theoretical distributions of risk for which treatments generate considerably more revenue than preventatitives is borne out for the actual distributions of risk analyzed in both examples. We will see that in the first case, the highest revenue is realized from charging a low price to a large portion of the population; in the second case, the highest revenue is realized from charging a high price to a small portion of the population.

For our first example, assume the risk of contracting a disease from one sexual partner is ϕ and normalize $\phi = 1$; thus, we approximate the risk of contracting the disease as a linear function of the number of sexual partners. Assume a unit mass of consumers with distribution of sexual partners as in NHANES (1999–2000), normalize the harm experienced conditional on contracting the disease to $h = 1$, and maintain all previous normalizations (that is, that medicines are costless to develop and produce, are perfectly effective, etc.). We will compute the revenue from a treatment and a preventatitive and then compare the two figures. Given the normalizations, the revenue from a treatment is equal to the fraction of the population that is expected to contract the disease, which in turn is equal to the mean number of sexual partners, 1.666. Thus the revenue from a treatment is 1.666. To compute the revenue from a preventatitive, we can compute (as in Table 1, see also the graphical illustration in Figure 4) the maximum price that induces each risk class to purchase (and strictly induces higher risk classes to purchase), and then find the highest revenue. Reading down the last column of Table 1, the highest revenue is gained from selling to the entire population that is sexually active at price 1, which yields 0.8991 in revenue. Thus, the ratio of revenue from a treatment to revenue from a preventatitive is 1.853, so a treatment would generate almost twice as much revenue as a preventatitive.

In our second example, we consider a study that directly provides estimates of projected HIV risk, Bell and Trevino (1999). The authors collected quite detailed information on

270 subjects living in poor Houston neighborhoods, including records of all the subjects' sexual acts over a given thirty day period. The authors used the data from this survey to parametrize an epidemiological model of HIV risk which combines risk behaviors, prevalence rates, and transmission probabilities. The 270 individuals in Bell and Trevino's sample are not representative of the U.S. population as a whole. In particular, 14 percent could be expected to develop HIV within ten years, an order of magnitude notably higher than the national average. Assuming a static population with no change in the prevalence of HIV within the population as well as no change in the risk level of new sexual partners over time, this model then allows them to compute an empirical distribution of the ten-year projected risk of contracting HIV for the given population. The resulting empirical distribution (based on data from Figure 1 in Bell and Trevino) is presented in the first two columns of Table 2.

Assuming a unit mass of consumers with the same distribution of HIV risk as in Table 2, normalizing the harm experienced conditional on contracting HIV to $h = 1$, and maintaining all the other previous normalizations, we can compute the potential revenue from an HIV preventatitive and an HIV/AIDS treatment. The revenue from a treatment equals the expected number of infected individuals times the avoided harm $h = 1$, or 0.1424. To compute the revenue from a preventatitive, we can compute (as in Table 2, see also the graphical illustration in Figure 5) the maximum price that induces each risk class to purchase (and strictly induces higher risk classes to purchase), and then find the highest revenue. Reading down the last column of the table, the highest revenue, 0.0694, is generated by charging a price that induces the 75 percent risk class and higher to purchase. Thus the ratio of treatment revenue to preventatitive revenue is 2.052, so a treatment would again generate more than twice as much revenue as a preventatitive.

Despite the fact that the distribution of HIV risk in Bell and Trevino's sample is likely to be less skewed than in the U.S. population as a whole, it is still somewhat skewed. Only nine percent of the mass of consumers have risks at or above 75 percent. Serving only these

high risk consumers with a preventatitive leaves a large mass of consumers from lower risk classes unserved, and thus leaves a great deal of unclaimed consumer surplus.

2.4 Multiple Sources of Consumer Heterogeneity

Our model has considered the case in which the only source of heterogeneity is in consumers' probability x of contracting the disease. In this sub-section, we consider the case in which there is also variation in some second source of heterogeneity y (for example, income or more generally willingness to pay for a unit reduction in probability of infection). In particular, we will consider applications to HIV/AIDS and tuberculosis.

If firms can perfectly price discriminate on the basis of y then the analysis above can be generalized by calculating the preventatitive and treatment revenue given the marginal distribution of x at each value of y and integrating over y. The qualitative conclusions will be similar to those above. On the other hand, if firms cannot discriminate on the basis of y, either because y is unobservable or because of problems with resale, if x and y are negatively correlated then the firm might prefer to develop a preventatitive rather than a treatment.

As two examples, let us examine the relative profits to a monopolist developer of either a preventatitive or a treatment for HIV/AIDS and tuberculosis if international price discrimination is impossible. Note that the real world falls somewhere between the extremes of no price discrimination and full price discrimination. Recent moves to ease imports of drugs into the US from Canada could weaken the ability of firms to price discriminate internationally.

Assume each consumer is characterized by two random variables (each of which is private information for the consumer): random variable $x_i \in [0,1]$, distributed according to the nontrivial cumulative distribution function $F(x_i)$, again represents the probability that consumer i will contract the disease; random variable $y_i \in [0,1]$, distributed according to the nontrivial cumulative distribution function $G(x_i)$, represents consumer i's willingness to

pay for a given reduction in probability of infection.

Consider as a potential population of consumers the entire world population, and treat all individuals within any given country as homogeneous, with the same income and chance of infection; note that an analogous analysis could extend our work to allow for distributions of x and y within each country. We use country-level data on population, per capita GNP, estimated number of HIV-positive individuals, and estimated cases of tuberculosis (TB) to approximate our two sources of consumer heterogeneity.[5] We approximate the risk of contracting the disease, x_i, as the fraction of people within a given country that are HIV-positive or infected with TB. For example, an estimated $1,500,000$ of the $11,689,010$ people living in Zimbabwe are HIV-positive, so we assign a risk attribute to people living in Zimbabwe of $x_i = 0.13$. We use as an approximation of willingness to pay, y_i, the country's per capita GNP. The correlation of x and y for HIV is -0.13; for TB is -0.12.

Consider first the firm's profit maximization problem if it develops a preventatitive. Because the firm is unable to price discriminate on the basis of either source of consumer heterogeneity, the firm is forced to charge a uniform price for the preventatitive. Normalizing the harm from the disease conditional on contracting it to 1, consumer i will buy the preventatitive if the price P_v is less than $x_i \cdot y_i$, which is consumer i's risk of contracting the disease times her willingness to pay. For convenience, let $z_i \in [0,1]$ denote the product $x_i \cdot y_i$ where z_i is distributed according to some cumulative distribution function $H(z_i)$. We can calculate the preventatitive revenue for each possible price by multiplying that price times the population of consumers who are weakly induced to purchase the preventatitive at that price: then there exists a cutoff type $\hat{z_v}$ such that consumer i weakly prefers to buy if and only if $x_i \cdot y_i \geqq \hat{z_v}$. The firm chooses $\hat{z_v}$, or equivalently chooses P_v, to maximize revenue,

[5] Population data is 1998 data from the 2000 World Bank World Development Indicators; per capita GNP data is 1998 data calculated with the World Bank Atlas method in 2000 US dollars from the 2000 World Bank World Development Indicators; HIV data is the estimated number of HIV-positive 0-to-49 year olds at the end of 1999 from the 2000 UNAIDS Epidemiological Fact Sheets by country; tuberculosis data is estimated incidence of tuberculosis in 1998 from the World Health Organizations' Global Tuberculosis Control 2000.

and the monopoly profit from a preventatitive in this case is

$$\Pi_v = \max_{\hat{z}_v \in [0,1]} \left\{ \hat{z}_v [1 - H(\hat{z}_v)] \right\}. \tag{5}$$

For the case of HIV, the firm would realize the highest revenue by charging the price that just induces individuals in the US to buy and strictly induces individuals in Switzerland, Swaziland, Namibia, the Bahamas, South Africa, and Botswana to purchase the preventatitive (see Figure ??).

Now consider the firm's profit maximization problem if it decides to develop a treatment. Any consumer who contracts the disease (i such that $d_i = 1$) would be willing to pay up to $h \cdot y_i$; maintaining that $h = 1$, consumer i will thus purchase a treatment if P_t is less than or equal to her willingness to pay. We calculate the treatment revenue for each possible price by multiplying that price times the population of infected individuals who are weakly induced to purchase the treatment at that price. There exists a cutoff willingness to pay $\hat{y}_t$ such that consumer i weakly prefers to buy if and only if $y_i \leq \hat{y}_t$. Hence, the maximum revenue (and equivalently, the maximum profit) from the treatment is

$$\Pi_t = \max_{\hat{y}_t \in [0,1]} \left\{ E(x_i) \cdot \hat{y}_t [1 - G(\hat{y}_t)] \right\}. \tag{6}$$

In our example of HIV, the firm would realize the highest revenue by charging the price that just induces individuals in France to buy and strictly induces individuals in sixteen other countries to purchase the treatment (see Table ??, Figure ??).[6] The expected revenues for an HIV preventatitive and an HIV/AIDS treatment yield a preventatitive-to-treatment-treatment revenue ratio in this case of 1.13 if price discrimination across countries is not possible.

[6] The sixteen countries strictly induced to purchase the treatment in this case are Finland, Netherlands, Belguim, Andorra, Sweden, Germany, Austria, Brunei Darussalam, Iceland, the US, Singapore, Japan, Denmark, Norway, Switzerland, and Luxembourg.

We can calculate the social benefit of the preventatitive and the treatment in this model by summing, over all countries which purchase the medicine, the averted harm from the disease for country i multiplied times the willingness to pay of country i. Note that the social benefit is not realized for people who do not take either the preventatitive or the treatment. In Table ?? we present calculations of the social benefits, private revenues, expected lives saved, and potential (that is, if the medicine were sold to all countries) social benefits for both the preventative and the treatment. For HIV/AIDS, the treatment achieves 0.76 of the social benefit that would be realized with a preventatitive, and a preventative would save 5.18 as many lives as a treatment (note that this model does not take into account the externalities associated with preventatives such as vaccines). Individuals in low-income countries would particularly benefit from vaccines.

A developer of a TB preventatitive would realize the highest revenue by charging the price that just induces individuals in India to pay and that strictly induces individuals in 115 other countries to purchase the preventatitive; the developer of a TB treatment would realize the highest revenue by charging the price that just induces individuals in India to pay and that strictly induces individuals in 138 other countries to purchase the treatment. The expected revenues for a TB preventatitive and a TB treatment yield a preventatitive-to-treatment-treatment revenue ratio in this case of 1.56. For TB, the treatment achieves 0.99 of the social benefit that would be realized with a preventatitive, and a preventative would save 1.09 as many lives as a treatment (see Table ??, Figures ??, ??).

3 Dynamic Model

In this section we show that treatments are more profitable than preventatitives even with a homogeneous population if preventatitives cause greater reductions in disease transmission than treatments. preventatitives (such as vaccines) typically reduce disease transmission

more than treatments for two reasons. First, people often spread the disease before receiving treatment. For example, much transmission of HIV is believed to take place during the first few months of an individual's infection;[7] during this "window period," viral loads (and thus transmission rates) are high,[8] but the individual is not producing enough antibodies to test positive on standard HIV tests, and thus will not seek treatments such as the Highly Active Antiretroviral Treatments (HAARTs). Second, treatments sometimes treat symptoms rather than actually curing the disease, so even if a patient receiving a treatment is experiencing no harm from the disease, he or she may still be a carrier.

To examine the effect of disease transmission on pricing and R&D expenditure decisions, it is useful to embed the economic model within a standard dynamic epidemiological model. We will consider a non-fatal disease since this simplifies modeling by allowing us to consider a constant population. Assume that people are born into the population at rate δ and that both infected and uninfected individuals die at rate δ as well. Let S, I, and V represent the fractions of the population that are susceptible, infected, and vaccinated. Normalizing the mass of the consumer population to unity, $S + I + V = 1$. If no preventatitive exists, $V = 0$ and $S + I = 1$.

The rate at which people become newly infected is βIS, where β depends on the rate at which susceptibles contact infecteds and the proportion of those contacts which lead to new infections. Let ξ denote the fraction of newborns who are vaccinated.[9] For now, we treat ξ as a parameter; later, we will solve for the equilibrium value of ξ and substitute that value back into this epidemiological model.

The rate of change of the susceptible population is equal to the birth rate times the

[7] Wawer et al. (2003) report that in their study approximately half of all HIV transmission was estimated to occur within the first five months of an individual's seroconversion (seroconversion usually takes place two to six weeks after acquisition of HIV).

[8] Wawer et al. (2003) show the rate of HIV transmission per coital act is highest in the first five months after seroconversion (0.0081 per coital act).

[9] Given the Poisson structure of the model, without loss of generality we can treat all vaccinations as if they are given to newborns.

non-vaccination rate, $1-\xi$, minus the loss of susceptibles to infection or death:

$$\dot{S} = \delta(1-\xi) - \beta IS - \delta S. \tag{7}$$

The rate of change of the infected population is

$$\dot{I} = \beta IS - \delta I \tag{8}$$

and the rate of change of the vaccinated population is

$$\dot{V} = \delta\xi - \delta V. \tag{9}$$

There is a trivial steady state in which $I^* = 0$ and $S^* = 1 - V^*$, but this is unstable for $\xi < 1-\delta/\beta$. Setting $\dot{S} = \dot{I} = \dot{V} = 0$ in equations (7) through (9) gives the non-trivial steady state

$$S^* = \delta/\beta, \quad V^* = \xi, \quad I^* = 1 - \xi - \frac{\delta}{\beta} \tag{10}$$

for $\xi < 1-\delta/\beta$. For brevity we will define $\lambda = \delta/\beta$. This term can be interpreted as the latent prevalence of healthy individuals in the steady state before a preventatitive is introduced, as can be seen by setting $\xi = 0$ in the equation for I^* in (10). With this notation, the steady-state rate at which new infections will occur if a preventatitive is developed is thus

$$\beta I_v^* S_v^* = \delta(1 - \xi - \lambda) \tag{11}$$

and the steady-state rate at which new infections will occur if a treatment is developed is

$$\beta I_t^* S_t^* = \delta(1 - \lambda). \tag{12}$$

We wish to consider a firm's incentives for developing either a preventatitive or a treatment. (We will show below in Proposition 9 that the firm will not choose to develop both.) Once developed, we assume that either medicine can be produced at zero cost. We suppose that a person taking a preventatitive does not contract the disease and is unable to transmit the disease to others. We assume that a single dose of a treatment perfectly relieves all symptoms permanently but still allows the treated individual to transmit the disease to others. These assumptions are clearly extreme, but results will be qualitatively similar as long as treatments interfere less with disease transmission than do preventatitives.

We will consider the case as the discount rate goes to zero so that consumers only care about their probability of contracting the disease (not when they will contract it) and the firm wants to maximize the steady-state flow of revenue; this allows us to abstract from transitional dynamics. Assume consumers are risk neutral, and define h to be the fixed amount a consumer will be willing to pay in order to avoid infection.

Revenue (also profit since production costs have been normalized to zero) Π_j will equal price P_j multiplied by quantity sold Q_j, where $j = v$ if a preventatitive is developed and $j = t$ if a treatment is developed. Let W_j denote social welfare. Let P_j^*, Q_j^*, Π_j^*, and W_j^* denote the equilibrium price, quantity, profit, and social welfare in the steady state, respectively.

We proceed by solving for the firm's profit-maximizing prices P_v^* and P_t^*, using these prices to compute the steady-state flow profits Π_v^* and Π_t^*, and then comparing these profits to determine which medicine is more profitable in the steady state. The results are contained in a series of propositions.

Proposition 7. In the steady state of the dynamic model with a treatment, equilibrium price is $P_t^* = h$, quantity is $Q_t^* = \delta(1-\lambda)$, flow profit is $\Pi_t^* = h\delta(1-\lambda)$ and flow welfare is $W_t^* = h\delta(1-\lambda)$.

The proof of Proposition 7 is straightforward. The firm sets a price extracting all the surplus from infecteds, $P_t^* = h$. Since we assumed that one dose of the treatment perma-

nently relieves all symptoms, only newly infected consumers will purchase the treatment. In equilibrium, all newly infecteds buy the treatment at price $P_t^* = h$, so by equation (12), $Q_t^* = \delta(1-\lambda)$. The resulting flow profit is thus $\Pi_t^* = P_t^* Q_t^* = h\delta(1-\lambda)$. The social benefit of the treatment is that all newly infected individuals are completely relieved of the harm from the disease h, so $W_t^* = h\delta(1-\lambda)$.

It is more difficult to derive a rational-expectations equilibrium in the preventatitive case because of the externality involved. The more consumers who are vaccinated, the lower the disease prevalence, thus reducing the incentives of consumers to be vaccinated. We will solve for the equilibrium of this system, drawn in Figure 3. The diagram is drawn for a given preventatitive price P_v; below we will solve for the profit maximizing P_v. The solid line AA' represents consumer demand for preventatitives as a function of the the overall infection level and the given price, $Q_v(I_v, P_v)$. As the line indicates, there is a cutoff level of infection prevalence $\hat{I}$ such that no consumer buys the preventatitive below this cutoff, all consumers buy the preventatitive above this cutoff, and consumers are indifferent exactly at this cutoff so that the fraction of newborns purchasing the preventatitive is indeterminate. The dotted line BB' represents the infection level as a function of the quantity of preventatitive consumed, $I_v(Q_v)$, which comes from the epidemiological model; in particular, this follows directly from the equation for I^* in (10). The equilibrium quantity as a function of price is given by the intersection of AA' and BB'. Note that an increase in price shifts AA' to the right, resulting in a lower intersection point and thus a lower equilibrium quantity—the familiar tradeoff for a monopolist. The next step is to compute the profit-maximizing price. The result of these calculations is provided by Proposition 8, proved in the Appendix.

Proposition 8. In the steady state of the dynamic model with a preventatitive, equilibrium price is $P_v^* = h(1-\sqrt{\lambda})$, quantity is $Q_v^* = \delta(1-\sqrt{\lambda})$, flow profit is $\Pi_v^* = h\delta(1-\sqrt{\lambda})^2$, and flow welfare is $W_v^* = h\delta(1-\sqrt{\lambda})$.

In view of Propositions 7 and 8, we can compare the firm's incentives to develop a

treatment versus a preventatitive:

Proposition 9. In the steady state of the dynamic model, the firm's profit is higher with a treatment than a preventatitive. The firm appropriates all the social surplus with a treatment but only a fraction $1-\sqrt{\lambda}$ with a preventatitive. The firm does not strictly prefer developing both a treatment and a preventatitive to developing a treatment alone.

The proof is a simply corollary of Propositions 7 and 8; the details are provided in the Appendix. The treatment allows the firm to appropriate 100 percent of the social benefits since consumers can be charged their maximum willingness to pay and there are no externalities. With a preventatitive, the firm does not serve all susceptibles since this would eradicate the disease and eliminate future demand. Instead, the preventatitive is priced such that only a fraction of susceptibles are served. The unvaccinated susceptibles obtain a positive externality from other's vaccinations, and this benefit is not appropriated by the firm.

As Proposition 9 states, since the firm appropriates all social surplus with a treatment, there is no additional benefit from also developing a preventatitive. This justifies our implicit assumption that the firm develops one or the other medicine but not both.

Analogous to the result in the static model of Section 2, we are able to obtain the result in the present dynamic model that the ratio of profit from a treatment to the profit from a preventatitive is unbounded. The key parameter is λ, which recall is interpreted as the initial proportion of healthy individuals prior to the introduction of the preventatitive. A disease that is initially quite rare can be represented by the limit as λ approaches one. Hence, the gap in revenue between a preventatitive and a treatment will in this case be especially large for rare diseases: for such diseases, a treatment is particularly profitable for the firm relative to a preventatitive, and the fraction of social benefits the preventatitive producer is able to appropriate is particularly small.

Proposition 10. In the limit as the initial prevalence of the disease approaches zero, the ratio of profit from treatment to preventatitive grows without bound, i.e., $\lim_{\lambda\to 1}(\Pi_t^*/\Pi_v^*)=\infty$, and the ratio of profit to social welfare from a preventatitive goes to zero, i.e., $\lim_{\lambda\to 1}(\Pi_v^*/W_v^*)=0$.

There is reason to think that, in practice, externality benefits may be quite large relative to direct benefits. For example, in a randomized evaluation of a project in Kenya, Miguel and Kremer (2003) find that school-based mass treatment with deworming drugs created substantial externalities among both untreated students in the treatment schools and among children in neighboring schools; the share of disease burden averted due to externalities in their study is estimated at about 76 percent.[10]

4 Government Purchases

The previous sections have focused on the case of pharmaceutical sales on private markets. However, at least in the case of preventatitives, governments are the main purchasers, not private parties. We argue in this section that our results are still applicable to the case of government procurement as long as price negotiations between the firm and the government are influenced by the threat point of what profits the firm would realize with private sales if negotiations with the government broke down.

Suppose the firm and government engage in Nash bargaining over the sale of medicine j. Assume they bargain after the firm has decided which medicine ($j = v$ for preventatitive, $j = t$ for treatment) to develop and has sunk its investment in R&D. For ease of comparison, we will assume that this sunk cost is the same for either medicine. Assume the government's objective is to maximize consumer surplus and the firm's is to maximize profit.[11]

Given these objectives, the "pie" over which the parties bargain equals social welfare at socially efficient prices, denoted $\tilde{W}_j$. Note the difference between $\tilde{W}_j$ and W_j defined earlier: $\tilde{W}_j$ is social welfare when the medicine is consumed at the socially efficient level, whereas W_j is social welfare given the amount of medicine that will be consumed at monopoly prices. Let

[10] See also the theoretical analysis of vaccine externalities in Boulier, Datta, and Goldfarb (2003).

[11] Assuming alternatively the government's objective is to maximize social welfare, with equal weights given to producer and consumer surplus, Nash bargaining would trivially result in all surplus being allocated to the firm.

Π_j be the monopoly profit and CS_j the consumer surplus from the private sale of medicine j at monopoly prices. Let Φ_j be the firm's threat point in Nash bargaining and Γ_j be the government's. Then the Nash bargaining formula yields the following expression for the firm's surplus:

$$\frac{1}{2}(\tilde{W}_j + \Phi_j - \Gamma_j). \tag{13}$$

It is plausible to assume that the firm's threat point is given by what it would earn if it sold to the private market rather than the government.[12] Under this assumption, $\Phi_j = \Pi_j$ and $\Gamma_j = CS_j$. Substituting these threat points into equation (13), we have that the firm prefers to develop a treatment to a preventatitive if and only if $(\tilde{W}_t + \Pi_t - CS_t)/2 > (\tilde{W}_v + \Pi_v - CS_v)/2$, or upon rearranging,

$$\Pi_t - \Pi_v > \tilde{W}_v - \tilde{W}_t - CS_v. \tag{14}$$

We have substituted $CS_t = 0$ in condition (13), consistent with the fact that the firm ends up extracting all consumer surplus in both the static model of Section 2 and the dynamic model of Section 3.

Condition (14) shows that, even if medicines are procured by the government, there is a wedge between social and private incentives that possibly distorts the firm's development decision. There is a range of cases in which $\tilde{W}_v > \tilde{W}_t$, so it is socially beneficial for the preventatitive to be developed, yet (14) holds so the firm instead develops the treatment. This range of cases may be broad for two reasons: as shown in Propositions 6 and 10, the ratio Π_t/Π_v is unbounded, and so the left-hand side of (14) may be large; furthermore, subtraction of the term CS_v reduces the right-hand side of (14) below $\tilde{W}_v - \tilde{W}_t$.

Our conclusions are essentially an instance of the familiar hold-up problem (Williamson 1975). The firm decides which medicine to develop prior to negotiating with the government.

[12] There are of course other possibilities. For example, the government could hypothetically refuse to grant approval for private sales of the medicine in the event of bargaining breakdown, implying $\Phi_j = 0$. However, at least in the United States (by far the largest single market), once approval is granted the U.S. government would not stop private sales of the product.

Recognizing that it does not appropriate all the surplus in bargaining, the firm may distort its decision to appropriate more surplus; thus the firm is concerned over how profitable the medicines are relative to each other in the threat point, i.e., on the private market.

The analysis can be repeated assuming that preventatitives are procured by the government but treatments are sold on the private market. The firm would then compare the Nash bargaining surplus from the preventatitive $(\tilde{W}_v + \Pi_v - CS_v)/2$ to the treatment profit Π_t. After rearranging and noting $W_v = \Pi_v + CS_v$, we see that the firm would prefer to develop a treatment to a preventatitive if and only if

$$\Pi_t - \Pi_v > \frac{1}{2}(\tilde{W}_v - W_v). \tag{15}$$

The right-hand side of (15) is the difference between social welfare given the socially efficient level of consumption and social welfare at the monopoly-price level of consumption, divided by two. The left-hand side is again the relative profit advantage of a treatment, which all our preceding results were directed toward showing can be large. Again, we have the result that the firm may be biased against developing a preventatitive even though preventatitive development may be more socially desirable.

One policy implication that emerges from this section is that there are advantages to the government bargaining with the firm as early as possible in the development process, since this will of course help protect the firm's R&D from hold up by the government and thus enhance investment. Our point here is that this will also encourage the firm to make the socially efficient decision regarding which medicine to develop. In the model, if the bargain takes place before the firm decides which medicine to develop, in equilibrium the firm will develop the preventatitive precisely when it is socially efficient to do so, i.e., when $\tilde{W}_v > \tilde{W}_t$. This provides one justification for advance purchase commitment programs for vaccines of the type described by Kremer (2001).

5 Conclusions

Numerous potential factors could induce firms to develop a treatment (administered after patients contract the disease) rather than a preventatitive (administered before), or vice versa, for a given disease. One or the other may involve "easier science," be cheaper to produce once developed, or have fewer or less severe side effects. The interests of both consumers and firms are likely to be aligned concerning all of these preceding factors: that is, consumers and firms are likely to agree that a cheaper treatment is better as is one with fewer side effects. In this paper, we identified more subtle issues that are present even if one abstracts away from all these preceding factors.

- Treatments emerge as better rent extraction tools than preventatitives if consumers vary in risk of contracting the disease. Because drug treatments are sold only to people who have contracted the disease, the firm has more information about individual consumer's valuations and can extract consumer surplus more efficiently with treatments than with preventatives. The revenue gap will be largest in the case of skewed distributions of consumer risk. We presented two examples (HIV/AIDS and tuberculosis) which illustrate that if there is a second source of consumer heterogeneity (such as income) that covaries negatively with risk of infection and if price discrimination is impossible, then preventatitives may be relatively more profitable than treatments.

- Treatments emerge as better rent extraction tools than preventatitives in a dynamic model if preventatitives are more likely than treatments to interfere with disease transmissions. Since the people who benefit from the positive externalities of vaccination do not compensate the firm for the benefits they receive from the preventatitive, the firm earns more revenue from treatments than from preventatitives. The revenue gap will be largest in the case of rare diseases.

We showed that in both the static and dynamic models, the firm can make arbitrarily higher revenue in percentage terms with treatments than with preventatitives. Fitting two actual estimates of the ex ante distribution of HIV risk—one a nationally-representative survey of HIV risk, the other a detailed survey of individuals in several poor Houston neighborhoods—into our theoretical framework, we demonstrated the empirical relevance of our theoretical results. In both samples, we calculated that the revenue-extraction properties of treatments would allow the firm to earn considerably more, around twice the revenue, compared to preventatitives.

Appendix

Proof of Proposition 3

Define

$$\hat{x}_v^* = \underset{\hat{x}\in[0,1]}{\operatorname{argmax}} \left\{ h\hat{x}[1 - F(\hat{x})] \right\}.$$

Then, in view of equations (3) and (4),

$$\begin{aligned} \Pi_t - \Pi_v &= h\int_0^1 x_i dF(x_i) - h\int_{\hat{x}_v^*}^1 \hat{x}_v^* dF(x_i) \\ &= h\int_0^{\hat{x}_v^*} x_i dF(x_i) + h\int_{\hat{x}_v^*}^1 (x_i - \hat{x}_v^*) dF(x_i). \end{aligned} \tag{A1}$$

Both terms in expression (A1) are nonnegative. There cannot be a measure one of consumers at $\hat{x}_v^*$ by maintained assumption. Thus there must be a positive measure on either a subset of $(0, \hat{x}_v^*)$, in which case the first term in (A1) is positive, or on a subset of $(\hat{x}_v^*, 1]$, in which case the last term in (A1) is positive. In either case, $\Pi_t - \Pi_v > 0$. Q.E.D.

Proof of Proposition 4

Curing the disease generates gross social welfare $hE(x_i)$ from an ex ante perspective. This is also the revenue from a drug treatment, and profit since costs have been normalized to zero, by equation (4). Hence the addition of a vaccine cannot increase the firm's profit. Q.E.D.

Proof of Proposition 5

A distribution of consumers into R risk classes involves parameters $\{m_r\}_{r=1}^R$ and $\{x_r\}_{r=1}^R$. These $2R$ parameters can be freely chosen to generate as high as possible a value of Π_t/Π_v subject to $m_r \in (0,1)$ for all $r = 1, \ldots, R$; $\sum_{r=1}^R m_r = 1$; and $0 \le x_1 \le \cdots \le x_R \le 1$. Let $\theta \in (0, 1/2)$. Define

$$m_r = \begin{cases} \theta^{r-1} & \text{if } r > 1 \\ 1 - \sum_{r=1}^{R-1} \theta^r & \text{if } r = 1. \end{cases} \tag{A2}$$

The definition of risk-class masses in equation (A2) produces a geometrically declining sequence. As is easily seen, this definition respects the constraints $m_r \in (0,1)$ for all $r = 1, \ldots, R$ and $\sum_{r=1}^R m_r = 1$. Next, we set the risk-class probabilities $\{x_r\}_{r=1}^R$. We will set them so that the firm makes the same revenue regardless of which risk class it decides to target with its vaccine pricing. Specifically, we will set $x_R = 1$ and define the rest, $\{x_r\}_{r=1}^{R-1}$, recursively by

$$hx_r \sum_{i=r}^R m_i = hx_{r+1} \sum_{i=r+1}^R m_i. \tag{A3}$$

The left-hand side of equation (A3) is the revenue (and profit) from charging a price hx_r and selling the vaccine to risk classes r and higher. The right-hand side is the revenue (and profit) from charging a price hx_{r+1} and selling to risk classes $r+1$ and higher. As is easily seen, our definition of $\{x_r\}_{r=1}^R$ respects the constraint $0 \leq x_1 \leq \cdots \leq x_R \leq 1$. From equation (4), we have $\Pi_t = \sum_{r=1}^R hm_r x_r$. By construction implicit in (A3), we have $\Pi_v = hx_1$; that is, it is weakly most profitable to charge hx_1 for the vaccine and sell to all consumers. Thus

$$\begin{aligned}
\frac{\Pi_t}{\Pi_v} &= \frac{\sum_{r=1}^R hm_r x_r}{hx_1} \\
&= m_1 + \sum_{r=2}^R \frac{m_r x_r}{x_1} \\
&= m_1 + \sum_{r=2}^R \frac{m_r}{m_r + \cdots + m_R} \\
&= 1 - \sum_{r=1}^{R-1} \theta^r + \sum_{r=2}^R \frac{\theta^{r-1}}{\theta^{r-1} + \cdots + \theta^{R-1}}.
\end{aligned}$$

We provided an argument previously for the first line. The second line holds by simple algebra. The third line holds since it is equally profitable to sell the vaccine to all consumers at price hx_1 or to consumers in risk classes r and above at price hx_r, so that $hx_1 = hx_r(m_r + \cdots + m_R)$, implying $x_r = x_1/(m_r + \cdots + m_R)$. The last line holds by substituting for $\{m_r\}_{r=1}^R$ from equation (A2). Taking limits,

$$\lim_{\theta \to 0} \left(\frac{\Pi_t}{\Pi_v} \right) = 1 - 0 + \sum_{r=2}^R 1 = R.$$

This shows that for any $\epsilon > 0$, and for the definitions of the parameters in (A2) and (A3), we can find $\theta > 0$ such that $\Pi_t/\Pi_v > R - \epsilon$. To prove $\Pi_t/\Pi_v \leq R$ for all distributions of consumers into R risk classes, note

$$\begin{aligned}
R\Pi_v &= R \max_{r \in \{1,\ldots,R\}} \left\{ hx_r \left(1 - \sum_{i=1}^{r-1} m_i \right) \right\} \\
&\geq R \max_{r \in \{1,\ldots,R\}} \{ hx_r m_r \} \\
&\geq \sum_{r=1}^R hx_r m_r \\
&= \Pi_t.
\end{aligned}$$

Hence $\Pi_t/\Pi_v \leq R$. Q.E.D.

Proof of Proposition 8

As stated in the text, for a given price P_v, the rational-expectations equilibrium vaccine quantity is given by the intersection of lines AA' and BB' in Figure 3. First we will compute the vaccine demand correspondence AA'. The probability a newborn will ever become infected if he or she is not vaccinated is $\beta I_v/(\beta I_v + \delta)$. Thus a consumer's maximum willingness to pay for the vaccine is $h\beta I_v/(\beta I_v + \delta)$. Consumers are indifferent between buying a vaccine and not if $P_v = h\beta I_v/(\beta I_v + \delta)$, an equation which can be inverted to yield the cutoff infection level $\hat{I} = \delta P_v/[\beta(h - P_v)]$. Thus AA' is given by

$$Q_v(I_v, P_v) = \begin{cases} 0 & \text{if } I_v < \hat{I}_v \\ [0,1] & \text{if } I_v = \hat{I}_v \\ 1 & \text{if } I_v > \hat{I}_v. \end{cases} \tag{A4}$$

Next, we will compute BB', the infection level from the epidemiological model. By the equation for I^* in (10), we have $I_v = 1 - \xi - \lambda$. But $Q_v = \delta\xi$, implying

$$I_v(Q_v) = 1 - \frac{Q_v}{\delta} - \lambda. \tag{A5}$$

Solving (A4) and (A5) simultaneously yields

$$Q_v = \delta\left[1 - \lambda\left(1 + \frac{P_v}{h - P_v}\right)\right]. \tag{A6}$$

Maximizing flow profit P_vQ_v with respect to P_v, where Q_v is given by (A6) yields a first-order condition, which can be expressed as

$$P_v^2 - 2P_vh + h^2(1 - \lambda) = 0.$$

This is a quadratic equation with two solutions: $P_v = h(1 + \sqrt{\lambda})$ and $P_v = h(1 - \sqrt{\lambda})$. The first solution exceeds h and thus would result in zero demand. We will thus use the second solution, $P_v^* = h(1-\sqrt{\lambda})$. By (A6), $Q_v^* = \delta(1-\sqrt{\lambda})$. Hence $\Pi_v^* = P_v^*Q_v^* = h\delta(1-\sqrt{\lambda})^2$. The flow social benefit from the vaccine, W_v^*, equals the foregone harm from the disease h times the flow of newborns δ times the reduction in the proportion of infecteds in the population. From the equation for I^* in (10), the proportion of the population that is infected in the steady state with a vaccine is $1 - \xi - \lambda$ and without a vaccine is $1 - \lambda$, the latter found by substituting $\xi = 0$ in (10). Thus,

$$\begin{aligned} W_v^* &= h\delta\xi \\ &= h\delta\frac{Q_v^*}{\delta} \\ &= h\delta(1 - \sqrt{\lambda}). \end{aligned}$$

Q.E.D.

Proof of Proposition 9

We have

$$\begin{aligned}\Pi_t^* &= h\delta(1-\lambda) \\ &= h\delta(1-\sqrt{\lambda})(1+\sqrt{\lambda}) \\ &> h\delta(1-\sqrt{\lambda})(1-\sqrt{\lambda}) \\ &= \Pi_v^*.\end{aligned}$$

The first line holds by Proposition 7, the second line by simple algebra, the third line by $\sqrt{\lambda} > 0$ and the fourth by Proposition 8. Thus $\Pi_t^*/W_t^* = 1$ but $\Pi_v^*/W_v^* = 1 - \sqrt{\lambda}$. To complete the proof, note 100 percent of gross consumer surplus is extracted by the drug treatment, so there is no additional benefit from also developing a vaccine. Q.E.D.

Proof of Proposition 10

To compute the first limit in the proposition,

$$\begin{aligned}\lim_{\lambda\to 1}\frac{\Pi_t^*}{\Pi_v^*} &= \lim_{\lambda\to 1}\frac{h\delta(1-\lambda)}{h\delta(1-\sqrt{\lambda})^2} \\ &= \lim_{\lambda\to 1}\frac{(1-\sqrt{\lambda})(1+\sqrt{\lambda})}{(1-\sqrt{\lambda})(1-\sqrt{\lambda})} \\ &= \lim_{\lambda\to 1}\frac{1+\sqrt{\lambda}}{1-\sqrt{\lambda}} \\ &= \infty.\end{aligned}$$

The first line holds by the expressions for profits in Propositions 7 and 8 and the remainder by simple algebra. The second limit in the proposition is $\lim_{\lambda\to 1}(\Pi_v^*/W_v^*) = \lim_{\lambda\to 1}(1-\sqrt{\lambda}) = 0$, where the first equality holds by Proposition 9. Q.E.D.

References

Bell, D. C. and R. A. Trevino. (1999) "Modeling HIV Risk,"Journal of Acquired Immune Deficiency Syndromes. 22: 280–287.

Boulier, B., T. S. Datta, and R. S. Goldfarb. (2003) "Vaccination Externalities," George Washington University mimeo.

Geoffard, P-Y., and T. Philipson. (1997) "Disease Eradication: Private vs. Public Vaccination," American Economic Review. 87: 222–230.

Kim-Farley, R. et al. (1992) "Global Immunization," Annual Review of Public Health. 13: 223–37.

Kremer, M. (2001) "Creating Markets for New Vaccines: Part I: Rationale, Part II: Design Issues," in Innovation Policy and the Economy. Adam B. Jaffe, Josh Lerner, and Scott Stern, eds. Cambridge: MIT Press, pp. 35–118.

Malueg, D. A. (1993) "Bounding the Welfare Effects of Third-Degree Price Discrimination," American Economic Review. 83: 1011–1021.

Miguel, E. and M. Kremer. (2003) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," forthcoming, Econometrica.

Moeti, Matshidiso. (2003) Press release for the World Health Organization, WHO workshop in Zimbabwe, 9 July.

Rosenberg, E. (1999) "Drug Makers Shy from Work on AIDS Vaccine," San Francisco Examiner. March 16.

Thomas, P. (2001) Big Shot: Passion, Politics, and the Struggle for an AIDS Vaccine. New York: Public Affairs.

Thomas, P. (2002) "The Economics of Vaccines," Harvard Medical International (HMI) World. September/October.

UNAIDS. (2000) UNAIDS Epidemiological Fact Sheets by Country.

United States Center for Disease Control. (2000) National Health and Examination Survey (NHANES) 1999–2000.

Varian, H. R. (1989) "Price Discrimination," in R. Schmalensee and R. D. Willig, eds., Handbook of Industrial Organization. New York: North-Holland, 597–654.

Wawer, M. J. et al. (2003) "HIV-1 Transmission per Coital Act, by Stage of HIV Infection in the HIV+ Index Partner, in Discordant Couples, Rakai, Uganda," Tenth Conference on Retroviruses and Opportunistic Infections, Boston, abstract 40.

Williamson, O. (1975) Markets and Hierarchies: Analysis and Antitrust Implications. New York: Free Press.

World Bank. (2000) World Development Indicators 2000. Washington, DC.

World Health Organization. (2000) WHO Global Tuberculosis Control 2000. Geneva.

World Health Organization. (2001) Macroeconomics and Health: Investing in Health for Economic Development. Geneva.

Table 1: Calculations of Vaccine Revenue for NHANES Sample

Number of Partners (n)	Fraction of Population in Risk Class	Maximum Price Inducing Purchase ($= n$)	Quantity Sold	Vaccine Revenue
0	0.1007	0	1.0000	0.0000
1	0.7232	1	0.8991	0.8991
2	0.0845	2	0.1759	0.3518
3	0.0369	3	0.0914	0.2742
4	0.0162	4	0.0545	0.2180
5	0.0128	5	0.0383	0.1915
6	0.0075	6	0.0255	0.1530
7	0.0008	7	0.0180	0.1260
8	0.0033	8	0.0172	0.1376
9	0.0008	9	0.0139	0.1251
10	0.0029	10	0.0131	0.1310
12	0.0012	12	0.0102	0.1224
13	0.0004	13	0.0090	0.1170
14	0.0004	14	0.0086	0.1204
15	0.0025	15	0.0082	0.1230
19	0.0004	19	0.0057	0.1083
20	0.0025	20	0.0053	0.1060
27	0.0004	27	0.0028	0.0756
30	0.0004	30	0.0024	0.0720
50	0.0004	50	0.0020	0.1000
100	0.0004	100	0.0016	0.1600
111	0.0004	111	0.0012	0.1332
150	0.0004	150	0.0008	0.1200
255	0.0004	255	0.0004	0.1020

Table 2: Calculations of Vaccine Revenue for Bell and Trevino Sample

Ex ante Risk Class	Fraction of Population in Risk Class	Maximum Price Inducing Purchase	Quantity Sold	Vaccine Revenue
0.0000	0.5852	0.0000	1.0000	0.0000
0.0625	0.1296	0.0625	0.4148	0.0259
0.1250	0.0667	0.1250	0.2852	0.0356
0.1875	0.0444	0.1875	0.2185	0.0410
0.2500	0.0185	0.2500	0.1741	0.0435
0.3125	0.0037	0.3125	0.1556	0.0486
0.3750	0.0185	0.3750	0.1519	0.0569
0.4375	0.0185	0.4375	0.1333	0.0583
0.5000	0.0185	0.5000	0.1148	0.0574
0.5625	0.0037	0.5625	0.0963	0.0542
0.6250	0.0000	0.6250	0.0926	0.0579
0.6875	0.0000	0.6875	0.0926	0.0637
0.7500	0.0148	0.7500	0.0926	0.0694
0.8125	0.0037	0.8125	0.0778	0.0632
0.8750	0.0148	0.8750	0.0741	0.0648
0.9375	0.0148	0.9375	0.0593	0.0556
1.0000	0.0444	1.0000	0.0444	0.0444