Why is There No AIDS Vaccine?

Michael Kremer

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

Christopher M. Snyder

George Washington University

October 6, 2004

Abstract: We argue that differences in timing of drug and vaccine consumption will lead firms to be biased against developing vaccines. Vaccines are sold before consumers are infected, when they still have private information regarding their infection risk, whereas drugs are sold after consumers are infected, when those with positive valuation have no private information on infection risk. Calibrations suggest that, for sexually transmitted diseases, for which infection risk is highly heterogeneous across consumers, producer surplus from drugs may exceed that from vaccines by a factor of four. Consistent with the model, empirical tests suggest vaccines are particularly unlikely to be developed for sexually transmitted diseases. Biases against vaccines are exacerbated by the durability of vaccines and by the interaction between the timing of vaccine and drug consumption and the temporary protection of intellectual property rights. We extend the analysis to allow for government procurement and for income heterogeneity among consumers. Given that antiretroviral drugs are difficult to deliver in the poor countries where most people infected with HIV/AIDS live, biases against developing a vaccine raise enormous public health concerns.

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 Emmanuelle Auriol, Chris Avery, Bryan Boulier, James Dana, Esther Duflo, Glenn Ellison, Corinne Langinier, David Malueg, David McAdams, Sendhil Mullainathan, Michael Schwarz, Andrew Segal, Lars Stole, and seminar participants at Dartmouth, Harvard, M.I.T., Northeastern, Northwestern, U.C.L.A. School of Public Health, University of Pennsylvania, University of Rochester, University of Toronto, the IAEN Symposium on the Economics of AIDS/HIV in Developing Countries (Barcelona), the International Industrial Organization Conference (Boston), the NBER Summer Institute, and the IDEI Conference on Markets for Pharmaceuticals and the Health of Developing Nations for helpful comments, and Ruben Enikolopov, Heidi Williams, and Dan Wood for excellent research assistance. Snyder thanks the George Washington University Facilitating Fund and the Stigler Center for Study of Economy and State at the University of Chicago for financial support.

Why is There No AIDS Vaccine?

Abstract: We argue that differences in timing of drug and vaccine consumption will lead firms to be biased against developing vaccines. Vaccines are sold before consumers are infected, when they still have private information regarding their infection risk, whereas drugs are sold after consumers are infected, when those with positive valuation have no private information on infection risk. Calibrations suggest that, for sexually transmitted diseases, for which infection risk is highly heterogeneous across consumers, producer surplus from drugs may exceed that from vaccines by a factor of four. Consistent with the model, empirical tests suggest vaccines are particularly unlikely to be developed for sexually transmitted diseases. Biases against vaccines are exacerbated by the durability of vaccines and by the interaction between the timing of vaccine and drug consumption and the temporary protection of intellectual property rights. We extend the analysis to allow for government procurement and for income heterogeneity among consumers. Given that antiretroviral drugs are difficult to deliver in the poor countries where most people infected with HIV/AIDS live, biases against developing a vaccine raise enormous public health concerns.

JEL codes: O31, L11, I18, D42

1 Introduction

> 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.
>
> —Patricia Thomas, author of *Big Shot: Passion, Politics, and the Struggle for an AIDS Vaccine* (Thomas 2001), quoted from Thomas (2002)

More than 40 million people are infected with HIV/AIDS, 95 perfect of whom live in developing countries. Because antiretroviral drugs are difficult to deliver in the poorest countries, vaccines arguably offer the best hope for defeating the epidemic (International AIDS Vaccine Initiative 2002).[1] Yet private investment in HIV/AIDS vaccine research remains minimal relative to both the health burden of the disease and to investments in antiretroviral drug research.[2] This paper explores whether economic factors could create gaps between social and private incentives to invest in vaccines relative to drugs that might help explain this gap in investment. Although our analysis focuses on the case of HIV/AIDS, much of our work is also applicable to other sexually transmitted diseases and, more broadly, to other diseases for which there is substantial heterogeneity in risk of infection.

Thomas' (2002) view that firms prefer drugs to vaccines because drugs are administered more frequently appears to be widely held (e.g., see also Rosenberg 1999). Yet from the perspective of neoclassical economics, this explanation seems odd. In the benchmark case of a risk-neutral, rational consumer with no credit constraints, the consumer would be willing to pay the expected present value of the stream of benefits in an up-front lump sum for a vaccine, and thus it might seem that vaccines and drugs should yield equivalent revenues if they are equally efficient technologically.

Consumer myopia or other forms of irrationality of course may lead drugs to be more profitable than

[1] Unlike vaccines, drugs require diagnosis, often must be taken on a long-term basis, and frequently have side effects that require monitoring by highly trained medical personnel, who are scarce in the poorest countries. Only 50,000 of the 30 million people with HIV/AIDS in Africa are using antiretroviral therapies (Moeti 2003), while three quarters of the world's children receive a standard package of vaccines (Kim-Farley et al. 1992).

[2] The International AIDS Vaccine Initiative (2002) estimates total investment in HIV/AIDS vaccine R&D at between $430 and $470 million, only between $50 and $70 million of which has come from private industry. As of this writing at least 20 antiretroviral drugs have been approved by the U.S. Food and Drug Administration. Huff (2003) cites the total R&D investment for the most recently approved antiretroviral drug (T-20, or Enfuvirtide) at $600 million. DiMasi, Hansen, and Grabowski (2003) estimate that an average of $802 million in R&D investment is required to get a new medicine from lab to patient.

vaccines. Drugs will also be more profitable if they are cheaper to develop, are more effective cures, and have fewer side effects. However, in this paper we show revenue equivalence can break down even in the benchmark case, because developers of the two medicines differ in their ability to capture the social value of their innovation due to differences in the timing of the administration of vaccines and drugs.

If consumers differ in their ex ante infection risk due to differences, for example, in the number of sexual partners, monopolists will be able to extract more surplus with drugs than vaccines. This result hinges on the interaction between the heterogeneity in infection risk and the differential timing of the administration of the medicines. Vaccines are administered before the disease is contracted, when consumers still have private information regarding their infection risk; drugs are sold after the disease is contracted, when consumers with positive valuation no longer have private information regarding their infection risk. The reduction in private information in moving from vaccines to drugs allows the firm to extract more surplus with drugs.

A simple example illustrates this point. Suppose that out of 100 people, 90 have a 10 percent chance of contracting a disease and 10 have a 100 percent chance. Let the harm from the disease be $100. For simplicity, assume consumers are risk-neutral, and thus are willing to pay $10 for each 10 percent reduction in their chance of getting the disease and $100 to be cured if they contract the disease. Suppose the medicines are perfectly effective, have no side effects, and are costless to manufacture. If the firm develops a drug, it sells to all people who contract the disease at a price of $100. In expectation, 19 consumers contract the disease (all 10 high-risk consumers, along with 9 low-risk consumers). So expected drug revenue is $1,900, which corresponds to the social value of the product. In contrast, if the firm develops a vaccine, it could either charge $100 and sell only to the 10 high-risk consumers, or charge $10 and sell to all 100 consumers. Either way, the firm's vaccine revenue is $1,000, only about half the revenue from a drug and only about half the social value of the product.

In Section 3, we prove that for any distribution of infection risk with a nontrivial amount of consumer heterogeneity, a drug yields more revenue than a similarly effective vaccine. The ratio of drug to vaccine revenue is less than two for left-skewed distributions, equal to two for uniform distributions, greater than two for right-skewed distributions, and can be arbitrarily high for highly skewed distributions of infection risk.

Empirically, distributions of numbers of sexual partners, and hence disease risk, are extremely skewed. In Section 4 we calibrate our model with data on rates of partner change in the United States. Our calibration suggests drug revenue could exceed vaccine revenue by more than a factor of four.

In Section 5, we provide evidence that, consistent with the model's prediction, drugs are significantly more likely, and vaccines significantly less likely, to have been developed for sexually transmitted diseases.

We then consider a series of extentions to the basic model. Since governments are large purchasers of pharmaceuticals in many countries, in Section 6 we allow for government procurement. We argue that, if the prices the government pays are influenced by the threat point of profits the firm could realize on the private market if bargaining breaks down, to the extent that vaccines are less profitable than drugs on the private market they will also be less profitable when sold to the government. This suggests a potential rationale for committing in advance to purchase vaccines at a prespecified price or for subsidizing vaccine R&D more than drug R&D.

In Section 7, we address the complication that heterogeneity among consumers in willingness to pay for a drug conditional on being infected (e.g., due to income) may prevent a drug monopolist from appropriating all consumer surplus. We show that, if income covaries negatively with risk of infection and firms cannot price discriminate based on income, vaccines may be relatively more profitable than drugs. Calibrations suggest that if firms' ability to engage in international price discrimination broke down, incentives to develop HIV/AIDS drugs could fall below incentives to develop vaccines.

In Section 8, we examine implications of the differing durability of vaccines and drugs for rent extraction. Vaccines provide durable protection against disease, whereas antiretroviral drugs provide only temporary relief. We confirm the intuition from standard analyses of the durable-good-monopoly problem (see, e.g., Coase 1972; Stokey 1981; Bulow 1982; and Gul, Sonnenschein, and Wilson 1986) that in the presence of heterogeneous consumers durability leads to a commitment problem. This is yet another factor biasing firms against developing vaccines. It turns out that vaccines are slightly different from standard durable goods, so new formal analysis is needed. High-demand vaccine consumers are at high risk of contracting the disease; if they contract the disease, the vaccine cannot cure them, reducing their incentive to wait for lower vaccine prices in the future.

In Section 9, we consider the impact of differences in the timing of drug and vaccine use when drug

and vaccine developers have temporary market power but must compete with each other and, after some delay, must compete with "me-too" products and generics. We show that competition can exacerbate the bias against vaccines. Drug developers are able to capture significant rents during the temporary period in which they have market power by serving the initial stock of infected consumers. Rents are difficult to capture with vaccines because vaccines cannot be used to treat the initial stock of infected consumers. Vaccines can only be used by subsequent generations, who will not be willing to pay much for vaccines if they anticipate entry of cheap generic drugs in the future.

Our work is related to the industrial organization literature on monopoly pricing when consumers gradually learn their demands. Lewis and Sappington (1994) and Courty (2003) assume consumers are initially identical, whereas we assume consumers have private information about their infection risk ex ante. Courty and Li (2000) compare optimal ex ante and ex post schemes under general conditions, where ex ante schemes are allowed to involve refunds. Refunds are impossible for vaccines because, once the vaccine is administered, the benefit is inalienable from the consumer. Clay, Sibley, and Srinagesh (1992) and especially Miravete (1996) are closest to our work. Our application calls for a specific mapping from ex ante private values into ex post types, whereas Miravete considers general functional forms for the mapping. The specificity in this one dimension allows us to examine general distributions of ex ante infection risk rather than the particular class of beta distributions examined by Miravete, and to establish bounds on the profit ratio both in the limit and as a function of skewness of the infection risk, all of which are new results in the literature. Our analysis of social welfare in Section 3, empirical analysis in Sections 4 and 5, and theoretical extensions in Sections 6 through 9 are new as well.

In a companion paper (Kremer, Snyder, and Williams 2004), we examine another reason firms can appropriate more consumer surplus with drugs than with vaccines. Vaccines are more likely than drugs to interfere with disease transmission. We build an integrated economic and epidemiological model and find that the revenue gap between drugs and vaccines, and the ratio of social-to-private value, will be largest in the case of rare diseases, and indeed can be arbitrarily large in percentage terms for sufficiently rare diseases. Thus, holding constant the total burden of disease, firms will find developing vaccines for the common but less serious diseases like the flu more profitable than for rarer but more deadly diseases. Since HIV/AIDS is rare in the high-income countries that account for the bulk of pharmaceutical revenue,

the model suggests that firms will be able to capture a greater fraction of the social value of drugs than of vaccines.

Finally, this paper is also related to recent empirical work on the effect of market size (Acemoglu and Linn 2003) and public policies (Finkelstein 2004) in determining pharmaceutical R&D investments for a given product. Here, we present a theoretical model, calibrations, and empirical evidence suggesting that R&D incentives across products are distorted, that these distortions can be large, and that they have affected the development of medicines for sexually transmitted diseases. To the extent these distortions have inhibited HIV/AIDS vaccine research, the welfare consequences are potentially enormous.

2 Model

Suppose a monopoly pharmaceutical manufacturer, called the firm, has the choice of developing a vaccine or a drug. For the purposes of this model, we will define a vaccine as a medicine administered as a preventative measure before a disease is contracted and define a drug as a medicine administered after a disease has been contracted.[3]

The firm chooses whether to develop a vaccine, a drug, or both. Let $k_j \in [0, \infty)$ be the present discounted value of the fixed cost of developing medicine j, where $j = v$ for the vaccine and $j = d$ for the drug. 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 drug may be administered later in a consumer's life than a vaccine, and so the nominal cost of the drug may be discounted more heavily than the vaccine, but such discounting is reflected in c_j since it is expressed as a present discounted value. Let $e_j \in [0, 1]$ be the efficacy of medicine j, that is, the probability that medicine j prevents the consumer from experiencing harm from the disease. Let $s_j \in [0, 1]$ be the expected present discounted harm of side effects from medicine j, that is, the probability that a consumer experiences side effects multiplied by the present discounted value of the harm from the side effects conditional on experiencing them. Assume the events that the medicine is ineffective or produces side effects are independent for a given consumer, and each are independent across

[3] We recognize that not all medicines fit neatly in these definitions. For example, some vaccines, called therapeutic vaccines, boost the immune systems of individuals who are already infected, and thus would be technically classified as drugs for the purposes of our model. For another example, statins function as both cholesterol-reducing drugs and as heart-disease preventatives, and thus could be considered a hybrid case for the purposes of our model.

consumers. Let $p_j \in [0, \infty)$ be the present discounted value of the price the firm receives for medicine j.[4] For $j = v, d, b$, where b represents the firm's developing both medicines, let π_j be producer surplus, $\Pi_j = \pi_j - k_j$ be profit, CS_j be consumer surplus, and $W_j = CS_j + \Pi_j$ be social welfare. Let $\tilde{W}_j$ be social welfare in the benchmark in which prices are set by a social planner at marginal cost.

Before purchasing any medicine, consumer i learns his or her infection risk $x_i \in [0, 1]$. Assume x_i is a random variable with 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 information for the consumer; the firm only knows the distribution from which x_i is drawn.[5] This assumption captures 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/AIDS, but such behaviors would be difficult for a firm to monitor accurately enough to be able to charge a discriminatory price. Although our focus here is on HIV/AIDS, this type of heterogeneity is relevant for other diseases: for example, frequenting mosquito-infected tropical regions increases the chances of contracting malaria but, again, may be difficult to monitor accurately.

Define $\Phi(\hat{x}) = \Pr(x_i \geq \hat{x}) = \int_{\hat{x}}^{1} dF(x_i)$, implying that, if the distribution of infection risk is continuous, $\Phi(\hat{x}) = 1 - F(\hat{x})$, while if the distribution is discrete or mixed with an atom at $\hat{x}$, $\Phi(\hat{x}) = 1 - F(\hat{x}) + \Pr(\hat{x})$. Define the expectations operator $E(x_i) = \int_0^1 x_i \, dF(x_i)$.

Whether or not consumer i contracts the disease is represented by Bernoulli random variable σ_i, where $\sigma_i = 1$ indicates i contracts the disease, an event that occurs with probability x_i, and $\sigma_i = 0$ indicates i does not contract the disease, an event that occurs with probability $1 - x_i$. The key difference between a vaccine and a drug thus hinges on when the medicine is administered relative to the realization of σ_i. A vaccine is administered before σ_i is realized and a drug is administered after. Throughout most of the paper, we will assume the firm cannot commit to prices ex ante. This means that the drug price will extract all of the surplus of a consumer who contracts the disease. The commitment assumption is without loss

[4] We will assume 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. The results would be identical assuming a *caveat venditor* regime in which the firm bears liability for harm. 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.

[5] Analysis of the case in which x_i is publicly observable but the firm cannot discriminate on x_i is identical.

of generality in the monopoly case since the optimal drug price without commitment is the same as the optimal price with commitment. The assumption of no commitment serves merely to simplify the proofs.

Suppose consumers are risk-neutral. If a consumer contracts a disease and has not had a vaccine or does not receive a drug, he or she experiences harm $h \in [0, \infty)$ in present discounted value terms. Normalize the mass of consumers to unity. To rule out trivial cases, assume

$$e_j h - s_j > c_j \quad \text{for } j = v, d. \tag{1}$$

The assumption in (1) ensures that the producer surplus from serving a consumer with the highest possible infection risk $x_i = 1$ is positive for both medicines. If (1) did not hold for medicine j, it is immediate that the firm would not develop the medicine. Finally, let D be the total social burden of the disease, i.e., $D = hE(x_i)$, a term we will use to normalize our welfare measures in the subsequent analysis.

The next proposition, proved in the Appendix, provides expressions for the firm's profits which we will use to determine which medicine it will develop in equilibrium.

Proposition 1. *The firm's profit from developing a vaccine alone is*

$$\Pi_v = \max_{p_v \in [0,\infty)} \{(p_v - c_v)\Phi(\hat{x}(p_v))\} - k_v \tag{2}$$

from developing a drug alone is

$$\Pi_d = (e_d h - s_d - c_d)E(x_i) - k_d \tag{3}$$

and from developing both medicines is

$$\begin{aligned} \Pi_b = \max_{p_v \in [0,\infty)} \Bigg\{ (e_d h - s_d - c_d) \left[\int_0^{\hat{x}(p_v)} x_i \, dF(x_i) + (1 - e_v) \int_{\hat{x}(p_v)}^1 x_i \, dF(x_i) \right] \\ + (p_v - c_v)\Phi(\hat{x}(p_v)) \Bigg\} - (k_v + k_d), \end{aligned} \tag{4}$$

where $\hat{x}(p_v) = (p_v + s_v)/(e_v h)$.

In equation (2), Π_v is the result of a standard monopoly pricing problem. At a price of p_v, a consumer with infection risk $\hat{x}(p_v)$ is indifferent between purchasing the vaccine and not. The vaccine producer earns markup $p_v - c_v$ for the mass of consumers $\Phi(\hat{x}(p_v))$ with infection risk $x_i \geq \hat{x}(p_v)$. To understand the expression for Π_d in (3), note that the firm sells the drug ex post at a price that extracts the consumer's

entire ex post surplus $p_d^* = e_d h - s_d$. The drug producer earns markup $p_d^* - c_d$ for the mass of consumers $E(x_i)$ who become infected. Finally, to understand the expression for Π_b in (4), consider the two terms in the maximand. The firm sells the vaccine at price p_v to those with infection risk $x_i \geq \hat{x}(p_v)$, generating producer surplus equal to the second term in the maximand. The firm sells the drug ex post at a price that extracts a consumer's entire ex post surplus $p_d^* = e_d h - s_d$ to those who contract the disease, both those with infection risk $x_i < \hat{x}(p_v)$ who did not purchase the vaccine and those with $x_i \geq \hat{x}(p_v)$ who purchased the vaccine but for whom the vaccine was ineffective. The producer surplus from the drug equals the first term in the maximand. Because the drug is priced to extract a consumer's entire ex post surplus, the presence of the drug does not affect the consumer's vaccine consumption decision, so the cutoff $\hat{x}(p_v)$ has the same functional form whether the vaccine is developed alone as in (2) or together with the drug as in (4).[6]

We can use the profit expressions from Proposition 1 to characterize which medicine the firm develops in equilibrium. It develops a vaccine alone if $\Pi_v > \max(\Pi_d, \Pi_b, 0)$, a drug alone if $\Pi_d > \max(\Pi_v, \Pi_b, 0)$, both if $\Pi_b > \max(\Pi_v, \Pi_d, 0)$, and neither if $\max(\Pi_v, \Pi_d, \Pi_b) < 0$.

Comparative statics with respect to k_j, c_j, e_j, and s_j are straightforward. *Ceteris paribus*, the firm prefers to develop medicine j, either alone or together with the other medicine, if medicine j is cheap to develop (k_j is low), cheap to produce a dose (c_j is low), involves mild side effects (s_j is low), and is an effective cure (e_j is high).[7]

Examination of equations (2) and (3) reveals some inherent factors in favor of drugs. The generalized marginal cost of administering medicine j—generalized to include both the cost of producing a dose c_j and the cost of side effects s_j—is borne with certainty by each consumer who is vaccinated whether or not

[6]The proof of Proposition 1 allows for the full range of complicated mechanisms to sell vaccines. The proof shows that the simple mechanism of selling the vaccine at a linear price is optimal. The assumption that the firm cannot commit ex ante to a drug price restricts the possible mechanisms that can be used to sell drugs to the one we consider—a linear price that extracts an infected consumer's entire ex post surplus. However, it can be shown that allowing more complicated mechanisms that would be feasible with commitment would not increase the firm's profit.

[7]The model does not exhaust the list of factors that might lead the firm to prefer vaccines over drugs or vice versa. It is straightforward to extend the model to consider alternative factors. For example, if consumers have diminishing marginal utility of other consumption, vaccines would become relatively more profitable, since they would provide insurance to consumers for which consumers would pay a premium. This premium may be substantial, since consumers would presumably be willing to pay a high fraction of income for a drug, and hence would be relatively risk-averse. If consumers face per-period liquidity constraints, the constraint may bind less with drugs since the total payment with drugs may be spread out in installments (with a payment for each separate drug treatment), whereas the total payment for the vaccine would need to be paid in a lump sum at the time the vaccine is administered.

they would eventually have contracted the disease, but is only borne by consumers who actually contract the disease with a drug. Hence the cost of administering a drug will tend to be lower *ceteris paribus*. On the other hand, it could be argued that other inherent factors favor vaccines. Since people often learn they have a disease only after suffering some harm from symptoms, whereas a vaccine, if effective, can prevent the appearance of any symptoms. To capture this factor, one could increase e_v relative to e_d in the model, favoring vaccines.

All of the factors from the previous paragraph and the comparative static effects from the paragraph before that influence private and social R&D incentives the same way. Indeed, in the special case of homogeneous consumers, we can make a stronger statement: Proposition 2 states that if consumers are homogeneous, private and social R&D incentives are identical.

Proposition 2. *Assume x_i takes on a single, known value in the population of consumers, implying there is no heterogeneity in the distribution of infection risk. Then the firm's choice of whether to develop a vaccine, drug, both, or neither is identical to a social planner's.*

The proposition follows immediately from the fact that, with homogeneous consumers, the monopolist can extract 100 percent of consumer surplus and thus fully internalize social welfare. In the next section, we argue that heterogeneity in infection risk drives a wedge between private and social incentives, leading to a bias against vaccines.

3 Distribution of Infection Risk

The analysis is simplest in the case in which $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$: both medicines are costless to produce and administer, have no side effects, and are perfectly effective. Our positive analysis will focus on this case throughout the remainder of the paper, although the key normative propositions will hold for general parameters.

Given $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$, developing the medicine with a lower development cost k_j is socially efficient. By Proposition 2, if consumers are homogeneous, the monopolist develops the lower-k_j medicine. If consumers are heterogeneous in their infection risk, there may be a bias against developing a vaccine in that the monopolist chooses to develop the drug over the vaccine even though $k_v < k_d$. We will adopt the following formal measure of what we mean by this "bias" against vaccines in

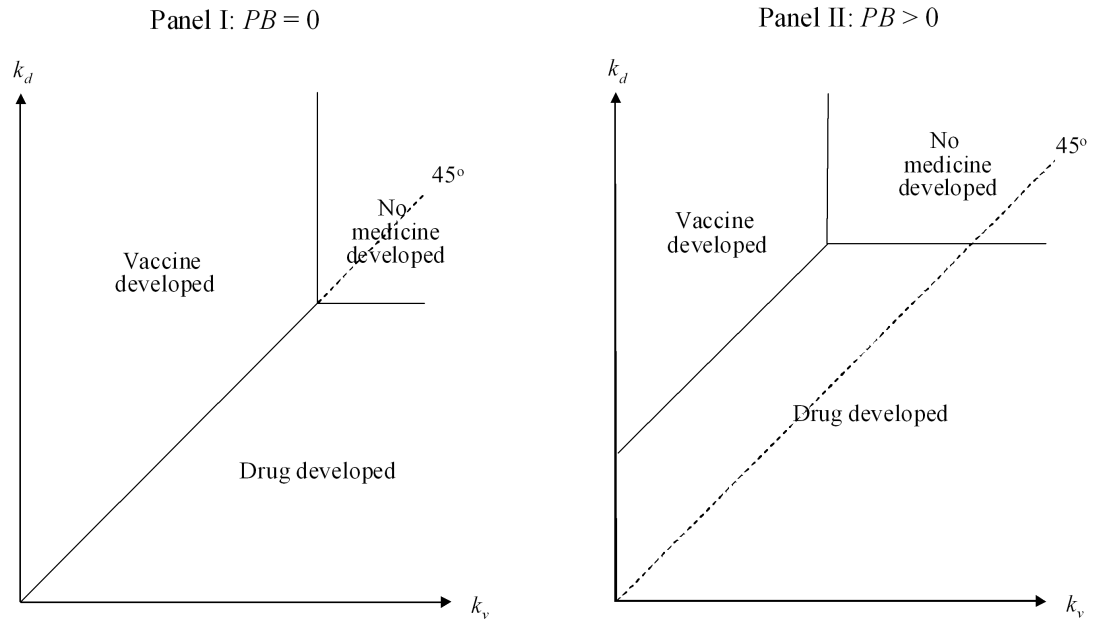

Figure 1: *PB* as a measure of bias in the firm's private incentives against developing a vaccine. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$.

firms' private incentives:

$$PB = \sup_{(k_v,k_d)\in[0,\infty)^2} \left\{ \left(\frac{k_d - k_v}{D} \right) \mathbf{1}\big(\Pi_d > \max(\Pi_v, 0)\big) \right\} \tag{5}$$

where $\mathbf{1}$ is the indicator function. In words, *PB* is an upper bound on how much more developing a drug could cost than a vaccine, but the firm still develops the drug.[8] It is expressed as a percentage of the total social burden of the disease, D. The idea behind *PB* is that, if one had an uninformed prior over the space of free parameters (k_v, k_d), if $PB = 0$, there would be no bias in favor of drugs in that the set of parameters for which a drug is developed has the same measure as the set for which a vaccine is developed (Figure 1, Panel I). If $PB > 0$, the set of parameters for which a drug is developed has a larger measure than the set for which a vaccine is developed (Figure 1, Panel II). Another nice feature of *PB* is that, given our parametric assumptions, there is a simple mapping between it and the ratio of producer

[8] An analogous measure of the bias against drugs could be defined. Proposition 4 implies that this analogous measure would never be positive.

surplus π_v/π_d, as the next proposition shows.

Proposition 3. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Then $PB = 1 - (\pi_v/\pi_d)$.*

By Proposition 3, we can equivalently consider *PB* or the ratio π_v/π_d to analyze the firm's bias against vaccines.

As argued above, if there is no heterogeneity in the distribution of infection risk, the monopolist develops the lower-k_j medicine, which in turn implies $PB = 0$. The next proposition, proved in the Appendix, shows that, if there is any heterogeneity in infection risk, then $PB > 0$; and so there is a bias against vaccine development.

Proposition 4. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Assume there is nontrivial heterogeneity in the distribution of infection risk; i.e., at least two distinct subintervals of $(0, 1]$ have positive measure. Then $\pi_d > \pi_v$, or, equivalently, $PB > 0$. That is, the firm earns more producer surplus from a drug than a vaccine, and is thus biased against developing the vaccine. The firm weakly prefers not to develop both medicines and strictly prefers not to if $k_v > 0$.*

Proposition 4 implies that if there is heterogeneity in infection risk, then the set of parameters for which different medicines are developed is given by Figure 1, Panel II. The locus of equal fixed costs (the 45 degree line) lies below the region in which vaccines are developed, so it follows that, for equal fixed costs, either a drug is developed or, if fixed costs are sufficiently high, no medicine is developed. Only if k_d is sufficiently high relative to k_v will a vaccine be developed.

Figure 2 provides simple graphical arguments that can be used to establish Proposition 4. Substituting the parametric assumptions from the statement of the proposition into equation (2) and noting $\pi_v = \Pi_v + k_v$, we have $\pi_v = \max_{p_v \in [0,\infty)} \{p_v \Phi(p_v/h)\}$. Now $\Phi(p_v/h)$ is the demand curve for a vaccine. Figure 2 graphs the corresponding inverse demand curve. One can see that π_v is the area of the largest shaded rectangle that can be inscribed under the inverse demand curve. Substituting the parametric assumptions from the statement of the proposition into (3) and noting $\pi_d = \Pi_d + k_d$, we have $\pi_d = hE(x_i)$, which can be shown, integrating by parts, to be equal to the whole area under the inverse demand curve. No matter how the rectangle is inscribed, and no matter the shape of the curve, the area of the rectangle will be less than the area under the whole curve, so $\pi_d > \pi_v$.

We have shown that the firm earns more producer surplus from drugs than from vaccines and in that sense is biased against vaccines, raising the question of how much more producer surplus drugs provide.

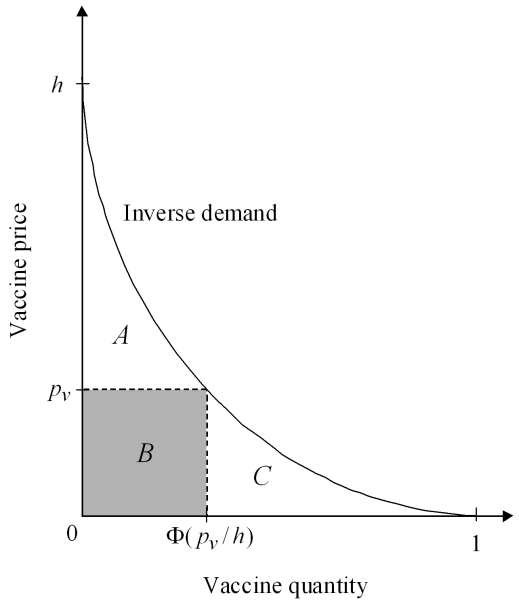

Figure 2: Geometric comparison of producer surplus from vaccines and drugs. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$.)

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. The next proposition, proved in the Appendix, shows that the number of risk classes determines a tight upper bound on the amount the profit from a drug exceeds that from a vaccine, and this proposition will serve as a useful building block for subsequent results.

Proposition 5. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. For any $\epsilon > 0$, there exist distributions of consumers in R risk classes such that $\pi_v/\pi_d < 1/R + \epsilon$. That is, we can find distributions of consumers in R risk classes such that the producer surplus from a vaccine can be made arbitrarily close to $1/R$ times the producer surplus from a drug or, equivalently, PB can be made arbitrarily close to $1 - 1/R$. Moreover, $1/R$ is a lower bound on π_v/π_d.*

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, and so on up to selling to the highest risk class alone at price hx_R.

A corollary of Proposition 5 is that there exist distributions of consumer types such that the producer surplus from vaccines is arbitrarily smaller than that from drugs. This can be seen by taking the limit as R approaches infinity in the proposition. Stated formally, we have the following proposition.

Proposition 6. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. There exist distributions of consumers such that π_v/π_d can be made arbitrarily close to zero or, equivalently, PB can be made arbitrarily close to its theoretical upper bound of 1.*

Proposition 5 has another straightforward corollary, to 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 10 percent chance of contracting the disease and 10 of whom have a 100 percent chance) is such a case. As noted in the Introduction, producer surplus from a vaccine is only 53 percent of that from a drug in this example. Proposition 5 implies that the producer surplus from a vaccine can be as little as half that from a drug in the two-type case, but no less. The example from the Introduction approaches this bound of one half, 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 1 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 example in the Introduction that the harm from the disease is \$100, that producer surplus from a drug is \$1,990 while producer surplus from a vaccine is \$1,000, only slightly more than half as much.

The two-type case provides important insights into the settings in which firms will strongly prefer drugs to vaccines. First, our results suggest that the gap in producer surplus between vaccines and drugs will be especially large in the case of skewed distributions of consumer risk: distributions in which a large segment of the population has a very small probability of contracting the disease and a small segment of the population has a large probability of contracting the disease will create the largest relative incentives for the firm to develop drugs. Proposition 7 provides a more formal statement of the relationship between skewness of the infection-risk distribution and the ratio of producer surplus π_v/π_d.

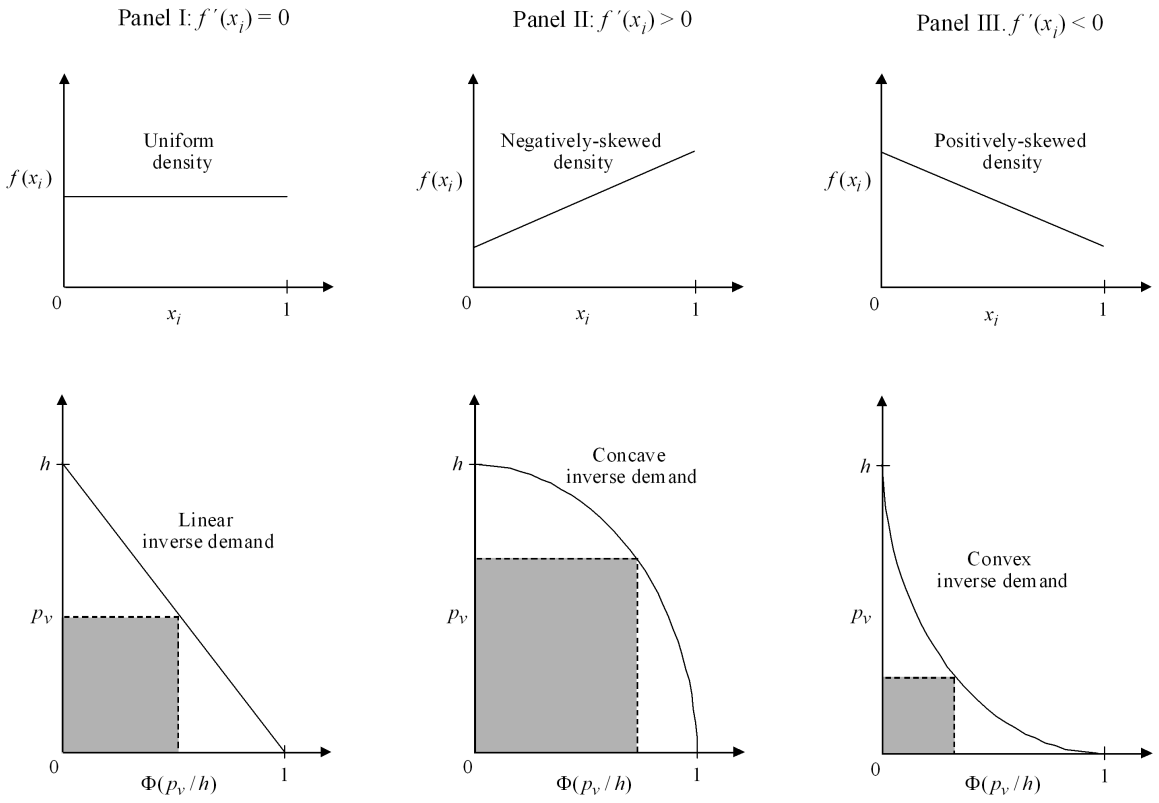

Figure 3: Ratio of producer surpluses depends on skewness of density and curvature of inverse demand. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$.)

Proposition 7. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Let $f(x_i)$ be the density function associated with consumer types x_i. Assume $f(x_i)$ is differentiable. If $f'(x_i) = 0$ (implying x_i is uniformly distributed), then $\pi_v/\pi_d = 1/2$. If $f'(x_i) > 0$ (a sufficient condition for right-skewness), then $\pi_v/\pi_d > 1/2$. If $f'(x_i) < 0$ (a sufficient condition for left-skewness), then $\pi_v/\pi_d < 1/2$.*

The proof, following from results in Malueg (1993), is illustrated in Figure 3. The case $f'(x_i) = 0$ is drawn in Panel I of the figure. If $f'(x_i) = 0$, then x_i is uniformly distributed and has no skewness. The associated inverse demand curve $\Phi(p_v/h)$ can easily be shown to be linear. By calculations similar to the proof of Proposition 1 in Malueg (1993), the area of the largest rectangle that can be inscribed under the curve (π_v in our setting) is half of the area under the curve (π_d in our setting), so $\pi_v/\pi_d = 1/2$. If $f'(x_i) > 0$ as in Panel II of the figure, then the distribution of x_i is left-skewed. The associated inverse demand can be shown to be concave. By Malueg (1993) (and by inspection of Panel II), the area of the largest rectangle that can be inscribed under the inverse demand curve is more than half the area under the

inverse demand curve, so $\pi_v/\pi_d > 1/2$. If $f'(x_i) < 0$ as in Panel III of the figure, then the distribution of x_i is right-skewed, and the associated inverse demand can be shown to be convex. By Malueg (1993) (and by inspection of Panel III), the area of the largest rectangle that can be inscribed under the inverse demand curve is less than half the area under the curve, so $\pi_v/\pi_d < 1/2$. In sum, in the baseline case with x_i following a uniform distribution and thus having no skewness, the producer surplus from vaccines is half that from drugs. Right-skewness increases the bias against vaccines.

Propositions 4 through 7 dealt with the positive question of how the shape of the distribution of infection risk affects the firm's bias against developing a vaccine. We now turn to the normative question of the potential social cost of this bias. The next proposition, proved in the Appendix, states there is socially too little incentive to develop a vaccine relative to a drug.

Proposition 8. *The firm never develops a vaccine unless it is socially efficient to do so. There exist cases in which the firm develops a drug but it would have been socially efficient to develop a vaccine. The results are true whether in the benchmark the social planner chooses medicines but the monopolist chooses prices (so social welfare is W_j) or the social planner chooses both medicines and prices (so social welfare is $\tilde{W}_j$).*

The proof of the proposition is fully general, holding for arbitrary parameters c_j, s_j, and e_j, not just the normalized values $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$.

Having demonstrated that the bias against vaccines is socially inefficient, we turn to an analysis of the conditions under which the social cost is large and how large it can possibly be. To this end, we introduce a measure of social cost that is analogous to our index of the bias in private incentives *PB*:

$$SB = \sup_{(k_v,k_d)\in[0,\infty)^2} \left\{ \left(\frac{W_v - W_d}{D} \right) \mathbf{1}\big(\Pi_d > \max(\Pi_v, 0)\big) \right\}. \tag{6}$$

In words, *SB* is the greatest possible social cost of the bias against vaccines, equal to the loss of welfare if the firm develops a drug when a vaccine would have provided more social welfare. It is expressed as a percentage of the total social burden of the disease D. *SB* ranges from zero for no social cost to a maximum value of 1. The next proposition provides a simple formula for *SB*.

Proposition 9. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Then $SB = CS_v/D$.*

The proof of Proposition 9 provided in the Appendix is more general than stated above. The proposition

continues to hold if we relax our parametric assumptions and redefine *SB* to be the supremum where parameters c_j, s_j, e_j are freely varied in addition to k_j, $j = v, d$. This is because the supremum over these freely varying parameters happens to be attained for the assumed values $c_j = s_j = 0$ and $e_j = 1$. This remark also applies to the other propositions involving *SB*, namely Propositions 10 and 11, below.

Proposition 10. *If Φ is linear, SB $= 1/4$. If Φ is concave, SB $\in [0, 1/3]$. If Φ is convex, SB $\in [1/8, 5/8]$. The bounds are tight in that distributions of infection risk can be constructed such that SB comes arbitrarily close to the bounds.*

Proposition 10 implies that in the benchmark case of a uniform distribution of infection risk, a case argued earlier (Figure 3, Panel I) leads to a linear Φ and thus a linear inverse demand curve, the social cost of the firm's bias toward drugs can be as much as $1/4$ of the total social burden of the disease. If there is positive-skewness in the distribution of infection risk, leading to a convex Φ and thus a convex inverse demand curve, the social cost of the firm's bias against vaccines can be as much as $5/8$ (62.5 percent) of the total social burden of the disease.

4 Calibrations for Sexually Transmitted Diseases

In this section, we calibrate the model for sexually transmitted diseases in general and also for the specific case of HIV/AIDS. We estimate the underlying distribution of infection risk using data on sexual behavior in the U.S. population. The 1999–2000 National Health Examination Survey (U.S. Centers for Disease Control 2000), which we will refer to as NHANES, provides nationally representative data on the lifetime number of sexual partners broken down by the individual's gender and sexual orientation.[9] The distribution of lifetime sexual partners is highly skewed: the median is 4, but the mean is 13.2. Skewness in the distribution of lifetime sexual partners induces skewness in the distribution of infection risk in our calibrations, which in turn leads to a large gap between the producer surplus from a vaccine and a drug.

In our first calibration, we assume the same probability of transmission per partner independent of gender or sexual orientation. We assume the simplest possible mapping, a linear mapping, from lifetime sexual partners to infection risk. Figure 4 graphs the resulting inverse demand curve. The skewed

[9] Due to data limitations, we assume individuals are exclusively heterosexual or homosexual. Data limitations prevent us from accounting for another source of heterogeneity, intravenous drug use, in the calibrations.

distribution of infection risk produces a highly convex inverse demand curve. Recall π_v is given by the area of the largest rectangle that can be inscribed under the curve (the shaded rectangle in the figure) and π_d by the area under the curve. The vertical axis was truncated to make the graph more readable, hiding some of the area under the curve. Still, it is apparent that π_v is much less than π_d. To be precise, $\pi_v/\pi_d = 0.23$. It follows that $PB = 0.77$. Using the formula from Proposition 9, it can be shown $SB = 0.44$.

Our second calibration also applies to sexually transmitted diseases in general. We maintain all the assumptions from the previous paragraph but change the mapping from lifetime sexual partners into infection risk. We will replace the linear mapping with a mapping due to Kaplan (1990). The virtue of the Kaplan over the linear mapping is that it allows for concavity that is consistent with available medical evidence (see, e.g., Winkelstein et al. 1987 for a study of the mapping for homosexual males in San Francisco). More sophisticated models would require parameters for which we do not have good proxies. Following Kaplan (1990), suppose there is a constant probability β of being infected from each partner regardless of past contact history. Then a person with n sexual partners would have probability $1-(1-\beta)^n$ of ever contracting the disease. We will take $\beta = 0.06$ percent, equal to an estimate of the current HIV/AIDS prevalence rate in the United States, which according to UNAIDS (2004) is 0.6 percent, times the average per-partner transmission rate, which following Roskstroh et al. (1995) we will take to be 10 percent. Figure 5 graphs the resulting inverse demand curve. One can compute $\pi_v/\pi_d = 0.24$, $PB = 0.76$, and $SB = 0.56$.

Our last calibration is directed toward the specific case of HIV/AIDS. An additional possible source of heterogeneity with HIV/AIDS beyond the heterogeneity in the rate of partner change is that the prevalence rate is much higher for a small subgroup, homosexual males, than for the general population. The NHANES data allow us to estimate the distribution of lifetime sexual partners for homosexual males separately from that for the rest of the population. We will use the Kaplan model to map lifetime sexual partners into infection risk. For homosexual males, we will take $\beta_h = 4.3$ percent, equal to the estimated HIV/AIDS prevalence rate among homosexual males in the United States of 14 percent times the per-partner transmission rate for homosexual males.[10] Royce *et al*. (1997) estimate the male-to-male transmission rate

[10] These prevalence rates can be computed as follows. Let P_h (respectively, P_o) be the population of homosexual males (others) and β_h (β_o) be the prevalence rate among homosexual males (others). P_h and P_o can be estimated from the NHANES data.

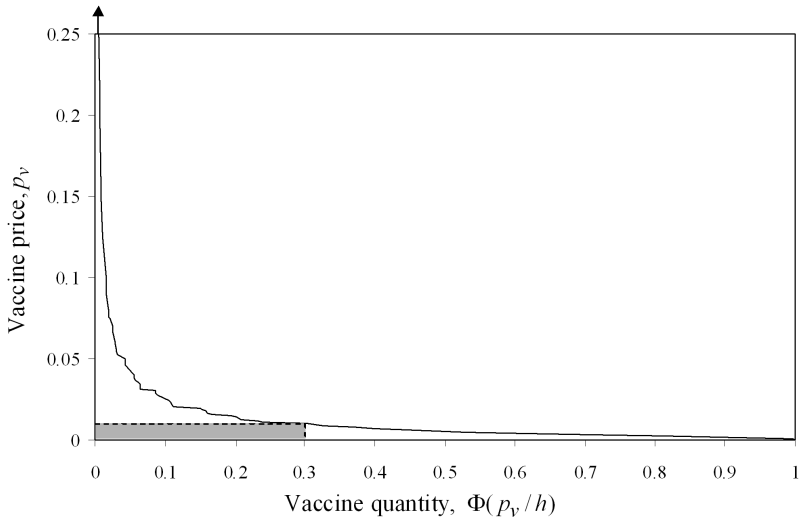

Figure 4: Inverse demand curve for sexually transmitted-disease calibration with probability of infection assumed linear in lifetime number of sexual partners. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$ and $h = 1$. To aid visualization, the vertical axis has been truncated from $p_v = 1$ to $p_v = 0.25$.)

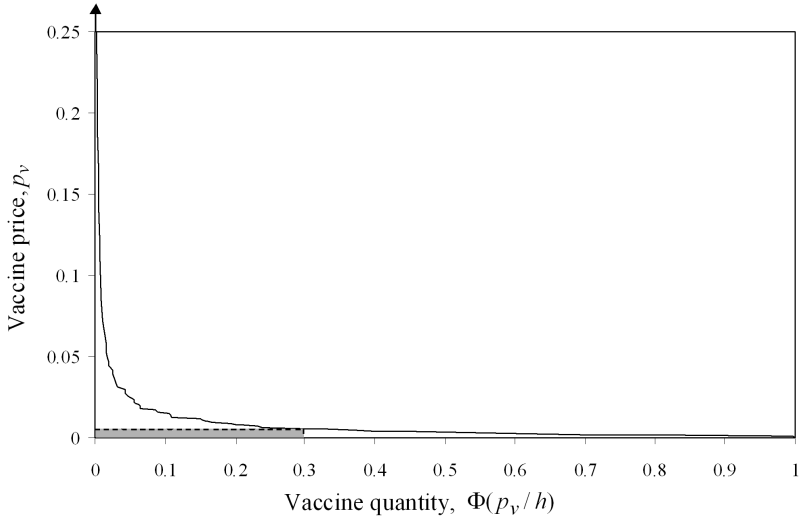

Figure 5: Inverse demand curve for sexually transmitted-disease calibration using Kaplan model of probability of infection. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$ and $h = 1$. To aid visualization, the vertical axis has been truncated from $p_v = 1$ to $p_v = 0.25$.)

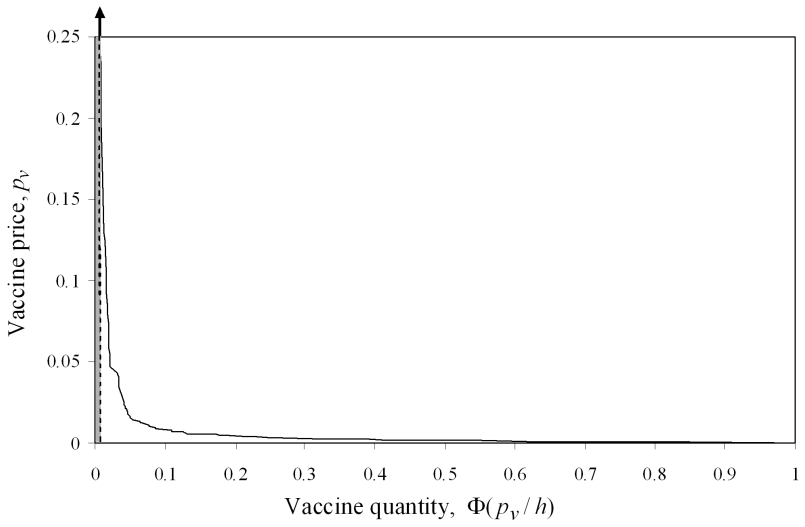

Figure 6: Inverse demand curve for HIV/AIDS calibration. Probability of infection given by the Kaplan model. Transmission rate for homosexual males allowed to differ from rest of population. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$ and $h = 1$. To aid visualization, the vertical axis has been truncated from $p_v = 1$ to $p_v = 0.25$.)

at three times the male-to-female transmission rate, so we will take this rate to be 30 percent. For others in the population, we will take $\beta_o = 0.028$ percent, equal to the estimated HIV/AIDS prevalence rate for non-homosexual males in the United States of 0.28 percent (see footnote 10) times the average per-partner transmission rate of 10 percent from Rockstroh et al. (1995). Figure 6 graphs the resulting inverse demand curve. One can compute $\pi_v/\pi_d = 0.30$, $PB = 0.70$, and $SB = 0.02$. This calibration contrasts with the first two in that, rather than selling the vaccine at a low price to a large segment of the population, the profit-maximizing strategy is to sell to a small segment of high-risk homosexual males at a very high price. The profit ratio is similar to that in the first two calibrations, but the maximum possible social distortion, *SB*, is much smaller at 2 percent. *SB* is so small in the last calibration because the vaccine is sold to so few consumers and thus does not generate much social welfare in the last calibration.

In sum, the results are similar across the calibrations. In all three, producer surplus from a vaccine

Given the 0.6 percent figure for HIV/AIDS prevalence from UNAIDS (2004), β_h and β_o must satisfy $(\beta_h P_h + \beta_o P_o)/(P_h + P_o) = 0.6$. Assuming the ratio of AIDS cases between two subgroups equals the ratio of HIV/AIDS cases provides another equation: $(\beta_h P_h)/(\beta_o P_o) = 480,509/396,765$, where 480,509 is the cumulative number of reported AIDS cases resulting from homosexual male contact as of 2002 and 396,765 the number of cases due to other causes (U.S. Centers for Disease Control 2004). Solving these two equations simultaneously gives the values of β_h and β_o in the text.

would be less than a third that from a drug. The cost of developing a drug could be higher than that of a vaccine by more than 70 percent of the total social burden of the disease and the firm could still have an incentive to develop the drug. The social cost of this bias against vaccines varies across the calibrations but can be more than half the total social burden of the disease. Thus, heterogeneity in the actual distribution of infection risk appears to be sufficient to make the bias against vaccines, which we identified in theory, large enough to be of practical concern.

5 Empirical Tests Based on Sexual Transmission

The model suggests that distortions toward drugs and away from vaccines are greater for diseases with highly skewed distributions of infection risk than for diseases for which people have similar infection risks. We do not have data on the distribution of infection risk for many diseases. A rough distinction can be drawn, however, between sexually transmitted and other diseases. Section 4 showed that the distribution of one of the main risk factors for sexually transmitted diseases, number of partners, has considerable variance and skewness in the population. For other diseases, especially diseases spread through the air like influenza, individuals' risk is likely to be more homogeneous. The fact that sexual transmission is at best an imperfect indicator of heterogeneity of infection risk is likely to reduce the power of our empirical tests that follow.

We test the prediction that heterogeneity in the distribution of infection risk, as proxied by sexual transmission, is associated with a lower likelihood of vaccine development and a higher likelihood of drug development using data on diseases and their associated medicines. The main data came from the list of Nationally Notifiable Infectious Diseases for the United States (U.S. Centers for Disease Control 2003). This list is maintained by state and national public health officials; it includes diseases whose spread is considered to need monitoring in the United States. The dataset was supplemented by the Maryland Department of Health list of common diseases (Maryland Department of Health 2003), which includes some common but non-notifiable diseases such as influenza and rotavirus. For this list of diseases, information was compiled from various public health sources (the U.S. National Institutes of Health, Department of Health and Human Services, and Centers for Disease Control) on the types of medicines available to treat

the disease and whether sexual contact is the most common means of transmission.

Table 1: Summary statistics

Transmission mode	Number of diseases in data set		
	Total	For which vaccine developed	For which drug developed
Sexually transmitted	10	1	9
Non-sexually transmitted	65	25	44
Total	75	26	53

Note: Categories may not sum to totals because some diseases have both a drug and a vaccine and some diseases have neither.

Table 4 in the Appendix contains the dataset, consisting of 75 diseases. Table 1 provides summary statistics. Consistent with the model, virtually all sexually transmitted diseases (9 out of 10, 90 percent) have a drug treatment but no vaccine. For non-sexually transmitted diseases, the division is more even between those with a vaccine (25 of 65, 38 percent) and with a drug treatment (44 of 65, 68 percent).

To provide a measure of statistical significance for these differences, we ran ordinary least squares regressions in which the dependent variable is an indicator for the medicine developed and the main variable of interest is an indicator for whether the disease is sexually transmitted.[11] The regressions include fixed effects for type of organism (parasite, virus, fungus, bacterium), because the medical literature suggests the technological difficulty of developing vaccines and drugs varies among these organisms. For example, unlike viruses, most bacterial infections can be treated with antibiotics. The results are reported in Table 2. Vaccines are significantly less likely to be developed for sexually than non-sexually transmitted diseases (a coefficient of –0.365, significant at the 5 percent level). Drugs are significantly more likely to be developed for sexually than non-sexually transmitted diseases (a coefficient of 0.340, significant at the 5 percent level). The fixed effects for type of organism come in as expected, with vaccines significantly more likely, and drugs significantly less likely, to be developed for viruses than the other diseases.

The results are generally stronger if we restrict attention to the subsample of 63 diseases for which some

[11] We ran a linear probability model (i.e., ordinary least squares regression) rather than alternative (e.g., probit, logit) specifications because sexual transmission is a perfect predictor of vaccine development for bacterial diseases, leading to a problem in estimating these alternatives. Probit results that pool bacterial and viral diseases in a single category, thereby avoiding the problem of perfect prediction, produced results similar to those reported in Table 2.

Table 2: Probability of developing medicines

	All diseases				Diseases having some medicine		
Dependent variable:	Vaccine developed (1)	Drug developed (2)	Both developed (3)	Neither developed (4)	Vaccine developed (5)	Drug developed (6)	Both developed (7)
Sexually transmitted	-0.365^{**}	0.340^{**}	-0.239^{*}	-0.214^{*}	-0.498^{***}	0.199	-0.299^{**}
	(0.148)	(0.134)	(0.139)	(0.127)	(0.142)	(0.180)	(0.149)
Virus	0.611^{***}	-0.477^{***}	0.130	-0.004	0.757^{***}	-0.600^{***}	0.157
	(0.157)	(0.143)	(0.148)	(0.135)	(0.165)	(0.113)	(0.174)
Fungus	-0.036	0.221	-0.077	-0.262	-0.094	-0.024	-0.118
	(0.282)	(0.257)	(0.266)	(0.243)	(0.274)	(0.187)	(0.289)
Bacterium	0.310^{***}	0.049	0.244^{*}	-0.115	0.312^{**}	-0.058	0.254
	(0.149)	(0.136)	(0.140)	(0.128)	(0.155)	(0.106)	(0.163)
Constant	0.036	0.779^{***}	0.077	0.262^{**}	0.094	1.024^{***}	0.118
	(0.137)	(0.125)	(0.129)	(0.118)	(0.145)	(0.098)	(0.153)
R^2	0.244	0.317	0.100	0.062	0.385	0.479	0.127

Notes: Ordinary least squares regressions. Regressions (1) through (4) use 75 observations; (5) through (7) use 63 observations. Omitted disease fixed effect is parasite. Significantly different from zero in a two-tailed test at the *10 percent level, **5 percent level, ***1 percent level.

Table 3: Probability of developing medicines, controlling for adult onset of disease

	All diseases				Diseases having some medicine		
Dependent variable:	Vaccine developed (1)	Drug developed (2)	Both developed (3)	Neither developed (4)	Vaccine developed (5)	Drug developed (6)	Both developed (7)
Sexually transmitted	-0.201	0.325^{*}	-0.145	-0.269^{*}	-0.327^{*}	0.114	-0.213
	(0.178)	(0.166)	(0.171)	(0.157)	(0.178)	(0.123)	(0.190)
Adult onset	-0.291	0.026	-0.166	0.099	-0.300	0.150	-0.150
	(0.189)	(0.172)	(0.177)	(0.162)	(0.192)	(0.132)	(0.205)
Virus	0.598^{***}	-0.476^{***}	0.122	0.000	0.741^{***}	-0.591^{***}	0.150
	(0.156)	(0.144)	(0.148)	(0.136)	(0.163)	(0.112)	(0.175)
Fungus	-0.051	0.222	-0.085	-0.257	-0.113	-0.015	-0.128
	(0.280)	(0.259)	(0.266)	(0.244)	(0.271)	(0.187)	(0.290)
Bacterium	0.316^{**}	0.048	0.247^{*}	-0.117	0.309^{**}	-0.056	0.253
	(0.148)	(0.136)	(0.140)	(0.128)	(0.153)	(0.106)	(0.164)
Constant	0.051	0.778^{***}	0.085	0.257^{**}	0.113	1.015^{***}	0.128
	(0.136)	(0.126)	(0.129)	(0.119)	(0.144)	(0.099)	(0.154)
R^2	0.270	0.317	0.112	0.067	0.411	0.490	0.134

Notes: See the notes to Table 2.

medicine has been developed [Table 2, columns (5) through (7)]. The model suggests there are reasons orthogonal to the heterogeneity in the infection risk distribution, reasons such as a low disease burden h, that it may not be worthwhile to develop any medicine. Excluding diseases for which no medicine has been developed can thus be expected to reduce the noise in the regressions. The one exception to the appearance of generally stronger results is the effect in column (6) of sexual transmission on the probability a drug is developed, which remains positive but becomes statistically insignificant.

Other factors could affect the relative cost of developing medicines for sexually transmitted diseases or the social burden of these diseases. One additional factor for which we have data is average age of disease onset. This variable will help address a concern that vaccines are most useful for childhood diseases because vaccines are usually administered in childhood, whereas sexually transmitted diseases are usually contracted by adults. Table 3 reports regressions including an indicator for whether the disease typically strikes adults. As expected because of its correlation with sexual transmission, the inclusion of the adult-onset indicator reduces the magnitude and significance of the sexual transmission coefficients. However, sexually transmitted diseases are still significantly more likely to have a drug treatment developed in the column (2) regression including all diseases (a coefficient of 0.325, significant at the 10 percent level) and are still significantly less likely to have a vaccine developed in the column (5) regression including the subset of diseases for which a medicine has been developed (a coefficient of –0.327, significant at the 10 percent level). The adult-onset indicator comes in as expected, making vaccines less likely and drug treatments more likely, but is not significant in any regression.

6 Government Purchases

In the remainder of the paper, we consider various extensions to our model of Section 2. Thus far, we have focused on the case of pharmaceutical sales on private markets. For many vaccines, however, 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 the 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 after the firm has decided which medicine 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.[12]

Given these objectives, the "pie" over which the parties bargain equals the potential social welfare from optimal use of the medicine (i.e., marginal-cost pricing), which recall is denoted $\tilde{W}_j$. Let T_j^f be the firm's threat point in Nash bargaining and T_j^g be the government's. Then the Nash bargaining formula yields the following expression for N_j, the firm's equilibrium surplus: $N_j = (\tilde{W}_j + T_j^f - T_j^g)/2$. Assuming parties' threat points are given by what they would earn if it the medicine were sold on the private market rather than to the government, we have $T_j^f = \Pi_j$ and $T_j^g = CS_j$.[13] Substituting these threat points into the Nash bargaining formula,

$$N_j = \frac{1}{2}(\tilde{W}_j + \Pi_j - CS_j). \tag{7}$$

To assess whether introducing government procurement solves the problem of the social cost of a bias against vaccines, we will introduce a bit more notation. Define

$$SBG = \sup_{(k_v,k_d)\in[0,\infty)^2} \left\{ \left(\frac{\tilde{W}_v - \tilde{W}_d}{D} \right) \mathbf{1}\big(N_d > \max(N_v, 0)\big) \right\}. \tag{8}$$

SBG is the greatest possible social cost of the bias against vaccines as a percentage of the total social burden of the disease. It is the analogue of *SB* from equation (6), except *SBG* is for the case of sales to the government rather than sales on the private market. The next proposition, proved in the Appendix, implies that allowing for government purchases does not eliminate the social cost of the bias against vaccines; indeed, by some measures it can exacerbate it.

Proposition 11. *SBG* $\geq$ *SB*.

The proof follows from the fact that with government procurement, the government sets efficient prices. Thus, there is more social welfare at stake ($\tilde{W}_j$ rather than W_j), and a larger potential distortion from the

[12] 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 the allocation of all surplus to the firm.

[13] There are other possibilities for threat points. For example, the government could hypothetically refuse to grant approval for private sales of the medicine in the event of a bargaining breakdown, implying $T_j^f = 0$. However, at least in the United States (by far the largest single market), once approval is granted the government would not stop private sales of the product.

wrong choice of medicine, than when medicines are sold on the private market at monopoly prices.[14]

It can be verified that Proposition 11 holds in the calibrations from Section 4. In the first calibration, $SBG = 0.72 > 0.44 = SB$. In the second calibration, $SBG = 0.78 > 0.56 = SB$. In the third calibration, $SBG = 0.51 > 0.02 = SB$.

The conclusion that government procurement does not eliminate the social cost of the bias against vaccines is essentially an instance of the familiar hold-up problem (Klein, Crawford, and Alchian 1978). The firm decides which medicine to develop before negotiating with the government. 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 about how profitable the medicines are relative to each other in the threat point, that is, on the private market.

In the model, there are advantages to having the government commit to prices before firms invest, because this will help protect the firm's R&D from hold up by the government and thus enhance investment. This point has been made before. What the analysis in this section makes clear is that precommitment to prices will also encourage the firm to make the socially efficient decision regarding which medicine to develop. In the model, if the government can set prices before the firm makes this decision, it can induce the firm to develop the vaccine precisely when it is socially efficient to do so, i.e., when $\tilde{W}_v > \tilde{W}_d$. This provides another justification for advance purchase commitment programs for vaccines of the type described by Kremer (2001).

7 Multiple Sources of Consumer Heterogeneity

This section considers the case in which consumers vary not only in probability x of contracting the disease but also in a second dimension, y, willingness to pay for a unit reduction in probability of infection. Variation in income provides a natural source of variation in y.[15]

If firms can perfectly price discriminate on the basis of y, the analysis from Section 3 can be generalized

[14] The fact that the maximum social cost of bias against vaccine (as measured by *SB* and *SBG*) is higher with government procurement does not imply that government procurement increases social cost for all distributions of infection risk. For instance, using the numerical example from the Introduction, it can be shown that if $k_d - k_v \in (450, 900)$, the firm makes the socially inefficient choice to develop a drug if medicines are sold on the private market but not if there is government procurement.

[15] In contemporaneous research, Kessing and Nuscheler (2002) study monopoly pricing of a vaccine when income is the sole source of consumer heterogeneity.

by calculating the vaccine and drug revenue given the marginal distribution of x at each value of y and integrating over y. The qualitative conclusions will be similar to those in Section 3. On the other hand, if firms cannot discriminate on the basis of y, either because y is unobservable or because of problems with resale, we can generate cases in which our previous results are reversed, and the firm prefers to develop a vaccine rather than a drug. In particular, the cases arise when x and y are negatively correlated.

Assume each consumer i has two pieces of private information: random variable $x_i \in [0, 1]$, continuing to represent the probability that i will contract the disease, and random variable $y_i \in [0, h]$, representing i's willingness to pay for a given reduction in probability of infection. Let $F(x_i, y_i)$ be the joint distribution function, $F_X(x_i)$ and $F_Y(y_i)$ be the marginal distribution functions, and $F_{X|Y}(x_i|y_i)$ and $F_{Y|X}(y_i|x_i)$ be the conditional distribution functions for x_i and y_i.

Assume the firm cannot discriminate on x_i or y_i. Maintain the parametric assumptions from Section 3: $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$.

Consider the vaccine producer's profit-maximization problem. Let $z_i = x_i y_i$ be consumer i's risk of contracting the disease times her willingness to pay, and let $G(z_i)$ be the cumulative distribution function associated with z_i. Using this notation, and recalling our parametric normalizations, consumers buy the vaccine if $z_i \geq p_v$, implying the demand for the vaccine is $\Gamma(p_v)$, where $\Gamma(p_v) = \int_{p_v}^{h} dG(z_i)$. Hence

$$\Pi_v = \max_{p_v \in [0,\infty)} \{p_v \Gamma(p_v)\} - k_v. \tag{9}$$

Next consider the drug producer's profit maximization problem.[16] Conditional on contracting the disease, consumer i would be willing to buy the drug as long as his/her willingness to pay y_i exceeds the price p_d. Integrating over the mass of consumers satisfying the condition $y_i \geq p_d$ implies that demand for the drug is $E_{X|Y}(x_i|y_i \geq p_d)\Phi_Y(p_d)$, where $E_{X|Y}(\cdot)$ is the expectation taken with respect to the

[16] To simplify the discussion, we ignore the possibility of developing both medicines. It can be shown that with multiple sources of heterogeneity,

$$\Pi_b = \max_{(p_v,p_d)\in[0,h]^2} \left\{ p_v E_Y \left(E_{X|Y} \left(x_i \middle| y_i, x_i > \frac{p_v}{\min(y_i, p_d)} \right) \right) + p_d E_Y \left(E_{X|Y} \left(x_i \middle| y_i, x_i \leq \frac{p_v}{\min(y_i, p_d)} \right) \right) \right\}.$$

conditional distribution $F_{X|Y}$ and $\Phi_Y(p_d) = \int_{p_d}^{h} dF_Y(y_i)$. Hence

$$\Pi_d = \max_{p_d \in [0,\infty)} \left\{ p_d E_{X|Y}(x_i | y_i \geq p_d) \Phi_Y(p_d) \right\} - k_d. \tag{10}$$

We saw in Proposition 4 that when there was nontrivial heterogeneity in infection risk alone, $\pi_d > \pi_v$. With multiple sources of heterogeneity, π_v and π_d can no longer be unambiguously ranked. Roughly speaking, the amount of consumers' private information embodied in (9)—a measure of the firm's difficulty in extracting rent from consumers—depends on the joint distribution of x_i and y_i, whereas the amount of consumers' private information embodied in (10) depends only on the marginal distribution of y_i since x_i has been integrated out. Whether one or the other embodies less private information depends on whether there is less private information in a joint or marginal distribution. If x_i and y_i are independent, integrating one of the sources of private information out, as in (10), will reduce the amount of private information. The result from Proposition 4, $\pi_d > \pi_v$, is recovered, as the following proposition, proved in the Appendix, states.

Proposition 12. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Assume there is non-trivial heterogeneity in the distribution of infection risk. If the distributions of x_i and y_i are independent, then $\pi_d > \pi_v$.*

On the other hand, there exist cases in which there is less private information in the joint distribution than the marginal distribution of y_i, in particular if x_i and y_i are negatively correlated. Then $\pi_v > \pi_d$.

We will demonstrate a case in which $\pi_v > \pi_d$ in a calibration using data on the distribution of HIV/AIDS prevalence and the distribution of income across countries. We will assume that price discrimination is impossible across countries. If we assumed alternatively that price discrimination across countries were possible, consumer heterogeneity would effectively disappear from the calibration for the reasons noted above. Besides providing us with a convenient numerical illustration, the assumption of no price discrimination across countries is relevant for policy. Although firms currently have at least a partial ability to price discriminate across countries, recent challenges to international price discrimination suggest this ability may be impaired in the future.[17]

Consider then the market as consisting of the entire world population. Treat all individuals within

[17] The calibration is thus relevant for the ongoing policy debate about facilitating the gray market in pharmaceuticals. See, for example, Cramps and Hollander (2003) for an analysis of parallel pharmaceutical importation.

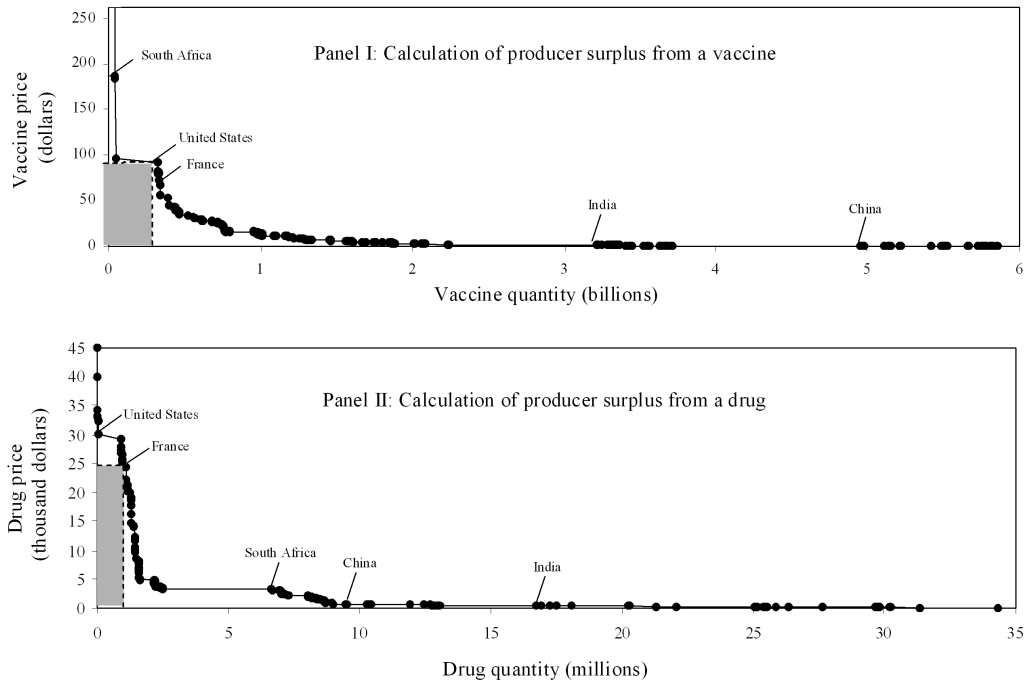

Figure 7: Comparison of producer surplus from an HIV/AIDS vaccine to that from a drug in international example with income heterogeneity and no price discrimination. (Drawn assuming $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Axes have been scaled so a unit of area represents the same producer surplus in both panels.)

any given country as homogeneous, with the same income and chance of infection; the analysis could be extended to allow for distributions of x_i and y_i within each country. We use country-level data on per-capita GNP, population, and estimated number of HIV-positive individuals to approximate our two sources of consumer heterogeneity.[18] We approximate the risk of contracting the disease, x_i, as the fraction of people within a given country that are HIV-positive. The correlation of x_i and per capita GNP y_i across countries for HIV/AIDS is significantly negative at -0.13, raising the possibility that $\pi_v > \pi_d$.

Figure 7, Panel I, shows the inverse demand curve for an HIV/AIDS vaccine, derived by plotting price on the vertical axis against the sum of the countries' populations whose expected benefit $z_i = x_i y_i$ exceeds this price. The producer surplus from an HIV/AIDS vaccine equals the area of the largest rectangle that can be inscribed under this curve. The firm maximizes profit by charging the price that just

[18] Population data are 1998 data from World Bank (2000); per-capita GNP data are 1998 data calculated with the World Bank Atlas method in 2000 U.S. dollars from World Bank (2000); HIV data are the estimated number of HIV-positive 0-to-49 year olds at the end of 1999 by country from UNAIDS (2000).

induces consumers in the United States to buy and strictly induces consumers in Switzerland, Swaziland, Namibia, the Bahamas, South Africa, and Botswana to purchase the vaccine. Figure 7, Panel II, shows the inverse demand curve for a drug, derived by plotting price on the vertical axis against the sum of countries' populations whose benefit y_i exceeds this price times x_i, the risk of HIV/AIDS in the country. Maximizing profit is equivalent to finding the rectangle of largest area inscribed under the curve. The profit-maximizing drug price just induces consumers in France to buy and strictly induces consumers in 16 other countries to buy. The axes on the two panels of Figure 7 have been scaled so that a unit of area in both represents the same revenue. The rectangle for the vaccine is slightly larger: $\pi_v/\pi_d = 1.13$.

While we have demonstrated a possible case in which $\pi_v > \pi_d$, this does not impair the practical significance of our main results concerning a bias against vaccines. First, the calibration required price discrimination to be impossible across countries, contrary to present policy. Second, while π_v is slightly larger than π_d in the calibration, W_v is much larger than W_d. Thus, there exist configurations of k_j, $j = v, d$, such that a drug is developed when it would have been socially efficient to have developed a vaccine. The social cost of this distortion can be shown to be more than 9 percent of the total social burden of the disease in this calibration. This conclusion is starker if we take an alternative measure of social welfare that weighs lives saved equally regardless of countries' incomes. The ratio of expected lives saved, L_d/L_v, equals 0.19. Thus, while the firm's incentive to develop a drug is roughly of the same order of magnitude as a vaccine, the drug would only save a fifth of the number of lives.

8 Vaccine Producers as Durable-Good Monopolists

Vaccines tend to be more "durable" medicines than drugs: a single vaccine dose can sometimes provide lifetime protection, while drugs sometimes offer only temporary protection. Antiretroviral drugs for HIV/AIDS are an extreme example, requiring daily doses. It is well-known that a durable-good monopolist serving a continuum of heterogeneous consumers faces a commitment problem (see, e.g., Coase 1972; Stokey 1981; Bulow 1982; and Gul, Sonnenschein, and Wilson 1986). We will show that a vaccine producer also faces a commitment problem, another effect biasing firms against vaccines relative to drugs. Vaccines turn out to be an unusual commodity, introducing some unique economic effects, discussed below,

to the standard durable-good-monopoly problem.

We will return to the model of Section 2 and modify it by assuming the fixed set of consumers live for two periods. Let $\delta \in [0, \infty)$ be the discount factor. At the start of the game, consumers learn their types x_i, distributed according to $F(x_i)$, where types are now reinterpreted to be the per-period hazard of contracting the disease. Once infected, a consumer suffers harm h every period thereafter unless treated. In each period, a consumer decides whether to purchase the vaccine (if available), then learns whether he has contracted the disease, then decides whether to purchase a drug (if available). A vaccine is 100 percent effective if administered before the disease is contracted, and its preventative effects last for the rest of the game. It is ineffective if administered after the disease is contracted. A drug relieves 100 percent of the harm h if administered after the disease is contracted, but this benefit does not last for future periods. Normalize the other parameters as in Section 3: $c_j = s_j = 0$ for $j = v, d$. We will measure the extent of the durable-good problem by comparing Π_j^{nc}, the profit from medicine j given that the firm cannot commit to prices across periods, to Π_j^c, the profit if it can commit, and thus there is no durable-good problem.

It is easy to compute Π_d^{nc} since the drug is sold at consumers' reservation value in both periods: $p_{d1}^{nc} = p_{d2}^{nc} = h$. We have

$$\Pi_d^{nc} = (1+\delta)h \int_0^1 x_i \, dF(x_i) + \delta h \int_0^1 x_i(1-x_i) \, dF(x_i) - k_d \tag{11}$$

$$= h \int_0^1 (x_i + 2\delta x_i - \delta x_i^2) \, dF(x_i) - k_d. \tag{12}$$

The first term on the right-hand side of (11) is the profit from consumers who contract the disease in the first period. They can be charged h in both periods, generating a discounted revenue stream $(1+\delta)h$ each. The second term is the profit from consumers who contract the disease in the second period. The integrand is adjusted by $1 - x_i$ since only that fraction of each type remain uninfected in the second period. The lack of commitment does not affect drug profits because the optimal price with commitment also equals h. Hence $\Pi_d^{nc} = \Pi_d^c$.

The profit from a vaccine is more complicated. In the no-commitment case we have

$$\Pi_v^{nc} = p_{v1}^{nc} \Phi(\hat{x}^{nc}(p_{v1}^{nc})) + \delta p_{v2}^{nc} \int_{p_{v2}^{nc}/h}^{\hat{x}^{nc}(p_{v1}^{nc})} (1 - x_i) \, dF(x_i) - k_v, \tag{13}$$

where

$$p_{v1}^{nc} = \underset{p_{v1}}{\operatorname{argmax}} \left\{ p_{v1}\Phi(\hat{x}^{nc}(p_{v1})) + \delta p_{v2}^{e}(p_{v1}) \int_{p_{v2}^{e}(p_{v1})/h}^{\hat{x}^{nc}(p_{v1})} (1 - x_i)\, dF(x_i) \right\} \tag{14}$$

$$p_{v2}^{e}(p_{v1}) = \underset{p_{v2}}{\operatorname{argmax}} \left\{ p_{v2} \int_{p_{v2}/h}^{\hat{x}^{nc}(p_{v1})} (1 - x_i)\, dF(x_i) \right\} \tag{15}$$

$$\hat{x}^{nc}(p_{v1}) = \frac{p_{v1} - \delta p_{v2}^{e}(p_{v1})}{(1+\delta)h - \delta p_{v2}^{e}(p_{v1})} \tag{16}$$

$$p_{v2}^{nc} = p_{v2}^{e}(p_{v1}^{nc}). \tag{17}$$

In the second period, the firm was not able to commit to a price. Thus, p_{v2}^{nc} is chosen to maximize continuation profits as in equation (15). This gives the second term in the expression for vaccine profit (13). In the first period, the marginal consumer with type $\hat{x}^{nc}$ is indifferent between buying in the first period and waiting until the second period. If he buys the vaccine in the first period, his surplus is simply $-p_{v1}$ since all harm is avoided. If he waits, with probability $\hat{x}^{nc}$ he contracts the disease and suffers discounted stream of harms $(1+\delta)h$, and with probability $1 - \hat{x}^{nc}$ he does not contract the disease and buys the vaccine in the second period, yielding discounted surplus $-\delta p_{v2}$. Since he is indifferent,

$$-p_{v1} = -(1+\delta)h\hat{x}^{nc} - \delta p_{v2}(1 - \hat{x}^{nc}),$$

which, upon rearranging and substituting $p_{v2}^{e}(p_{v1})$ for p_{v2} yields equation (16). The first term on the right-hand side of (13) is the profit generated by all consumers with types greater than the marginal consumer buying at price p_{v1}^{nc}, which as equation (14) indicates, is set to maximize profit recognizing the dependence of the second-period price on it. Finally, equation (17) adds the condition that price expectations must be consistent in equilibrium.

As stated above, the commitment problem does not affect drug profits: $\Pi_d^{nc} = \Pi_d^c$. The commitment problem at least weakly reduces vaccine profit because in the commitment case the firm can always commit to mimicking the price path from the no-commitment case. Hence $\Pi_v^{nc} \leq \Pi_v^c$. In the Appendix, we complete the proof of the following proposition by providing an example in which $\Pi_v^{nc} < \Pi_v^c$.

Proposition 13. *Assume $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. The commitment problem associated with the "durable" medicine, vaccines, increases the bias against developing vaccines versus drugs. That is,*

$\Pi_d^{nc} = \Pi_d^c$ *but* $\Pi_v^{nc} \leq \Pi_v^c$*, where examples can be provided in which the latter inequality is strict.*

The durable-good problem analyzed here would increase the gap between drug and vaccine profit the less durable are potential drug treatments and the more durable are potential vaccines for a disease. Assuming that most potential vaccines would provide a lifelong cure, the relevant issue is the durability of drug treatments across diseases. As noted above, existing antiretroviral drugs for HIV/AIDS are about as nondurable as possible, requiring daily doses. Thus, there is no problem committing to high prices for such HIV/AIDS drugs since the high-value consumers return to the market each day. There would be a commitment problem with an HIV/AIDS vaccine providing lifelong protection.

We noted at the beginning of the section that the analysis for vaccines differs in the details from the standard durable-good problem. High demanders of vaccines have less benefit from waiting until future periods for prices to drop. High demanders of vaccines have a high risk of contracting the disease. Once the disease is contracted, the vaccine is ineffective as a cure. Hence the commitment problem is less severe for vaccines than standard durable goods. More formally, for a given price p_{v1}, the cutoff type $\hat{x}^{nc}(p_{v1})$ is higher for vaccines, given in equation (16), than for standard durable goods, which can be shown to be $[p_{v1} - \delta p_{v2}^e(p_{v1})]/h$.[19]

9 Competing Firms

The previous sections have focused for simplicity on the case of a monopoly pharmaceutical manufacturer. Modeling competition is more difficult than monopoly because, among other reasons, there is no one canonical oligopoly model to start from. In this section, we show that bias against vaccines may persist, indeed may be exacerbated, in a plausible model of competition in the industry, in which firms developing a new medicine return an effective monopoly only temporarily.

To allow for generic entry, we will extend the model of Section 2 to an overlapping-generations setting. In period 0, $N \geq 2$ firms (branded manufacturers) sequentially decide whether to expend fixed cost k_j and develop medicine j or not to enter. Each period $t = 1, 2, \ldots$ thereafter, the old generation from $t-1$ (O_{t-1}) dies, the young generation from $t-1$ (Y_{t-1}) becomes old (O_t), and a young generation (Y_t) with

[19] Other complications arise because of the interaction between the durable-good problem and disease epidemiology. We discuss these issues further in Kremer, Snyder, and Williams (2004).

distribution of infection risk $F(x_i)$ is born. Consumers have the following lifecycle: young consumers first learn of their infection risk, decide whether or not to be vaccinated if a vaccine is available, and then turn old; old consumers contract the disease or not, decide whether or not to buy a drug if infected, and then die. Old consumers suffer harm h if they contract the disease and do not purchase a drug or the drug is ineffective. As in Section 2, we will allow for general parameter values e_j, c_j, s_j, and k_j for $j = v, d$. We will abuse notation slightly and take h, c_j, and s_j to be current values rather than present discounted values as in Section 2. Let $\delta \in [0, 1]$ be the per-period discount factor.

The first firm to develop a medicine enjoys temporary patent protection, preventing other branded manufacturers or generic entrants from copying it. For simplicity, assume the effective length of a patent is one period and, after that, either the medicine goes off patent or faces competition from "me-too" drugs that have been designed around the patent.[20] After medicine j goes off patent, a fringe of generic manufacturers enter, and price falls to marginal cost c_j. Although we have allowed for $N \geq 2$ potential entrants, at most two will enter the market in period 0 since there are only two different medicines: patents prevent further branded entry in period 0. Thus, we can restrict attention to a first and second mover. Note patents prevent the second mover from developing the same medicine as the first mover in period 0; it must either produce the other medicine or not enter.

To derive the equilibrium of this model, note that if the first mover develops a drug, the present value of its profit stream simply equals Π_d from equation (3) whether the second mover produces a vaccine or does not enter. The first mover earns this Π_d by serving the infected in generation O_1. It earns zero flow profit serving subsequent generations because of generic entry. Note that the second mover also earns Π_d if it develops a drug given the first mover does not.

The first mover's profit from developing a vaccine depends on what the second mover does. If the second mover does not enter, the present value of the first mover's profit stream has the same functional form as Π_v from equation (2), but where the cutoff type changes from $\hat{x}(p_v) = (p_v + s_v)/(e_v h)$ to $\hat{x}(p_v) = (p_v + s_v)/(\delta e_v h)$. Label this profit Π_{v0}. The first mover earns this Π_{v0} from selling to consumers in generation Y_1. The discount factor δ inserted in the new formula for $\hat{x}(p_v)$ reflects the fact that the benefit to consumers in generation Y_1 from being vaccinated is the harm avoided in the next period

[20] Roughly speaking, the implication is that the patent's effective length is about equal to the average time a person takes to contract the disease conditional on eventually contracting it, a reasonable assumption for HIV/AIDS.

when they become generation O_2; this benefit thus needs to be discounted by δ. The first mover earns zero flow profit serving subsequent generations because of generic entry.

If the second mover instead develops a drug, the first mover's profit from a vaccine is lower because consumers in generation Y_1 anticipate cheap generic drugs will be available when they become generation O_2. The present value of the first mover's profit stream again has the same functional form as Π_v in equation (2), but now the formula for the cutoff type is

$$\hat{x}(p_v) = \frac{p_v + s_v}{\delta e_v[c_d + s_d + (1 - e_d)h]}. \tag{18}$$

Label this profit Π_{vd}. Equation (18) comes from equating the surplus the marginal vaccine consumer in generation Y_1 obtains if he/she buys the vaccine to that if he/she waits until the next period and buys the drug at price c_d if he/she becomes infected. Equation (18) accounts for the fact that a vaccinated consumer has the option of taking the drug the next period if the vaccine turns out to be ineffective. Again, the first mover earns zero flow profit serving subsequent generations because of generic entry.

The next proposition characterizes the subgame-perfect equilibrium entry decisions.

Proposition 14. *If $\Pi_{vd} > \Pi_d > 0$, the first mover develops a vaccine and the second mover a drug. If $\Pi_d > \Pi_{vd} > 0$, the first mover develops a drug and the second mover a vaccine. If $\Pi_d > 0 > \Pi_{vd}$, the first mover develops a drug and the second mover does not enter. If $\Pi_{v0} > 0 > \Pi_d$, the first mover develops a vaccine and the second mover does not enter. If $0 > \max(\Pi_d, \Pi_{v0})$, neither firm enters.*

If we ignore knife-edge cases $\Pi_d = 0$, $\Pi_{v0} = 0$, and $\Pi_{vd} = 0$, we can summarize Proposition 14 neatly as follows: a drug is developed (either alone or together with a vaccine) if and only if $\Pi_d > 0$; a vaccine is developed (either alone or together with a drug) if and only if (a) $\Pi_{vd} > 0$ or (b) $\Pi_{v0} > 0 > \Pi_d$.

Proposition 14 follows from the asymmetry between the development decisions for vaccines and drugs. The drug is developed if and only if its stand-alone profit is positive. The entry of a vaccine is immaterial because the vaccine does not compete directly for customers in generation O_1 because vaccines cannot treat consumers who are already infected. The conditions under which a vaccine is developed are more restrictive. Development of a drug does affect vaccine profit because it is a substitute product for the generation Y_1 customers the vaccine profits from serving. This competition effect is in fact amplified because the generation Y_1 consumers will not only have access to the original drug but also will benefit

from competition between that drug and generic drugs that follow the development of the original.

The next proposition, proved in the Appendix, formalizes the notion that the competition model adds a new effect biasing firms in favor of drugs and against vaccines.

Proposition 15. *Compare the present model involving competition among multiple firms capable of developing vaccines and drugs to a model in which the entry decisions are made by a single firm, retaining the assumption that the development of medicine j leads to generic entry in that medicine the next period. The existence of multiple competing firms enlarges the set of parameters for which a drug is developed but reduces the set of parameters for which a vaccine is developed.*

The logic behind the proof is that a monopolist would internalize the negative externality drugs exert on vaccines that arises because medicines are substitute products. There exist cases in which a monopolist would not develop the drug in order to keep vaccine profit high, while a competitive firm develops the drug since it does not care about the vaccine producer's profits, and in some of these cases entry of the drug deters the other competitive firm from developing a vaccine.

Proposition 15 states that competition generates a new effect biasing firms against vaccines in favor of drugs. This competition effect may be socially costly, as shown by the next proposition, proved in the Appendix.

Proposition 16. *In the competitive model, social welfare never falls with a reduction in the cost of developing a vaccine, k_v, but may fall with a reduction in the cost of developing a drug, k_d.*

A reduction in the cost of developing a drug increases the incentive to develop a drug, which may deter the entry of vaccines, even some vaccines that generate more social surplus than the drug. Competition between vaccines and drugs is asymmetrically tougher on vaccines, so vaccines do not have a similar competitive effect on drugs.

10 Conclusions

In this paper, we argued that differences in the timing of the administration of drugs and vaccines allow drug manufacturers to extract more rent from consumers, driving a wedge between private and social incentives to invest in vaccine R&D. If consumers are heterogeneous in their risk of infection, a monopolist can extract more revenue from vaccines sold before the risk of infection is realized than from drugs sold when there

is no longer heterogeneity in infection risk among consumers with positive valuation. The "durability" of the medical benefits from vaccines makes it still harder for vaccine developers to appropriate consumer surplus. Finally, when inventors are rewarded with market power that is only temporary—introducing the possibility of competition between vaccines and drugs and competition from later generic entry—a further asymmetry develops as future generic drug production constrains vaccine pricing, but drug pricing is unaffected by competition from vaccines.

We showed that, in theory, the ratio between drug and vaccine producer surplus can be arbitrarily large in the presence of infection-risk heterogeneity. We performed several empirical exercises to evaluate the importance of this effect in practice. Calibrating our model to an estimated risk distribution for sexually transmitted diseases in the United States gave an estimated drug-vaccine revenue gap of more than four-fold, reflecting the skewness of the underlying distribution of number of lifetime sexual partners. As an empirical test of the model, using data on infectious diseases, we regressed dummies for whether drugs or vaccines have been developed on a dummy for whether the disease is sexually transmitted and other controls. We found vaccines are significantly less likely, and drugs significantly more likely, to have been developed for sexually transmitted than non-sexually transmitted diseases. These results provide support for our theory: sexually transmitted diseases are likely to have more skewed risk distributions than non-sexually transmitted diseases, and our theory suggests that diseases with relatively skewed risk distributions should be more likely to have drug treatments than vaccines.

We extended the analysis to allow for alternative government procurement policies, dynamic consumption decisions, and competing manufacturers. While the analysis becomes more complicated, none of these extensions forces the bias against vaccines to disappear, and each can indeed introduce new effects that may exacerbate it.

To the extent that distortions in pharmaceutical markets bias R&D investments toward drugs and away from vaccines, developing countries would be particularly adversely affected. Although antiretroviral drugs are keeping a high proportion of HIV/AIDS-infected individuals in high-income countries alive, it is much more difficult for this technology to diffuse to low-income countries due to poor health infrastructure coupled with lower levels of health spending. The development of an HIV/AIDS vaccine is arguably key to curbing the epidemic, and the market distortions we discuss may be a significant obstacle to vaccine

development. This suggests a case for subsidies to vaccine R&D beyond those for pharmaceutical R&D in general or for committing to vaccine pricing in advance of R&D.

Appendix

Proof of Proposition 1: First we will compute Π_v. Consumers have unit demand for the vaccine. Thus, by Theorem 4 of Harris and Raviv (1981), the optimal mechanism involves selling the vaccine at a linear price p_v. The marginal consumer purchases the vaccine if his or her loss of surplus if he buys is weakly less than that if he does not buy: $p_v + (1 - e_v)hx_i + s_v \leq hx_i$, or, rearranging, $x_i \geq \hat{x}(p_v)$, where $\hat{x}(p_v) = (p_v + s_v)/(e_v h)$ is the cutoff defined in the statement of the proposition. Since consumers have unit mass,

$$\begin{aligned} \Pi_v &= \max_{p_v \in [0,\infty)} \{(p_v - c_v) \Pr(x_i \geq \hat{x}(p_v))\} - k_v \\ &= \max_{p_v \in [0,\infty)} \{(p_v - c_v)\Phi(\hat{x}(p_v))\} - k_v \end{aligned}$$

verifying equation (2).

Next, we will compute Π_d. The firm cannot commit to a drug price ex ante, so sets p_d to extract all the ex post surplus from a consumer who contracts the disease. Equivalently, the loss of surplus if the infected individual buys the drug equals his or her loss of surplus if he or she does not: $p_d + (-e_d)h + s_d = h$, or, rearranging, $p_d^* = e_d h - s_d$. Since consumers have unit mass,

$$\begin{aligned} \Pi_d &= (p_d^* - c_d) \int_0^1 x_i \, dF(x_i) - k_d \\ &= (e_d h - s_d - c_d) E(x_i) - k_d \end{aligned}$$

verifying equation (3). It can be shown that, even if the firm could commit to a pricing mechanism ex ante, its profit would be no higher than it would earn from the optimal mechanism if it had commitment power.

Finally, we will compute Π_b. Suppose the firm charges p_v ex ante for the vaccine. Since it cannot commit to a drug price, p_d is set to extract all the surplus from an infected individual. As computed in the previous paragraph, $p_d^* = e_d h - s_d$. Since they obtain no net surplus from purchasing the drug, the presence of the drug does not affect a consumer's decision to purchase the vaccine ex ante. Consumer i purchases the vaccine if $x_i \geq \hat{x}(p_v)$. The quantity of the drug sold is

$$\int_0^{\hat{x}(p_v)} x_i \, dF(x_i) + (1 - e_v) \int_{\hat{x}(p_v)}^1 x_i \, dF(x_i) \tag{A1}$$

where the first term is the mass of consumers who did not buy the vaccine and become infected and the second term is the mass who bought the vaccine but the vaccine was ineffective for them. The producer surplus from the drug equals $p_d^* - c_d = he_d - s_d - c_d$ times the quantity in (A1). The producer surplus from the vaccine is, following the calculations in the first paragraph, $(p_v - c_v)\Phi(\hat{x}(p_v))$. Adding the producer surpluses together and subtracting the fixed cost of developing both medicines $k_v + k_d$ gives the expression for Π_d in equation (4). Theorem 4 of Harris and Raviv (1981) establishes that the optimal mechanism for selling the vaccine indeed involves a linear price p_v as we have implicitly assumed in this paragraph. *Q.E.D.*

Proof of Proposition 3: We have

$$\begin{aligned} PB &= \sup_{(k_v,k_d)\in[0,\infty)^2} \left\{ \left(\frac{k_d - k_v}{D} \right) \mathbf{1}\big(\Pi_d > \max(\Pi_v, 0)\big) \right\} \\ &= \sup_{(k_v,k_d)\in[0,\infty)^2} \left\{ \left(\frac{k_d - k_v}{D} \right) \mathbf{1}(k_d - k_v < \pi_d - \pi_v) \right\} \\ &= \frac{\pi_d - \pi_v}{D} \\ &= 1 - (\pi_v/\pi_d). \end{aligned}$$

The second line holds by substituting $\Pi_j = \pi_j - k_j$ for $j = v, d$. (In the second line, it should also be noted that k_v can be chosen to be sufficiently small so that $\Pi_v \geq 0$. Hence the term $\Pi_d > \max(\Pi_v, 0)$ reduces to $\Pi_d > \Pi_v$.) The third line is an algebraic manipulation. The last line holds since $\pi_d = D$. To verify $\pi_d = D$, note that substituting $c_d = s_d = 0$ and $e_d = 1$ into equation (3) implies $\Pi_d = hE(x_i)$, in turn implying $\pi_d = hE(x_i)$. But $D = hE(x_i)$ as well. *Q.E.D.*

Proof of Proposition 4: Let $c_j = s_j = 0$ and $e_j = 1$ for $j = v, d$. Substituting these values into equation (2), noting $\pi_v = \Pi_v + k_v$, and rearranging, we have

$$\begin{aligned} \pi_v &= \max_{p_v\in[0,\infty)} \left\{ p_v \int_{p_v/h}^{1} dF(x_i) \right\} \\ &= h \int_{\hat{x}^*}^{1} \hat{x}^* \, dF(x_i) \end{aligned}$$

where $\hat{x}^* = \text{argmax}_{\hat{x}\in[0,1]} \left\{ h \int_{\hat{x}}^{1} \hat{x} \, dF(x_i) \right\}$. The second line holds by the change of variables $\hat{x} = p_v \, h$. Substituting the parameter normalizations into equation (3), noting $\pi_d = \Pi_d + k_d$, and rearranging, we have $\pi_d = h \int_0^1 x_i \, dF(x_i)$. Thus,

$$\begin{aligned} \pi_d - \pi_v &= h \int_0^1 x_i \, dF(x_i) - h \int_{\hat{x}^*}^{1} \hat{x}^* \, dF(x_i) \\ &= h \int_0^{\hat{x}^*} x_i \, dF(x_i) + h \int_{\hat{x}^*}^{1} (x_i - \hat{x}^*) \, dF(x_i). \end{aligned} \tag{A2}$$

Both terms in expression (A2) are nonnegative. There cannot be a measure one of consumers at $\hat{x}^*$ by maintained assumption. Thus, there must be a positive measure on either a subset of $(0, \hat{x}^*)$, in which case the first term in (A2) is positive, or on a subset of $(\hat{x}^*, 1]$, in which case the last term in (A2) is positive. In either case, $\pi_d - \pi_v > 0$.

Finally, we analyze the firm's incentive to develop both medicines. Now $\Pi_b \leq D - (k_v + k_d)$. Further, as argued above in the proof of Proposition 3, $\Pi_d = D - k_d$. Hence $\Pi_b \leq \Pi_d - k_v \leq \Pi_d$, with strict inequality if $k_v > 0$. *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 low as possible a value of π_v/π_d subject to $m_r \in (0, 1)$ for all $r = 1, \ldots, R$; $\sum_{r=1}^R m_r = 1$; and $0 \leq x_1 \leq \cdots \leq x_R \leq 1$. Let $\theta \in (0, 1/2)$.

Define

$$m_r = \begin{cases} \theta^{r-1} & \text{if } r > 1 \\ 1 - \sum_{r=1}^{R-1} \theta^r & \text{if } r = 1. \end{cases} \tag{A3}$$

The definition of risk-class masses in equation (A3) 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 preventative pricing. Specifically, we will set $x_R = 1$ and define the rest, $\{x_r\}_{r=1}^{R-1}$, recursively by

$$hx_r \sum_{i=r}^{R} m_i = hx_{r+1} \sum_{i=r+1}^{R} m_i. \tag{A4}$$

The left-hand side of equation (A4) 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 (3), we have $\pi_d = \sum_{r=1}^{R} hm_r x_r$. By construction implicit in (A4), 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,

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

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 preventative 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 (A3). Taking limits, $\lim_{\theta \to 0}(\pi_d/\pi_v) = 1 - 0 + \sum_{r=2}^{R} 1 = R$, or, equivalently, $\lim_{\theta \to 0}(\pi_v/\pi_d) = 1/R$. This shows that for any $\epsilon > 0$, and for the definitions of the parameters in (A3) and (A4), we can find $\theta > 0$ such that

$\pi_v/\pi_d < 1/R + \epsilon$. To prove $\pi_v/\pi_d \geq 1/R$ for all distributions of consumers into R risk classes, note

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

Hence $\pi_v/\pi_d \geq 1/R$. *Q.E.D.*

Proof of Proposition 8: For a drug, $\Pi_d = W_d = \tilde{W}_d$. For a vaccine, letting p_v^* be the argmax associated with the value function in equation (2),

$$\begin{aligned} \Pi_v &= \int_{\hat{x}(p_v^*)}^{1} (p_v^* - c_v)\, dF(x_i) - k_v \\ &\leq \int_{\hat{x}(p_v^*)}^{1} (e_v hx_i - s_v - c_v)\, dF(x_i) - k_v \\ &= W_v. \end{aligned}$$

the first line holds by equation (2). The second line holds since $x_i \geq \hat{x}(p_v^*)$ for all x_i in the integrand, implying $e_v hx_i - s_v \geq p_v^*$. The third line holds by definition. Since $W_v \leq \tilde{W}_v$, $\Pi_v \leq \tilde{W}_v$. Note these calculations hold for general parameter values, not just the normalized values $c_j = s_j = 0$ and $e_j = 1$.

Putting these facts together, if $W_d > W_v$, then $\Pi_d = W_d > W_v \geq \Pi_v$. Similarly, if $\tilde{W}_d > \tilde{W}_d$, then $\Pi_d = \tilde{W}_d > \tilde{W}_v \geq \Pi_v$. Thus, if it is socially efficient to develop a drug (by either measure W_j or $\tilde{W}_j$, $j = v, d$), the firm will develop a drug in equilibrium.

To provide a case in which $W_v > W_d$ but $\Pi_d > \Pi_v$, suppose x_i is uniformly distributed on $[0, 1]$; $k_j = 1/8$ for $j = v, d$; $c_j = s_j = 0$ for $j = v, d$; $h = 1$; $e_v = 1$; and $e_d = 5/8$. For a drug, we have $\Pi_d = e_d E(x_i) - k_d = (5/8)(1/2) - 1/8 = 3/16 = W_d = \tilde{W}_d$. For a vaccine, $\Pi_v = \max_{p\in[0,\infty)} \{p_v \Phi(\hat{x}(p_v))\} - k_v = \max_{p\in[0,\infty)} \{p_v(1 - p_v)\} - k_v = 1/4 - 1/8 = 1/8$; $p_v^* = 1/2$; $W_v = \int_{p_v^*}^{1} x_i\, dx_i - k_v = 3/8 - 1/8 = 1/4$; $\tilde{W}_v = E(x_i) - k_v = 1/2 - 1/8 = 3/8$. Thus, $\Pi_d = 3/16 > 2/16 = \Pi_v$, but $W_v = 4/16 > 3/16 = W_d$, and $\tilde{W}_v = 6/16 > 3/16 = \tilde{W}_d$. *Q.E.D.*

Proof of Proposition 9: To provide a more general proof of the proposition, we will redefine *SB* as the supremum over all the parameters rather than just the fixed costs. Let ω be a vector containing particular values of the parameters k_j, c_j, s_j, and e_j for $j = v, d$. Let Ω be the set of all ω. Define

$$SB = \sup_{\omega\in\Omega} \left\{ \left(\frac{W_v - W_d}{D} \right) \mathbf{1}\big(\Pi_d > \max(\Pi_v, 0)\big) \right\} \tag{A5}$$

Then

$$\begin{aligned} SB &= \sup_{\omega \in \Omega} \left\{ \left(\frac{CS_v + \Pi_v - \Pi_d}{D} \right) \mathbf{1}(\Pi_d > \Pi_v) \right\} \\ &= \sup_{\omega' \in \Omega'} \left\{ \left(\frac{CS_v}{D} \right) \right\}. \end{aligned}$$

The first line holds since $W_v = CS_v + \Pi_v$ and since, as argued in the proof of Proposition 8, $W_d = \Pi_d$. (In the first line, it should also be noted that k_v can be chosen to be sufficiently small so that $\Pi_v \geq 0$. Hence the term $\Pi_d > \max(\Pi_v, 0)$ reduces to $\Pi_d > \Pi_v$.) The second line requires some work to prove. We are allowed to choose k_j, $j = v, d$, freely to maximize the expression in the supremum. Set $k_d - k_v = \pi_d - \pi_v - \epsilon|\pi_d - \pi_v|$ for $\epsilon > 0$. Hence $\Pi_d - \Pi_v = \pi_d - \pi_v - (k_d - k_v) = \epsilon|\pi_d - \pi_v|$. By taking ϵ to be arbitrarily close to zero, we can force Π_d to approach Π_v arbitrarily closely from above and get the expression in the supremum arbitrarily close to CS_v/D. The supremum is taken with respect to parameter vectors $\omega' \in \Omega'$, the space of all parameters except k_j, $j = v, d$.

We next choose these other parameters to maximize CS_v/D. Now

$$CS_v = \int_{\hat{x}(p_v^*)}^{1} (e_v h x_i - s_v - p_v^*)\, dF(x_i) \tag{A6}$$

where p_v^* is the argmax of the value function in (2). The right-hand side of (A6) can be shown to be increasing in e_v and decreasing in c_v and s_v. Thus, we will set $e_v = 1$ and $c_v = s_v = 0$. *Q.E.D.*

Proof of Proposition 10: To provide a more general proof, we will adopt the definition of *SB* from equation (A5). In view of the formula for *SB* from Proposition 9, it is apparent that h is just a scale factor that divides out of CS_v/D, so normalize $h = 1$ without loss of generality. After substituting $h = 1$, $c_j = s_j = 0$ $j = v, d$, and $e_j = 1$ $j = v, d$, CS_v/D can be depicted using Figure 2. Abusing notation, let A, B, and C be the areas of the indicated regions on the graph. Then $CS_v = A$ and $D = A + B + C$, implying $CS_v/D = A/(A + B + C)$.

Suppose Φ is linear, implying inverse demand for the vaccine is linear. Then it is weakly concave, and Proposition 1 of Maleug (1993) implies $B \geq 2A$. It is also weakly convex, and Proposition 1 of Maleug (1993) implies $B \leq 2A$. Therefore, $B = 2A$. Elementary geometry then implies $C = A$. Hence $CS_v/D = A/(A + 2A + A) = 1/4$.

Suppose Φ is concave implying inverse demand for the vaccine is concave. Then Proposition 1 of Malueg (1993) implies $B \geq 2A$. Hence

$$\begin{aligned} \frac{A}{A+B+C} &\leq \frac{A}{A+B} \\ &\leq \frac{A}{A+2A} \\ &= 1/3. \end{aligned}$$

Therefore, $SB \leq 1/3$. To show this bound is tight, consider an example with

$$\Phi(\hat{x}) = \begin{cases} 2 - 2\hat{x} & \hat{x} \in (1/2, 1] \\ 1 & \hat{x} \in [0, 1/2] \end{cases}$$

resulting in the inverse demand curve drawn in Figure 8. It is evident that $2A = B$, $C = 0$, and so $CS_v/D = A/(A + 2A) = 1/3$, implying $SB = 1/3$.

By definition, $SB \geq 0$. To show this lower bound is tight for concave Φ, take $\Phi(\hat{x}) = 1$, resulting in the inverse demand curve drawn in Figure 9. It is evident that $A = 0$, and so $CS_v/D = 0$, implying $SB = 0$.

The proof for the case in which Φ is convex is more complicated. First, we will derive the lower bound on *SB*. It can be shown that we can restrict attention to Φ of the following form without loss of generality:

$$\Phi(\hat{x}) = \begin{cases} 0 & \hat{x} \in [b_1, 1] \\ \dfrac{b_1 - \hat{x}}{m_1} & \hat{x} \in \left[\dfrac{b_2 - b_1}{m_2 - m_1}, b_1\right) \\ \dfrac{b_2 - \hat{x}}{m_2} & \hat{x} \in \left[0, \dfrac{b_2 - b_1}{m_2 - m_1}\right) \end{cases}$$

resulting in the inverse demand curve drawn in Figure 10. The inverse demand curve is a linear spline with two segments, ℓ_1 and ℓ_2, where $m_i \geq 0$ is the absolute value of the slope and $b_i \geq 0$ is the vertical intercept of ℓ_i extended. The parameters b_i and m_i will be specified so that a weakly largest rectangle that can be inscribed under the inverse demand curve, B, hits the inverse demand curve along segment ℓ_1. It can be shown that $A = b_1^2/8m_1$ and $B = b_1^2/4m_1$. To minimize $CS_v/D = A/(A + B + C)$, we will maximize C subject to the constraint that B remain inscribed as it is and is not replaced by a rectangle inscribed so its corner touches segment ℓ_2. This implies $b_2^2/4m_1 = b_1^2/4m_1$. The further requirement that ℓ_2 intersect point $(0, 1)$ implies $b_2 = m_2 = b_1^2/m_1$. Tedious calculations show $A/(A + B + C) = (b_1 + m_1)/8m_1$. This expression is minimized for $b_1 = 0$. In the limit as $b_1 \to 0$, $A/(A + B + C) \to m_1/8m_1 = 1/8$. Hence $SB \geq 1/8$. Since we established this bound by construction, it is tight.

To obtain an upper bound on *SB* in the case Φ is convex, we can perform similar calculations on the inverted analogue of Figure 10. One can show $A/(A+B+C) \leq (5m_1-3b_1)/8m_1)$, which is maximized for $b_1 = 0$. Hence $A/(A + B + C) \leq 5/8$, implying $SB \leq 5/8$. Since we established this bound by construction, it is tight. *Q.E.D.*

Proof of Proposition 11: To provide a more general proof of the proposition, analogous to equation (A5) we will redefine *SBG* as the supremum over all the parameters rather than just the fixed costs:

$$SBG = \sup_{\omega \in \Omega} \left\{ \left(\frac{\tilde{W}_v - \tilde{W}_d}{D} \right) \mathbf{1}\big(N_d > \max(N_v, 0)\big) \right\}. \tag{A7}$$

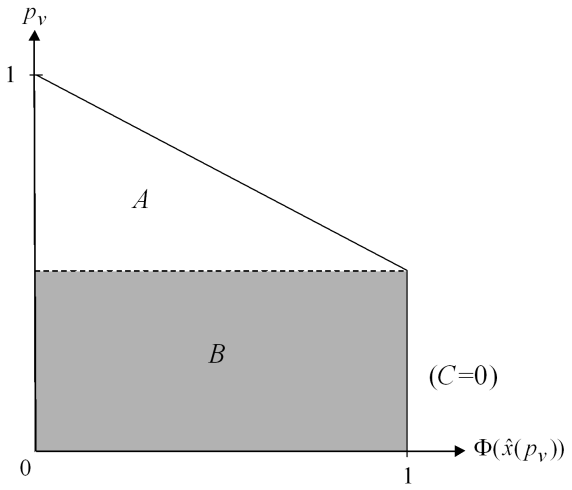

Figure 8: Example used to prove tightness of bound $SB \leq 1/3$ when Φ is concave

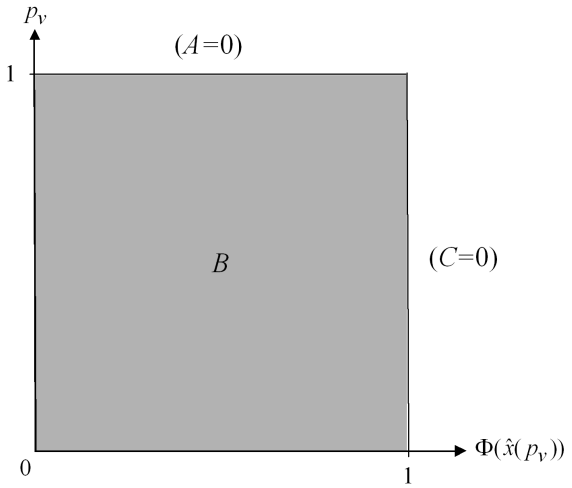

Figure 9: Example used to prove tightness of bound $SB \geq 0$ when Φ is concave

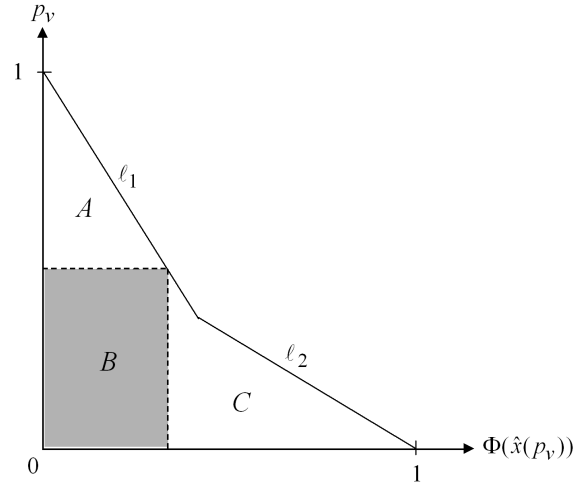

Figure 10: Example used to prove $SB \geq 1/8$ when Φ is convex

We have

$$\begin{aligned}
SBG &= \sup_{\omega\in\Omega}\left\{\left(\frac{\tilde{W}_v-\tilde{W}_d}{D}\right)\mathbf{1}(\tilde{W}_d+\Pi_d-CS_d>\tilde{W}_v+\Pi_v-CS_v)\right\}\\
&= \sup_{\omega\in\Omega}\left\{\left(\frac{1}{D}\right)(\tilde{W}_v-\Pi_v+\pi_v-\tilde{W}_d+\Pi_d-\pi_d+k_d-k_v)\right.\\
&\qquad\left.\times\,\mathbf{1}((\tilde{W}_d-\Pi_d+2\pi_d-CS_d-\tilde{W}_v+\Pi_v-2\pi_v+CS_v)/2>k_d-k_v)\right\}\\
&= \sup_{\omega'\in\Omega'}\left\{\frac{1}{2D}(\tilde{W}_v-\Pi_v-\tilde{W}_d+\Pi_d-CS_d+CS_v)\right\}\\
&= \sup_{\omega'\in\Omega'}\left\{\frac{1}{2D}(\tilde{W}_v-\Pi_v+CS_v)\right\}\\
&= \sup_{\omega'\in\Omega'}\left\{\frac{CS_v}{D}+\frac{\tilde{W}_v-W_v}{2D}\right\}\\
&\geq SB.
\end{aligned}$$

The first line holds by substituting for N_j from equation (7) into the formula for *SBG* in (8). (In the first line, it should also be noted that k_v can be chosen to be sufficiently small so that $\tilde{W}_v+\Pi_v-CS_v\geq 0$, implying the term $N_d>\max(N_v,0)$ reduces to $N_d>N_v$ or, equivalently, $\tilde{W}_d+\Pi_d-CS_d\geq\tilde{W}_v+\Pi_v-CS_v$.) The second line holds by substituting $\Pi_j=\pi_j-k_j$. The third line holds by judicious choice of free parameters k_j. The details of the argument are analogous to a similar step in the proof of Proposition 9 and are omitted. Note that the supremum is taken with respect to parameter vectors $\omega'\in\Omega'$, the space of all parameters except k_j, $j=v,d$, the same notation as used in the proof of Proposition 9. The fourth line holds since $\tilde{W}_d=\Pi_d$ (see, e.g., the proof of Proposition 8), implying $CS_d=0$. The fifth line holds since $W_v=\Pi_v+CS_v$. The last line holds since $\tilde{W}_v-W_v\geq 0$, implying $SBG\geq\sup_{\omega'\in\Omega'}\{CS_v/D\}=SB$, where the last equality was shown in the proof of Proposition 9. *Q.E.D.*

Proof of Proposition 12: Suppose the distributions of x_i and y_i are independent. Then

$$\begin{aligned}
\pi_v &= \max_{p\in[0,\infty)}\left\{\int_0^1\left[\int_{p/x_i}^h p\,dF_Y(y_i)\right]dF_X(x_i)\right\}\\
&\leq \int_0^1\max_{p\in[0,\infty)}\left\{\int_{p/x_i}^h p\,dF_Y(y_i)\right\}dF_X(x_i)\\
&= \int_0^1\max_{p'\in[0,\infty)}\left\{\int_{p'}^h p'x_i\,dF_Y(y_i)\right\}dF_X(x_i)\\
&= \int_0^1 x_i\max_{p'\in[0,\infty)}\left\{\int_{p'}^h p'\,dF_Y(y_i)\right\}dF_X(x_i)\\
&= E(x_i)\max_{p'\in[0,\infty)}\left\{p'\Phi_Y(p')\right\}\\
&= \pi_d.
\end{aligned}$$

The first and last lines hold by using the independence assumption in the formulae (9) and (10) and noting $\pi_j = \Pi_j + k_j$, $j = v, d$. The rest of the steps are algebraic manipulations. The inequality in the second line is strict if there is nontrivial heterogeneity in the distribution of x_i. *Q.E.D.*

Proof of Proposition 13: The proof is completed by providing an example in which $\Pi_v^{nc} < \Pi_v^c$. Suppose there are 100 consumers with a 50 percent chance of contracting the disease and 100 with a 10 percent chance. Let $\delta = 1$ and $h = 100$. The optimal commitment strategy is to sell to the high-risk consumers only in the first period at a price of 125, which extracts all their expected surplus (with probability 0.5, they contract the disease in the first period, leading to a stream of harms of 200; with probability 0.5×0.5, they contract the disease in the second period, leading to harm 100, for a total expected harm of 125). Hence, $\Pi_d^c = 12,500$.

Under no commitment, it turns out to be optimal to sell the vaccine to all high-risk types at a price of 105 in the first period and all low-risk types at a price of 10 in the second period. The first-period price of 105 makes the high-risk types just indifferent between buying in the first period and waiting to buy at the lower price in the second. Hence, $\Pi_v^{nc} = 11,400 < \Pi_v^c$. *Q.E.D.*

Proof of Proposition 15: Compare the model of Section 9 involving competition between drugs and vaccines, which we will label Model 1, to the monopoly model laid out in the statement of Proposition 15, which we will label Model 2. We begin by proving two facts that will be useful later in the proof. Fact 1 is that Π_b, the monopolist's profit from developing both medicines, equals $\Pi_d + \Pi_{vd}$. Conditional on developing both, the monopolist's optimal pricing strategy is to charge a drug price maximizing profit from sales to generation O_1, yielding marginal profit Π_d, and charging a vaccine price that maximizes profit from sales to generation Y_1 given generics will enter the drug market, yielding marginal profit Π_{vd}. Fact 2 is that $\Pi_b \leq \Pi_d + \Pi_{v0}$. This holds because $\Pi_{v0} \geq \Pi_{vd}$ because of the negative externality between vaccines and drugs due to their substitutability.

Suppose the parameters are such that a drug is not developed in equilibrium in Model 1. By Proposition 14, $\Pi_d < 0$. (We will ignore knife-edged cases such as $\Pi_d = 0$ throughout the proof for simplicity. It is easily seen that the proof holds for these cases as well.) But $\Pi_d < 0$ implies $\Pi_b < \Pi_{v0}$ by Fact 2, in turn implying $\max(\Pi_d, \Pi_b) < \max(\Pi_{v0}, 0)$, and so a drug would not be developed in equilibrium in Model 2.

Suppose the parameters are such that a vaccine is developed in equilibrium in Model 1. By Proposition 14, either (a) $\min(\Pi_d, \Pi_{vd}) > 0$ or (b) $\Pi_{v0} > 0 > \Pi_d$. If (a) holds, then by Fact 1, $\Pi_b = \Pi_d + \Pi_{vd} > \Pi_d > 0$. Thus, $\max(\Pi_{v0}, \Pi_b) > \max(\Pi_d, 0)$. Thus a vaccine is developed in equilibrium in Model 2. If (b) holds, then again $\max(\Pi_{v0}, \Pi_b) > \max(\Pi_d, 0)$, and so a vaccine is developed in equilibrium in Model 2.

The proof is completed by constructing a case in which a drug is developed in equilibrium in Model 1 but a vaccine is developed in equilibrium in Model 2. Let consumers be homogeneous, with $x_i = 1$ for all i. Let $\delta = e_v = 1$. Let $c_j = s_j = 0$ for $j = v, d$. Let $k_d < e_d h$ and $k_v \in \big((1-e_d)h, (1-e_d)h + k_d\big)$. It can be shown that $\Pi_d = e_d h - k_d > 0$, $\Pi_{v0} = h - k_v$, and $\Pi_{vd} = (1-e_d)h - k_v < 0$. By Proposition 14, since $\Pi_d > 0$ and $\Pi_{vd} < 0$, a vaccine alone is developed in equilibrium in Model 1. Since $k_v < (1-e_d)h + k_d$, $\Pi_{v0} > \Pi_d$. Hence $\Pi_{v0} > \Pi_d > \Pi_d + \Pi_{vd} = \Pi_b$, where the last step holds by Fact 1. Thus, a vaccine alone is developed in equilibrium in Model 2. *Q.E.D.*

Proof of Proposition 16: All of the direct and indirect effects of reducing k_j on social welfare are non-positive except possibly for one: the possibility of deterring entry by the other medicine. Proposition 14

implies that a drug will be developed if $\Pi_d > 0$, independent of the vaccine's entry decision, and thus independent of k_v. So reducing k_v weakly increases social welfare.

The proof is completed by demonstrating a case in which a reduction in k_d reduces social welfare. Let consumers be homogeneous, with $x_i = 1$ for all i. Let $e_v = 1$. Let $c_j = s_j = 0$ for $j = v, d$. Let $k_v \in \big((1 - e_d)h, h\big)$. We will compare the case in which k_d is high, namely $k_d \in (e_d h, \infty)$, to a case in which k_d is low, namely $k_d = 0$. In the first case, $\Pi_d = e_d h - k_d < 0$. Further, $\Pi_{v0} > 0$. Hence by Proposition 14, a vaccine alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{\delta h}{1 - \delta} - k_v. \tag{A8}$$

In the second case, $\Pi_d = e_d h - k_d = e_d h > 0$. Further, $\Pi_{vd} = (1 - e_d)h - k_v < 0$. Hence by Proposition 14, a drug alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{e_d h}{1 - \delta} - k_d. \tag{A9}$$

The limit as $\delta \to 1$ of the ratio of expression (A8) to (A9) equals $1/e_d$. Thus, for δ sufficiently close to one, both k_d and social welfare are higher in the first than the second case. *Q.E.D.*

Table 4: Data set used in Section 5

Disease	Organism	Sexually transmitted	Medicine developed	Typical Age of onset
anthrax	bacterium	no	both	all
botulism	bacterium	no	drug	all
brucellosis	bacterium	no	drug	all
campylobacteriosis	bacterium	no	drug	all
chancroid	bacterium	yes	drug	adult
chlamydia trachomatis	bacterium	yes	drug	adult
cholera	bacterium	no	both	all
coccidioidomycosis	fungus	no	drug	all
conjunctivitis (“pink eye”)	bacterium	no	drug	all
coxsackievirus (“hand, foot, mouth disease”)	virus	no	none	child
cryptosporidiosis	parasite	no	none	all
cyclosporiasis	parasite	no	drug	child
cytomegalovirus	virus	yes	drug	all
diphtheria	bacterium	no	both	all
ebola	virus	no	none	all
ehrlichiosis	bacterium	no	drug	all
encephalitis	bacterium	no	vaccine	all
enterohemorrhagic escherichia coli	bacterium	no	none	all
erythema infectiosum (“fifth disease”)	virus	no	none	all
genital warts	virus	yes	drug	adult
genital herpes	virus	yes	drug	adult
giardiasis	parasite	no	drug	child
gonorrhea	bacterium	yes	drug	adult
haemophilus influenzae type b (hib)	bacterium	no	both	all
hansen disease (leprosy)	bacterium	no	drug	all
hantavirus pulmonary syndrome	virus	no	none	all
head lice	parasite	no	drug	all
hemolytic uremic syndrome	bacterium	no	none	all
hepatitis A	virus	no	vaccine	all
hepatitis B	virus	yes	vaccine	all
hepatitis C	virus	no	drug	adult
histoplasmosis	fungus	no	drug	all
HIV/AIDS	virus	yes	drug	all
impetigo	bacterium	no	drug	child
influenza	virus	no	vaccine	all
kawasaki	unknown	no	drug	child
legionellosis (“legionnaire disease”)	bacterium	no	drug	all
leptospirosis	bacterium	no	drug	all
listeriosis	bacterium	no	drug	all
lyme disease	bacterium	no	both	all
malaria	parasite	no	drug	all
measles	virus	no	vaccine	child
meningococcal disease	virus	no	both	all
mononucleosis	virus	no	none	all
mumps	virus	no	vaccine	child
mycobacterium marinum	bacterium	no	drug	adult
mycoplasma	bacterium	no	drug	child
pertussis (“whooping cough”)	bacterium	no	both	all

Table 4 continued

Disease	Organism	Sexually transmitted	Medicine developed	Typical age of onset
pinworm	parasite	no	drug	child
plague	bacterium	no	drug	all
pneumococcal disease	bacterium	no	both	all
poliomyelitis	virus	no	both	all
psittacosis	bacterium	no	drug	all
Q fever	bacterium	no	drug	all
rabies	virus	no	both	all
ring worm	parasite	no	drug	child
rocky mountain spotted fever	bacterium	no	drug	all
rubella	virus	no	vaccine	child
salmonellosis	bacterium	no	none	all
scabies	parasite	yes	drug	adult
scarlet fever	bacterium	no	drug	child
shigellosis	bacterium	no	none	all
smallpox	virus	no	both	all
streptococcal disease	bacterium	no	both	all
syphilis	bacterium	yes	drug	all
tetanus	bacterium	no	both	all
toxic-shock syndrome	bacterium	no	none	adult
toxoplasmosis	parasite	no	drug	all
trichinosis	parasite	no	none	child
tuberculosis	bacterium	no	both	all
tularemia	bacterium	no	both	all
typhoid fever	bacterium	no	both	all
varicella (“chicken pox”)	virus	no	both	child
vibrio vulnificus illness	bacterium	no	drug	all
viral gastroenteritis (“rotavirus”)	virus	no	vaccine	child
yellow fever	virus	no	vaccine	all

Notes: Kawaski syndrome is listed above, but is omitted from the data set of 75 observations used in the regressions because the type of organism causing it is unknown. In determining the type of medicine developed for various diseases we followed the medical literature rather than making a personal judgment based on definitions in the model. Sources for the data provided in the text.

References

Acemoglu, D. and J. Linn. (2003) "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," National Bureau of Economic Research (NBER) Working Paper no. 10038.

Bulow, J. (1982) "Durable-Goods Monopolists," *Journal of Political Economy* 90: 314–332.

Clay, K. B., D. S. Sibley, and P. Srinagesh. (1992) "Ex Post vs. Ex Ante Pricing: Optional Calling Plans and Tapered Tariffs," *Journal of Regulatory Economics* 4: 115–138.

Coase, R. (1972) "Durability and Monopoly," *Journal of Law and Economics* 15: 143–149.

Courty, P. (2003) "Ticket Pricing Under Demand Uncertainty," *Journal of Law and Economics* 46: 627–652.

Courty, P. and H. Li. (2000) "Sequential Screening," *Review of Economic Studies* 67: 697–717.

Cramps, C. and A. Hollander. (2003) "The Pricing of Pharmaceuticals Facing Parallel Imports," Institut d'Economie Industrielle working paper.

DiMasi, J., R. Hansen, and H. Grabowski. (2003) "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics* 22: 151–185.

Finkelstein, A. (2004) "Health Policy and Technological Change: Evidence from the Vaccine Industry," forthcoming, *Quarterly Journal of Economics*.

Gul, F., H. Sonnenschein, and R. Wilson. (1986) "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory* 39: 155–190.

Harris, M. and A. Raviv. (1981) "A Theory of Monopoly Pricing Schemes with Demand Uncertainty," *American Economic Review* 71: 347–365.

Huff, B. (2003) "Risky Business: The Case of T-20," *Bulletin of Experimental Treatments for AIDS*. San Francisco AIDS Foundation.

International AIDS Vaccine Initiative. (2002) "Delivering an AIDS Vaccine," World Economic Forum Briefing Document.

Kaplan, E. H. (1990) "Modeling HIV Infectivity: Must Sex Acts Be Counted?" *Journal of Acquired Immune Deficiency Syndromes* 3: 55–61.

Kessing, S. and R. Nuscheler. (2002) "Monopoly Pricing in the Market for Vaccines," Freie Universität Berlin working paper.

Kim-Farley, R. and the Expanded Programme on Immunization Team. (1992) "Global Immunization," *Annual Review of Public Health* 13: 223–37.

Klein, B., R. A. Crawford, and A. A. Alchian. (1978) "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics* 21: 297–326.

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.

Kremer, M., C. Snyder, and H. Williams. (2004) "Vaccines: Integrated Economic and Epidemiological Models," mimeo, Harvard University.

Lewis, T. R. and D. E. M. Sappington. (1994) "Supplying Information to Facilitate Price Discrimination," *International Economic Review* 35: 309–327.

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

Maryland Department of Health. (2003) "Community Health Administration Disease Fact Sheets." Downloaded January 16, 2004 from http://edcp.org/html/cdindex.html.

Miravete, E. (1996) "Screening Consumers Through Alternative Pricing Mechanisms," *Journal of Regulatory Economics* 9: 111–132.

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

Rockstroh, J. K., S. Ewig, T. Bauer, G. Lüchters, J. Oldenberg, E. Bailly, R. Kaiser, K. E. Schneweis, H. H. Brackmann, J. Dengler, and T. Sauerbruch. (1995) "Male-to-Female Transmission of HIV in a Cohort of Hemophiliacs—Frequency, Risk Factors and Effect of Sexual Counseling," *Infection* 23: 29–32.

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

Royce, R. A., A. Seña, W. Cates, and M. S. Cohen. (1997) "Sexual Transmission of HIV," *New England Journal of Medicine* 336: 1072–1078.

Stokey, N. (1981) "Rational Expectations and Durable Goods Pricing," *Bell Journal of Economics* 12: 112–128.

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) *Epidemiological Fact Sheets by Country*.

UNAIDS. (2004) *USA: Epidemiological Fact Sheets on HIV/AIDS and Sexually-Transmitted Infections*.

U. S. Centers for Disease Control. (2000) *National Health and Examination Survey (NHANES) 1999–2000*. Washington, D.C.

U. S. Centers for Disease Control. (2003) "Summary of Notifiable Diseases—United States, 2001," *Morbidity and Mortality Weekly Report* 50: 1–108.

U. S. Centers for Disease Control. (2004) "Basic Statistics," National Center for HIV, STD, and TB Prevention, Division of HIV/AIDS Prevention. Downloaded July 13, 2004 from http://www.cdc.gov/-hiv/stats.htm.

Winkelstein, W., Jr., D. M. Lyman, N. Padian, R. Grant, M. Samuel, J. A. Wiley, R. E. Anderson, W. Lang, J. Riggs, and J. A. Levy. (1987) "Sexual Practices and Risk of Infection by the Human Immunodeficiency Virus: The San Francisco Men's Health Study," *Journal of the American Medical Association* 257: 321–325.

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