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

Dartmouth College

June 2006

Abstract: For diseases such as HIV/AIDS, for which individuals vary considerably in infection risk, a pharmaceutical manufacturer's ability to extract consumer surplus differs depending on whether the firm produces a vaccine or a drug. Vaccines are sold before consumers are infected, when consumers still have private information regarding their infection risk, whereas drugs are sold after a consumer's infection state is realized. If consumers vary only in infection risk, drug revenue will exceed vaccine revenue. If consumers also vary in income, relative revenues are determined by the joint distribution of infection risk and income and by the manufacturer's ability to sell advance insurance contracts. Simulations based on the distribution of infection risk and income in the U.S. population suggest that revenue from a HIV/AIDS drug may be twice that from a vaccine. We extend the analysis to allow for competition among manufacturers and for negotiated government prices.

JEL codes: O31, L11, I18, D42

Contact information: Kremer: Department of Economics, Harvard University, Littauer Center 207, Cambridge MA 02138; email: mkremer@fas.harvard.edu. Snyder: Department of Economics, Dartmouth College, 301 Rockefeller Hall, Hanover NH 03755; email: Christopher.M.Snyder@dartmouth.edu.

Acknowledgments: The authors would like to thank Emmanuelle Auriol, Chris Avery, Bryan Boulier, James Dana, Esther Duflo, Glenn Ellison, Amy Finkelstein, Corinne Langinier, David Malueg, David McAdams, Sendhil Mullainathan, Robert Porter, Michael Schwarz, Andrew Segal, Lars Stole, and seminar participants at the American Enterprise Institute, Dartmouth, Harvard, M.I.T., Northeastern, Northwestern, Princeton, Stanford, 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 Institutes on Health and Aging and on Innovation and the Global Economy, and the IDEI Conference on Markets for Pharmaceuticals and the Health of Developing Nations for helpful comments, to David Blanchflower for generously sharing his data, 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: For diseases such as HIV/AIDS, for which individuals vary considerably in infection risk, a pharmaceutical manufacturer's ability to extract consumer surplus differs depending on whether the firm produces a vaccine or a drug. Vaccines are sold before consumers are infected, when consumers still have private information regarding their infection risk, whereas drugs are sold after a consumer's infection state is realized. If consumers vary only in infection risk, drug revenue will exceed vaccine revenue. If consumers also vary in income, relative revenues are determined by the joint distribution of infection risk and income and by the manufacturer's ability to sell advance insurance contracts. Simulations based on the distribution of infection risk and income in the U.S. population suggest that revenue from a HIV/AIDS drug may be twice that from a vaccine. We extend the analysis to allow for competition among manufacturers and for negotiated government prices.

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 percent of whom live in developing countries. Antiretroviral drugs are not reaching the majority of people in the poorest countries, and vaccines arguably offer the best hope for defeating the epidemic.[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, vaccines and drugs should yield equivalent revenues if they are equally technologically efficient. A risk-neutral, rational consumer with no credit constraints would be willing to pay the expected present value of the stream of benefits in an up-front lump sum for either product.

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. In 2003, only 50,000 of the 30 million people with HIV/AIDS in Africa were 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. However, in this paper we show revenue equivalence can break down even in the benchmark case, because developers of the two products 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.

To see the logic, consider first a simple case in which consumers differ only in their infection risk (for example because they differ in the number of their sexual partners). Vaccines are administered before the disease is contracted, when consumers still have private information about their infection risks. Drugs are sold after the disease is contracted and consumers no longer have private information regarding their infection risks. The reduction in consumer heterogeneity in moving from vaccines to drugs allows the firm to extract more surplus with drugs.

To illustrate 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 products 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 if consumers are heterogeneous only in infection risk, a drug yields more revenue than a similarly effective vaccine. The drug/vaccine revenue ratio equals two for a uniform distribution of infection risk, is less than two for left-skewed distributions, and is greater than two for right-skewed distributions, indeed arbitrarily high for sufficiently skewed distributions.

In Section 4, we generalize the analysis by allowing variation in consumer income (and thus willingness to pay) as well as infection risk. We show that if infection risk is increasing in income or independent of income, a drug will still yield more revenue than a similarly effective vaccine. The correlation between infection risk and income must be sufficiently negative for vaccine revenue to exceed drug revenue. Even

then, a drug manufacturer may be able to recover all the vaccine revenue by selling future drug access (through an insurance contract, for example) to consumers ex ante, before their infection status is realized. If such insurance contracts are feasible, drugs never generates less revenue than similarly effective vaccines because drugs provide the manufacturer with more options (can be sold either ex ante or ex post) than vaccines (only sold ex ante).

In Section 5 we simulate vaccine and drug revenue in our model using empirical data both from within the United States and across countries. Empirically, the distribution of the number of sexual partners, and hence infection risk, is extremely skewed, favoring drug revenue; offsetting this is the fact that income is negatively correlated with infection risk. Simulations based on U.S. data in Section 5.1 suggest that revenue from an HIV/AIDS drug could exceed that from a vaccine by between two and four times. Price discrimination is currently possible across countries, but, under the assumption that the ability to engage in international price discrimination broke down, simulations in Section 5.2 offer the possibility that drug revenue could fall below vaccine revenue

We then consider a series of extentions to the basic model. In Section 6.2, we examine the case in which drug and vaccine developers have only temporary market power and so must compete against each other and, after some delay, against 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. In Section 6.3, motivated by the fact that governments are large purchasers of pharmaceuticals in many countries, we examine 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 a product yields greater revenue on the private market it will also yield greater revenue when sold to the government.

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, simulations in Sections 5, and theoretical extensions in Sections 4 and 6 are new as well.

As discussed further in the conclusion, we explore additional empirical and theoretical analyses in two related papers. The working paper version of this paper (Kremer and Snyder 2004) provides suggestive empirical evidence consistent with the theoretical model here. A companion paper (Kremer, Snyder, and Williams 2006) examines another reason why firms may be able to appropriate more consumer surplus with drugs than with vaccines: vaccines may be more likely than drugs to interfere with disease transmission, creating a positive externality that is difficult for the firm to capture.

It is of course impossible to know whether a vaccine would have been developed in the absence of the distortions against vaccines that we examine. The scientific challenges behind the development of an AIDS vaccine are tremendous. Efforts are hampered by the diversity of HIV subtypes as well as scientists' current lack of understanding of which anti-HIV immune responses are required to generate protective immunity. Nonetheless, many leading scientists are optimistic that the development of an effective AIDS vaccine would be possible with sufficient resources despite the scientific challenges (World Health Organization 2006). Ideally, public policy would match pharmaceutical manufacturers' private incentives and social incentives across products and states of the world so that, however the technological opportunities for the development of HIV/AIDS vaccines and drugs unfold, manufacturers could be counted on to pursue socially efficient strategies. We argue that standard intellectual-property-rights (IPR) institutions will not do this. If standard IPR institutions create a good match between private and social incentives for drug development, then the bias we identify would suggest private incentives would be inadequate for vaccines;

if they create a good match for vaccine development, then incentives for drug development would be excessive. Thus, there is value in considering public policies that target the distortion across vaccines and drugs identified in this paper. In the conclusion, we discuss some candidates (such as vaccine subsidies and advance-purchase commitments) for such policies.

2. Model

A monopoly pharmaceutical manufacturer has the choice of developing a vaccine or a drug. For the purposes of this model, we will define a vaccine as a product administered as a preventative measure before a disease is contracted and define a drug as a product administered after a disease has been contracted.[3] To simplify the presentation, we will initially consider the case in which vaccines and drugs are perfectly effective, have no side effects, and are costless to manufacture and administer. As discussed in Section 6.1, the key results hold when these assumptions are relaxed. The firm's only cost is the present discounted value of the fixed cost of developing product j, denoted $k_j \in [0, \infty)$, where $j = v$ for the vaccine and $j = d$ for the drug. Let $p_j \in [0, \infty)$ be the present discounted value of the price the firm receives for product j. Let π_j be producer surplus, $\Pi_j = \pi_j - k_j$ be profit, CS_j be consumer surplus, $W_j = CS_j + \Pi_j$ be equilibrium social welfare, and $\tilde{W}_j$ be first-best social welfare (i.e., social welfare when the product's price is set to marginal cost) from product j. Using notation that drops the subscript j for products, let W be equilibrium social welfare given the firm's equilibrium choice of product, and let $\tilde{W}$ be first-best social welfare given the first-best choice of product.

Assume consumers are risk neutral. Before purchasing any product, consumer i learns his or her infection risk, $x_i \in [0, 1]$, i.e., the probability he or she contracts the disease. Assume x_i is a random variable with cumulative distribution function $F(x_i)$. Normalizing the mass of consumers to unity, the mass of consumers with infection risk as least as great as some value $\hat{x}$ is denoted $\Phi(\hat{x}) = \int_{\hat{x}}^{1} dF(x_i)$. The mean infection risk in the population is $E(x_i) = \int_0^1 x_i \, dF(x_i)$. Assume that the firm cannot price discriminate

[3]Not all products 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.

based on x_i.[4] This assumption can be justified either by assuming that x_i is private information (for example, regarding consumers' sexual behavior or intravenous drug use, conducted in private) or that x_i is public information but the firm is prevented from discriminating on the basis of it by, for example, political factors or the difficulty of controlling resale. The firm is assumed to know the distribution of x_i in the population.

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. In this and the next section, we will assume that consumers all would pay the same amount to avoid harm h, but in Section 4 we will generalize the analysis to allow consumers to be heterogeneous in willingness to pay. Let $D = hE(x_i)$ be the total social burden of the disease, a term we will use to normalize our welfare measures in the subsequent analysis.

We next turn to a preliminary analysis of which product the firm chooses to develop. If the firm develops a vaccine, consumers purchase before becoming infected. A consumer with infection risk p_v/h would be indifferent between purchasing the vaccine at price p_v and not.[5] The vaccine producer thus sells to the mass of consumers $\Phi(p_v/h)$ with infection risk $x_i \geq p_v/h$, implying the profit from developing a vaccine is

$$\Pi_v = \max_{p_v \in [0,\infty)} [p_v \Phi(p_v/h)] - k_v. \tag{1}$$

If the firm develops a drug, on the other hand, the consumer purchases after becoming infected. The profit from developing a drug is

$$\Pi_d = hE(x_i) - k_d. \tag{2}$$

Equation (2) holds because the drug is optimally sold at a price that extracts the consumer's entire ex post surplus $p_d^* = h$; the drug is purchased by the mass of consumers who become infected, $E(x_i)$. The firm develops a vaccine if $\Pi_v > \max(\Pi_d, 0)$, a drug if $\Pi_d > \max(\Pi_v, 0)$, and neither if $\max(\Pi_v, \Pi_d) < 0$.[6]

[4] This assumption is only relevant for vaccine pricing. Ex post, when drugs are sold, realized infection replaces infection risk as the payoff-relevant state variable. Drug prices are automatically conditioned on the payoff-relevant state variable because drugs are only bought by infected individuals.

[5] Arguments along the lines of Theorem 4 of Harris and Raviv (1981) establish that a simple linear price p_v is optimal among the set of potentially complicated mechanisms that might be used to sell the vaccine.

[6] The remaining strategy—the firm develops both products—can be ignored in the analysis because it is weakly dominated given products are perfectly safe, effective, and costless to manufacture. The working paper version of this paper (Kremer and Snyder 2004) affords the firm the option to develop both products in a model with general values of parameters for side effects, efficacy, and marginal cost. The key results continue to hold.

3. Distribution of Infection Risk

If consumers are homogeneous, then there is no wedge between private and social R&D incentives, and the first best is obtained in equilibrium, as the following proposition states.

Proposition 1. *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. In equilibrium the firm makes the first-best product choice and produces the first-best quantity of this product.*

The proposition follows immediately from the fact that the monopolist can extract 100 percent of the surplus from homogeneous consumers with either product and thus fully internalizes social welfare.[7]

Heterogeneity in consumers' infection risks will drive a wedge between private and social R& D incentives. In the model, the firm cannot perfectly price discriminate based on infection risk and so is no longer able to extract 100 percent of consumer surplus with a vaccine. Producer surplus from a vaccine, π_v, will thus fall below producer surplus from a drug, π_d, as Proposition 2, proved in the Appendix, states.

Proposition 2. *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 $\pi_v < \pi_d$.*

Figure 1 sketches a simple graphical proof of Proposition 2. Producer surplus from a vaccine, π_v, equals the area of the largest rectangle that can be inscribed under inverse demand curve $\Phi(p_v/h)$, while π_d equals the area under the whole 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 $\pi_d > \pi_v$.

The result from Proposition 2 that $\pi_v < \pi_d$ has consequences for social welfare because it leaves room for cases in which the firm prefers to develop the drug even though the vaccine is cheaper to develop ($k_v < k_d$) and hence would be developed in the first best. The measure of such cases is what we mean by the firm's "bias" against vaccines. The lower is π_v relative to π_d, the greater the firm's bias against vaccines. The producer-surplus ratio π_v/π_d provides a convenient index of the bias against vaccines because this ratio can be linked to the potential social cost of this bias, as Proposition 3, proved in the Appendix, formalizes.

[7]The firm may no longer have first best incentives for product development if we depart from the monopoly assumption by allowing patent races, finite patent lives, rent-dissipating competition, etc. Section 6.2 discusses some of these issues further.

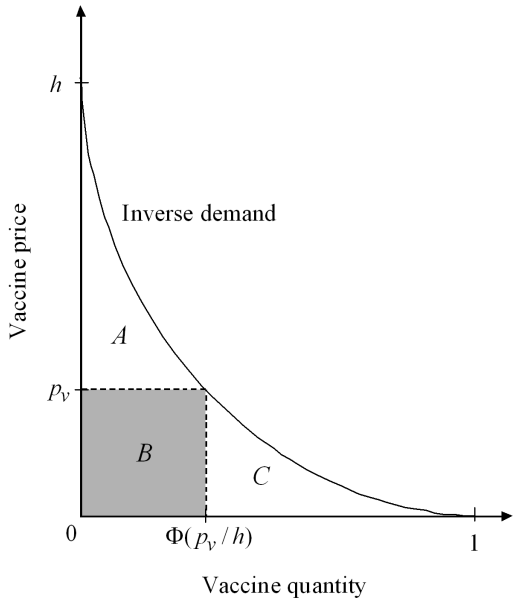

Figure 1: Geometric comparison of producer surplus from a vaccine and a drug.

Proposition 3. *The difference between first-best social welfare, $\tilde{W}$, and equilibrium social welfare, W, as a percentage of the total disease burden, D, has a tight upper bound given by $1 - \pi_v/\pi_d$. Formally,*

$$\sup_{(k_v,k_d)\in[0,\infty)^2} \left[\frac{\tilde{W} - W}{D}\right] = 1 - \frac{\pi_v}{\pi_d}.$$

Proposition 2 states that the firm will be biased against vaccines if there is heterogeneity in infection risk, raising the theoretical question of how large this bias can possibly be. The next proposition, proved in the Appendix, states that in the case in which consumers fall into discrete risk classes, the number of risk classes determines a tight lower bound on the relative producer surplus from a vaccine.

Proposition 4. *Distributions of consumers into R risk classes can be constructed such that π_v/π_d can be made arbitrarily close to $1/R$, a lower bound on π_v/π_d.*

The Introduction offered an example with two risk classes (90 consumers with a 10 percent chance of contracting the disease and 10 with a 100 percent chance) in which $\pi_v/\pi_d = 0.53$. The fact that this result was close to 1/2 was no accident: an implication of Proposition 4 is that π_v/π_d can be driven down as low as, but no lower than, 1/2 in examples with two-risk classes.

An immediate consequence of Proposition 4 is that π_v/π_d can be there exist distributions of consumer

types such that the bias against vaccines can be made arbitrarily large in percentage terms. This can be seen by taking the limit as R approaches infinity in the proposition.

Proposition 5. *There exist distributions of consumers such that π_v/π_d can be made arbitrarily close to zero.*

As the intuition from the two-type example provided in the Introduction suggests, the bias against vaccines is especially large when a large segment of the population has a very small probability of contracting the disease and a small segment of the population has a high probability. Translated in more general terms, the bias against vaccines should be expected to be largest when the distribution of infection risk is skewed. Proposition 6 provides a formal statement of the relationship between skewness of the infection-risk distribution and the ratio of producer surplus π_v/π_d.

Proposition 6. *Let $f(x_i)$ be a differentiable density function associated with consumer types x_i. 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 is illustrated in Figure 2. 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 $\Phi(p_v/h)$ turns out to be linear. Standard results imply that the area of the largest rectangle that can be inscribed under a linear demand curve is half of the area under the curve, 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 turns out to be concave. As the figure shows, 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 is convex. As the figure shows, 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.

Some empirical implications can be drawn from an examination of Propositions 1 through 6. The extent of heterogeneity in infection risk varies across diseases. The bias against vaccines which we have examined will be greatest for diseases having substantial heterogeneity in infection risk—and especially

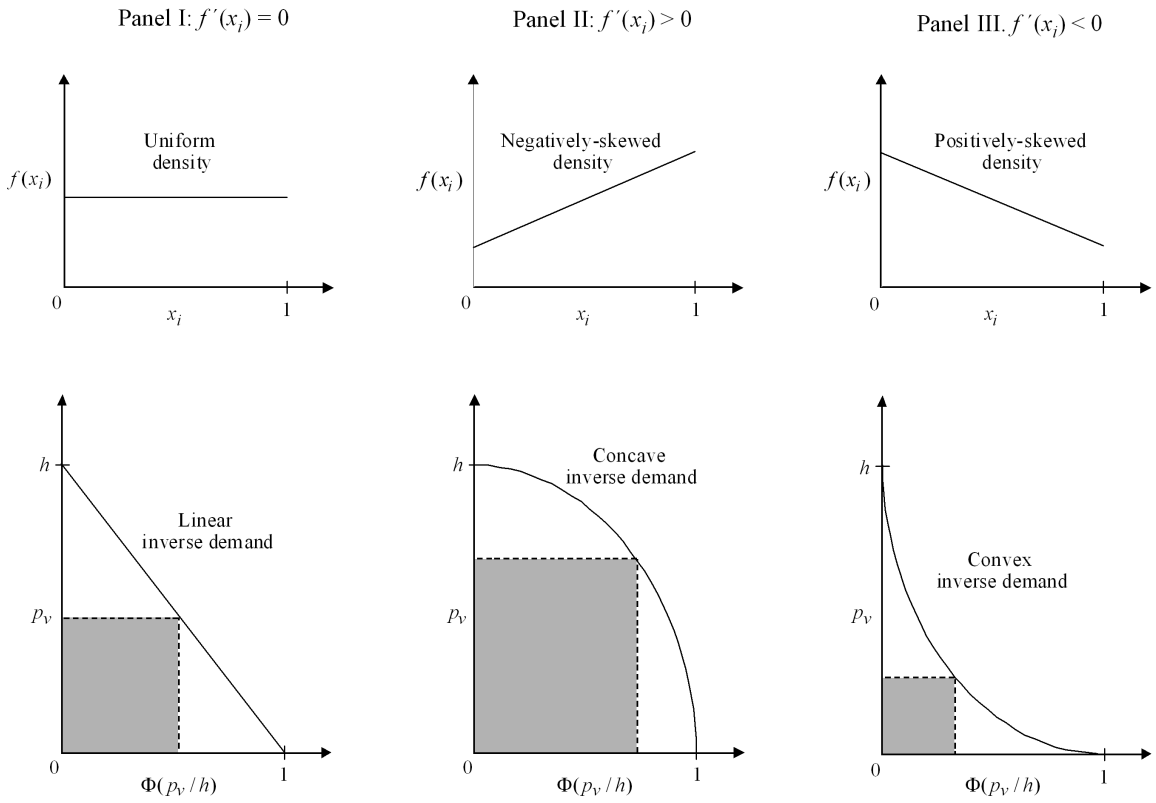

Figure 2: Ratio of producer surpluses depends on skewness of density and curvature of inverse demand.

with skewness in this risk distribution, with the largest risks concentrated in a small segment of the population. Sexually transmitted diseases including AIDS might be expected to fall into this category because of the heterogeneity evidenced in sexual behavior and skewness evidenced in the distribution of the number of sexual partners. The simulations in Section 5 will provide more detail on the shape of the distribution of infection risk for sexually transmitted diseases. On the other hand, for diseases with less heterogeneity in infection risk—such as airborne diseases—the factors we examine will create less bias against vaccines relative to drugs.

We conclude the section by drawing out the social-welfare implications of the analysis. The next proposition, proved in the Appendix, states that there is socially too little incentive to develop a vaccine relative to a drug.

Proposition 7. *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.*

Proposition 7 holds whether social efficiency is measured by first-best social welfare ($\tilde{W}_j$) or equilibrium social welfare (W_j). The main social-welfare implications of Propositions 1 through 6 should also be emphasized. Proposition 5 implies that $1-\pi_v/\pi_d$ can approach one, implying that the potential social cost of the bias against vaccines can be as large as the entire disease burden D itself. Proposition 6 implies that the potential social cost of the bias against vaccines can be as much as half the disease burden for uniformly distributed disease risk, less for left-skewed distributions, and more for right-skewed distributions.[8] In sum, the bias against vaccines can be socially costly in the model, this social cost can be quite large, and will be particularly large for right-skewed distributions of infection risk.

4. Income Heterogeneity

This section considers the more general case in which consumers vary not only in probability x_i of contracting the disease but also in a second dimension, willingness to pay to avoid harm from the disease, y_i. Variation in income provides a natural source of variation in y_i.[9]

If firms can perfectly price discriminate on the basis of y_i, the analysis from Section 3 can be generalized by calculating the vaccine and drug revenue given the marginal distribution of x_i at each value of y_i and integrating over y_i. The qualitative conclusions will be similar to those in Section 3. On the other hand, if firms cannot discriminate on the basis of y_i, either because y_i is unobservable or because of problems with resale, we can generate cases in which the firm prefers to develop a vaccine rather than a drug. As we will see, the latter cases will arise when x_i and y_i are negatively correlated over some region.

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. Let $z_i = x_i y_i$ be consumer i's risk of contracting the disease times her willingness to pay, and let $F_Z(z_i)$ be the cumulative distribution function associated with z_i. Assume the firm cannot discriminate on x_i, y_i, or z_i.

[8] The working-paper version of this paper (Kremer and Snyder 2004) develops tight upper and lower bounds on social cost under various assumptions about the distribution of disease risk.

[9] Kessing and Nuscheler (2002) study monopoly pricing of a vaccine when income is the sole source of consumer heterogeneity.

Consider the vaccine producer's profit-maximization problem. Consumers buy the vaccine if $z_i \geq p_v$, implying the demand for the vaccine is $\Phi_Z(p_v)$, where $\Phi_Z(p_v) = \int_{p_v}^{h} dF_Z(z_i)$. Hence

$$\Pi_v = \max_{p_v \in [0,\infty)} [p_v \Phi_Z(p_v)] - k_v. \tag{3}$$

Next consider the drug producer's profit maximization problem. Conditional on contracting the disease, consumer i would be willing to buy the drug as long as his or 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 conditional distribution $F_{X|Y}$ and where $\Phi_Y(p_d) = \int_{p_d}^{h} dF_Y(y_i)$. Hence

$$\Pi_d = \max_{p_d \in [0,\infty)} \left[p_d E_{X|Y}(x_i|y_i \geq p_d)\Phi_Y(p_d)\right] - k_d. \tag{4}$$

We saw in Proposition 2 that if infection risk is the only source of heterogeneity, $\pi_d > \pi_v$. 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 (3)—a measure of the firm's difficulty in extracting surplus from consumers—depends on the joint distribution of x_i and y_i, whereas the amount of consumers' private information embodied in (4) depends only on the marginal distribution of y_i since x_i has been integrated out. Which expression 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 (4), will reduce the amount of private information. Similarly, if y_i is an increasing function of x_i, then there will be less private information in the marginal than the joint distribution. In either case, the result from Proposition 2, $\pi_d > \pi_v$, is maintained, as the following proposition, proved in the Appendix, states.

Proposition 8. *Assume there is heterogeneity in the distribution of infection risk among vaccine consumers. If y_i is an increasing function of x_i or y_i is independent of x_i, then $\pi_d > \pi_v$.*

Although, as just shown, adding independently distributed income heterogeneity cannot reverse the bias against vaccines, it will reduce the bias as the next proposition, proved in the Appendix, shows.

Proposition 9. *Adding income heterogeneity that is distributed independently from the heterogeneity in infection risk causes π_v/π_d to fall at least weakly (strictly for continuous distributions).*

Intuitively, the addition of income variation has less of an impact on the overall heterogeneity of vaccine demand than drug demand because consumer demand for a vaccine involves the multiplication of income variation with infection-risk variation ($z_i = x_i y_i$), and the combination of these two independent sources of variation has a homogenizing effect on consumer valuations.

There exist cases in which adding income heterogeneity reverses the producer-surplus inequality so that $\pi_v > \pi_d$. This can happen when x_i and y_i are negatively correlated in the relevant region, so that there is less private information in the joint distribution than the marginal distribution of y_i. For example, suppose $z_i = \bar{z}$ for all i, an extreme case of negative correlation between x_i and y_i since $x_i = x_i y_i = \bar{z}$ implies $x_i = \bar{z}/y_i$. In this case the demand for vaccines would be homogeneous across consumers, allowing a vaccine monopolist to extract all social welfare—the entire disease burden D. A drug monopolist, on the other hand, cannot fully extract D if there is nontrivial heterogeneity in y_i. The analysis resembles that of Section 3 with the roles of vaccines and drugs reversed and heterogeneity in y_i substituted for heterogeneity in x_i. Indeed, the entire analysis in Section 3, suitably reinterpreted, applies to the case $z_i = \bar{z}$. For example, Proposition 2 implies $\pi_v > \pi_v$; Proposition 5 implies that distributions of y_i can be constructed such that vaccines generate arbitrarily higher producer surplus than drugs; Proposition 6 implies that vaccines would generate twice the producer surplus of drugs if the distribution of y_i were uniform; and so forth.

We have so far ignored a factor that can reverse the conclusions of the previous paragraph and, depending on the institutional environment, may guarantee that drugs always generate at least as much revenue as similarly effective vaccines. The factor relates to the fundamental asymmetry of timing between when vaccines and drugs are taken. Vaccines must be sold before infection status is realized. Drugs are taken after infection status is realized, but depending on the institutional environment it may be possible to sell future drug access (through an insurance contract, for example) to consumers before their infection status is realized. If such insurance contracts are feasible, drug manufacturers effectively have the option to imitate vaccines, and hence can always earn at least as much as from a similarly effective vaccine. The results of this section should then be reinterpreted as indicating when the manufacturer would prefer to

sell the drug ex ante versus ex post. If such insurance contracts are infeasible, then the results of this section need no reinterpretation. In particular, according to the previous paragraph, there are cases in which vaccine revenue exceeds drug revenue.

5. Simulations for HIV/AIDS

We have shown that whether firms are biased toward drugs or vaccines depends on the joint distribution of infection risk and income, and that the bias may be arbitrarily large. In this section, we simulate vaccine and drug revenue in the model using data on the distribution of infection risk, first within the United States and then across countries.

5.1. U.S. Market

The United States is by far the world's largest pharmaceutical market and is widely seen as the driver of R&D decisions. Data on infection risk are not available, but several surveys report information on risk factors such as numbers of sexual partners. As a check on the robustness of the results, we will try several different approaches to mapping the relationship between observed characteristics and infection risk and employ data from two different surveys.

Our first simulations use the 1989–2004 General Social Survey (GSS), which provides nationally representative data on the lifetime number of sexual partners broken down by the individual's gender and sexual orientation and the partners' genders.[10] The distribution of lifetime sexual partners is highly skewed: the median is 3 but the mean is 10.7. Skewness in the distribution of lifetime sexual partners induces skewness in the distribution of infection risk in our simulations, which in turn leads to a large gap between the producer surplus from a vaccine and a drug.

Column (1) of Table 1 contains the results from the simulations that use GSS data and that account for infection risk heterogeneity but not income heterogeneity. The model in the first row of the table involves a simple linear mapping from lifetime sexual partners to infection risk with a constant probability

[10] We are grateful to David Blanchflower for providing us with the cleaned version of the GSS data used in, among other studies, Blanchflower and Oswald (2004). Income is based on the family income variable interpolated as the median of the bands or, for top-coded observations, 1.25 times the top code. Other top-code factors produced essentially identical results. Income is converted into 2004 dollars using the Consumer Price Index. We label "lifetime sexual partners" the response to the survey question asking the number of sexual partners since age 18.

Table 1: Vaccine/Drug Producer Surplus Ratio in Simulations for the U.S. Market

	No income heterogeneity			Income heterogeneity
Survey:	GSS	GSS	NHANES	GSS
Ages in sample:	All	35–40	All	All
	(1)	(2)	(3)	(4)
Linear model	0.253	0.260	0.227	0.496
Kaplan model				
$\beta = 0.06\%$	0.252	0.265	0.246	0.504
β varies by sexual orientation	0.362	0.457	0.350	0.557
β varies by sexual orientation and race	0.297	0.355	0.367	0.571
β varies by sexual orientation and race; with IV drug users	0.375	0.402	0.371	0.571
Observations	17,255	2,478	2,457	15,827

of transmission per partner. Figure 3 graphs the resulting inverse demand curve for this simulation. The skewed 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, $\pi_v/\pi_d = 0.253$. As shown in the figure, the firm's optimal strategy in this simulation turns out to be to sell the vaccine at a high price to a small segment of high-risk individuals.

In the second row of simulations, we replace the simple linear model with a model due to Kaplan (1990), in which a person with n sexual partners has probability $1-(1-\beta)^n$ of ever contracting the disease, where β is the probability of contracting the disease from any given partner. We 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 Rockstroh et al. (1995) we take to be 10 percent. The estimated figure for π_v/π_d, 0.252, is quite similar to that from the linear model. Results are insensitive to varying β by one third in either direction.

The third row of simulations use GSS data to estimate infection risk for homosexual and bisexual

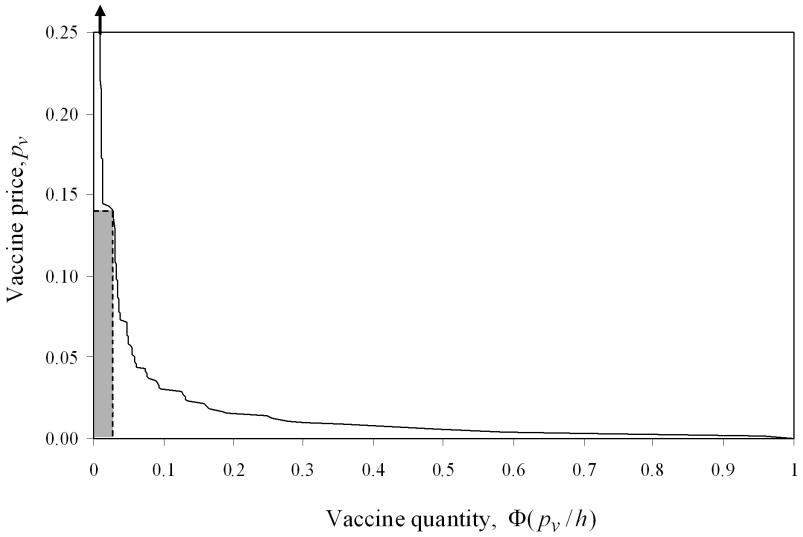

Figure 3: Inverse demand curve for simulation in which probability of infection assumed linear in lifetime number of sexual partners. (To aid visualization, the vertical axis has been truncated from $p_v = 1$ to $p_v = 0.25$.)

males separately from that for the rest of the population.[11] For the male partners of males, we scale the parameter $\beta = 0.06$ percent by 36.75, equal to the estimated prevalence of HIV/AIDS among homosexual males relative to the general population.[12] We further scale the parameter up by a factor of three to reflect the estimate from Royce et al. (1997) that HIV/AIDS is three times more likely to be passed between males than from males to females. For the rest of the sample, we scale β by 0.58, equal to the prevalence of HIV/AIDS among the population that is not homosexual male relative to the prevalence in the general population (including homosexual males). The figure for π_v/π_d in the fifth row of column (1), 0.362, is higher than those from the previous simulations in column (1). The difference arises because the firm's optimal pricing strategy is to sell to a small segment of higher-risk individuals (homosexual males with many partners in the simulation); although this higher-risk segment is small, there is sufficient

[11] Given the small number of bisexual males in the GSS sample, 0.2 percent, the results do not depend on how the transmission rates for their male and female partners are treated (we allow for differential rates) and indeed are similar if bisexual males are omitted from the calculations.

[12] The scaling factor of 36.75 for homosexual males is computed by starting with percentage of people living with HIV/AIDS in 2004 who contracted the disease from male-to-male contact—199,085 out of 462,792 cases in the 35 reporting states according to the U.S. Centers for Disease Control (2006a)—and dividing by the percentage of homosexual males in the population, estimated to be 1.2 percent in our GSS data. The scaling factor for the remainder of the population that are not homosexual males is computed similarly.

concentration of individuals in the segment that when coupled with the higher infection risk leads vaccines to be more profitable than in the previous simulations.

The fourth row of simulations in Table 1 adds variation in infection risk by race, important because the prevalence of HIV/AIDS among blacks, for example, is estimated to be nearly eight times that for whites. We take the β parameter, adjusted for variation in infection risk by sexual orientation described in the previous paragraph, and further scale it by 2.55 for blacks, 0.324 for whites, and 1.00 for Hispanics.[13] Implicit in this scaling is the assumption that an individual matches with partners of the same race. Accounting for race in addition to sexual orientation reduces the concentration of very-high-risk consumers, causing the producer surplus ratio to fall from 0.362 to 0.297.

The last row of simulations in Table 1 accounts for intravenous (IV) drug users, an important source of HIV infection in the United States. The GSS does not contain information on IV drug use, but rough estimates of the number of IV drug users and their infection risk can be inferred from other data sources. A study of HIV prevalence among IV drug users in drug treatment centers across the United States (U.S. Centers for Disease Control 2006b) provides information on the distribution of IV drug users' infection risk.[14] Coupled with an estimate of the total number of HIV cases due to IV drug use from U.S. Centers for Disease Control (2006a), we can back out the total number of IV drug users in different infection-risk categories and append simulated observations to the GSS data to represent the population of IV drug users. Adding IV drug users to the simulations in this way causes π_v/π_d to rise from 0.297 to 0.375 since adding IV drug users increases the concentration of high risk consumers to which the firm targets the vaccine. The measurable impact of IV drug users on our simulations is an artifact of the assumption that all consumers share the same willingness to pay for products: when we account for IV drug users' likely low ability to pay in column (4) of Table 1, the addition of IV drug users to the simulation will no longer affect π_v/π_d.

Columns (2) and (3) provide robustness checks. Column (2) repeats the simulations from column (1) for a single age cohort, 35 to 40 year olds. At the cost of a smaller sample size, the simulations address the potential concern that number of sexual partners may have different meanings for people in different age cohorts because older cohorts have had a longer time to accumulate partners and also lived in environments

[13] See footnote 12 for calculations, also based on statistics from the U.S. Centers for Disease Control (2006a).

[14] HIV prevalence averaged 18 percent across drug treatment centers but was heterogeneous, higher on the West than the East Coast, ranging from 1 percent in a Los Angeles to 36 percent in New York City.

with different sexual norms. The producer-surplus ratio π_v/π_d increases across simulations from column (1) to (2), for example from 0.253 to 0.260 for the linear model. Column (3) repeats the simulations from column (1) using a different data source for infection risk: the 2003–2004 National Health Examination Survey (U.S. Centers for Disease Control 2005), or NHANES. The resulting producer surplus ratios in column (3) are quite close to their analogues in column (1).

Column (4) of Table 1 repeats the simulations from column (1) allowing for heterogeneity in income in addition to heterogeneity in infection risk, assuming that price discrimination based on income is impossible and that willingness to pay to avoid harm from the disease (y_i) is proportional to income. An individual's demand for a vaccine equals his or her infection risk x_i (computed as described for the previous set of simulations) multiplied by y_i. Producer surplus from a vaccine is calculated as the rectangle of maximum area under this inverse demand curve. The demand curve for a drug is constructed by ordering consumers by y_i and then stepping off the expected drug quantity x_i each consumer would buy at this reservation price. Focusing on the first four entries in column (4)—we will discuss the last entry separately—and comparing them with the corresponding entries in column (1), we see that accounting for heterogeneity in income cuts the bias against vaccines about in half but does not reverse the bias. Even though the bias against vaccines is lower in the first four entries of column (4) compared to column (1), the simulations in column (4) still suggest that the producer surplus from drugs is about twice that from vaccines.

The last simulation in column (4) adds IV drug users in the manner described previously. Since we do not have information on income for IV drug users, we take their income to be the U.S. Federal poverty line for individuals ($9,827 in 2004).[15] Although IV drug users' infection risk is orders of magnitude higher than the typical consumers', their ability to pay is sufficiently low that they drop out of the firm's revenue base and thus do not affect the estimate of π_v/π_d, which remains at 0.571.

In sum, the simulations suggest that the underlying distribution of infection risk and income in the U.S. population would lead a firm to be biased against an HIV/AIDS vaccine and toward a drug. In simulations with the richest specification of infection risk, drug revenue was about three times vaccine revenue if income heterogeneity was not accounted for and about twice vaccine revenue if income heterogeneity was accounted for. Recalling that $1 - \pi_v/\pi_d$ measures the maximum potential social cost of the bias against

[15] The poverty line is likely an upper bound on IV drug users' income. Any multiple from 0 to 1.25 times the poverty line produced the same result for π_v/π_d as reported in the table; higher multiples up to three led to a lower value of π_v/π_d.

vaccines, the simulations suggests that the social cost of this bias can be as much as half to two-thirds of the total disease burden, D.

5.2. International Market

Firms currently have considerable ability to price discriminate across countries, but there is an active policy debate on whether this ability should be curtailed—for example, in the contexts of parallel trade for pharmaceuticals within the European Union (Cramps and Hollander 2003) or re-importation of Canadian pharmaceuticals in the United States (Pecorino 2002). The simulation in this section contributes to the debate by indicating that the abolition of international price discrimination would differentially affect the profitability of vaccines and drugs. The simulation also illustrates the theoretical possibility raised in Section 4 that the bias against vaccines can be reversed if infection risk x_i and willingness to avoid harm (as proxied by income y_i) are sufficiently negatively correlated and drug access cannot be sold before infection status is realized. It should be remembered that the simulations, because they assume no price discrimination across countries, are for a counterfactual case.

We consider the market as consisting of the entire world population and treat all individuals within 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.[16] We approximate 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 4 shows the inverse demand curve for an HIV/AIDS vaccine in the upper panel and for a drug in the lower panel. The demand curves are derived as explained in the previous subsection. The firm maximizes vaccine profit by charging the price that just 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. The profit-maximizing drug price just induces consumers in France to buy and

[16] 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).

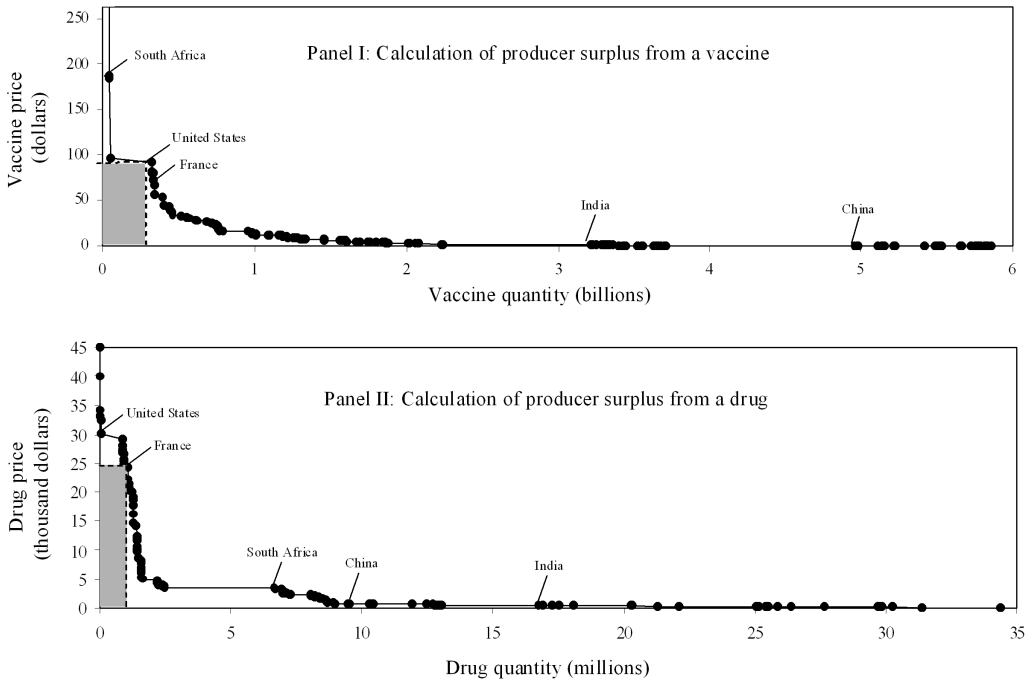

Figure 4: Comparison of producer surplus from an HIV/AIDS vaccine to that from a drug in international example with income heterogeneity and no price discrimination. (Axes scaled so that a unit of area represents the same producer surplus in both panels.)

strictly induces consumers in 16 other countries to buy. The axes on the two panels of Figure 4 have been scaled so that a unit of area in both represents the same revenue. The rectangle for the vaccine is slightly larger: $\pi_v/\pi_d = 1.13$.

The analysis suggests that impeding international price discrimination would diminish revenue from an HIV/AIDS drug more than from a vaccine, and in the extreme could reduce drug revenue below vaccine revenue if drug access cannot be sold before infection status is realized. Nonetheless, even in the unlikely case of a policy that abolished international price discrimination entirely, there would be an important sense in which the bias against vaccines would persist. Although producer surplus from a vaccine is 1.13 times that from a drug in our simulation, at equilibrium prices, social surplus from a vaccine is 1.31 times larger than from a drug, and nearly five times as many lives would be saved from a vaccine as from a drug. This is because it is privately optimal for the firm to target a drug only to high income countries. The deadweight loss from monopoly pricing is much larger with drugs than vaccines. Hence, the firm

might develop a drug even if a vaccine would yield greater social surplus and save many more lives.

6. Extensions

6.1. Generalizing Product Characteristics and Consumer Behavior

The key welfare results from Section 3 continue to hold if we relax the simplifying assumptions that products are perfectly safe and effective and costless to manufacture and administer. Let $c_j \in [0, \infty)$ be the the present discounted value of the marginal cost of manufacturing product $j \in \{v, d\}$ and administering it to a consumer. Let $e_j \in [0, 1]$ be the efficacy of product j—the probability that product j prevents the consumer from experiencing harm from the disease. Let $s_j \in [0, 1]$ be the expected harm of side effects from product j—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.[17] The Appendix provides the formal restatement of the key welfare results for general parameter values (Proposition 12) along with a proof.

It is straightforward to see how changes in parameters c_j, e_j, and s_j would affect the firm's decision regarding which product to develop: *ceteris paribus*, the firm prefers to develop the product that is more effective, has fewer side effects, and is cheaper to manufacture and administer. Consideration of general parameter values reveals some factors inherently favoring drugs (for example, vaccines expose all consumers to side effects whether or not they would eventually have contracted the disease, but only affect consumers who actually contract the disease with a drug) and some inherently favoring vaccines (for example, vaccines can prevent the appearance of any symptoms, whereas drugs may be administered only after consumers learn they have a disease because they have suffered some harm from symptoms). These factors will affect social and private product development incentives similarly, unlike the factors we focus on in this paper.

[17] Adding consumer risk aversion would increase the attractiveness of vaccines, since vaccines function as insurance against disease risk. If drugs can be sold before infection status is realized, then the insurance function of vaccines can be mimicked with drugs, and drugs would always generate at least as much revenue as similarly effective vaccines. Adding a per-period liquidity constraint for consumers would increase the attractiveness of 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.

6.2. Competing Firms

Thus far we have focused on the case of a monopoly pharmaceutical manufacturer. Modeling competition is more complicated than monopoly among other reasons because there is no one canonical oligopoly model from which to start. In this section, we show that competition can lead to an additional bias against vaccines in a plausible oligopoly model in which the patent system only provides temporary monopoly power to a firm that develops a new product, after which there is generic entry.

To allow for generic entry, we extend the model of Section 2 to an overlapping-generations setting. In period 0, N firms with the research capacity to develop new products sequentially decide whether to expend fixed cost k_j and develop one product 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 distribution of infection risk $F(x_i)$ is born. To simplify the analysis, we will focus on one source of heterogeneity, infection risk, and abstract away from other sources of heterogeneity such as income. Consumers have the following life cycle: 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. We will allow for general values of marginal cost c_j, efficacy e_j, and side effects s_j for product $j \in \{v, d\}$ as in Section 6.1 in order to allow for cases in which vaccines and drugs both enter the market. Let $\delta \in [0, 1]$ be the per-period discount factor.

The first firm to develop a product enjoys patent protection for one period.[18] After product j goes off patent, a fringe of generic manufacturers enter, and price falls to marginal cost c_j. Besides delaying generic entry, the patent prevents others of the N research-capable firms from developing the same product.[19] Thus, we can restrict attention to at most a first and second mover, which must develop different products.

In this model, competition between a vaccine and a drug is asymmetric. Competition from a vaccine does not reduce the profits of the drug patenter. The drug patenter makes its profits from sales to the infected among the initial old generation O_1. It is too late for these consumers to be vaccinated, and they

[18] The assumption of one period of patent protection roughly means that a patent's length equals the average time a person takes to contract the disease conditional on eventually contracting it, a reasonable assumption for HIV/AIDS.

[19] Even if a second firm were able to invent a "me-too" substitute around the first firm's patent for product j, in equilibrium the second firm would not develop the "me-too" product if competition between them were intense enough to reduce producer surplus below the development cost k_j.

will die before generic drugs become available. On the other hand, competition from a drug does reduce the profits of the vaccine patenter. The vaccine patenter makes its profits from sales to the initial young generation Y_1. The drug is a substitute product for these consumers: rather than buying the vaccine, they can wait to see if they become infected and buy the drug. This competition effect is amplified because the generation Y_1 consumers will not only have access to the patented drug but also will benefit from competition between that drug and generic drugs that follow, driving drug prices to marginal cost.

To derive the equilibrium of this model, first consider the firm's profit from developing a drug. Let Π_d be the single-period monopoly profit from a drug. Extending (2) to allow for general parameter values, it can be shown that $\Pi_d = (e_d h - s_d - c_d)E(x_i) - k_d$. In the competition model, the firm earns Π_d as well, whether its rival produces a vaccine or does not enter. The firm earns this Π_d by serving the infected in generation O_1. It earns zero flow profit serving subsequent generations because of generic entry.

A firm's profit from developing a vaccine depends on what its rival does. If its rival does not enter, the present value of its profit stream, denoted Π_{v0}, has the same functional form as Π_v from equation (1), but where the cutoff type indifferent between buying and not changes from $\hat{x}(p_v) = p_v/h$ to $\hat{x}(p_v) = (p_v + s_v)/(\delta e_v h)$. The vaccine developer 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 when they become generation O_2. The vaccine developer earns zero flow profit serving subsequent generations because of generic entry. If the rival develops a drug rather than not entering, the vaccine developer's profit 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 vaccine developer's profit stream, denoted Π_{vd}, again has the same functional form as Π_v in equation (1), but now the formula for the cutoff type is

$$\hat{x}(p_v) = \frac{p_v + s_v}{\delta e_v[c_d + s_d + (1 - e_d)h]}. \tag{5}$$

Equation (5) 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 (5) 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 vaccine developer earns zero

flow profit serving subsequent generations because of generic entry.

Entry decisions in the subgame-perfect equilibrium can be characterized as follows. 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. Ignoring knife-edge cases $\Pi_d = 0$, $\Pi_{v0} = 0$, and $\Pi_{vd} = 0$, equilibrium entry decisions can be neatly summarized: 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) $\Pi_{v0} > 0 > \Pi_d$.

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

Proposition 10. *The existence of $N \geq 2$ competing firms in the model enlarges the set of parameters for which a drug is developed and reduces the set of parameters for which a vaccine is developed compared to a model in which a single research-capable firm makes both sequential development decisions.*

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

The competition effect identified in Proposition 10 may be socially costly, as the next proposition states.

Proposition 11. *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.*

The intuition behind the result is that a reduction in k_d 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. As noted, competition between vaccines and drugs is asymmetrically tougher on vaccines, so vaccines do not have a similar competitive effect on drugs.

6.3. Government Purchases

Thus far, we have focused on the case of pharmaceutical sales on private markets. In many cases, however, governments are important or even dominant purchasers. Our results can be extended 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 product j after the firm has decided which product to develop and has sunk its investment in R&D. Supposing the government's objective is to maximize consumer surplus, the firm's Nash-bargaining surplus is

$$n_j = \frac{1}{2}[(\tilde{W}_j + k_j) + \pi_j - CS_j], \tag{6}$$

a combination of the first-best "pie" toward which parties bargain, $\tilde{W}_j - k_j$, plus the firm's threat-point surplus from selling product j on the private market, π_j, minus the government's surplus in this threat point, CS_j. Substituting $W_j = \pi_j - k_j + CS_j$ into (6), we have $n_j = \pi_j + (\tilde{W}_j - W_j)/2$, implying that the firm's objective function with government procurement is the sum of its objective function with private procurement π_j and a second term, reflecting incremental social surplus. The presence of this second term may mitigate the firm's bias against the product that extracts less surplus on the private market but need not eliminate the bias and indeed may even exacerbate it.

The fact that government procurement need not eliminate bias in the firm's incentives is an instance of the familiar hold-up problem (Klein, Crawford, and Alchian 1978). The firm decides which product to develop before negotiating with the government. Recognizing that it does not appropriate all the surplus in bargaining, the firm may distort its decision in order to appropriate more surplus. The literature on the hold-up problem focuses on distortions at the intensive margin of how much to invest; in our setting, the hold-up problem also leads to a distortion at the extensive margin of which product to develop.[20] Removing both extensive- and intensive-margin distortions provides another justification for advance purchase commitment programs for vaccines of the type described by Kremer and Glennerster (2004).

[20] Stole and Zwiebel (1996), among others, identify a different extensive-margin distortion resulting from the hold-up problem, in their case a distortion in the firm's technology choice.

Rather than purchasing on behalf of all consumers, the government may just provide the product to certain segments of the population, for example, to the poor through a program such as Medicaid. Such programs are more complicated to analyze than government purchases for all consumers, but similar effects arise. Assuming that the firm and government engage in Nash bargaining over the supply of product j to all consumers below a certain income threshold (say 75 percent of the U.S. poverty line, the threshold for Supplemental Security Income eligibility) and that the firm sells to the rest of the consumers as usual on the private market, we can perform simulations analogous to those in Table 1 to determine the effect of the government program. In the last simulation in Table 1 (β varies by sexual orientation and race, including IV drug users and income heterogeneity), the producer-surplus ratio, π_v/π_d, was found to be 0.571 in the absence of any government program; in the presence of the Medicaid program outlined here, the surplus ratio (now a ratio of Nash-bargaining surpluses) rises slightly to 0.607. The government program has the effect of homogenizing the population, making the firm relatively more inclined to develop a vaccine, although the firm's bias against vaccines persists.

7. Conclusion

In this paper, we argued that differences in the timing of when drugs and vaccines are taken affect the firm's ability to extract consumer surplus. Thus the wedge between private and social R&D incentives will be different for drugs than for vaccines. If consumers vary only in their infection risk, a monopolist can extract less revenue from vaccines, which are sold before consumers learn their infection status, than from similarly effective drugs, which are sold after consumers learn their infection status there is no heterogeneity among those with positive valuation. If consumers vary in both income and infection risk, vaccine revenue may exceed drug revenue, but only if the correlation between income and infection risk is sufficiently negative and only if the firm is unable to offer contracts for drugs sold in advance of consumers learning their infection status. If such advance contracts are feasible, for example through insurance programs, then drugs never generate less revenue than similarly effective vaccines.

Adding competition to the model—competition between a vaccine and a drug, as well as later generic entrants—introduces an additional bias against vaccines. Future generic drug production constrains vaccine pricing, but drug pricing is unaffected by competition from vaccines. Adding government procurement

reduces but does not eliminate the gap between private and social incentives for product development. The firm cares about the outcome on the private market because this is its threat point in negotiations with the government.

We simulated revenues from vaccines and drugs in our model using estimates of the distribution of number of sexual partners and other risk factors in the United States. In our richest specification of infection risk, drug revenue was found to be about three times vaccine revenue, owing to the skewness in the distribution of lifetime sexual partners. Additionally accounting for income heterogeneity in the population reduced but did not reverse the bias against vaccines as theory predicts given the fact that infection risk and income are nearly independently distributed in our sample. Drug revenue was still found to be about twice vaccine revenue after accounting for income heterogeneity.

A companion paper (Kremer, Snyder, and Williams 2006) examines another reason why firms may be able to appropriate more surplus with drugs than with vaccines: vaccines are more likely 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.

In Kremer and Snyder (2004) we present some suggestive empirical evidence consistent with the theoretical model and simulations presented in this paper. Using data on diseases and their associated medicines, 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. In regressions of dummies for whether drugs or vaccines have been developed on dummies for whether the disease is sexually transmitted (and other controls), we find that vaccines are significantly less likely (and drugs significantly more likely) to have been developed for sexually transmitted than non-sexually transmitted diseases. Given the small sample size and colinearity between sexual transmission and adult onset of disease, however, it is impossible to rule out competing explanations.

The scientific challenges involved in producing an AIDS vaccine are daunting, so we do not maintain that the market distortions identified in this paper necessarily account for the absence of a vaccine. Nor can we say for sure whether there has been socially too little or too much spent on R&D for an AIDS vaccine. There are a host of other factors not modeled here, from patent-race effects to research spillovers, that may affect overall R&D incentives even if they do not the differential incentives to engage in vaccine versus drug R&D. In the absence of clear information on the efficient level of R&D, we argue that public policy should be designed to match private and social incentives to develop vaccines and drugs as closely as possible across the range of potential states of the world and information sets of market participants. We have argued that this will not be achieved under current institutions. To the extent that distortions in pharmaceutical markets bias R&D investments toward drugs and against 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, the vast majority of individuals in the poorest countries are not benefitting from these technologies; the development of an HIV/AIDS vaccine is arguably key to curbing the epidemic. The market distortions against vaccine development we discuss could potentially be corrected through subsidies to vaccine R&D beyond those for pharmaceutical R&D in general, or through commitments to purchase vaccines if they are developed (Kremer and Glennerster 2004).

Appendix

Proof of Proposition 2: Substituting $\pi_v = \Pi_v + k_v$ and $\Phi(p_v/h) = \int_{p_v/h}^1 dF(x_i)$ into equation (1) and making the change of variables $\hat{x} = p_v h$ yields $\pi_v = h \int_{\hat{x}^*}^1 \hat{x}^* \, dF(x_i)$, where

$$\hat{x}^* = \underset{\hat{x} \in [0,1]}{\operatorname{argmax}} \left[h \int_{\hat{x}}^1 \hat{x} \, dF(x_i) \right]. \tag{A1}$$

Substituting $\pi_d = \Pi_d + k_d$ and $E(x_i) = \int_0^1 x_i \, dF(x_i)$ into equation (2) yields $\pi_d = h \int_0^1 x_i \, dF(x_i)$. Thus,

$$\begin{aligned} \pi_d - \pi_v &= h \int_0^1 x_i \, dF(x_i) - h \int_{\hat{x}^*}^1 \hat{x}^* \, dF(x_i) && \text{(A2)} \\ &= h \int_0^{\hat{x}^*} x_i \, dF(x_i) + h \int_{\hat{x}^*}^1 (x_i - \hat{x}^*) \, dF(x_i). && \text{(A3)} \end{aligned}$$

Both terms in expression (A3) 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 (A3) is positive, or on a subset of $(\hat{x}^*, 1]$, in which case the last term in (A3) is positive. In either case, $\pi_d - \pi_v > 0$. *Q.E.D.*

Proof of Proposition 3: We have

$$\begin{aligned} \sup_{(k_v,k_d)\in[0,\infty)^2} \left(\frac{\tilde{W} - W}{D} \right) &= \max_{j,\ell \in \{v,d\}} \left\{ \sup_{(k_v,k_d)\in[0,\infty)^2} \left[\left(\frac{\tilde{W}_\ell - W_j}{D} \right) \mathbf{1}(\Pi_j = \max(\Pi_v, \Pi_d)) \right] \right\} && \text{(A4)} \\ &= \max \Bigg\{ \sup_{(k_v,k_d)\in[0,\infty)^2} \left[\left(\frac{\tilde{W}_v - W_v}{D} \right) \mathbf{1}(\Pi_v \geq \Pi_d) \right], \\ &\qquad \sup_{(k_v,k_d)\in[0,\infty)^2} \left[\left(\frac{\tilde{W}_v - W_d}{D} \right) \mathbf{1}(\Pi_d \geq \Pi_v) \right] \Bigg\}, && \text{(A5)} \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Equation (A4) holds by definition of $\tilde{W}$ and W. To see equation (A5), note that if a drug is developed in the first best, then $W_d = D - k_d = \tilde{W}_d = \tilde{W} \geq W_v$. Thus if $\ell = d$, then $j = d$ as well. But then $\tilde{W}_d - W_d = 0$, implying that the term in braces in (A4) equals zero for $\ell = d$. We will see below that the term in braces in (A4) is non-negative for $\ell = v$, so we can restrict attention to maximizing the term in braces in (A4) over $\ell = v$, which leaves the two possible terms in braces in equation (A5). Manipulating the first braced term from equation (A5):

$$\begin{aligned} \sup_{(k_v,k_d)\in[0,\infty)^2} \left[\left(\frac{\tilde{W}_v - W_v}{D} \right) \mathbf{1}(\Pi_v \geq \Pi_d) \right] &\leq \sup_{(k_v,k_d)\in[0,\infty)^2} \left(\frac{\tilde{W}_v - W_v}{D} \right) && \text{(A6)} \\ &= \sup_{(k_v,k_d)\in[0,\infty)^2} \left[\frac{(D - k_v) - (\pi_v + CS_v - k_v)}{D} \right] && \text{(A7)} \\ &= 1 - \frac{\pi_v}{\pi_d} - \frac{CS_v}{\pi_d}. && \text{(A8)} \end{aligned}$$

Condition (A6) follows from $\mathbf{1}(\Pi_v - \Pi_d) \leq 1$, (A7) from the definitions of $\tilde{W}_v$ and W_v, and (A8) from simple algebra. Manipulating the second braced term from equation (A5):

$$\sup_{(k_v,k_d)\in[0,\infty)^2}\left[\left(\frac{\tilde{W}_v - W_d}{D}\right)\mathbf{1}(\Pi_d \geq \Pi_v)\right] = \sup_{(k_v,k_d)\in[0,\infty)^2}\left[\left(\frac{k_d - k_v}{D}\right)\mathbf{1}(\pi_d - k_d \geq \pi_v - k_v)\right] \quad \text{(A9)}$$

$$= \frac{\pi_d - \pi_v}{D} \quad \text{(A10)}$$

$$= 1 - \frac{\pi_v}{\pi_d}. \quad \text{(A11)}$$

Equation (A9) holds by substituting the definitions of $\tilde{W}_v$, W_d, Π_d, and Π_v and simplifying. Equation (A10) holds by noting that the greatest value of $k_d - k_v$ subject to the constraint $\pi_d - \pi_v \geq k_d - k_v$ equals $\pi_d - \pi_v$. Equation (A11) follows from dividing numerator and denominator through by π_d and noting $D/\pi_d = 1$ since the firm can extract 100 percent of social welfare with a drug so that $\pi_d = D$. Since $CS_v \geq 0$, equation (A11) at least weakly exceeds (A8). Equation (A11) is non-negative by Proposition 2. Hence (A5) equals (A11). *Q.E.D.*

Proof of Proposition 4: 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 $0 \leq x_1 \leq \cdots \leq x_R \leq 1$. The idea will be to set the masses of the R risk classes $\{m_r\}_{r=1}^R$ so that they decline geometrically. The probabilities $\{x_r\}_{r=1}^R$ will be set such that the firm earns almost the same producer surplus whether it sells a vaccine 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. Producer surplus from a vaccine will thus be about $1/R$ that from a drug.

Let $\theta \in (0, 1/2)$. Define

$$m_r = \begin{cases} \theta^{r-1} & \text{if } r > 1 \\ 1 - \sum_{r=1}^{R-1}\theta^r & \text{if } r = 1. \end{cases} \quad \text{(A12)}$$

The definition of risk-class masses in equation (A12) 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

$$hx_r \sum_{i=r}^R m_i = hx_{r+1} \sum_{i=r+1}^R m_i. \quad \text{(A13)}$$

The left-hand side of equation (A13) 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 (2), we have $\pi_d = \sum_{r=1}^R hm_r x_r$. By construction implicit in (A13), 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,

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

We provided an argument previously for the first line. The second line holds 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 (A12). 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 (A12) and (A13), 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 7: For a drug, $\Pi_d = W_d = \tilde{W}_d$. Since the firm extracts all social surplus with a drug, the firm always develops a drug if it is socially efficient (by either social-welfare measure W_d or $\tilde{W}_d$) to do so.

For 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 8: Suppose y_i is independent of x_i. Then

$$\pi_v = \max_{p\in[0,\infty)} \left\{ \int_{p/h}^1 \left[\int_{p/x_i}^h p \, dF_Y(y_i) \right] dF_X(x_i) \right\} \tag{A14}$$

$$\leq \int_{p/h}^1 \max_{p\in[0,\infty)} \left[\int_{p/x_i}^h p \, dF_Y(y_i) \right] dF_X(x_i) \tag{A15}$$

$$= \int_{p/h}^1 \max_{p'\in[0,\infty)} \left[\int_{p'}^h p' x_i \, dF_Y(y_i) \right] dF_X(x_i) \tag{A16}$$

$$\leq E(x_i) \max_{p'\in[0,\infty)} [p' \Phi_Y(p')] \tag{A17}$$

$$= \pi_d. \tag{A18}$$

Equations (A14) and (A18) hold by applying the independence condition to the formulae (3) and (4) and noting $\pi_j = \Pi_j + k_j$, $j = v, d$. The rest of the steps are algebraic manipulations. The inequality in (A15) is strict if there is nontrivial heterogeneity in the distribution of x_i.

Suppose $y_i = \mu(x_i)$, where μ is an increasing function. Let p_v^* be the optimal vaccine price. Vaccine demand equals $\Phi_Z(p_v^*) = \Phi_Y(\hat{y}_i)$ for $\hat{y}_i$ given by the solution to $\mu^{-1}(\hat{y}_i)\hat{y}_i = p_v^*$. Hence $\pi_v = p_v^* \Phi_Y(\hat{y}_i) = \mu^{-1}(\hat{y}_i)\hat{y}_i \Phi_Y(\hat{y}_i)$. Turning to producer surplus from a drug,

$$\pi_d \geq \hat{y}_i \int_0^1 \int_{\hat{y}_i}^h x_i \, dF(x_i, y_i) \tag{A19}$$

$$\geq \hat{y}_i \int_0^1 \int_{\hat{y}_i}^h \mu^{-1}(\hat{y}_i) \, dF(x_i, y_i) \tag{A20}$$

$$= \mu^{-1}(\hat{y}_i)\hat{y}_i \Phi_Y(\hat{y}_i) \tag{A21}$$

$$= \pi_v. \tag{A22}$$

Equation (A19) holds because the producer surplus at the optimal drug price π_d at least weakly exceeds producer surplus from a drug sold at price $\hat{y}_i$. Equation (A20) holds because μ^{-1} is an increasing function, so $x_i \geq \mu^{-1}(\hat{y}_i)$ for $y_i \geq \hat{y}_i$. Equation (A21) is a straightforward calculation. Equation (A22) follows from the previous calculations regarding vaccine producer surplus. The inequality in (A20) is strict if there is nontrivial heterogeneity in the distribution of x_i for vaccine consumers. *Q.E.D.*

Proof of Proposition 9: Let π'_v and π'_d be producer surpluses in the model with no income heterogeneity and π'_v and π'_d be producer surpluses when income heterogeneity which is independently distributed from infection-risk heterogeneity has been added to the model. Then

$$\begin{aligned} \pi'_v &= p^*_z \Pr(z_i \geq p^*_z) && \text{(A23)} \\ &\geq p^*_x p^*_y \Pr(x_i y_i \geq p^*_x p^*_y) && \text{(A24)} \\ &\geq p^*_x p^*_y \Pr(x_i \geq p^*_x) \Pr(y_i \geq p^*_y) E(x_i)/E(x_i) && \text{(A25)} \\ &= \pi_v \pi'_d / \pi_d, && \text{(A26)} \end{aligned}$$

where $p^*_z = \operatorname{argmax}_p[p \Pr(z_i \geq p)]$, $p^*_x = \operatorname{argmax}_p[p \Pr(x_i \geq p)]$, and $p^*_y = \operatorname{argmax}_p[p \Pr(y_i \geq p)]$. Equation (A23) follows from equation (3). Condition (A24) follows because p^*_x, as an argmax, produces a higher value for $p \Pr(z_i \geq p)$ than $p^*_x p^*_y$. Condition (A25) follows since $\Pr(x_i y_i \geq p^*_x p^*_y) \geq \Pr(x_i \geq p^*_x) \Pr(y_i \geq p^*_y)$. Equation (A26) follows because $\pi_v = p^*_x \Pr(x_i \geq p^*_x)$ by equation (1), $\pi_d = hE(x_i)$ by equation (2), and $\pi'_d = E(x_i) p^*_y \Pr(y_i \geq p^*_y)$ applying the independence assumption to equation (4). Conditions (A23)–(A26) together imply $\pi_v/\pi_d \leq \pi'_v/\pi'_d$. If the distributions of x_i and y_i are continuous, then the inequality in (A25) would be strict. *Q.E.D.*

Proposition 12. *The key welfare results from Section 3 continue to hold for general values of the parameters $c_j \in [0, \infty)$, $e_j \in [0, 1]$, and $s_j \in [0, \infty)$.*

i. 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.

ii. $1 - \pi_v/\pi_d$ provides a tight upper bound on social cost $\sup_{k_j, c_j, e_j, s_j}[(\tilde{W} - W)/D]$.

iii. There exist parameters $c_j \in [0, \infty)$, $e_j \in [0, 1]$, and $s_j \in [0, \infty)$ and distributions of infection risk such that π_v/π_d can be made arbitrarily close to zero.

Proof: To prove part (i), a drug is always developed if it is socially efficient to do so because a drug extracts 100 percent of social surplus. The proof of Proposition 7 provides a case in which a drug is developed but it would have been socially efficient to develop a vaccine. The proof of part (ii) is similar to Proposition 3 with the added fact that the supremum is generated by setting $c_j = s_j = 0$ and $e_j = 1$ for $j \in \{v, d\}$, the values that happen to be assumed in Proposition 3. Part (iii) follows immediately from Proposition 5. *Q.E.D.*

Proof of Proposition 10: Compare the present model involving competition between drugs and vaccines, which we will label Model 1, to the monopoly model laid out in the statement of the proposition, 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 products, 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. According to the paragraph preceding the proposition, we must have $\Pi_d < 0$. (We 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. According to the paragraph preceding the proposition, 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 $\Pi_{vd} = (1 - e_d)h - k_v < 0$. According to the paragraph preceding the proposition, since $\Pi_d > 0 > \Pi_{vd}$, 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 11: 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 product. In the text, we established 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, $\Pi_{v0} > 0$. But, as noted in the text of Section 6.2, $\Pi_{v0} > 0 > \Pi_d$ implies that a vaccine alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{\delta h}{1 - \delta} - k_v. \tag{A27}$$

In the second case, $\Pi_d = e_d h - k_d = e_d h > 0$. Further, $\Pi_{vd} = (1 - e_d)h - k_v < 0$. But, as noted in the text of Section 6.2, $\Pi_d > 0 > \Pi_{vd}$ implies that a drug alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{e_d h}{1 - \delta} - k_d. \tag{A28}$$

The limit as $\delta \to 1$ of the ratio of expression (A27) to (A28) 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.*

References

Blanchflower, D. G. and A. J. Oswald. (2004) "Well-Being Over Time in Britain and the USA," *Journal of Public Economics* 88: 1359–1386.

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.

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.

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. and R. Glennerster. (2004) *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases*. Princeton: Princeton University Press.

Kremer, M. and C. M. Snyder. (2004) "Why Are Drugs More Profitable Than Vaccines?" National Bureau of Economic Research working paper no. 9833.

Kremer, M., C. M. Snyder, and H. Williams. (2006) "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.

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.

Pecorino, P. (2002) "Should the US Allow Prescription Drug Reimports from Canada?" *Journal of Health Economics* 21: 699–708.

Rockstroh, J. K., et al. (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., et al. (1997) "Sexual Transmission of HIV," *New England Journal of Medicine* 336: 1072–1078.

Stole, L. A. and J. Zwiebel. (1996) "Organizational Design and Technology Choice under Intrafirm Bargaining," *American Economic Review* 86: 195–222.

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. (2005) *National Health and Examination Survey (NHANES) 2003–2004*. Washington, D.C.

U. S. Centers for Disease Control. (2006a) "Cases of HIV Infection and AIDS in the United States, 2004," *HIV/AIDS Surveillance Report*, vol. 16. Table 8: "Estimated Numbers of Persons Living with HIV/AIDS, by Year and Selected Characteristics, 2001-2004—35 Areas with Confidential Name-Based HIV Infection Reporting" downloaded March 25, 2006 from www.cdc.gov/hiv/topics/-surveillance/resources/reports/2004report/default.htm.

U. S. Centers for Disease Control. (2006b) *HIV Prevalence Trends in Selected Populations in the United States: Results from National Serosurveillance, 1993–1997*. Table 3: "HIV Prevalence Among Injection Drug Users Entering Drug Treatment Centers, by Metropolitan Area and Sex, 1993–1997" downloaded April 14, 2006 from www.cdc.gov/hiv/pubs/hivprevalence/selected.htm.

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

World Health Organization. (2006) "The Scientific Challenges of an HIV/AIDS Vaccine." Immunization, Vaccines, and Biologicals Division. Downloaded March 17, 2006 from http://www.who.int-/immunization/topics/hiv/en/index2.html.