Abstract: If consumers have heterogeneous disease risk, and firms are unable to price discriminate based on this risk, their ability to extract consumer surplus will differ between preventatives and treatments, potentially distorting R&D decisions. If consumers vary only in disease risk, revenue from treatments always exceeds revenue from preventatives—by an unbounded proportion if the disease-risk distribution is sufficiently skewed. The bias against preventatives can be reversed if consumers also vary in income and income covaries negatively with disease risk. If governments and manufacturers Nash bargain over bulk purchases after products are developed with the threat point of firms making sales on the private market, then R&D biases persist, but if they can bargain ex ante, firms will face efficient R&D incentives. Calibrations based on the joint distribution of income and infection risk for sexually transmitted infections in the U.S. population suggest that firms may only earn half the revenue from a vaccine as from a drug for HIV, but the disadvantage nearly disappears for HPV, which is common enough that its risk distribution does not exhibit much skewness. Empirical tests are consistent with the model’s predictions that vaccines are particularly unlikely to be developed for diseases with substantial infection-risk heterogeneity.

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.

*A previous version of this paper was circulated under the title “Why Is There No AIDS Vaccine?”. The authors would like to thank Emmanuelle Auriol, Chris Avery, Bryan Boulier, Jim Dana, Esther Dufo, Glenn Ellison, Amy Finkelstein, Jon Hamilton, 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, RAND, Stanford, U.C.L.A., University of Pennsylvania, University of Rochester, University of Toronto, the IAEN Symposium on the Economics of AIDS/HIV in Developing Countries (Barcelona), the IDEI Conference on Markets for Pharmaceuticals and the Health of Developing Nations, the International Industrial Organization Conference (Boston), the NBER Summer Institutes on Health and Aging and on Innovation and the Global Economy, and the Southern Economic Association Conference (Charleston) for helpful comments, to David Blanchflower for generously sharing his data, and Lindsey Beckett, Ruben Enikolopov, Scott Lee, Casey Tang, 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.
1. Introduction

Many public health specialists and pharmaceutical observers believe that pharmaceutical firms can earn more revenue from developing products to treat disease than from developing products to treat disease (see e.g. Rosenberg 1999, Thomas 2002). A concomitant view is that governments should adopt policies to encourage the development of preventative health technologies. The view that industry R&D decisions are biased toward treatment, rather than prevention is particularly common in the case of HIV/AIDS, (see, for example, Thomas 2002) and helps explain why a consortium of national governments, together with foundations, support the International AIDS Vaccine Initiative.

We first present a simple model of rational, homogeneous consumers in which developers of preventative and treatments can both fully extract consumer surplus, and hence there is no wedge between private and social incentives to invest in preventatives relative to treatments.

We then show that if consumers are heterogeneous in infection risk, and if firms cannot effectively price discriminate based on this heterogeneity, firms selling preventatives to consumers may be able to extract a smaller proportion of consumer surplus than firms selling treatments.

Consider the following simple example. Suppose that out of 100 risk-neutral consumers, 90 have a 10% chance of contracting a disease while 10 have a 100% chance; they are otherwise identical. Let the harm from the disease be $100. The firm can develop a drug or a vaccine; both are costless to manufacture, are perfectly effective, and have no side effects. If the firm develops a drug, it can sell to all people who contract the disease at a price (equal to the avoided harm) of $100. In expectation, 19 consumers contract the disease: all 10 high-risk consumers, along with nine low-risk consumers. Thus, expected drug revenue is $1,900, which corresponds to the social value of the product. On the other hand, if the firm develops a vaccine, it could either sell to the 10 high-risk consumers at their expected harm of $100, or sell to all consumers at a price of $10, equal to the low-risk consumers’ expected harm. 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.

Section 3 proves a general result that if consumers are heterogeneous only in infection risk, a curative technology yields more revenue than a similarly effective preventative technology. The ratio of preventative to curative revenue equals 1/2 for a uniform distribution of infection risk, is greater than 1/2 for monotonic distributions that are negatively skewed, and is less than 1/2 for monotonic distributions that are positively skewed—indeed the ratio can be driven to zero for sufficiently positively skewed distributions.

How close to zero the ratio can be driven is limited by the prevalence of the disease in the population.
This point is easiest to see in the extreme case in which the disease is ubiquitous: if nearly everyone is expected to contract the disease, there is little scope for the distribution of infection risk to exhibit the dispersion required to generate a gap between preventatives and curatives revenue. We compute a tight lower bound on the ratio as a function of disease prevalence and show that this bound is strictly increasing. The empirical implication is that there may be little difference in the incentives to develop preventatives versus curatives for the most common diseases; diseases must be sufficiently rare for the factors we identify (such as positive skewness in the distribution of infection risk) to impair firms’ relative incentives to develop preventatives.

We then show that adding a second source of consumer heterogeneity—heterogeneity in income as a proxy for willingness to pay—can dampen or indeed reverse firms’ bias against preventatives. This result is stronger the more negative is the correlation between infection risk and income (and requires that the firm not be able to price discriminate based on income). In the extreme, if income varies exactly in inverse proportion to infection risk across consumers, all of our results from Section 3 are inverted: the firm can then capture all social surplus with a preventative but only a fraction with a curative technology.

In Section 5 we consider government procurement. If governments and firms Nash bargain over bulk purchases of preventatives and curatives after products are are developed, with the threat point of no government purchases and only direct sales to consumers, the bias against preventatives persists. However, if bargaining takes place ex ante it may be possible to match private and social R&D incentives. Some real-world mechanisms, such as advance market commitment (Kremer and Glennerster, 2004) and have elements which may help create these commitments.

Having established theoretical bounds on the ratio of preventative to treatment revenue, in Section 6 we calibrate where between these bounds the revenue ratio might fall in practical examples. Distributions of numbers of sexual partners are highly skewed, suggesting that the effects we examine might be important for sexually transmitted infections (STIs). Using U.S. data on the distribution of sexual partners to infer infection risk for STIs, we calibrate vaccine and drug revenue for HIV and HPV (human papillomavirus). These diseases present an interesting contrast because HPV is an order of magnitude more prevalent in the United States than HIV. The highly skewed distribution of sexual partners leads to a highly skewed distribution of HIV infection risk, in turn leading calibrated revenue from a vaccine to fall short of that from a drug by a factor of between two and four. Starting from the same highly skewed distribution of sexual partners, the prevalence of HPV limits how skewed its risk distribution can be. Calibrated vaccine revenue is much closer to drug revenue than for HIV.

Section 6 also provides a separate set of calibrations based on the joint worldwide distribution of
income and HIV risk across countries to shed light on how changes in firms’ ability to price discriminate internationally could potentially affect incentives to invest in HIV vaccines R&D relative to drug R&D. We find that if firms’ existing ability to differentiate prices across countries were eliminated, drug revenue could potentially fall below vaccine revenue.

Section 7 empirically tests the whether infection-risk heterogeneity affects the products that are developed for different diseases and how soon these products became available. We construct a unique dataset including proxies for heterogeneity in infection risk (e.g., STIs, disease concentration in certain subpopulations or regions or transmission through specialized vectors) for a cross-section of diseases. We find that disease-risk heterogeneity significantly reduces the probability of vaccine development—by over 25 percentage points—but has no effect on drug development, consistent with the theory from Section 3.

Section 8 extends the basic model, moving from a monopoly market structure to one in which a drug and vaccine manufacturer may potentially compete and may face competition from generic entrants after a period of patent protection. We argue that this creates an additional factor restricting private incentives for vaccine development relative to drug development compared to the relative social needs, of course in setting forth a new factor which may affect the relative profitability of drugs versus vaccines, we do not seek to deny the potential role of other factors. These factors include risk preferences, behavioral biases, regulatory policy, scientific and technological constraints, and epidemiological externalities (vaccines may limit transmission more than drugs), among many others. Unlike the factor we study, many of these other factors either are straightforward in their effect, are well understood in the literature, do not help explain a bias toward drugs identified by industry observers, or do not address the question of whether firms will invest in vaccines (leaving aside drugs) if the social value of investment exceeds the social cost.

While the scientific and technological difficulties of developing new products and manufacturing and delivery costs may differ between preventatives and treatments, this would not necessarily create a wedge between relative private and social incentives to invest in preventatives as opposed to treatments. In the important special case of infectious disease it is well understood that epidemiological externalities may limit pharmaceutical firms ability to capture social value, and this issue may be more acute for vaccines than drugs. A companion paper, Kremer, Synder, and Williams (2011), examines the determinants of the magnitude of these effects. Other papers that examine firm incentives in the presence of epidemiological externalities include Brito, Sheshinski, and Intrilligator (1991); Boulier (2006); Francis (1997); Geoffard and Philipson (1997); Gersovitz (2003); and Gersovitz and Hammer (2004, 2005).

The analysis in this paper applies to preventative drugs more generally rather than only to infectious diseases, and thus has implications for government policy that are analytically distinct from those found
in the literature on infectious disease externalities. For example, our analysis suggests that there may be underinvestment not only in preventatives for infectious disease, but also in products designed to prevent certain non-communicable diseases such as certain forms of cancer, for which infection risk is heterogeneous. The analysis also applies to other preventative, such as cholesterol-reducing drugs or heart-disease preventatives. Within the class of vaccines against infectious diseases, our analysis suggests biases against vaccines will be particularly severe heterogeneity in relative private and social incentives to develop products for diseases with skewed distributions of private risk, such as HIV. Behavioral factors could also affect willingness to pay for prevention rather than treatment. Here, we wanted to see how far we could get within a rational model.

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 as a function of skewness of the infection-risk distribution and as a function of disease prevalence, all of which are new results in the literature. Our analysis of social welfare in Section 3, calibrations and empirical work, and analysis of generic competition between preventatives and treatments are new as well.

Ideally, public policy would robustly match pharmaceutical manufacturers’ private incentives to develop products to their social value across states of the world so that, whatever technological opportunities for the development of prevention and treatment technologies unfold, manufacturers would have incentives to pursue socially efficient strategies. Our model suggests that existing institutions will not do this. If existing institutions create a good match between private and social incentives for development of new treatments, then the bias we identify would suggest private incentives would be inadequate for preventatives; if they create a good match for preventative development, then incentives for treatment development would be excessive.

This distortion is likely to be particularly severe for HIV. This provides a potential justification either
for subsidies to vaccine R&D, for example, through the International AIDS Vaccine Initiative, or for institutional mechanisms that help commit to pricing in advance, such as those found in the Pneumococcus Advance Market Commitment or the de facto operation of the Advisory Committee on Immunization Policy in the United States.

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. Note that so-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. Statins function as both cholesterol-reducing drugs and as heart-disease preventatives, and thus could be considered a hybrid case.

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. (Proposition 13 will show that the key results continue to 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 \( \pi_j \) be producer surplus (equivalently revenue in the case of costless production), \( \Pi_j = \pi_j - k_j \) be profit, \( CS_j \) be consumer surplus, \( WE_j = CS_j + \Pi_j \) be equilibrium social welfare, and \( WF_j \) 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 \( WE \) be equilibrium social welfare given the firm’s equilibrium choice of product, and let \( WF \) be first-best social welfare given the first-best choice of product.

Consider the case in which the firm sells directly to risk-neutral consumers. 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 \( x \) is denoted \( \bar{F}(x) = \int_x^1 dF(x_i) \). The mean infection risk in the population (also the realized disease prevalence in the absence of a vaccine) is \( \mu = \int_0^1 x_i dF(x_i) \). Assumes the firm knows the distribution of \( x_i \) in the population but cannot price discriminate across consumers based on \( x_i \).\(^1\)

\(^1\)Price discrimination can be ruled out if \( x_i \) is private information for consumers (for example, related to their sexual behavior...
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 $D = h\mu$ 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. The vaccine producer thus sells to the mass of consumers $\bar{F}(p_v/h)$ with infection risk $x_i \geq p_v/h$, implying the profit from developing a vaccine is

$$\Pi_v = \max_{p_v \in [0, \infty)} \left[ p_v \bar{F}(p_v/h) \right] - k_v. \quad (1)$$

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

$$\Pi_d = h\mu - k_d. \quad (2)$$

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 $\mu$ of consumers who become infected. The firm develops a vaccine if $\Pi_v > \max(\Pi_d, 0)$, a drug if $\Pi_d > \max(\Pi_v, 0)$, and neither if $\max(\Pi_v, \Pi_d) < 0$.\footnote{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. Section ?? allows for the possibility that both products are developed in an extension with general parameter values and potential competition between manufacturers.}

3. Equilibrium with heterogeneity in 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 there is no heterogeneity in the distribution of infection risk, so $x_i$ takes on a single, known value in the population of consumers. In equilibrium the firm makes the first-best product choice and produces the first-best quantity of the product.

The proposition follows immediately from the fact that the monopolist can extract 100% of the surplus from homogeneous consumers with either product and thus fully internalizes social welfare.\footnote{The firm may no longer have first best incentives for product development if we depart from the monopoly assumption by...}
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% of consumer surplus with a vaccine. Producer surplus from a vaccine, $\pi_v$, will thus fall below producer surplus from a drug, $\pi_d$, 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, $\pi_v$, equals the area of the largest rectangle that can be inscribed under inverse demand curve $\bar{F}(p_v/h)$, while $\pi_d$ 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 $\pi_v$ relative to $\pi_d$, the greater the firm’s bias against vaccines. The producer-surplus ratio $\pi_v/\pi_d$ (more precisely, one minus this ratio) provides a convenient index of the bias against vaccines because this ratio can be linked to the potential social cost of this bias, allowing patent races, finite patent lives, rent-dissipating competition, etc. Section ?? discusses some of these issues further.
as Proposition 3, proved in the Appendix, formalizes.

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

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

Proposition 2 states that the firm will be biased against vaccines if there is heterogeneity in infection risk, raising the 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 $\frac{\pi_v}{\pi_d}$ can be made arbitrarily close to $1/R$, a lower bound on $\frac{\pi_v}{\pi_d}$.

The Introduction offered an example with two risk classes (90 consumers with a 10% chance of contracting the disease and 10 with a 100% chance) in which $\frac{\pi_v}{\pi_d} = 0.53$. The fact that this result was close to 1/2 was no accident: an implication of Proposition 4 is that $\frac{\pi_v}{\pi_d}$ 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 there exist distributions of consumer types such that $\frac{\pi_v}{\pi_d}$ can be made arbitrarily small. This can be seen by taking the limit as $R$ approaches infinity in the proposition.

**Proposition 5.** There exist distributions of consumers such that $\frac{\pi_v}{\pi_d}$ can be made arbitrarily close to zero.

When is the bid likely to be large? 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 $\frac{\pi_v}{\pi_d}$.

**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 $\frac{\pi_v}{\pi_d} = 1/2$. If $f'(x_i) > 0$ (a sufficient condition for negative skewness), then $\frac{\pi_v}{\pi_d} > 1/2$. If $f'(x_i) < 0$ (a sufficient condition for positive skewness), then $\frac{\pi_v}{\pi_d} < 1/2$.

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
\( \bar{F}(p_v/h) \) 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 negatively skewed. The associated inverse demand is then 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 positively 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 \( \pi_v/\pi_d < 1/2 \).

We saw from Proposition 6 that the revenue ratio \( \pi_v/\pi_d \) is bounded below if the monotone infection-risk distribution is uniform or negatively skewed. Another lower bound on the revenue ratio can be obtained by focusing on the prevalence of the disease, which in the absence of a vaccine equals \( \mu \). Such a bound is empirically useful because prevalence is readily observable. Intuitively, if \( \mu \) is close to 1, most consumers’ infection risk must be close to 1, limiting how much heterogeneity there can be in the distribution of infection risk. Lower values of \( \mu \) allow for more heterogeneity in infection risk, but there are limits to this heterogeneity for any given value of \( \mu \).
Proposition 7. Take the prevalence of the disease in the absence of a vaccine, $\mu$, to be some constant in $[0, 1]$. A tight lower bound on $\pi_v/\pi_d$ is provided by the implicit solution for $B$ in

$$B[1 - \ln(B\mu)] = 1.$$ 

(3)

$B$ is strictly increasing in $\mu$, with $\lim_{\mu \to 0} B = 0$ and $\lim_{\mu \to 1} B = 1$.

Figure 3 graphs numerical solutions for $B$ as a function of $\mu$. The empirical implication of the figure is for the most common diseases, infection-risk heterogeneity cannot be an important factor in a firm’s decision to develop a vaccine versus a drug. For example, the figure shows that if the prevalence of the disease is above 0.74, it is mathematically impossible to generate enough infection-risk heterogeneity to drive $\pi_v/\pi_d$ below 1/2. The results from this section that heterogeneity and skewness in infection risk contribute to a bias against vaccines are more likely to apply to sufficiently rare diseases.

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 8. The firm never develops a vaccine unless it is socially efficient to do so. There exist cases in which the firm develops a drug but it would have been socially efficient to develop a vaccine.

Proposition 8 holds whether social efficiency is measured by first-best social welfare ($WF_j$) or equilibrium social welfare ($WE_j$). 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 negatively skewed distributions, and more for positively skewed distributions. In sum, the model with consumer heterogeneity in the single dimension of infection risk suggests that R&D decisions may be biased against vaccines and that the social loss from these biases can be quite large for positively skewed distributions of infection risk.

4. Income Heterogeneity

This section shows that the results may be attenuated or even reversed in 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, \(y_i\), for example due to variation in income.\(^5\) 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, negative correlations between \(x_i\) and \(y_i\) over some region empirically can generate cases in which the firm prefers to develop a vaccine rather than a drug. As we will see, this case may be more than a theoretical curiosity because infection risk and income are highly negatively correlated for HIV.

To see this, 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 income, which will serve as a proxy for \(i\)'s willingness to pay for a given reduction in probability of infection.\(^6\) 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 his or 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 \(z_i\).

Consider the vaccine producer's profit-maximization problem. Consumers buy the vaccine if \(z_i \geq p_v\),

\(^{5}\)Kessing and Nuscheler (2002) study monopoly pricing of a vaccine when income is the sole source of consumer heterogeneity.

\(^{6}\)Besides income, other sources of variation in willingness to pay include differences across consumers in the physical harm caused by the disease and differences in risk preferences (see, e.g., Cutler, Finkelstein, and McGarry 2008).
implying the demand for the vaccine is $\bar{F}_Z(p_v)$, where $\bar{F}_Z(p_v) = \int_{p_v}^{h} dF_Z(z_i)$. Hence

$$\Pi_v = \max_{p_v \in [0, \infty)} \left[ p_v \bar{F}_Z(p_v) \right] - k_v. \quad (4)$$

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 $E_{X|Y}(x_i|y_i \geq p_d) \bar{F}_Y(p_d)$, where $E_{X|Y}(\cdot)$ is the expectation taken with respect to the conditional distribution $F_{X|Y}$ and where $F_Y(p_d) = \int_{p_d}^{h} dF_Y(y_i)$. Hence

$$\Pi_d = \max_{p_d \in [0, \infty)} \left[ p_d E_{X|Y}(x_i|y_i \geq p_d) \bar{F}_Y(p_d) \right] - k_d. \quad (5)$$

We saw in Proposition 2 that if infection risk is the only source of heterogeneity, $\pi_d > \pi_v$. With multiple sources of heterogeneity, $\pi_v$ and $\pi_d$ can no longer be unambiguously ranked. Roughly speaking, the amount of consumers’ private information embodied in (4)—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 (5) 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 (5), 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 9.** Assume there is heterogeneity in the distribution of infection risk among vaccine consumers. If $y_i$ is an increasing function of $x_i$, or if $y_i$ 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 10.** Adding income heterogeneity that is distributed independently from the heterogeneity in infection risk causes $\pi_v / \pi_d$ 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.

When income and infection risk are negatively correlated in the relevant region, so that there is less private information in the joint distribution than the marginal distribution of $y_i$, profits from developing a vaccine can exceed those from developing a drug. This is easiest to see in the extreme case of negative correlation between $x_i$ and $y_i$ in which $z_i = \bar{z}$ for all $i$, implying $x_i = \bar{z}/y_i$. 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 from Section 3 carries over under a suitable reinterpretation, with that the roles of vaccines and drugs reversed and heterogeneity in $y_i$ is substituted for heterogeneity in $x_i$. Proposition 2 would then imply $\pi_v > \pi_d$; Proposition 5 would then imply that distributions of $y_i$ can be constructed such that vaccines generate arbitrarily higher producer surplus than drugs; Proposition 6 would then imply that vaccines would generate twice the producer surplus of drugs if the distribution of $y_i$ were uniform; and so forth.

5. Alternative Purchasing

Having laid the groundwork by first considering the cases of sales direct to consumers, in this section, we consider alternative purchasing, in particular the case in which firms can sell through an insurance contract or to the government as a bulk purchaser.

If it is possible to sell future drug access (through an insurance contract, for example) to consumers before their infection status is realized, 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 the previous sections, would then be reinterpreted as indicating when the manufacturer would prefer to sell the drug ex ante versus ex post.

In considering government purchases, we first consider the case in which governments and firms Nash bargain over bulk purchases after products are developed with the threat point if negotiations between the government and the firm break down consisting of the equilibrium with direct sales to consumers considered above. In this case, the distortions considered above to R&D incentives persist, albeit likely in attenuated form. If government, and firms can negotiate over bulk purchases, before products are developed it is possible to generate appropriate incentives to encourage firms to invest in developing vaccines and drugs. Institutions such as Advisory Committee on Immunization Practices (ACIP) in the United States and Advance Market Commitments internationally may be seen as helping governments commit to vaccine pricing and purchasing policy even before vaccines are purchased, this addressing the hold up problem.
associated with ex post negotiation.

Consider first the case without commitment. 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. Assume that the government’s objective is to maximize consumer surplus and that the (firm’s) Nash-bargaining surplus is

\[
n_j = \frac{1}{2}[(WF_j + k_j) + \pi_j - CS_j],
\]

a combination of the first-best “pie” toward which parties bargain, \( WF_j - k_j \), plus the firm’s threat-point surplus from selling product \( j \) on the private market, \( \pi_j \), minus the government’s surplus in this threat point, \( CS_j \). Substituting \( WE_j = \pi_j - k_j + CS_j \) into (6), we have \( n_j = \pi_j + (WF_j - WE_j)/2 \), implying that the firm’s objective function with government procurement is the sum of its objective function with private procurement \( \pi_j \) 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.\(^7\)

It is straightforward to see that if firms and governments bargain in advance of a product’s development instead of after, it is possible to reach efficiency. There are some practically relevant circumstances in which firms and governments bargain before products are developed. For example, the U.K. government negotiated for the advance purchase of a meningococcal vaccine. For another example, while the recommendations of the U.S. Advisory Committee on Immunization Practices (ACIP) are not legally binding, in practice they are almost always followed, and the ACIP has established a practice of conducting cost-effectiveness analyses prior to firm’s price setting and recommending that vaccines be purchased if the price falls below a threshold which would make them cost effective. Firms respond by pricing this threshold. This effectively commits to a price setting procedure in advance.

Finally the advance market commitment programs for vaccines of the type described by Kremer and

\(^7\)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.
Glennerster (2004) and implemented for pneumococcus vaccine by a group of bilateral donors foundations and the Gates Foundation. This group committed to help finance purchase of a pneumococcus vaccine covering strains of the disease common in developing countries at a price that would roughly correspond to something slightly under its social value under certain conservative assumptions. (GAVI 2009).

6. Calibrations for Sexually Transmitted Infections

Section 3 showed that heterogeneity in infection risk could lead toward a bias in favor of developing drugs rather than vaccines, while Section 4 argued that negative correlation between income and infection risk could potentially lead firms to favor vaccines over drugs. Since bias towards either vaccines or drugs is possible a priori, the direction and sign the bias depends on the empirical joint distribution of infection risk. This section calibrates the model using data on the joint distribution of infection risk and income for sexually transmitted infections, first within the United States and then across countries. We focus on sexually transmitted infections because available data can be used to infer the distribution of infection risk for them and because the risk distribution exhibits some of the features (skewness, rareness) that proved to be important in the theory.

More specifically, most of our calibrations will be for HIV. Using individual-level data for the U.S. market in Subsection 6.1, the calibrated revenue for an HIV vaccine is generally much lower than for a drug, only one quarter to one half as much, providing a potential contributing factor for the continued delay in developing HIV vaccines relative to drugs. These results contrast additional calibrations for HPV, a much more common disease than HIV and with an infection-risk distribution that is consequently less skewed. The calibrated revenue for an HPV vaccine is close to that for a drug, suggesting that firms may have less bias against developing HPV vaccines.

In Subsection 6.2 we move from U.S. to cross-country data on the joint distribution of HIV risk and income. These calibrations will allow us to explore the effect of international price discrimination in the pharmaceutical market on relative incentives to invest in vaccines. The calibrations suggest that restricting the scope for international price discrimination could potentially substantially reduce revenue from an HIV drug, possibly below that from a vaccine.

6.1. U.S. Market

The U.S. pharmaceutical market is by far the world’s largest and is widely seen as the driver of firms’ R&D decisions. Several surveys report information on risk factors for HIV and other sexually transmitted
Table 1: Vaccine/Drug Producer Surplus Ratio in Calibrations for the U.S. Market

<table>
<thead>
<tr>
<th>Survey</th>
<th>GSS</th>
<th>GSS</th>
<th>NHANES</th>
<th>GSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income heterogeneity:</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ages in sample:</td>
<td>All</td>
<td>35-40</td>
<td>All</td>
<td>All</td>
</tr>
</tbody>
</table>

(1) (2) (3) (4)

HIV calibrations

- HIV1: Linear model 0.253 0.260 0.227 0.496
- HIV2: Kaplan model with $\beta = 0.06\%$ 0.252 0.265 0.246 0.504
- HIV3: Kaplan model with $\beta$ varying by sexual orientation, race, IV drug use 0.316 0.369 0.371 0.571

HPV calibrations

- HPV1: Kaplan model with $\beta = 13.5\%$ 0.482 0.517 0.547 0.830

Observations

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17,255</td>
<td>2,478</td>
<td>2,457</td>
<td>15,827</td>
</tr>
</tbody>
</table>

Infections such as numbers of sexual partners. 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 calibrations use nationally representative data on the lifetime number of sexual partners broken down by the individual’s gender and sexual orientation and the partners’ genders from the 1989–2004 General Social Survey (GSS) to calibrate the model of Section 3.\(^8\) The distribution of lifetime sexual partners is highly positively skewed: the median is 3 but the mean is 10.7. This skewness induces skewness in the distribution of infection risk in our calibrations, which in turn leads to a large gap between the producer surplus from a vaccine and a drug.

Column (1) of Table 1 contains the results from calibrations that use GSS data and that account for infection risk heterogeneity but not income heterogeneity. The calibration labeled HIV1 involves a simple linear mapping from lifetime sexual partners to infection risk with a constant probability of transmission per partner. Figure 4 graphs the resulting inverse demand curve for this calibration. The positively skewed distribution of infection risk produces a highly convex inverse demand curve. Recall $\pi_v$ is given by the area of the largest rectangle that can be inscribed under the curve (the shaded rectangle in the figure) and $\pi_d$ by the area under the curve. It is apparent that $\pi_v$ is much less than $\pi_d$; to be precise, $\pi_v/\pi_d = 0.253$. As shown in the figure, the firm’s optimal strategy in this calibration turns out to be to sell the vaccine at

\(^8\)We use the cleaned version of the GSS data used in Blanchflower and Oswald (2004) among other studies. 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 the same 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.
Figure 4: Inverse demand curve for calibration 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 $p_v = 0.25$.)

a high price to a small segment of high-risk individuals.

In the row of calibrations labeled HIV2, 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 $\beta$ is the probability of contracting the disease from any given partner. We take $\beta = 0.06\%$, equal to an estimate of the current HIV prevalence rate in the United States, which according to UNAIDS (2004) is 0.6%, times the average per-partner transmission rate, which following Rockstroh et al. (1995) we take to be 10%. The estimated figure for $\pi_v/\pi_d$, 0.252, is quite similar to that from the linear model.\(^9\)

In the row of calibrations labeled HIV3, we allow the $\beta$ in the Kaplan model to vary by sexual orientation,\(^10\) race,\(^11\) and interavenous (IV) drug use.\(^12\) These are important sources of infection-risk

\(^9\)Results are insensitive to varying $\beta$ by one third in either direction.

\(^10\)For the male partners of males, we scale $\beta$ up in two stages. We first multiply by 36.8, the estimated prevalence of HIV among homosexual males relative to the general population, computed by taking the percentage of people living with HIV 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 Centers for Disease Control (2006a)—and dividing by the percentage of homosexual males in the population, estimated to be 1.2% in our GSS data. We further scale $\beta$ by a factor of three to reflect the estimate from Royce et al. (1997) that HIV is three times more likely to be passed between males than from males to females. For the rest of the sample, we scale $\beta$ by 0.58, equal to the prevalence of HIV among the population that is not homosexual male relative to the prevalence in the general population (including homosexual males). Given the small number of bisexual males in the GSS sample, 0.2%, 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.

\(^11\)We take the $\beta$ parameters which have been adjusted to reflect variation in infection risk by sexual orientation as described in the previous footnote, and further scale them by 2.55 for African American, 0.324 for whites, and 1.00 for Hispanics, estimated from statistics from the Centers for Disease Control (2006a). Implicit in this scaling is the assumption that an individual matches with partners of the same race.

\(^12\)The GSS does not report IV drug use, so we resort to other data sources. A study of HIV prevalence among IV drug users in U.S. drug treatment centers (Centers for Disease Control 2006b) found that HIV prevalence averaged 18% but varied across
heterogeneity in the population: our estimates suggest that HIV is over 60 times more prevalent among homosexuals than heterosexual males, eight times more prevalent among blacks than whites, and over 30 times more prevalent among IV-drug users than others. Although one might expect these additional potential sources of heterogeneity to reduce the relative profitability of vaccines, in fact $\pi_v / \pi_d$ increases from 0.252 to 0.316 in column (1). The firm ends up concentrating its sales of the vaccine among even higher-risk individuals compared to the previous calibration. Although sales fall, the vaccine price to these consumers can be increased enough that the overall profitability of vaccines rises.

Columns (2) and (3) provide robustness checks. Column (2) repeats the calibrations from column (1) for a single age cohort, 35 to 40 year olds. At the cost of a smaller sample size, the calibrations 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 with different sexual norms. The producer-surplus ratio $\pi_v / \pi_d$ increases slightly across calibrations from column (1) to (2), for example from 0.253 to 0.260 for the linear model. Column (3) repeats the calibrations from column (1) using a different data source for infection risk: the 2003–2004 National Health Examination Survey (Centers for Disease Control 2005), or NHANES. The resulting producer surplus ratios are close to their analogues in column (1).

Column (4) repeats the calibrations 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 $x_i$ 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. Comparing the results to 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 reduced, the calibrations in column (4) still suggest that the producer surplus from drugs is nearly twice that from vaccines.

It is also possible to calibrate the impact of government purchases. Assuming that the firm and government engage in Nash bargaining over the supply of product $j$ to all consumers below a certain cities, ranging from 1% in a Los Angeles to 36% in New York City. Coupled with an estimate of the total number of HIV cases due to IV drug use from 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. Since we do not have information on income for IV drug users, for the calibration in column (4) we take their income to be the U.S. poverty line for individuals ($9,827 in 2004). This is likely to overstate most IV drug users’ income, but any multiple from 0 to 1.25 times the poverty line produced the same result as in the table. At any of these low income levels, IV drug users cannot afford vaccines or drugs in the calibration.
income threshold (say 75% of the U.S. poverty line, the threshold for Supplemental Security Income eligibility) and that the firm sells to the rest of the consumers as usual on the private market, we can perform calibrations analogous to those in Table 1 to determine the effect of the government program. In the last calibration in Table 1 ($\beta$ varies by sexual orientation and race, including IV drug users and income heterogeneity), the producer-surplus ratio, $\pi_v/\pi_d$, 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.

As a counterpoint to the calibrations for HIV, Table 1 adds a set of calibrations for a much more common disease, HPV. These calibrations, labeled HPV1, are directly comparable to the HIV2 calibrations—both are Kaplan models with fixed values of $\beta$—but $\beta$ is increased from 0.06% to 13.5%. The ratio of vaccine to drug producer surplus is much greater for HPV than HIV across all four columns. Indeed, in the calibration in column (4), the ratio of 0.830 is quite close to 1. With a disease as prevalent as HPV, the infection risk cannot be very positively skewed, putting a bound on the discrepancy between vaccine and drug revenue, as shown in Figure 3.

6.2. 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 calibration in this section suggests that the abolition of international price discrimination would substantially reduce the profitability of drugs. The calibration also illustrates the 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 calibration, because it assumes no price discrimination across countries, is for a counterfactual case.

We consider the market as consisting of the entire world population and treat all individuals within

---

13 This value of $\beta$ is computed as the HPV prevalence rate times its transmission rate. Dunne et al. (2007) estimated the prevalence among U.S. women of the HPV strains classified as posing a high cervical-cancer risk as 15.2%. Dunne et al. estimated the prevalence of the four strains included in the Gardasil HPV vaccine as 3.4%, but the vaccine also offers cross-protection against other high-risk strains (Ault 2007). Data from from Hernandez et al. (2008) data imply an HPV transmission rate of 88.8%; of the 18 couples in which one partner had an HPV strain that the other did not at the beginning of their study, 16 ended up transmitting a strain to the other.
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 HIV prevalence to approximate our two sources of consumer heterogeneity.\footnote{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).} We approximate $x_i$ by the fraction of people within a given country that are HIV-positive and $y_i$ by per-capita GNP. The correlation between $x_i$ and $y_i$ across countries is significantly negative at $-0.13$, raising the possibility that $\pi_v > \pi_d$.

Figure 5 shows the inverse demand curve for an HIV 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 strictly induces consumers in 16 other countries to buy. The axes on the two panels of Figure 5 have been scaled so that a unit of area in both represents the same revenue. The rectangle for the vaccine is slightly
larger: \( \frac{\pi_v}{\pi_d} = 1.13 \).

The analysis suggests that impeding international price discrimination would diminish revenue from an HIV 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 calibration, 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.

7. Empirical Tests

The basic theory from Section 3 suggests that holding constant the burden of a disease, changing the distribution of infection risk by adding heterogeneity or positive skewness should reduce their incentives to develop a vaccine. If income heterogeneity is added to the model as in Section 4, however, heterogeneity in infection risk could potentially either increase or decrease incentives for vaccine development. The calibrations in the previous section suggest that in the case of STIs, heterogeneity in infection risk is likely to reduce incentives for vaccine development. Contrasting the implications for vaccines, the model provides no reason for skewness to affect drug development.

In this section, we present empirical tests using data on vaccines and drugs that have been developed over the last century for a sample of about 100 infectious-disease-causing microorganisms. Unfortunately, quantitative information on the distribution of infection risk is not systematically available for a cross-section of diseases. Instead, we develop several proxies for heterogeneity and positive skewness in infection risk and combine these proxies into a single indicator. The imperfect nature of these proxies is a source of measurement error that may reduce the power of our tests and/or add bias but they are the best measures we had available.

We use this indicator for the shape of the infection-risk distribution as a right-hand-side variable in a model of product (vaccine or drug) development. We use a linear probability model to study a 0–1 measure of whether a product has been developed for a disease. The presumption underlying the model is that lucrative products are more likely to be developed. Of course, many other factors are important determinants of product development, factors including the ease of the science involved, other cost factors,
Table 2: Descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaccine development indicator</td>
<td>91</td>
<td>0.29</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Drug development indicator</td>
<td>91</td>
<td>0.69</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Infection-risk heterogeneity</td>
<td>91</td>
<td>0.46</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Childhood onset</td>
<td>91</td>
<td>0.15</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Viral</td>
<td>91</td>
<td>0.43</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Prevalence (max. over period)</td>
<td>51</td>
<td>0.52</td>
<td>1.11</td>
<td>0</td>
<td>4.74</td>
</tr>
<tr>
<td>Prevalence (time varying)</td>
<td>6,528</td>
<td>0.16</td>
<td>0.54</td>
<td>0</td>
<td>4.74</td>
</tr>
</tbody>
</table>

Notes: Fewer observations for year of vaccine or drug development because descriptive statistics for subsample of diseases having that product developed. Prevalence measured in yearly cases per 1,000 U.S. population.

government subsidies, number of competing firms, etc. Indeed, we will study a number of these factors further in Section 8. We control for the type of organism causing disease (virus vs. bacterium) because it is believed to be relatively easier technologically to develop vaccines rather than drugs for viral disease. Our assumption is that other factors do not vary systematically across diseases and are captured by the error term.

The dataset was constructed by a team of research assistants including a senior medical student. A list of disease-causing organisms was taken from Harpavat and Nissim (2001), a widely-used teaching reference that covers the most clinically important organisms. This source provided summary information on type of organism (bacterium, virus, parasite, fungus), available treatments, whether children or adults are disproportionately affected, sexual and insect transmission, etc. Table 2 provides descriptive statistics for the dataset. Note that fewer observations are listed for the vaccine and drug development years because the descriptive statistics are conditional on the disease having that product developed for it.

The indicator for infection-risk heterogeneity deserves special comment because it is the regressor of central interest. This indicator is set to 1 if a discrete high-risk group could readily be defined from a review of the disease’s epidemiology and transmission patterns. Specifically, the indicator is set to 1 if

15 We limited attention to bacterial and viral diseases because all variation in the availability of products for other types of organisms (parasitic, fungal) would be captured by organism fixed effects.

16 This basic source was supplemented by the microbiology reference Mandell, Bennett, and Dolin (2010). Dates of product development were compiled from Mandell, Bennett, and Dolin (2010), the dates of vaccine development supplemented by public-health websites (Centers for Disease Control 2009, National Network for Immunization Information 2009, Immunization Action Coalition 2009, U.S. Food and Drug Administration 2009) and the dates of drug development by medical histories (Corey, Kürti, and Czakó 2007; Greenwood 2008). Historical data on disease prevalence was taken from the *Morbidity and Mortality Weekly Report* (various years).
the disease satisfies at least one of the following conditions:

- sexually transmitted;
- transmitted by animal contact;
- chiefly affects a small population of either hospitalized patients, immuno-compromised individuals, intravenous-drug users, or soldiers;
- organism has restricted ecological habitat (e.g., tropics for malaria).

For each disease, we construct a time series of prevalence by taking the number of reported U.S. cases each year from 1944 to 2007 from the *Morbidity and Mortality Weekly Report* (various years), expressed per 1,000 population. This information was only available for a subset of 51 “notifiable” diseases as defined by the Centers for Disease Control. Because diseases enter and exit the notifiable list over time, we interpolate and extrapolate missing years using a quadratic time trend for each disease. The resulting panel involves 6,528 disease-year observations. A single prevalence measure for use in cross-sectional regressions is computed by taking the maximum over non-missing years for each disease.\(^\text{17}\)

Table 3 reports the results from two specifications of a linear probability model, which regresses an indicator for product (vaccine or drug) availability on infection-risk heterogeneity using ordinary least squares. Results from alternative specifications (probit, logit) are quite similar. Consider the spare specification in columns (1)–(3) in which infection-risk heterogeneity is the only covariate. The –0.265 coefficient in the first row of column (1) indicates that vaccines are 26.5 percentage points less likely to have been developed for diseases with infection-risk heterogeneity, significant at the 1% level. The analogous coefficient in column (2) indicates that there is no statistically significant effect of infection-risk heterogeneity on drug development. The difference between the vaccine and drug coefficients in column (3) indicates that infection-risk heterogeneity reduces vaccine development 26.2 percentage points more than it does drug development, a difference significant at the 10% level.

The difference between the constant terms in column (3) indicates that vaccines are less common than drugs, the average disease being 28.6 percentage points less likely to have a vaccine than a drug, significant at the 1% level. This result may capture a host of factors besides heterogeneity in infection risk that may make vaccines harder to market than drugs such as tendencies for people to invest less on prevention or the greater epidemiological externalities from vaccines.

One concern with results is that our infection-risk heterogeneity may be proxying for more than just the shape of the risk distribution; it may be proxying for low overall disease burden, as diseases that

\(^{17}\)We use the historical maximum to address the problem that a product’s introduction may reduce the disease’s prevalence, inducing a correlation between the prevalence variable and the regression error. The maximum captures prevalence in the absence of a drug or vaccine. The results are similar using alternative prevalence measures such as the mean over the period rather than the maximum.
are transmitted through specialized vectors or concentrated in subpopulations may have an overall low prevalence. Virtually any theory would suggest that firms would have less of an incentive to develop products for low-burden diseases, and so a significantly negative coefficient on our proxy may not be a dispositive test of the particular theory in Section 3. This concern is partially addressed in the spare specification by focusing not on the negative coefficient in the vaccine regression in isolation but on a comparison of the vaccine to the drug regression. If infection-risk heterogeneity were proxying for low overall disease burden, one would expect to find a negative effect on drug development as well, but the coefficient on infection-risk heterogeneity in column (2) is close to 0. The result in column (3), which can be viewed as a difference-in-differences, indicates that our proxy is having a statistically significantly different effect on vaccine than on drug development.

The concern is further addressed by the rich specification, reported in columns (4)–(6), adding an explicit prevalence measure as well as other controls. The sample is restricted to the subset of 51 observations for which we have prevalence data. The results are if anything a bit stronger than in the spare specification, with infection-risk heterogeneity decreasing the probability of vaccine development by a statistically significant 40.0 percentage points, but having essentially no effect on drug development, resulting in a differential effect on vaccines vs. drugs reported in column (6) of 35.5 percentage points, now significant at the 5% level.

The additional controls in the rich specification are of some independent interest. Vaccines are significantly more likely to be developed for diseases that disproportionately affect children and drugs significantly less likely. This is consistent with the widespread practice of childhood immunization programs. Viral diseases show the same pattern, indicating that the technology of vaccine production is particularly suitable for viruses. The prevalence measure does not show up as important in any regression. One explanation is that the subsample in the rich specification, including as it does only diseases listed as notifiable by the Centers for Disease Control, already selects for high-burden diseases, so there may not be important variation left for a prevalence measure to capture.\footnote{Confirming this explanation, we ran a regression similar to the rich specification but retaining all 91 observations and including an indicator for CDC-notifiable diseases; this indicator was quite often large, positive and statistically significant. We prefer the reported specification because it involves a more homogeneous set of diseases and because the omitted CDC-notifiability indicator may be endogeneous, in particular if the CDC is more likely to require notification for disease that are part of immunization programs.}

We also ran a Cox proportional hazards model on the date of product development. It is not completely clear how to test for the interaction between infection risk heterogeneity and product type in this model. Ai and Norton (2003) have questioned the interpretation of interaction terms in nonlinear models such as a hazard model. Puhani (2008) shows that this provides the correct test in the special case of a treatment
Table 3: Impact of infection-risk heterogeneity on product development

<table>
<thead>
<tr>
<th>Variable</th>
<th>Vaccine developed</th>
<th>Drug developed</th>
<th>Difference</th>
<th>Vaccine developed</th>
<th>Drug developed</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3) = (1) − (2)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6) = (4) − (5)</td>
</tr>
<tr>
<td>Infection-risk heterogeneity</td>
<td>−0.265***</td>
<td>−0.003</td>
<td>−0.262*</td>
<td>−0.400***</td>
<td>−0.044</td>
<td>−0.355**</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.098)</td>
<td>(0.145)</td>
<td>(0.136)</td>
<td>(0.089)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>Childhood onset</td>
<td>0.408***</td>
<td>−0.242*</td>
<td>0.650***</td>
<td>0.204*</td>
<td>−0.693****</td>
<td>0.897****</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.122)</td>
<td>(0.130)</td>
<td>(0.121)</td>
<td>(0.116)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>Viral</td>
<td>0.408***</td>
<td>0.694***</td>
<td>−0.286***</td>
<td>0.491***</td>
<td>1.037****</td>
<td>−0.546***</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.067)</td>
<td>(0.101)</td>
<td>(0.123)</td>
<td>(0.043)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Prevalence (max. over period)</td>
<td>−0.022</td>
<td>0.011</td>
<td>−0.033</td>
<td>−0.025</td>
<td>(0.023)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.09</td>
<td>0.00</td>
<td>0.39</td>
<td>0.67</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

Notes: Ordinary least squares regressions in which dependent variable is an indicator for development of product. Bacterial is omitted organism category in rich specification. White (1984) heteroskedasticity-robust standard errors reported in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Effect of the “difference-in-differences” sort, but not necessarily more generally. In the spare specification, infection-risk heterogeneity cuts the hazard of vaccine development by more than two-thirds but does not reduce the hazard of drug development. In the rich specification of the hazard model, infection-risk heterogeneity leads to a similar reduction in the hazard of vaccine development as in the spare specification. There is now also some evidence it reduces the hazard of drug development (at the 10% level). The point estimate implies that infection-risk heterogeneity reduces the vaccine hazard by twice as much as the drug hazard, but due to large standard errors, the effect on the vaccine hazard is not statistically significantly different from that on the drug hazard.

Overall, our empirical tests are consistent with the idea that infection-risk heterogeneity reduces incentives to develop vaccines.

8. Temporary IPR protection

Thus far we have focused on the case of a monopoly pharmaceutical manufacturer. In this section, we first relax these assumptions and then show that competition can lead to an additional bias against vaccines.
in a plausible oligopoly model in which the patent system provides only temporary monopoly power to a firm that develops a new product, after which there is generic entry.

It is useful to first relax the assumption that all products are perfectly safe and effective and costless to manufacture. The key welfare results from Section 3 continue to hold, relaxing the simplifying assumptions that products are perfectly safe and effective and costless to manufacture and administer and allowing for more general product characteristics. Let $c_j \in [0, \infty)$ be 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. The Appendix provides the formal restatement of the key welfare results for general parameter values (Proposition 13) along with a proof.

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. Let $\delta \in [0, 1]$ be the per-period discount factor.

The first firm to develop a product enjoys patent protection for one period. 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. 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

---

19 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.

20 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 $k_j$. 

26
will die before generic drugs become available. On the other hand, competition from a drug does reduce
the profits of the vaccine patentee. The vaccine patentee 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
\(\Pi_d\) be the single-period monopoly profit from a drug. Extending (2) to allow for general parameter values,
it can be shown that
\[
\Pi_d = (e_d h - s_d - c_d)\mu - k_d.
\]
In the competition model, the firm earns \(\Pi_d\) as well, whether its rival produces a vaccine or does not enter. The firm earns this \(\Pi_d\) 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 \(\Pi_{v0}\), has the same functional form as \(\Pi_v\) 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 \(\Pi_{v0}\) from selling to consumers in generation
\(Y_1\). The discount factor \(\delta\) 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 \(\Pi_{vd}\),
again has the same functional form as \(\Pi_v\) in equation (1), but now the formula for the cutoff type is
\[
\hat{x}(p_v) = \frac{p_v + s_v}{\delta e_v |c_d + s_d + (1 - e_d) h|}.
\]
(7)
Equation (7) 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 (7) 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 11.** 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 11 may be socially costly, as the next proposition states.

**Proposition 12.** 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.

**9. Conclusion**

In this paper, we argued that differences in the timing of when drugs and vaccines are taken may affect firms’ ability to extract consumer surplus under direct sales to consumers. 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 price discriminate by income or offer insurance contracts for drugs sold in advance of consumers learning their infection status. Allowing firms to sell insurance contracts for their products creates a potentially valuable option for a drug manufacturer, which can choose to sell drug insurance ex ante (before infection) or continue just selling the drug ex post. The option is worthless for a vaccine manufacturer, whose product already functions like insurance because it is administered ex ante, before infection.

An extension incorporating competition between a vaccine and a drug as well as later generic entrants suggests an additional bias against vaccines. Future generic drug production constrains vaccine pricing, but drug pricing is unaffected by competition from vaccines.

If governments cannot commit to a pricing policy in advance of product development, bulk government procurement reduces but does not eliminate gaps 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.

A calibration using estimates of the joint distribution of income and disease risk in the United States suggests vaccine revenue would only be about half drug revenue for HIV but would almost equal drug revenue for HPV. The difference is that HIV is rare enough that the skewness in number of sexual partners generates skewness in HIV infection risk while HPV is so prevalent that it is mathematically impossible for HPV infection risk to exhibit much skewness. Calibrations for HIV revenue in the international market provide insight into the effects of banning price discrimination across countries. Such a ban may reduce incentives to develop HIV drugs.

As an empirical test of the model, using a novel dataset on infectious diseases, we regressed indicators for whether drugs or vaccines have been developed on a indicator for heterogeneity in infection risk, which we constructed from underlying proxies, along with other controls. In line with the basic theory, we found vaccines are significantly less likely to have been developed for diseases with heterogeneity in infection risk we found no similar effect for drugs.

A companion paper (Kremer, Snyder, and Williams, 2006) builds an integrated economic and epidemiological model of externalities from combatting infectious diseases and finds that the ratio of social-to-private benefit from prevention, 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 common but less serious diseases like the flu more profitable than for
rarer but more deadly diseases. Since HIV 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 HIV drugs than of HIV vaccines.

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. However, in the absence of clear information on the efficient level of R&D, 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. This will not be achieved under current institutions. Although antiretroviral drugs are keeping a high proportion of HIV-infected individuals in high-income countries alive, many in the poorest countries are not benefitting from these technologies; and the development of an HIV 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 \( F(p_v/h) = \int_{p_v/h}^1 f(x_i) \) into equation (1) and making the change of variables \( \tilde{x} = p_v h \) yields \( \pi_v = h \int_0^1 \tilde{x}^+ dF(x_i) \), where
\[
\tilde{x}^+ = \arg\max_{\tilde{x} \in [0,1]} \left[ h \int_0^1 \tilde{x} dF(x_i) \right].
\] (A1)

Substituting \( \pi_d = \Pi_d + k_d \) and \( \mu = \int_0^1 x_i dF(x_i) \) into equation (2) yields \( \pi_d = h \int_0^1 x_i dF(x_i) \). Thus,
\[
\pi_d - \pi_v
= h \int_0^1 x_i dF(x_i) - h \int_0^1 \tilde{x}^+ dF(x_i)
= h \int_0^1 x_i dF(x_i) + h \int_{\tilde{x}^+}^1 (x_i - \tilde{x}^+) dF(x_i).
\] (A2)

Both terms in (A3) are nonnegative. There cannot be a measure one of consumers at \( \tilde{x}^+ \) by maintained assumption. Thus, there must be a positive measure on either a subset of \( (0, \tilde{x}^+) \), in which case the first term in (A3) is positive, or on a subset of \( (\tilde{x}^+, 1] \), 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
\[
\sup_{\ell \in \{v,d\}} \left\{ \sup \left[ \frac{WF - WE}{D} \right] \times 1(\Pi_\ell = \max(\Pi_v, \Pi_d)) \right\}
\leq \sup \left[ \frac{WF_v - WE_v - WE_d}{D} \right] 1(\Pi_v \geq \Pi_d)
= \sup \left[ \frac{WF_v - WE_v}{D} \right] 1(\Pi_v \geq \Pi_d)
\] (A4)
\[
\sup \left[ \frac{WF_d - WE_d}{D} \right] 1(\Pi_d > \Pi_v).
\] (A5)

where \( 1(\cdot) \) is the indicator function and where the suprema are all taken over parameters \( (k_v, k_d) \in [0, \infty)^2 \). Equation (A4) holds by definition of \( WF \) and \( WE \). To see (A5), note that if a drug is developed in the first best, then \( WE_d = D - k_d = WF_d = WF \geq WE_v \). Thus if \( \ell = d \), then \( j = d \) as well. But then \( WF_d - WE_d = 0 \), 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 (A5). Manipulating the first braced term from (A5):
\[
\sup \left[ \frac{WF_v - WE_v}{D} \right] 1(\Pi_v \geq \Pi_d)
\leq \sup \left[ \frac{WF_v - WE_v}{D} \right]
= \sup \left[ \frac{(D - k_v) - (\pi_v + CS_v - k_v)}{D} \right]
\] (A6)

Condition (A6) follows from \( 1(\Pi_v - \Pi_d) \leq 1 \), (A7) from the definitions of \( WF_v \) and \( WE_v \), and (A8) from simple algebra. Manipulating the second braced term from equation (A5):
\[
\sup \left[ \frac{WF_d - WE_d}{D} \right] 1(\Pi_d \geq \Pi_v)
= \sup \left[ \frac{k_d - k_v}{D} \right] 1(\pi_d - k_d \geq \pi_v - k_v)
\] (A9)
\[
= \frac{\pi_d - \pi_v}{D}
\] (A10)

Equation (A9) holds by substituting the definitions of \( WF_v \), \( WE_d \), \( \Pi_d \), and \( \Pi_v \) and simplifying. Equation (A10) holds by noting that the greatest value of \( k_v - k_v \), 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 \( \pi_d \) and noting \( D/\pi_d = 1 \) since the firm can extract 100% of social welfare with a drug so that \( \pi_d = D \). Since \( CS_v \geq 0 \), (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 \( 2R \) parameters \( \{m_r\}_{r=1}^R \) and \( \{x_r\}_{r=1}^R \) satisfying the following feasibility conditions:
\[
m_r \in (0, 1) \text{ for all } r = 1, \ldots, R,
\] (A12)
\[
\sum_{r=1}^R m_r = 1,
\] (A13)
\[
0 \leq x_1 \leq \cdots \leq x_R \leq 1.
\] (A14)

We will choose these \( 2R \) parameters so that \( m_v/\pi_v \) is very close to \( 1/R \). We will do this by having the risk-class masses \( \{m_r\}_{r=1}^R \) decline geometrically and arranging the risk-class probabilities \( \{x_r\}_{r=1}^R \) so that the firm is indifferent between serving all consumers with a low vaccine price than serving a smaller group with higher prices.

Let \( \theta \in (0, 1/2) \). Define risk-class masses
\[
m_v = \begin{cases} 
\theta^{r-1} & \text{if } r > 1 \\
1 - \sum_{r=1}^{R-1} \theta^r & \text{if } r = 1.
\end{cases}
\] (A15)

It can be shown that this geometrically declining sequence respects constraints (A12) and (A13). We define the risk-class probabilities recursively as follows: set \( x_R = 1 \), and set
\[
hx_r \sum_{i=r}^R m_i = hx_{r+1} \sum_{i=r+1}^R m_i.
\] (A16)

for \( r = 1, \ldots, R - 1 \). The left-hand side of (A16) is the profit from charging a price \( hx_r \) and selling the vaccine to risk classes \( r \) and higher. The right-hand side is the profit from charging a price \( hx_{r+1} \) and selling to risk classes \( r + 1 \) and higher. It is easy to see that the risk-class probabilities respect constraint (A14). From equation (2), we have \( \pi_d = \sum_{r=1}^R \mu_r x_r \). By
construction implicit in (A16), we have $\pi_v = h x_1$; that is, it is weakly most profitable to charge $h x_1$ for the vaccine and sell to all consumers. Thus,

$$\frac{\pi_d}{\pi_v} = \frac{\sum_{r=1}^{R} h m_r x_r / h x_1}{m_1 + \sum_{r=2}^{R} m_r / m_1} = 1 - \sum_{r=1}^{R-1} \frac{\theta^r}{\theta} + \sum_{r=2}^{R} \theta^{r-1} + \ldots + \theta^{R-1}.$$  \hspace{1cm} (A19)

Equation (A17) follows from previous arguments. Equation (A18) holds since it is equally profitable to sell the vaccine to all consumers at price $h x_i$ or to consumers in risk classes $r$ and above at price $h x_r$, so that $h x_1 = h x_r (m_r + \ldots + m_R)$, implying $x_r = x_1 / (m_r + \ldots + m_R)$. Equation (A19) holds by substituting for $\{m_r\}_{r=1}^{R}$ from equation (A15). Taking limits, \[ \lim_{\theta \to 0} (\pi_v / \pi_d) = 1 - 0 + \sum_{r=2}^{R} 1 = R, \] 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 (A15) and (A16), 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

\[ R \pi_v = R \max_{r \in \{1, \ldots, R\}} \left[ h x_r \left( 1 - \sum_{i=1}^{r-1} m_i \right) \right] \geq R \max_{r \in \{1, \ldots, R\}} \left\{ h x_r, m_r \right\} \geq \sum_{r=1}^{R} h x_r m_r = \pi_d. \]

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

**Proof of Proposition 7:** Let $B$ be the value of the following minimization problem, labeled MIN1:

$$\min_{F} \left\{ \pi_v / \pi_d \right\}$$ \hspace{1cm} (A20)

subject to

$$\mu \geq m,$$ \hspace{1cm} (A21)

where $m$ is some constant in $[0, 1]$ and where the minimization is taken over the set of all functions $F$ satisfying the following three conditions:

$$\bar{F}(0) = 1,$$ \hspace{1cm} (A22)

$$\bar{F}(x_i) \in [0, 1] \text{ for all } x_i \in [0, 1],$$ \hspace{1cm} (A23)

$$\bar{F}(x_i) \text{ is nonincreasing.}$$ \hspace{1cm} (A24)

$B$ provides a tight lower bound on $\pi_v / \pi_d$ for a disease with a prevalence rate of at least $m \in [0, 1].$

We next establish several facts that will allow us to transform MIN1 into an equivalent minimization problem. First, integrating by parts shows

$$\mu = \int_0^1 x_i dF(x_i) = \int_0^1 \bar{F}(x_i) dx_i.$$ \hspace{1cm} (A25)

Second, we can show constraint (A21) binds. To do so, note that as the constraint is relaxed, the solution to MIN1 approaches 0 by Proposition 5. But $\pi_v / \pi_d$ approaches 0 for finite $\pi_d$ only if $\pi_v$ approaches 0. Furthermore, $\pi_v$ approaches 0 if and only if $\mu$ approaches 0, violating constraint (A21). Third, having established (A21) binds, we have $\pi_d = h \mu = h m r$.

Fourth, $\pi_v = h \max_{x \in [0, 1]} \{x \bar{F}(x)\}$. Substituting these four facts into MIN1 gives the equivalent problem, labeled MIN2:

$$\frac{1}{m} \min_{F} \left\{ \max_{x \in [0, 1]} \bar{F}(x) \right\}$$ \hspace{1cm} (A26)

subject to

$$\int_0^1 \bar{F}(x) dx \geq m,$$ \hspace{1cm} (A27)

where the minimization is again taken over the set of all functions $F$ satisfying (A22)–(A24).

We proceed to solve MIN2. Let $\bar{F}^*(x)$ be any solution to MIN2, and let $x^* = \max_{x \in [0, 1]} \{x \bar{F}^*(x)\}$. Because $x^*$ is a maximizer, $x \bar{F}^*(x) \leq x^* \bar{F}^*(x^*)$ for all $x \in [0, 1]$. Because $\bar{F}^*(x)$ is a solution to MIN2 and thus MIN1, it must generate a value of $B$ in objective function (A26), which upon rearranging implies $x^* \bar{F}^*(x^*) = B m$. Combining these equalities with condition (A23) implies, for all $x \in [0, 1]$,

$$\bar{F}^*(x) \leq \min \{1, B m / x\}.$$ \hspace{1cm} (A28)

Consider the function $\bar{F}^{**}(x)$ given by the right-hand side of (A28), i.e., $\bar{F}^{**}(x) = \min \{1, B m / x\}$. It can be verified that $\bar{F}^{**}$ yields $B$ as the value of the objective function (A26), that it respects constraint (A27), and that it satisfies conditions (A22)–(A24). Hence $\bar{F}^{**}$ must also be a solution to MIN2.

We argue that the constraint (A21) binds, implying that the equivalent constraint (A27) must also bind. Substituting $\bar{F}^{**}$ into (A27) treated as an equality yields

$$\int_0^1 \min \{1, B m / x\} dx = m.$$ \hspace{1cm} (A29)

which after integrating yields

$$B m [1 - \ln(B m)] = m.$$ \hspace{1cm} (A30)

Canceling terms and substituting $\mu = m$ from binding constraint (A21) gives the expression for $B$ in (3). \(Q.E.D.\)

**Proof of Proposition 8:** For a drug, $\Pi_d = WE_d = WF_d$. 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 $WE_d$ or $WF_d$) to do so.

For a case in which $\Pi_\nu > WE_d$ but $\Pi_d > \Pi_\nu$, suppose $x_i$ is uniformly distributed on $[0, 1]$; $k_j = 1/8$ for $j = v, d$; $c_j / s_j = 0$ for $j = v, d$; $h = 1$; $e_0 = 1$; and $e_d = 5/8$. For a drug, we have $\Pi_d = e_d \mu - k_d = (5/8)(1/2) - 1/8 = \ldots$.  

32
Proof of Proposition 9: Suppose \( y_i \) is independent of \( x_i \). Then \( \pi_v \) equals

\[
\max_{p \in [0, \infty)} \left\{ \int_{p/h}^h \left[ \int_{p/x}^h p \, dF_Y(y_i) \right] \, dF_X(x_i) \right\} \tag{A31}
\]

\[
\leq \int_1^1 \max_{p/h \in [0, \infty)} \left[ \int_{p/x}^h p \, dF_Y(y_i) \right] \, dF_X(x_i) \tag{A32}
\]

\[
= \int_1^{\infty} \max_{p/h} \left[ \int_{p/x}^h p \, dF_Y(y_i) \right] \, dF_X(x_i) \tag{A33}
\]

\[
\leq \mu \max_{p' \in [0, \infty)} \left[ p' \, F_Y(p') \right] \tag{A34}
\]

\[
= \pi_d. \tag{A35}
\]

Equations (A31) and (A35) hold by applying the independence condition to the formulae (4) and (5) and noting \( \tau_j = \Pi_j + k_j \), \( j = v, d \). The rest of the steps are algebraic manipulations. The inequality in (A32) is strict if there is nontrivial heterogeneity in the distribution of \( x_i \).

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

\[
\pi_d \geq \hat{y}_i \int_0^1 \int_0^{x_i} \, dF(x_i, y_i) \tag{A36}
\]

\[
\geq \hat{y}_i \int_0^1 \int_0^{g^{-1}(\hat{y}_i)} \, dF(x_i, y_i) \tag{A37}
\]

\[
= g^{-1}(\hat{y}_i)\hat{y}_i F_Y(\hat{y}_i) \tag{A38}
\]

\[
= \pi_v. \tag{A39}
\]

Equation (A36) holds because the producer surplus at the optimal drug price \( \pi_v \) at least weakly exceeds producer surplus from a drug sold at price \( \hat{y}_i \). Equation (A37) holds because \( g^{-1} \) is an increasing function, so \( x_i \geq g^{-1}(\hat{y}_i) \) for \( y_i \geq \hat{y}_i \). Equation (A38) is a straightforward calculation. Equation (A39) follows from the previous calculations regarding vaccine producer surplus. The inequality in (A37) is strict if there is nontrivial heterogeneity in the distribution of \( x_i \) for vaccine consumers. \( Q.E.D. \)

Proof of Proposition 10: Let \( \pi_v' \) and \( \pi_d' \) be producer surpluses in the model with no income heterogeneity and \( \pi_v \) and \( \pi_d \) be producer surpluses when income heterogeneity which is independently distributed from infection-risk heterogeneity has been added to the model. Then \( \pi_v' \) equals

\[
p_v^* \Pr(z_i \geq p_v^*) \tag{A40}
\]

\[
\geq p_v^* p_y^* \Pr(z_i \geq p_v^*) \Pr(y_i \geq p_y^*) \tag{A41}
\]

\[
\geq p_v^* p_y^* \Pr(z_i \geq p_v^*) \Pr(y_i \geq p_y^*) \tag{A42}
\]

\[
= \pi_v \pi_d'/\pi_d. \tag{A43}
\]

Equation (A40) follows from equation (4). Condition (A41) follows because \( \rho_v^* \), as an argmax, produces a higher value for \( p \Pr(z_i \geq p) \) than \( p_v^* p_y^* \). Condition (A42) follows since \( \Pr(y_i | y_i > p_v^*) \geq \Pr(y_i | y_i \geq p) \). Equation (A43) follows because \( \pi_v = p_v^* \Pr(x_i \geq p_v^*) \) by equation (1), \( \pi_d = h\mu \) by equation (2), and \( \pi_d' = \mu p_y^* \Pr(y_i \geq p_y^*) \) applying the independence assumption to equation (5). Conditions (A40)–(A43) 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 (A42) would be strict. \( Q.E.D. \)

Proof of Proposition 11: 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 \( \Pi_b \), the monopolist’s profit from developing both products, equals \( \Pi_d + \Pi_{d'} \). Conditional on developing both, the monopolist’s optimal pricing strategy is to charge a price maximizing profit from sales to generation \( O_i \), yielding marginal profit \( \Pi_{d'} \) and charging a vaccine price that maximizes profit from sales to generation \( Y_i \) given generics will enter the drug market, yielding marginal profit \( \Pi_d \). Fact 2 is that \( \Pi_b = \Pi_d + \Pi_{d'} \). This holds because \( \Pi_{d'} \geq \Pi_{d'} \) 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 \( \Pi_{d'} < \Pi_{d'} \) by Fact 2, in turn implying \( \max(\Pi_d, \Pi_{d'}) < \max(\Pi_{d'} , 0) \), 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) \( \min(\Pi_{d}, \Pi_{d'}) > 0 \) or (b) \( \Pi_{d'} > 0 \). If (a) holds, then by Fact 1, \( \Pi_b = \Pi_d + \Pi_{d'} > \Pi_d > 0 \). Thus, \( \max(\Pi_{d'}, \Pi_b) > \max(\Pi_{d}, 0) \). Thus a vaccine is developed in equilibrium in Model 2. If (b) holds, then again \( \max(\Pi_{d'}, \Pi_b) > \max(\Pi_{d}, 0) \), 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 $k_v \in ((1 - e_d) h, (1 - e_d) h + k_d)$. It can be shown that $\Pi_d = e_d h - k_d > 0$, $\Pi_{vd} = h - k_v$, 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 $k_v < (1 - e_d) h + k_d$, $\Pi_{vd} > \Pi_d$. Hence $\Pi_{vd} > \Pi_d > \Pi_d + \Pi_{vd} = \Pi_v$, 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 12: 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 $k_v \in ((1 - e_d) h, h)$. 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, $\Pi_{vd} > 0$. But, as noted in the text of Section ??, $\Pi_{vd} > 0 > \Pi_d$ implies that a vaccine alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{\delta h}{1 - \delta} - k_v.$$ (A44)

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 ??, $\Pi_{vd} > 0 > \Pi_{vd}$ implies that a drug alone is developed. The present discounted value of the stream of social welfare in equilibrium is

$$\frac{e_d h}{1 - \delta} - k_d.$$ (A45)

The limit as $\delta \to 1$ of the ratio of expression (A44) to (A45) equals $1/e_d$. Thus, for $\delta$ sufficiently close to one, both $k_i$ and social welfare are higher in the first than the second case.

Q.E.D.

Additional Proposition: The Appendix concludes with the statement and proof of an additional proposition referenced in the text.

Proposition 13. 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 $\sup_{k_j, e_j, s_j} ((WF - WE)/D)$.

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 $\pi_v/\pi_d$ 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% of social surplus. The proof of Proposition 8 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.
References


